ACTA UNIVERSITATIS AGRICULTURAE SUECIAE

SLU

DOCTORAL THESIS NO. 2024:38
FACULTY OF VETERINARY MEDICINE AND ANIMAL SCIENCE

Exploring facial expressions in horses

- biological and methodological approaches

JOHAN LUNDBLAD

Exploring facial expressions in horses

- biological and methodological approaches

Johan Lundblad
Faculty of Veterinary Medicine and Animal Science
Department of Animal Biosciences
Uppsala

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Uppsala 2024

Acta Universitatis Agriculturae Sueciae
2024:38

Cover: Illustration of the facial region of a horse. Copyright 2024 Nora Signér

ISSN 1652-6880
ISBN (print version) 978-91-8046-340-9
ISBN (electronic version) 978-91-8046-341-6
https://doi.org/10.54612/a.15rcen99qf

Swedish University of Agricultural Sciences,
Department of Animal Biosciences, Uppsala, Sweden

Print: SLU Grafisk service, Uppsala 2024

Exploring facial expressions in horses - biological and methodological approaches

Abstract

Facial expressions can be used as a tool to recognize pain and stress. However, these states are often intertwined, and sedative drugs may also influence pain recognition tools utilizing facial expressions. In order to investigate how the states of stress and sedation could affect facial expressions of pain, the Facial Action Coding System for horses was used in situations where nociception, stress and sedation were experimentally induced, both separately and in combinations, in a cross over design. The frequencies of the Action Units ”upper lid raiser”, “eye white increase”, “inner brow raiser”, “blinking”, “ear movements” and “nostril dilator” all increased during stressful managerial situations ($p<0.05$), supporting empirical descriptions of a stressed horse. No change in frequency of Action Units during sedation could be found in comparisons to baseline. A Partial Least Squares Discriminant Analysis which created two weighted components containing the frequency of Action Units, (t1 $Q^2>0.05$, t2 $Q^2<0.05$), could successfully discriminate the states of sedation, social isolation stress and nociception (pain) from each other. A combined state of nociception-isolation could not be discriminated from isolation without nociception. The combination of nociception-sedation did not differ from a neutral horse, suggesting that sedation conceals the frequency of Action Units of nociception. Blinking frequency increased in a sedated horse with nociception but not in a sedated horse without nociception ($p=0.011$). This finding needs further investigation. Lastly, dimensionality reduction methods were compared and their accuracy was measured. Highest accuracy was 76.9 % meaning that certain methods can introduce thresholds for automated evaluations, but methods are suggested to include intensity and temporal information of facial contractions as well as physiological- and behavioural measures for increased performance. In conclusion, the presence of stress, as well as sedation should be carefully considered when evaluating pain using facial expressions.

Keywords: stress, pain, sedation, facial expression, horse, methodology, EquiFACS, compound, dimensionality reduction, rater agreement

Exploring facial expressions in horses - biological and methodological approaches

Sammanfattning

Ansiktsuttryck kan användas som ett verktyg för att upptäcka smärta och stress. Dessa tillstånd bär många liknelser. Vidare kan sedering påverka vilka ansiktsuttryck som ses. För att undersöka hur dessa tillstånd kan påverka ansiktsuttryck vid smärta videofilmades hästar i situationer där nociception (smärta), stress och sedering var inducerat experimentellt, separat och i kombination. Videofilmerna analyserades med ett kodningssystem för ansiktsrörelser, Facial Action Coding System, utvecklat för hästar. Vid stress sågs en ökad frekvens av höjning av övre ögonlocket och mediala ögonmuskeln, visad ögonvita, blinkningar, öronrörelser samt dilatation av näsborrarna (p<0,05) vilket stödjer empiriska beskrivningar av stressade hästar. Ingen statistiskt signifikant ändring i frekvens av ansiktsrörelser sågs vid sedering. En Partial Least Squares Discriminant Analys genererade två viktade komponenter innehållande frekvensen av ansiktsuttryck (t1 Q^2>0,05, t2 Q^2<0,05) som kunde skilja på sedering, social isolering (stress) och nociception (smärta). När nociception och isolering kombinerades kunde ansiktsuttrycken inte skiljas från enbart isolering. Kombinationen av nociception och sedering visade ingen skillnad från en neutral häst, vilket tyder på att sedering minskar frekvensen av ansiktsuttryck hos en häst i smärta. Blinkfrekvensen ökade hos en sederad häst med smärta men inte hos sederade hästar utan smärta (p=0,011), ett fynd som kräver vidare forskning. Till sist utforskades metoder för dimensionsreduktion av dataseten och binära klassifikationstest mätte hur väl de presterade. Noggrannheten i dessa tester låg som mest på 76,9 %. Särskilda tröskelvärden kan därför introduceras om automatisk avläsning ska tillämpas men metoder bör inkludera intensitet och temporal distribution av ansiktsuttryck samt kompletteras med beteenden eller fysiologiska parametrar för bättre noggrannhet. Sammanfattningsvis så påverkar stress och sedering smärtavläsning med ansiktsuttryck och närvaron av dessa bör tas i beaktning när hästar smärtbedöms.

Keywords: stress, smärta, sedering, ansiktsuttryck, häst, metodologi, EquiFACS, sammansatta tillstånd, dimensionsreduktion, observatörers tillförlitlighet

Dedication

To Stig, Majken and Bertil.

Whose support made this thesis possible despite not being here to see it finished.

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Lundblad J., Rashid M., Rhodin M., Haubro Andersen P. (2021). Effect of transportation and social isolation on facial expressions of healthy horses. PLoS One, 16(6): e0241532.

II. Lundblad J., Rhodin M., Hernlund E., Bjarnestig H., Hidén Rudander S., Haubro Andersen P. (2024). Facial expressions during compound interventions of nociception, conspecific isolation, and sedation in horses. *(Manuscript)*

III. Lundblad J., Ask K., Hernlund E., Rhodin M., Haubro Andersen P. (2024). A framework for analyzing and interpreting Facial Action Coding System data in horses. *(Manuscript)*

1. Introduction

The horse has played a significant role throughout the span of human history, in uses ranging from agriculture and warfare to recreation and sport (Ransom & Kaczensky 2016). Today, the horse is still a source of joy, purpose and well-being for many people. Like other animals bred and kept by humans, ethical and welfare aspects of horse-keeping need to be considered, since there are many situations in the modern world in which welfare can be challenged such as housing, management or the use of horses in sport.

Animal welfare can be described in terms of five domains (Mellor & Beausoleil 2015), where the first three domains mainly involve physical or survival needs. The fourth domain mainly refers to the possibility for the animal to perform natural behaviours. The fifth domain, which concerns the mental welfare of the horse, involves "freedom from fear and distress" and "freedom from pain, injury and disease". While this looks simple in theory, freedom from these negative welfare states is largely an unresolved issue. Although Mellor & Beausoleil (2015) discussed the need to further expand the five domains model to include positive welfare states, recognition of negative welfare states is still an important issue. The five domains approach primarily focuses on physical measures on a population level, which ensures that correct decisions are made regarding the care and use of animals. However, it is known that experiences involving negative welfare components are highly individual (Ijichi et al. 2014). Quality-of-life centred appraisal of animal welfare, especially in the context of compromised animal welfare on individual level, has been suggested as the optimal way forward in determining the motivations and needs of the individual animal, *e.g.* during evaluation of pain and stress in the horse (Muir 2013).

In the clinical setting, methodologies for recognition of negative states in horses are still lacking, especially states relating to a low degree of pain (de

Grauw & van Loon 2016). This could be due to a number of confounding factors, including the fact that the range of procedures commonly performed in an equine clinic produce great variation in the number of behaviours displayed (Torcivia & McDonnell 2021). Moreover, the presence of an observer during evaluation of discomfort may affect the outcome of the evaluation (Torcivia & McDonnell 2020). Other important factors include the influence of other biological systems, *e.g.* stress or effect of drugs, but remain rather unexplored. Because horses are non-verbal, the causes and impact of factors that cause negative welfare impairment are challenging to identify. However, identification of such factors is an important task, because a horse in pain needs to be handled differently than a horse in discomfort due to other affective states, *e.g.* stress may have a multitude of aetiologies which could address the rider or stable habits rather than treatment of pain.

In the remainder of this chapter, a number of these concepts are introduced and summarised, as background to the research questions addressed in the thesis. These concepts primarily refer to biological systems, some of which may involve an emotional component. However, since the biology of emotions is not yet fully understood (Cabanac 2002), the emotions of the horse are not addressed in this thesis. Instead, the possible impact of the different states on the horse as a whole is discussed, with the focus on detecting and categorising differences in facial expressions in horses experiencing pain, stress and sedation, and combinations of these.

The physiology of relevant body systems (nociception, pain, stress, anxiousness, fear) and their interrelations are described in section 1.1. In subsequent sections, the focus turns to facial expressions and how they can be assessed in the horse, in order to shed light on relevant research questions.

1.1 The physiology of body systems

1.1.1 Nociception

Nociception refers to the body system which processes and transmits information originating from a noxious stimulus (Gaynor & Muir 2014). Information on a noxious stimulus is generally carried through afferent neurons, called nociceptive neurons. Peripheral nerve endings (nociceptors) register the nociceptive input by detecting changes in many types of tissues. These changes, such as heat, cold, chemical changes or pressure, may all contribute information to these highly flexible nerve endings in order to transduce a noxious stimulus into a signal. This signal activates a separate pathway from other sensory systems, with the primary goal of activating the pain pathway. Nociceptors are present in a wide array of tissues, including the viscera. However, the focus in this thesis is on the nociceptive pathway of somatic tissues, since this thesis partly includes pain inductions. In mammals, the nociceptors in somatic tissues, such as the skin and muscles, consist of two types of fibres: Aδ- and C-fibres (Smith & Lewin 2009). The C-fibres are unmyelinated, and thus slower, while the Aδ-fibres are lightly myelinated and transmit at higher speed, resulting in a sharper and more intense sensory experience than that provided by the C-fibres.

Although much remains to be determined with regard to how nociceptors are activated, many receptors within the membrane of the nerve endings have been identified and transient receptor channels (TRP) in particular have been shown to play a large role in transduction. For example, heat is transduced through TRPV1 receptors, whose activation allows the neuron membrane to depolarise and forward the signal. Similarly, TRPM8 receptors react to cold, while acid-sensing ion channels react to chemical stimulants (Basbaum et al. 2009). Mechanical transduction has several proposed mechanisms of transduction. These include high-threshold mechanosensitive muscle receptors, which respond to chemical alterations within muscles (Graven-Nielsen & Mense 2001), and more direct forms of transduction, such as depolarisation of the membrane due to cell damage. The threshold for activating these nociceptors varies and may be modified locally by signal substances including, but not limited to, prostaglandin, nerve growth factor, free protons or adenosine triphosphate (Julius & Basbaum 2001). These modulatory properties of nociceptors are important to note, since inflammation or previous injury can affect the excitability of nociceptors and thus may have an impact on the perception of pain.

Farther along in the afferent direction of the nociceptive pathway, the signal reaches the central pain pathway through the dorsal root ganglia and crosses over the dorsal spinal horn into the anterolateral system, from which, through third-order neurons, it is conveyed to several parts of the brain. The discriminatory and intensity-determining parts of the signal are mediated through the thalamus and then reach the primary somatosensory cortex, while the affective and behavioural parts of the system are activated through the amygdala, cingula and insula cortex, and periaqueductal grey (PAG) (Augustine et al. 2024).

1.1.2 Pain

When the nociceptive input is perceived in the somatosensory cortex, a pain experience occurs. The concept of pain is often mentioned in different contexts and is a complex system which affects several body systems, resulting in hormonal, behavioural and emotional changes. The International Association for the Study of Pain (IASP) defines pain as:

> "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" (Raja et al. 2020)

In most cases, an association with a sensory input and experience, together with emotional processing of the input, are present for an individual to experience pain. This means that a nociceptive input in itself does not infer pain, *e.g.* the patient may be anaesthetised and thus not perceive the nociceptive input even though physiological responses occur. In the remainder of this thesis, the word 'pain' mostly refers to nociceptive pain, due to the experimental pain modality used, even though there are multiple types of pain that do not stem from activation of nociceptors. This definition is done mainly with the experimental induction methods in mind. The distinction between nociception and pain is debated by scientists, but it is thought that nociception is only a part of the explanation of perceived pain, since the same nociceptive input can elicit an extraordinary range in the degree of pain perceived by healthy subjects (Fillingim 2017).

This phenomenon is not a strange, when the physiology of pain modulation is considered. Modulation occurs through descending pathways, mainly after activation of PAG, which plays a central role in the descending pathway (Rainville 2002). Activation of PAG triggers the release of serotonin, which is transported to the dorsal root of the spinal ganglion where it binds to inhibitory interneurons. On binding to serotonin, these neurons release endogenous opioids that have an inhibitory effect on incoming Aδ- and C-fibres, diminishing the sensory input. This system can be activated by a multitude of factors that are mainly determined by the individual subject experiencing the pain. For example, memory (Reicherts et al. 2016), sex, race, expectations (Rainville et al. 1999), controllability (Salomons et al. 2007) and social modulation (Benedetti et al. 2014) are all factors that can influence the modulation of pain. Thus, it is important to take these factors into account when evaluating pain. In humans, previously used pain evaluation tools range from verbal-based tools (Melzack & Torgerson 1971) to numerical scales (Wong & Baker 1988), following an early investigation of clinical pain assessment tools (Hardy et al. 1947). Several new approaches for evaluating pain in horses using scales have attracted the interest of the scientific community in recent years. No single physiological parameter has the capacity to assess different types of pain under clinical conditions (Gleerup & Lindegaard 2016), and therefore use of behaviours, sometimes in combination with facial expressions or physiological parameters, has become common in pain assessments. Behavioural tools are highly flexible and are often used within known contexts in the clinic, such as in post-surgical evaluations (Pritchett et al. 2003). However, it is important to note that knowledge of context is necessary in order to draw conclusions about pain from these scales, since similar behaviours may stem from other causes, *e.g.* stress (Young et al. 2012).

1.1.3 The stress response

The word 'stress' carries many meanings, depending on context. For humans, the concept 'I feel stressed' indicates confusion with stress as an emotion. Certain emotional and stress neural pathways share a commonality (Wang & Saudino 2011), and a stress response may elicit an emotional response (Lazarus 1991). However, emotions causing stress may also be a reality, making any causality between these impossible to determine (Dantzer 2001). Moreover, there is still lack of consensus within the scientific collective on what exactly constitutes an emotion in humans (Cabanac 2002). Therefore, delving further into subjective emotional states during stress in animals is rather unwise, and was far beyond the scope of this thesis. Instead, in this thesis, stress was considered to be a neuroendocrine body system and an adaptive response to a harmful or potentially harmful stimulus of an unpredictable or uncontrollable nature (Koolhaas et al. 2011) that invokes the regulatory capabilities of the neuroendocrine functions and behaviours in order to maintain homeostasis. Emotional involvement is highly probable in most mammals, but difficult to quantify without verbal reporting or specific consideration in the study design. Stress in horses is a topic of interest for many research disciplines and for clinical and welfare applications, and has thus been studied in connection with common management procedures such as road transportation (Smith et al. 1996), which is considered to induce stress. Other studies have explored the effects of social isolation which, due to the social structure of feral horses, also induces a stress response (van Dierendonck et al. 2005).

Stress induces both physiological and behavioural responses. Physiologically, stress activates the sympathetic part of the autonomous nervous system and the neuroendocrine system. Specifically, two distinct neuroendocrine pathways are activated during stress: the hypothalamic-pituitary-adrenocortical (HPA) axis and the sympatho-adrenomedullar system. The HPA axis is activated when the hypothalamus receives a signal to activate. The hypothalamus then produces corticotropin-releasing hormone, which in turn causes the pituitary gland to release adrenocorticotropic hormone, thyroid-stimulating hormone, vasopressin and growth hormone. This cascade causes a number of metabolic and haemodynamic changes necessary to put the body in ‘fight-or-flight mode’. Adrenocorticotropic hormone has its effect on the adrenal cortex, with its main end-product being release of corticosteroids into the blood stream (Hurcombe 2011). Activation of the adrenal medulla may instead come directly from the sympatho-adrenomedullar system, through efferent sympathic innervation directly to the medulla of the adrenal gland, stimulating the release of catecholamines. The stress response may be clinically detectable as elevated heart and respiratory rate, blood pressure and temperature (Buchanan 2000), as well as increase in blood parameters such as serum cortisol or adrenaline (Ayala et al. 2012). These responses occur together with a number of short-term and long-term behavioural changes (Buchanan 2000). In horses, these behaviours are generally associated with restlessness, *e.g.* excessive pawing, throwing of the head, vocalisation, defecation, pacing or other nervous behaviours such as quivering or attempts at fleeing (Harewood & McGowan 2005; Young et al. 2012). Dilated nostrils, backwards-orientated ears and repetitive mouth movements (yawning, tongue movements) are facial expressions that may be present (Young et al. 2012). Blink frequency, which has been proposed as a tool to measure stress in horses, also increases during stress (Mott et al. 2020).

1.1.4 Anxiousness and fear

Fear and anxiety are not part of a physiological body system and are rather classified as emotions, but both are considered to have a negative impact on animal welfare (Carey & Fry 1995) and in clinical work (Reid et al. 2017). The two concepts are generally discussed in parallel, since some overlap occurs (Öhman 2008). The main difference is that fear generally refers to a response to an actual threat, while anxiety concerns the emotional response arising due to worry about a potential threat. Determining whether an animal perceives a threat as real or potential requires complex experimentation. In this thesis, perceived threat was assessed indirectly as a cause of negative impact on quality of life in the horse. Gaining information on emotional states is difficult without verbal communication, which is important to recognise when drawing conclusions regarding evaluation of negative states in animals. However, certain approximations can be used, *e.g.* current understanding of animal behaviours in relation to a model designed to test for anxiety (Bourin et al. 2007). In horses, proxies for anxiety overlap with some of the metrics of stress, but may still involve novel object testing (Visser et al. 2001) or other behavioural analyses (Reid et al. 2017). Another way to test for anxiety is to measure responses arising on administering anxiolytic drugs. There are several anxiolytic drugs available for different animal species. In horses, acepromazine (a derivate of phenothiazine) and benzodiazepine are commonly used. Benzodiazepine is mainly used during induction of anaesthesia and as a sedative for foals (Shini 2000), while acepromazine is used as pre-medication to anaesthesia (Driessen et al. 2011) and an anxiolytic and sedative during common clinical procedures (López-Sanromán et al. 2015). Acepromazine in itself does not provide analgesia but, due to the excitatory effects of anxiety on pain, can work as a good complement to other analgesic drugs (Sellon 2015). It can be administered intravenously or intramuscularly and the effect reaches a maximum within 30-60 minutes, with recovery from sedation starting within 60-90 minutes (Pequito et al. 2012). It has a half-life of 150-180 minutes in blood (Schneiders et al. 2012; Knych et al. 2018).

1.1.5 Interrelations of pain, stress and anxiety systems

Although pain, stress and anxiety are often studied alone, the systems are closely related to each other and share some commonalities. For instance, pain in itself is considered a stressor (Gaynor & Muir 2014). Stress in itself may modulate the perception and experience of pain, either by elevating it (Jennings et al. 2014) or diminishing it (Butler & Finn 2009), which means that stress has a major impact on how the individual animal expresses pain. Consequently, physiological parameters reflecting pain and stress may share commonalities. For example, heart rate, respiratory rate and blood pressure increase in the presence of both pain and stress (Rietmann et al. 2004). Similarly, behaviours such as restlessness, pawing and head movements may be included in ethograms developed for different causes *e.g.* pain (Gleerup & Lindegaard 2016) and stress (Young et al. 2012).

Because of the intricacy of these body systems and the involvement of emotions, causality of these states may not be wholly clear. This is especially important when using physiological markers and behavioural methods in unspecific contexts. Increased heart rate does not necessary infer a negative experience such as anxiety or fear, since simple exercise causes an increase (Hörnicke et al. 1977). Cortisol release displays diurnal variation (Hoffis et al. 1970) and may be affected by pathologies (Rietmann et al. 2004). Similarly, anxiousness may cause stress, but presence of stress does not infer an anxious horse and an anxious horses may experience pain differently than a calm one (Chayadi & McConnell 2019).

1.2 Facial expressions

Facial expressions are one of the most extensive communication tools in the human species (Straulino et al. 2023). Facial expressions originate from three different pathways, physiological, communicative and emotional (Waller et al. 2020). They consist of voluntary or involuntary contraction of muscles in the facial area and all contractions in the face is activated by different branches of the facial nerve (König & Liebich 2020). An early study by Duchenne (1876) combining the use of transcutaneous stimulation of individual human facial muscles and photography of the face revealed that emotions and voluntary contraction of facial muscles have separate pathways. This raised awareness of the role of facial expressions in communication of internal states, an aspect that has since been studied extensively, *e.g.* as regards recognition of pain in patients unable to report on their own pain verbally (Wong & Baker 1988). The non verbal part has since been recognised in the updated IASP definition of pain, which underwent some revisions in 2020 and now states that:

> "Verbal description is only one of several behaviours to express pain; inability to communicate does not negate the possibility that a human or a nonhuman animal experiences pain" (Raja et al. 2020).

For humans, this is especially important with regard to patients with speaking impairments who may be unable to communicate their pain verbally, but certainly experience the same degree of pain as those able to speak. This surprisingly late definition was long due for another group of patients unable to verbalise their pain, namely neonatal infants. Until the 1980s, the standard anaesthesia protocol for open heart surgery in small infants was nothing more than muscle relaxants, such as benzodiazepines, which do not have any analgesic properties (see section 1.1.4), and nitrous oxide gas (Castaneda et al. 1974). Luckily, the need for non-verbal assessment of pain quickly gained attention and it was mainly due to the need for pain recognition in infants that use of facial expressions to detect pain started to attract attention in the clinical world (Baker & Wong 1987).

This paradigm change in attitudes to pain detection has also affected animal sciences. In horses, anaesthesia was partly considered as a means of fixation of the animal for certain procedures in the past, but as understanding of pain grew, anaesthesia protocols were amended to accommodate better analgesia procedures (Stevenson 1963). Regarding pain recognition, physiological parameters and subjective assessment were formerly the mainstays in pain perception in horses. However, facial expressions have begun to play a more prominent role since the facial expression of pain displayed by human infants was proven to be present in animals as well (Langford et al. 2010). In fact, the morphology of facial expressions has seemed to remain consistent between mammal species over the course of evolution, despite marked differences in the morphology and anatomy of their faces (Chambers & Mogil 2015).

Studies began exploring facial expressions in horses in greater depth, including facial expressions of pain (Dalla Costa et al. 2014; Gleerup et al. 2015; van Loon & Van Dierendonck 2015). It is now recognised that assessment of facial expressions can be a valuable, but underused, tool for assessment of animal welfare (Descovich et al. 2017). Clinical problems with pain detection include the fact that horses hide painful behaviours because of their status as prey animals (Coles 2016; Coles et al. 2018; Torcivia & McDonnell 2020). It is therefore suggested that evaluation of pain should be performed remotely, which Torcivia & McDonnell (2020) did manually. However, this is time-consuming and interesting findings has emerged with the advent of data-driven artificial intelligence and computer vision approaches was that facial expressions show distinct qualities that can be evaluated through automated recognition using video surveillance (Andersen et al. 2021; Boneh-Shitrit et al. 2022; Feighelstein et al. 2022).

1.2.1 Anatomical basis for facial expressions in horses

The facial muscles, which consist of an attachment from bone to skin or circular muscles, are activated through innervation of *nervus facialis* and may contract independently of each other (König & Liebich 2020). There are a number of facial muscles present in the face, with these performing different functions. For example, the facial muscles concerning mimicry functions are generally the superficial muscles of the external part of the eyes, ears, nostrils and muzzle (Table 1), while large and deeper muscles have more profound functions such as mastication or feed intake. Certain muscles may have pure physiological or behaviourally motivational functions, such as blinking for the extra orbital muscles or rotation of ears to better pick up sounds. Horses has 17 separate muscle contractions in the face, which is surprisingly many. Furthermore, despite the large differences in facial morphology compared with *e.g.* human anatomy, many of the superficial muscles of the equine head share many similarities with those of other mammals (Wathan et al. 2015).

Table 1. Important muscles for facial mimicry in horses (summary based on Wathan et al. (2015) and Constantinescu (2018)).

Area	Muscle	Function
Extra orbital eye (Upper face)	*m. orbicularis oculi*	Sphincter muscle around the eye aperture, closes and opens the eye
	m. levator anguli occuli medialis	Raises the medial part off the upper eyelid
	m. malaris	Weak palpebral muscle of the eye
Nostrils, mouth and chin (Lower face)	*m. orbicularis oris*	Sphincter muscle around the lips, closes and opens the mouth aperture
	m. caninus	Pulls the lateral nostril wing
	m. levator nasolabialis	Pulls the upper lip and nostril towards orbita
	m. levator labii superioris	Raises the upper lip
	m. mentalis	Tenses the lower lip
	m. depressor labii inferioris	Depresses the lower lip
	m. zygomaticus	Pulls the corner of the lip towards the facial crest
	m. buccinator	Profound muscle of the cheek
	m. dilator naris apicalis	Dilates the nostril in medial direction
	m. lateralis nasi	Diffusely dilates the nostril caudally and dorsally
External ear	*mm. auriculares rostrales*	Raises the ears rostrally
	mm. auriculares caudales	Rotates the ears caudally
	mm. auriculares ventrales	Pulls the ears ventrally and caudally

1.2.2 Neurobiology of contractions of facial muscles

As mentioned, facial expressions stem from activation of branches of *n. facialis*. Activation of these efferent nerves proceeds via two pathways: descending pathways from the motor cortex and brainstem, which are responsible for voluntary movements, and descending pathways from the limbic system and hypothalamus, which account for facial movement in expression of emotion (Müri 2016). Convergence of these two systems into common motor neurons has been detected in patients with partial facial paralysis (Trosch et al. 1990). When those patients were asked to show their teeth, the smile was symmetrical, but when they responded to something funny, the smile was asymmetrical (Trosch et al. 1990). Interestingly, humans are able to pick up on the difference between a volatile smile and an emotional smile, *e.g.* subjects in one study were able to differentiate between an actor producing a smile and a person genuinely laughing (Poole & Craig 1992). This suggests that the interpretation is hardwired into our genes and that small differences have a great impact in interpretation of facial communication. The system is also rather complex and, although innervation *of n. facialis* is mainly referred to as efferent, some studies suggests that the mechanism may actually be more complicated. In studies where subjects were given specific step-by-step instructions on how to produce a certain emotional expression in the face, without knowledge of which emotion is being expressed, it has been found that adopting the expression described also gives rise to the underlying emotional state (Augustine et al. 2024). This complicates analysis of cause and effect regarding facial expressions and indicates that some emotions could emerge as feedback from efferent input from muscles.

1.3 Measuring facial expressions

1.3.1 Grimace scales

The mouse grimace scale was the first to be developed for the purpose of recognising pain in animals and was developed for the laboratory environment (Langford et al. 2010). Since then, a number of grimace scales have been developed for other species of animals (McLennan et al. 2019). Grimace scales focus on specific areas of the face of the animal and evaluate whether certain conditions are met, *i.e.* whether a grimace consisting of several simultaneous contractions of facial muscles, forming a morphological change in the face, is present. Scores typically range from 0-2 and are summarised to provide the observer with a total score for the subject. The Horse Grimace Scale (HGS), developed in 2014 for horses undergoing routine castration, uses samples from video surveillance in horses and was developed in a similar way to scales for other species (Dalla Costa et al. 2014). This scale was later been evaluated with regards to emotional impact as well (Dalla Costa et. al. 2017), where emotions were deemed to have little impact on the scale. Independently, Gleerup *et al.* (2015) developed a description of the Equine Pain Face, based on experimentally induced nociceptive stimulus through capsaicin-induced activation of the HRPV1-recptor and activation of mechanoreceptors by the use of ischemic induction in the foreleg. Results from the trial was later used to develop an ethogram based on the facial findings and common behaviours associated with pain (Gleerup & Lindegaard 2016). At around the same time, van Loon & Van Dierendonck (2015) developed the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP), using direct observations on colic-affected horses in the university clinic in combination with common behaviours associated with abdominal pain, highlighting important facial features. Components of the different grimace scales developed for horses are summarised in Table 2.

Table 2. Summary of morphological features indicative of pain face in the horse

Scale item	**Horse Grimace Scale (Dalla Costa et al. 2014)**	**Equine Pain Face (Gleerup et al. 2015)**	**Equine Utrecht University Scale for Facial Assessment of Pain (van Loon & Van Dierendonck 2015)**
Ears	Stiffly backwards ears	Asymmetrical or low position	Backwards and no response to sounds
Eyelids	Orbital tightening	Withdrawn or intense stare	Opened or tightened eyelids. Sclera visible
Area above eye	Tension above eye	Angled eye	-
Muscle tonus	Strained chewing muscles and mouth, pronounced chin	Tension of muzzle and mimic muscles	Fasciculations in muscle tone, corners of mouth lifted
Nostrils	Strained and flattened profile	Square-like shape	More opened, flaring, and possible audible breathing

1.3.2 Facial Action Coding System

The Facial Action Coding System (FACS), developed by Ekman *et al.* (2002), sought to standardise the coding of facial expressions in humans. Over time, systems were developed for several other mammals, including horses (Wathan et al. 2015). The system is built up of standardised codes, called Action Units (AU) or Action Descriptors (AD), that are assigned to a visible change in the face. Each AU is attributed a muscular basis and are coded based on visual landmarks and changes in appearance in related to that specific muscle contraction. The ADs instead refer to general changes in the face that involve multiple active muscles or where the underlying basis cannot be identified. The purpose of the system is to provide a tool from which facial expressions can be measured without any presumption about their meaning beforehand. The system has now been standardised for multiple species and its use allows for comparisons within and between species, which is especially relevant when drawing conclusions regarding the evolution of facial expressions (Waller et al. 2020). All existing animal systems, derived from FACS developed for humans by Ekman *et al*. (2002), all require training to use (McLennan et al. 2019). The equine system (EquiFACS; Table 3) was developed using video material of naturally occurring facial expressions in horses, which were correlated to superficial muscles on the basis of dissection (Wathan et al. 2015).

Because of the nature of facial movements where facial expressions are not static and instead subject to dynamic changes within a very short period (Parr et al. 2005), temporal and contextual information should be considered when measuring FACS data. For example, Rashid et al. (2020) found that the likelihood of observing multiple co-occurring AUs in videos of horses experiencing pain, compared with a baseline, was far more likely when using videos rather than still images. Those authors identified the AUs *chin raiser* (17) and *half-blink* (47) during pain, as well as *ears backwards* (EAD104) and *nostril dilator* (AD38). A recent study by Ask et al. (2024) found the AUs *half-blink* (47), *lower lip depressor* (16) and *lips part* (25), combined with single ear movement, in horses with experimentally induced arthritis. Other studies have employed EquiFACS to compare facial expressions in different situations, *e.g*. Ricci-Bonot & Mills (2023) studied modelled frustration and disappointment in horses and found that *blinking* (AU145), *nostril lift* (AUH113), *tongue show* (AD19) and *chewing* (AD81) were present during disappointment. EquiFACS has also been employed to study micro-expressions in horses (Tomberg et al. 2023).

Table 3. Action Units and Action Descriptors used in the Facial Action Coding System for horses (EquiFACS; Wathan et al. 2015)

Action Unit	**Description**
EAD101	*Ears forward*
EAD102	*Ear adductor*
EAD103	*Ear flattener*
EAD104	*Ear rotator*
AU101	*Inner brow raiser*
AU143	*Eye closure*
AU145	*Blink*
AU47	*Half blink*
AU5	*Upper lid raiser*
AD1	*Eye white increase*
AU10	*Upper lip raiser*
AU12	*Lip corner puller*
AU113	*Sharp lip puller*
AUH13	*Nostril lift*
AU16	*Lower lip depressor*
AU17	*Chin raiser*
AU18	*Lip pucker*
AU122	*Upper lip curl*
AU24	*Lip presser*
AU25	*Lips part*
AU26	*Jaw drop*
AU27	*Mouth stretch*
AD160	*Lower lip relax*
AD19	*Tongue show*
AD29	*Jaw thrust*
AD30	*Jaws sideways*
AD133	*Blow*
AD38	*Nostril dilator*

2. Aims

Negative welfare states in horses, such as pain and stress, share many similarities but require different approaches to prevent. Because of this, how to detect and distinguish them is important. Use of sedatives in a clinical environment may also influence the expression and evaluation of pain. Since it is difficult to discriminate between the exact cause of changes in facial activity during clinical- or welfare evaluations, experimental methods were employed in this thesis with the aim of identifying facial expressions during the above mentioned states from a clinical- and welfare point-of-view. An additional aim was to provide an overview of methods for increasing the quality of such measurements.

Using EquiFACS, the overall aims in this thesis were to:

- Investigate how facial expressions change during short-term stress.
- Evaluate the effect of stress from other sources than nociception, on facial expressions when evaluating pain.
- Evaluate the effect of a non-analgesic, anxiolytic and sedative drug (acepromazine) on facial expressions associated with nociception.
- Investigate if the induced states of pain, stress, and sedation could be discriminated from each other using facial expressions.
- Explore how to analyse high dimensional FACS data from longer time intervals and propose a framework for data handling.

3. Hypotheses

The hypothesis for this thesis was that stress and sedation could produce measurable facial expressions and that these can impact the recognition of pain in horses. The following specific hypotheses were tested:

- Facial expressions when horses are experiencing stress from isolation and transportation, are present and measurable using EquiFACS.

- Facial expressions produced from a horse in pain are also present when the horse is exposed to stressors from a non-nociceptive origin.

- Facial expressions from states of nociception, stress and sedation can be used to differentiate between the states.

- Compound states of stress-pain or sedation-pain generate expressions similar to those of the non-compound states.

- Statistical methods that measure temporal overlap of Action Units shows better performance in selecting relevant Action Units for analysis than other commonly used methods.

- Statistical methods for dimensionality reduction show differences and choice of method affects the conclusions which can be drawn.

4. Methods

Results and conclusions presented in this thesis are based on data obtained in multiple experiments. The methods used in these experiments are summarised in this chapter. Data used in Papers I and II were obtained in experiments conducted and reported in those papers. Data used in Paper III were obtained in the experiments reported in Papers I and II and taken from studies by Ask et al. (2024) and Rashid et al. (2020). The experiments in Papers I and II were performed in accordance with a study protocol approved by the Ethics Committee for Animal Experiments in Uppsala, Sweden (permission number 5.8.18-10767/2019).

4.1 Study design

4.1.1 Transportation and social isolation (Paper I)

The effects of transportation and social isolation, two stressful management procedures, were scrutinised in an observational set-up and an experimental set-up (Paper I). In the observational set-up, horses were observed when being loaded into standard trailers designed for road transport. Facial expressions were recorded using video surveillance during the road transport and as baseline measurements before and after transport. The experimental set-up was conducted at the research stables at the Swedish University of Agricultural Sciences (SLU), where video material for analysing facial expressions was recorded in individual research boxes during a social isolation intervention that consisted of the horses being left alone in the stable without contact with a conspecific. Inferences on facial expressions were made in relation to a baseline recorded before the intervention (Paper I).

4.1.2 Nociception, isolation, sedation and compound interventions (Paper II)

In order to analyse the differences between discrete nociception, isolation, sedation interventions and compound interventions of these, a cross-over design was implemented in Paper II (Figure 1). The order of the interventions between horses was randomised and stratified to counteract the effect of order of interventions. Isolation and its compound interventions were carried out in the morning when the horses expected to be brought out, causing a disruption in the controllability of their situation. Nociception (ischemic mechanoreceptor activation), sedation (intravenous injection of Acepromazine) and compound nociception-sedation interventions were performed in the evening. All interventions were standardised in terms of duration and in terms of the people present, time of day and handling of the horses.

4.1.3 Methodology analysis (Paper III)

In Paper III, which mainly explored methodological approaches, data from Papers I and II were used together with other published EquiFACS data (Rashid et al. 2020; Ask et al. 2024). The data taken from Rashid et al. (2020) were from horses in which pain was experimentally induced by activation of mechanoreceptor, using ischemic pressure over the foreleg. The data taken from Ask et al. (2024) related to horses in experimentally induced pain consisting of lipopolysaccharide (LPS)-induced arthritis in the tarsocrural joint. EquiFACS data from that experiment were recorded during two different intensities, which were judged based on a composite pain scale and movement asymmetry measured using an optical motion capture system (Ask et al. 2024). From this study, annotations from the highest pain intensity were analysed.

4.2 Study populations

4.2.1 Paper I

Two study groups were formed, representing the origin of the horses used and the interventions in which they participated (due to ethical legislation). A population of privately owned (PRI) horses made up a group that was included only in the transportation intervention. A group of university owned

(UNI) horses underwent both the transportation intervention and the social isolation intervention. Horses kept at the university but subjected to the same procedures and experimental design as the PRI horses (*i.e.* no social isolation component) were included in the PRI group. Inclusion criteria for all horses were that they were considered healthy by their owner or caretaker at the time of the experiment and not subject to veterinary treatment which could be suspected to be still present at that time. For the PRI group, horses were managed at the discretion of their regular caretaker and baseline data were recorded in an area where the caretaker deemed the horse to be calm. For the UNI group, the horses were kept in their regular research stables and followed the routines to which they were accustomed. They were kept in indoor boxes (3 m x 4 m) at night and outside on pasture or in paddocks during daytime. These horses had previously been transported once a year to summer pasture, but were otherwise not accustomed to transportation, while horses in the PRI group had varying degrees of experience of transportation. During the social isolation trials, UNI horses were moved to boxes with video cameras at least 16 hours before experimentation. Horses were moved in pairs according to already existing social structure.

4.2.2 Paper II

The horses participating in Paper II were all of the same breed and were of similar colour, weight, and height. These horses were kept at the same university facilities as described in Paper I, following the same routines. For at least three days before the study, the horses were habituated to other boxes in the same facility that had remote controlled video-cameras installed. Horses were stabled in pairs in view of each other according to existing social structure. Each pair of horses had the same feeding and housing routine and was handled by their regular caretaker during the habituation period. Participating horses underwent clinical examinations by a veterinarian and were considered healthy before experimentation.

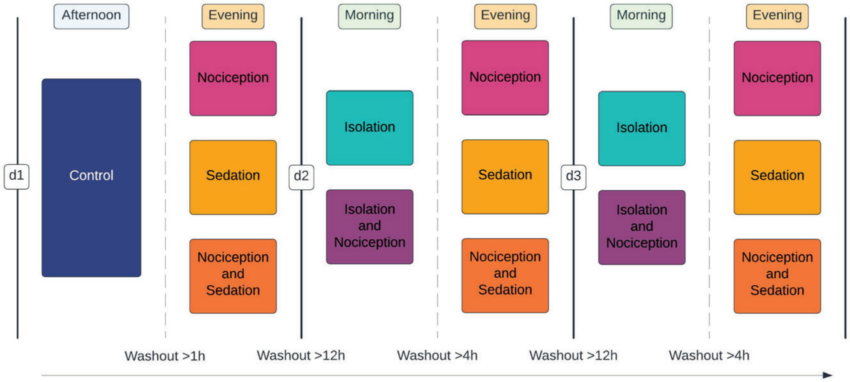


Figure 1. Schematic diagram of the cross-over design used in the experiment in Paper II.

4.3 Recording of facial expressions

Video for analysis of facial expressions were recorded using video surveillance equipment. For video recording in the research facility, eight video surveillance cameras (WDR EXIR Turret Network Camera, HIKVISION, Hangzhou, China) were mounted, in the corners of the two boxes that were used for experimentation. Nine standard fluorescent lights were mounted in the ceiling of the boxes to provide extra illumination for the videos. Video recording in the on-road transportation trailer was performed using commercial point-and-click action cameras (GoPro Hero 3+ Silver Edition, GoPro Hero 7 Black, Gopro Inc., San Mateo, California, USA). Both the surveillance cameras and the action cameras recorded at resolution 1080p and at 30 frames per second.

4.4 Video annotation

Video recordings from the interventions and baselines in the box were exported and processed through face-finder software (Rashid et al. 2018), which identified sections of the video where the face was present for analysis and calculated the probability of a face actually being present in the frame (Figure 2). For the isolation stress and baseline recordings in Paper I, 30 seconds with the highest probability of a face actually being present were cut out for analysis. The videos recorded during the on-road transportation were manually inspected and cut. Ten minutes after the twenty minute long intervention started in Paper II, the 30 seconds with the highest probability of a face being present were cut out for analysis. Relevant parts of the video were blurred in order to blind observers to the intervention they analysed. For Paper I, video material from one camera was extracted. For Paper II, video material from all four cameras was extracted and synchronised.

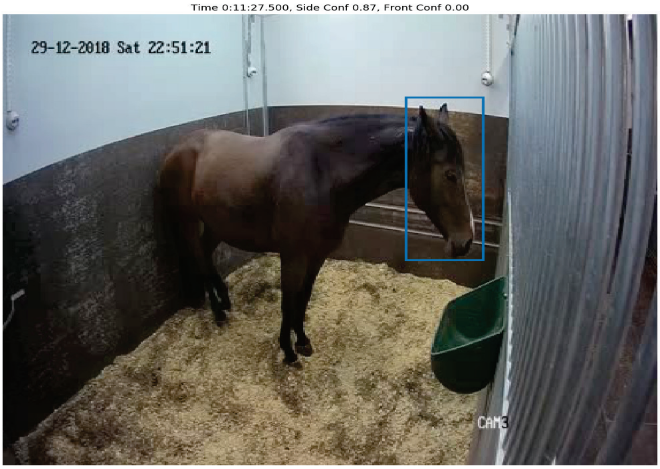


Figure 2. Illustration of the output face-finding software used in interventions and baselines in the box. In this case, the software was 87 % certain that the blue box contained a face of a horse.

The video recordings were exported to the open software ELAN, version 5.4 (Paper I) and version 6.0 (Paper II) (Lausberg & Sloetjes 2009). The software allowed for frame-by-frame or slow-motion annotation of EquiFACS codes over a timeline while watching the video segment as the observer was annotating the codes (Figure 3). Onset to offset was recorded for the codes on the timeline and exported for analysis. For Paper II, the annotators had access to all four camera angles and could use several of them to make their annotations (Figure 3).

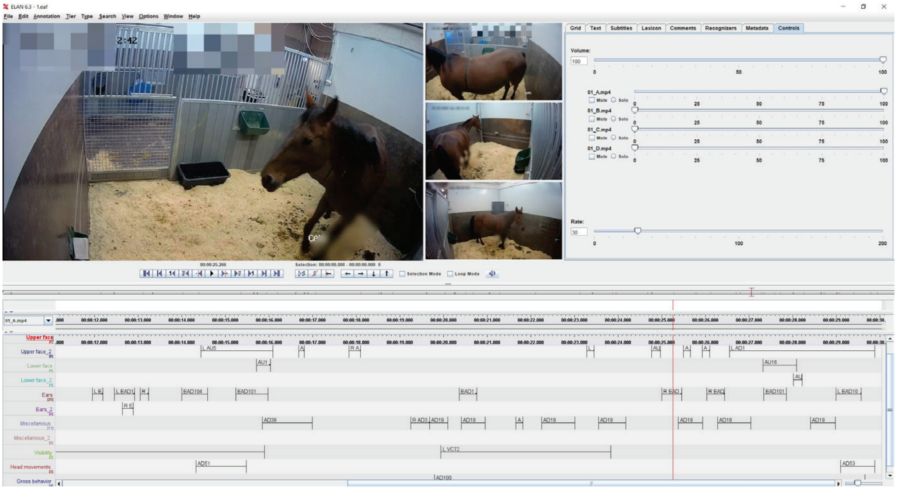


Figure 3. Illustration of the interface of ELAN while annotating video recordings from four angles.

4.5 Physiological measures

Heart rate was measured to confirm a physiological response in the social isolation and transportation interventions. R-R beat intervals were recorded using a body-mounted heart rate monitor for equines (Polar Wearlink, Polar Electro OY, Kempele, Finland) and exported through Polar ProTrainer Equine Edition (Polar Electro OY, Kempele, Finland) for outlier removal using the built in algorithm for moving avarage.

Serum cortisol was analysed during the social isolation interventions in Paper II. Blood samples were taken before and after the discrete and compound interventions. Since cortisol level shows diurnal variation (Zolovick et al. 1966), samples were taken at the same time of day before and after the interventions, and results were reported as delta (δ) at the same time between the two days. All blood samples were taken through venous puncture into clot-activating serum tubes. The samples were left for at least 30 minutes in a refrigerator for the clot to separate, after which they were centrifuged at 5000 RPM for 10 minutes. Plasma and serum were then aliquoted and frozen at -80 °C until analysis. Serum cortisol was analysed in two replicates, using a commercial immunoassay instrument (Immulite 2000XPi, Siemens Healthcare Diagnostics) with reagent Veterinary Cortisol, lot 127.

4.6 Dimensionality reduction

Since FACS extensively record all facial expressions, dimensionality reduction of the width of the dataset and selection of AUs for conventional statistical methods were required before inference was tested. In Paper I, selection of relevant AUs to analyse was carried out using the Human FACS Interpretation (HFI) method developed by Kunz et al. (2019) and the co-occurrence method first reported by Rashid et al. (2020). The HFI method selected FACS codes that were present in more than 5% of the observations (arbitrary threshold) and whose frequency increased between the stressful intervention and the baseline. Using directed graphs, the co-occurrence of selected AUs was determined based on temporal overlap within the video clips, presenting codes that occurred together more often in the stressful videos than in the baseline. The size of the window where codes could overlap was within the range 2-30 seconds, at $\alpha = 0.5$. Paired t-tests for mean values were used to test for significance of differences.

In Paper II, which sought to identify AUs that were discriminatory between several interventions (categorical), partial least squares discriminant analysis (PLS-DA) (Nguyen & Holmes 2019) was performed in SIMCA version 17 (Sartorius Stedim Data Analytics AB, Umeå, Sweden) using standard settings for that analysis. Interventions were set as predictor variables to supervise the algorithm to find variables with the greatest discriminatory power. Weights and scores were analysed using scatterplots. Standardised variable importance for projection (VIP) values were calculated for each AU by summation of PLS loading weights, corrected by the sums of squares. This calculation was done in SIMCA. Action Units with VIP values >0.8 were deemed important for further analysis.

Selection of AUs was also investigated in Paper III, where all of the above methods were applied on a merged dataset from multiple studies. In addition, methods for network analysis described by Mielke et al. (2022) were compared. Using the *NetFACS* package developed by Mielke et al. (2022) in R software (*R: The R Project for Statistical Computing*), Bayesian network analysis of single combinations of AUs were assessed to identify significant codes.

4.7 Accuracy tests

Accuracy, sensitivity and specificity were mainly investigated in Paper III and were calculated using standard contingency tables. The experimental inductions were deemed as true positive and negative conditions. AUs selected from the different dimensionality reduction methods were included as measures of a positive or negative test. Two thresholds were investigated (α), 0.25 and 0.5, representing the quotas of selected AUs present in the sample in order to count as a positive test.

Tests of accuracy were also conducted for data in Paper I, mainly from a machine learning point-of view using a machine learning classifier. A Linear Support Vector Machine was trained on the data with the purpose of classify the stressful interventions in comparison to baseline. Outcome parameters after test on one of the datapoints who were excluded from the training (called Leave-one-out) were calculated based on the selected AUs from both co-occurrence and HFI methods from Paper I.

4.8 Agreement analyses

Inter-rater agreement was used to measure how well annotators agreed with each other in Papers I and II, while intra-rater agreement was used to measure how consistent an annotator was with their own scoring in Paper II. Agreement in both cases was calculated using Wexler's method (Paper I), as described by Ekman et al. (2002), using the formula:

$$Agreement = \frac{Frequency\ of\ annotations\ agreed\ upon\ \times 2}{Annotations\ rater\ 1 + Annotations\ rater\ 2}$$

In Paper II, where agreement was calculated based on fewer observations, intra-class correlation (ICC) was used to compute agreement using the package *psych* in R (William Revelle 2023). Agreement wase tested as strict agreement compared with single raters using absolute values, *i.e.* ICC type 1 (Koo & Li 2016). In Paper III, both methods were compared to each other using the data from Paper I and Paper II.

4.9 Statistical analyses

Statistical testing was conducted in R (*R: The R Project for Statistical Computing*). In Paper I, descriptive statistics and paired t-tests on inference of the average frequency and duration for each AU were computed in base R.

In Paper II, generalised linear mixed models (GLMM) were constructed based on zero-inflated Poisson distributions of the frequency data. These models were created using the package *glmmTMB* (Brooks et al. 2017). To account for individual variation, horse was set as a random factor and, for the purpose of inference testing, interventions were set as a fixed factor. Estimated marginal means were calculated using the *emmeans* package (Lenth 2023), with p-value correction by multivariate t distribution.

For the physiological values, two-tailed paired t-tests were used to determine inference for the heart rate data in Papers I and II and two-tailed Wilcoxon signed rank tests were used for significance testing in analysis of serum cortisol in Paper II, using base R. The level of significance was set to $p<0.05$ in all statistical analyses. Based on the dimensionality reduction in Paper III, sensitivity and specificity were calculated using standard equations and contingency tables. A result was considered positive when presence of a certain quota of selected AUs was detected in a sample (30-second video recording).

5. Results

5.1 Facial expressions during interventions

In the studies on which this thesis is based (Papers I-III), facial activity exhibited distinct patterns across interventions. Of the interventions, facial activity was generally highest during isolation (mean of 27.1 annotations). The number of annotations during the interventions of nociception (mean of 24.4 annotations) was lower than at baseline (mean of 28.7 annotations). As expected, the lowest number of facial activity annotations was recorded during pharmacological sedation, with a mean of 20.2 annotations.

Stressful interventions induced a multitude of facial expressions, mainly during transport-induced physiological stress. In Table 4, facial expressions which increased during social isolation and transportation compared with baseline ($p<0.05$) are compared with those recorded during nociceptive pain in other studies using EquiFACS. Some similarities were detected between studies and interventions, mainly in terms of *ear movement* (EAD), *blink* (AU145) and *nostril dilator* (AD38), presence of which increased irrespective of which intervention tested. Some AUs, such as *chin raiser* (AU17) and *half-blink* (AU47), were only present during the nociception intervention, *i.e.* they were absent from the managerial inductions. AUs, *upper lid raiser* (AU5), *tongue show* (AD19) and *inner brow raiser* (AU101) was exclusively present during stressful interventions although they only appeared in the combined measurement of isolation and transportation, and not in the intervention isolation alone, in either Paper I or II. Introduction of sedation did not cause any significant changes in terms of AUs expressed compared to baseline (Paper II).

Table 4. Statistically significant Action Units which were selected by the co-occurrence method (alpha 0.5 and observation window size of 2 seconds; Paper I, Rashid et al. and Ask et al.) and by the PLS-DA (Paper II) which increased in frequency in different interventions compared with baseline.

Action Unit	**Isolation and transportation (I)**	**Isolation (II)**	**Nociception (II)**	**Pain (Rashid et.al)**	**Pain (Ask et al.)**
Inner brow raiser	✓ (p=0.042)				
Lips part	✓ (p<0.001)				
Tongue show	✓ (p<0.001)				
Nostril dilator	✓ (p<0.001)	✓ (p<0.001)		✓ (p<0.001)	
Ears	✓ (p<0.001)	✓ (p<0.001)		✓ (p<0.01)	✓ (p < 0.001)
Half-blink				✓ (p<0.01)	✓ (p < 0.001)
Blink	✓ (p<0.001)	✓ (p=0.001)	✓ (p=0.048)		✓ (p < 0.001)
Upper lid raiser	✓ (p<0.001)				
Chin raiser				✓ (p<0.001)	
Chewing		-	-	✓ (p<0.001)	

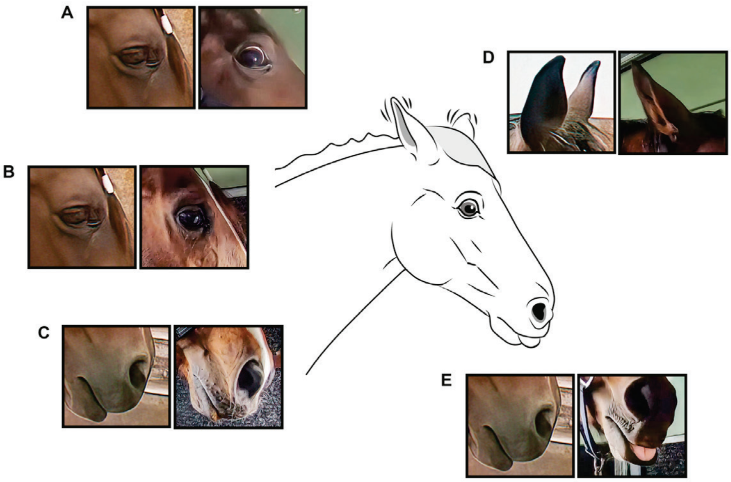


Figure 4. Illustration of the e Action Units in a stressed horse (right) compared to a baseline (left) (a) upper lid raiser, (b) inner brow raiser, (c) nostril dilator, (d) ear movements and (e) tongue show.

5.2 Facial expressions during compound states

The approach of measuring each intervention individually might be preferred in situations where the situation is well controlled, but during clinical and welfare evaluations, multiple and compound states may occur. In Paper II, the same subjects were exposed to different stimuli, as described in Chapter 4. A scatter plot of the weights in the PLS-DA analysis is shown in Figure 5, where EquiFACS codes (variables) are plotted together with indicator variables of intervention over the two components of the PLS-DA explaining the most variability. Only the first component (t1) was significant (Q^2>0.05). The precise location of variables in the weights scatter plot is less important than their location in relation to each other. Variables and indicator variables which are scattered in close proximity of each other are closely associated within the dataset. Proximity of a variable to an indicator variable (EquiFACS code close to an intervention) indicates an association between the FACS code and the intervention, while those far apart and located in diagonally different quadrants are inversely related. The results of statistical testing of important variables (VIP values above 0.8) are presented in Table 5.

Overall, the weights plot demonstrated that the frequency of AUs present during the experimentally induced interventions could be used to discriminate between nociception, isolation and sedation as separate experiences (Figure 5). This is indicated by the location of the indicator variables (coloured dots), which represent the sum of frequency of each AU within the intervention multiplied by the weight of each AU. The weights represent the impact that this specific AU had on the model (its discriminatory power), as further discussed in section 5.3, where the importance of the AU is used as a dimensionality reduction method. Nociception was closely associated with head movements up and down (AD53-54), together with the facial displays *inner brow raiser* (AU101), *sharp lip puller* (AU113) and *nostril lift* (AUH13). The similar code *nostril dilator* (AD38), which codes for a dilation both medially and laterally instead of only laterally, and ear movements (EAD101 and 104), *blinking* (AU145), *upper lid raiser* (AU5) and *eye white increase* (AD1) were associated with isolation stress. The projection on the weights scatter plot also indicated that the isolation intervention and sedation intervention were inversely related (due to their relative position in Figure 5) and that *chin raiser* (AU17) and *eye closure* (AU143) were closely associated with sedation.

The discriminatory power decreased for the compound interventions, mainly represented by the location of the indicator variables for the compound interventions nociception-isolation and nociception-sedation. While nociception-isolation could be discriminated from the control (baseline), it was close to the isolation intervention in terms of frequency of FACS codes, indicating that the same AUs which were discriminatory for isolation were also discriminatory for the compound intervention involving isolation. The opposite was seen for the nociception-sedation intervention, *i.e.* nociception and sedation could be discriminated from each other using FACS, but the compound intervention could not be discriminated from the control.

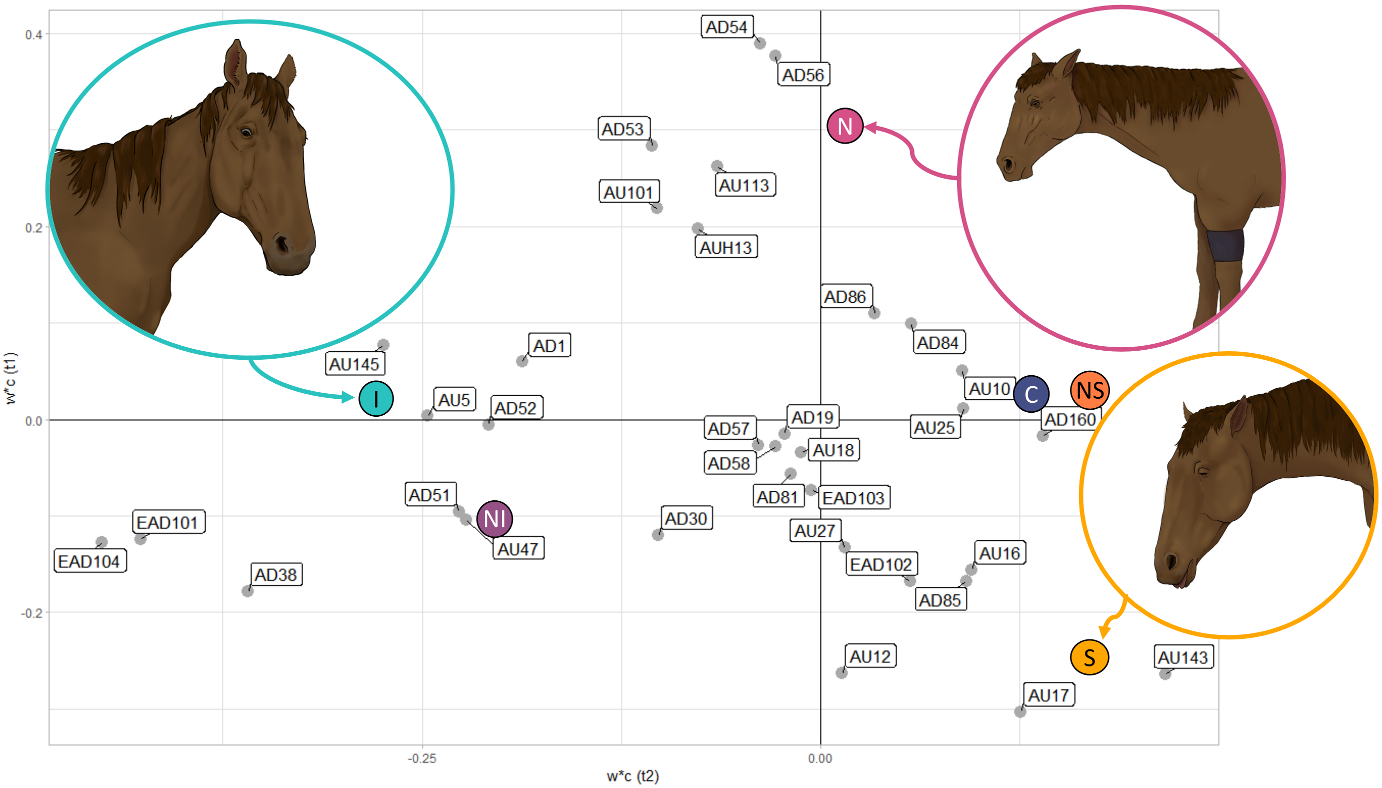


Figure 5. Weight scatter plot (w*c(t1), w*c(t2)) of EquiFACS codes (grey) and indicator variables of the interventions. (isolation (I), sedation (S), nociception (N), compound nociception and isolation (NI), compound nociception and sedation (NS), control (C)).

Table 5. Estimated ratios of specific EquiFACS codes between baseline (control) and different interventions calculated using the ZI-GLMM model, where a bold type a significant difference. SE = standard error.

AU/AD	**Contrast**	**Ratio**	**SE**	***p-value***
EAD104	Nociception	1.33	0.33	0.636
EAD104	Sedation	1.14	0.31	0.986
EAD104	Sedation and Nociception	0.54	0.17	0.165
EAD104	**Isolation**	**2.38**	**0.49**	**<0.001**
EAD104	**Isolation and Nociception**	**2.41**	**0.49**	**<0.001**
EAD101	Nociception	1.34	0.35	0.671
EAD101	Sedation	1.04	0.28	1.000
EAD101	Sedation and Nociception	0.65	0.19	0.423
EAD101	**Isolation**	**2.45**	**0.55**	**<0.001**
EAD101	**Isolation and Nociception**	**2.60**	**0.58**	**<0.001**
AD38	Nociception	3.64	3.04	0.251
AD38	Sedation	4.19	3.37	0.164
AD38	Sedation and Nociception	5.04	4.04	0.101
AD38	**Isolation**	**19.18**	**14.39**	**<0.001**
AD38	**Isolation and Nociception**	**20.47**	**15.30**	**<0.001**
AU145	**Nociception**	**1.60**	**0.30**	**0.048**
AU145	Sedation	1.05	0.23	0.999
AU145	**Sedation and Nociception**	**1.75**	**0.33**	**0.011**
AU145	**Isolation**	**1.92**	**0.35**	**0.001**
AU145	**Isolation and Nociception**	**2.36**	**0.41**	**<0.001**
AU101	Nociception	2.32	0.84	0.070
AU101	Sedation	1.45	0.62	0.800
AU101	Sedation and Nociception	0.66	0.36	0.865
AU101	Isolation	1.43	0.56	0.755
AU101	Isolation and Nociception	1.13	0.45	0.996

5.3 Methods for dimensionality reduction

Those Action Units deemed most important by the four different dimensionality reduction methods, were compared and summarized in Figure 6. All methods selected those AUs that showed the greatest difference compared with their respective baseline, except for PLS-DA (all), which selected the AUs most likely to discriminate between all interventions. Overall, PLS-DA selected the most AUs, while the HFI method consistently selected the fewest. Although there were differences in the array of AUs selected, many AUs occurred over multiple interventions and were selected by multiple methods, while only a few AUs were exclusive to a certain method or a certain intervention.

More specifically, *chin raiser* (AU17) and *lip pucker* (AU18) were selected across almost all interventions and methods, while there were some similarities within interventions for AUs such as *eye closure* (AU143), which was selected by several methods within sedation, but not within pain or stress. Some individual methods picked AUs exclusively for one intervention. *Upper lid raiser* (AU10) and *lips part* (AU25) were selected exclusively for stress by the PLS-DA method (intervention-specific) and *lip presser* (AU24) was selected exclusively for pain interventions by the co-occurrence and NetFACS methods. Some likenesses were also present between interventions for the same methods: *mouth stretch* (AU27) was selected across all interventions by NetFACS, *eye white increase* (AD1) by PLS-DA and *inner brow raiser* (AU101) by HFI.

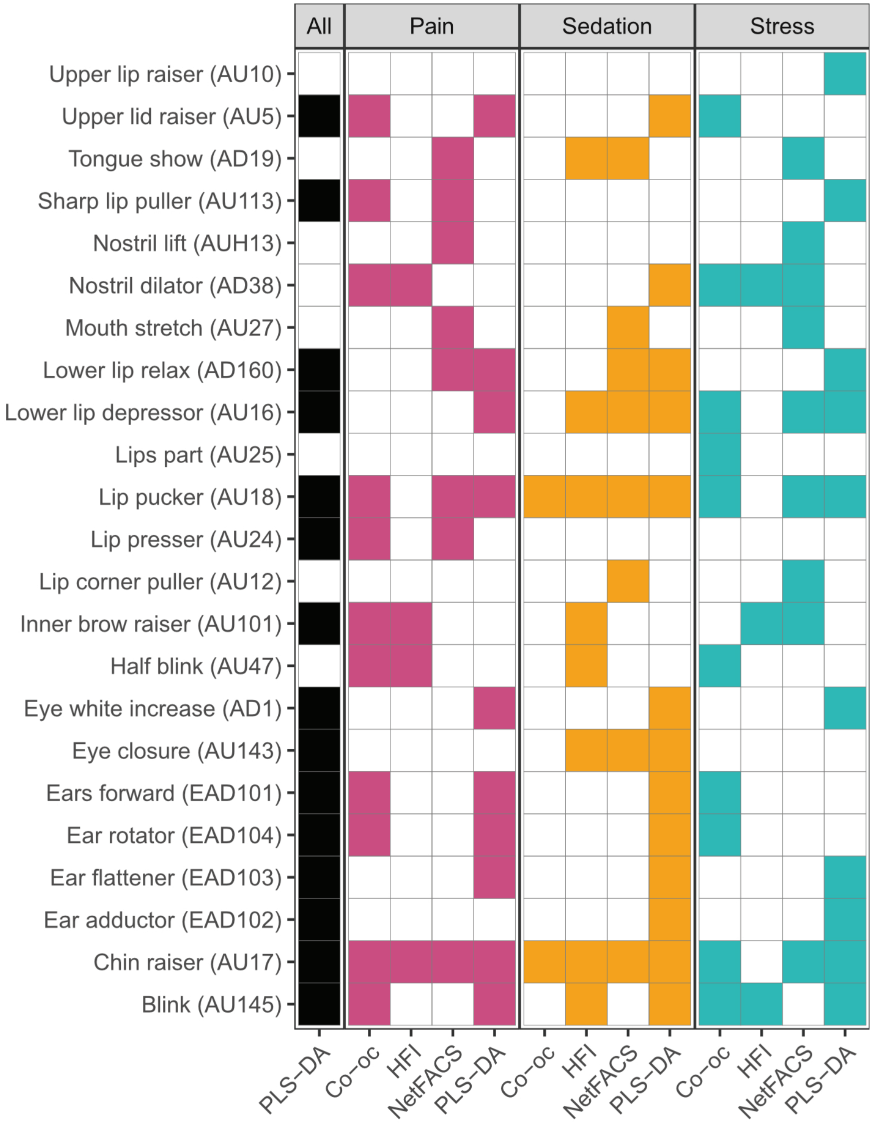


Figure 6. Action Units deemed most important for analysis by the different dimensionality reduction methods compared in this thesis (PLS-DA, co-occurrence, HFI, NetFACS). For the PLS-DA and additional column (black) represents Action Units selected for discrimination between all interventions.

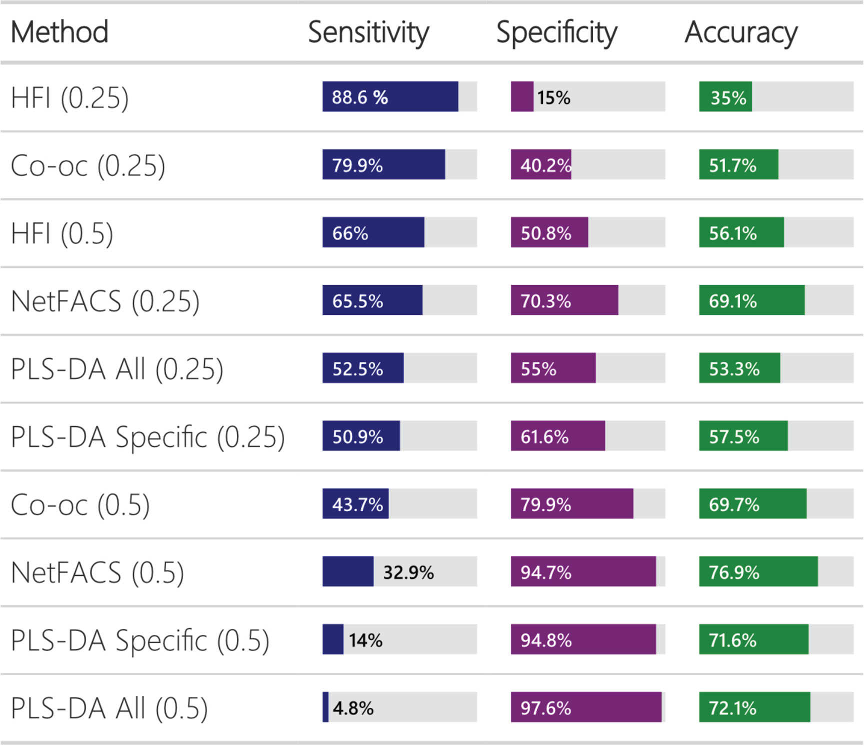
5.4 Method accuracy

Parameters of accuracy for the machine learning method in Paper I, representing accuracy in determining stress from baseline, are presented in Table 6. Performance was best when using frequency as predictor and adding duration to the classification performance improved the accuracy slightly. The best accuracy for stressful interventions was achieved using the HFI method for both duration and frequency combined. This method also had the highest sensitivity. The co-occurrence method did not perform better than performance when using the whole dataset.

Table 6. Results of leave-one-out classification analysis based on data from Paper I (HFI and co-occurrence method).

	Frequency	**Max duration**	**Both**
HFI method			
Positive Predictive Value (Precision)	75.56 %	66.67 %	75.00 %
Sensitivity (Recall)	89.47 %	68.42 %	94.74 %
Accuracy	77.27 %	62.12 %	78.79 %
Co-occurrence method			
Positive Predictive Value (Precision)	73.68 %	66.67 %	73.91 %
Sensitivity (Recall)	73.68 %	68.42 %	89.47 %
Accuracy	69.70 %	62.12 %	75.76 %
Without selection of AUs			
Positive Predictive Value (Precision)	71.79 %	62.86 %	76.74 %
Sensitivity (Recall)	73.68 %	57.89 %	86.84 %
Accuracy	68.18 %	56.06 %	77.27 %

In Figure 7, all interventions are considered. At the higher of the two thresholds tested (α=0.5), the HFI and co-occurrence methods had the highest sensitivity, while the PLS-DA method with α=0.5 threshold had the highest specificity (Figure 7, see also Paper III). The sensitivity seemed to be inversely related to the specificity and no method reached the empirical practical minimum of 1.5 for a clinical test, indicating that no method achieved remarkable performance. In general, sensitivity was higher at the lower threshold (α=0.25) and specificity was higher at the higher threshold (α=0.5) (Figure 7). The exception was the HFI method, which had higher sensitivity than other methods when using the lower threshold. Accuracy, a measure of how many times the method was correct, usually had values somewhere between those for sensitivity and specificity (Figure 7). It can been taken as an overall descriptor of the clinical performance of the methods.



Method	Sensitivity	Specificity	Accuracy
HFI (0.25)	88.6 %	15%	35%
Co-oc (0.25)	79.9%	40.2%	51.7%
HFI (0.5)	66%	50.8%	56.1%
NetFACS (0.25)	65.5%	70.3%	69.1%
PLS-DA All (0.25)	52.5%	55%	53.3%
PLS-DA Specific (0.25)	50.9%	61.6%	57.5%
Co-oc (0.5)	43.7%	79.9%	69.7%
NetFACS (0.5)	32.9%	94.7%	76.9%
PLS-DA Specific (0.5)	14%	94.8%	71.6%
PLS-DA All (0.5)	4.8%	97.6%	72.1%

Figure 7. Average sensitivity, specificity and accuracy of the PLS-DA, co-occurrence, HFI, NetFACS methods. Thresholds used for percentage of annotations present in the sample, which constituted a positive test, are shown in brackets.

5.5 Rater agreement

In general, rater agreement was good. In Paper I, inter-rater agreement between the three coders according to Wexler’s methods was on average 0.75 (rater 1-2: 0.76; rater 2-3: 0.76; rater 1-3: 0.71), where a value of 1 indicates perfect agreement between raters. In Paper II, the inter-rater agreement (ICC) was slightly better, with a confidence interval of 0.73-0.81 and an estimated value of 0.78 ($p<0.0001$). Consistency (ICC) within coders in Paper II was excellent, with estimated values above 0.90 ($p<0.0001$).

5.6 Heart rate and serum cortisol

In Paper I, heart rate increased significantly between baseline interventions and both the managerial interventions ($p<0.01$) with an increase of between 23-53 bpm compared to baseline. In Paper II, heart rate increased significantly ($p<0.05$) for all interventions except nociception. Serum cortisol, measured in delta from the same time period the previous day, increased for the isolation intervention and the compound nociception-isolation (isolation V=243, $p=0.007$; nociception-isolation V=224, $p=0.009$).

6. Discussion

Determination of facial expressions has been suggested as a welfare tool to detect negative welfare states (Descovich *et al*. 2017). One state that is important to detect in a welfare context is the presence of pain. Studies to date have examined facial expressions of horses in pain during clinical or experimental interventions (Dalla Costa et al. 2014; Gleerup et al. 2015; van Loon & Van Dierendonck 2015; Ask et al. 2024). In such interventions, the context of the situation is partly or fully known. There is control over the nociceptive input (time, intensity) during an experimental situation, but the presence of clinical (spontaneous) pain are usually only inferred from clinical deduction, such as examination before and after surgery or by analgesic testing. The experimental situation offers a satisfying level of ground truth, whereas clinical pain studies are less controllable regarding construct validity for the study (certainty that pain is in fact present in the horse). To an increasing degree, hospital-based pain scales and the use of facial expressions are being applied outside the clinical environment, *i.e.* in populations where the probability of pain is much lower. Examples include use of pain scales during evaluation of ridden horses during (stressful) events (Dyson & Ellis 2022). Bearing in mind that facial expressions may stem from a multitude of factors, including attention (Wathan & McComb 2014), voluntary contraction (Augustine et al. 2024), social communication (Waller et al. 2020) and other sources of affect (Holstege 1992), attention to these factors is important and they need to be addressed in much greater detail.

This thesis presents the first approach to describe facial expressions associated with stress and sedation in horses using the Facial Action Coding System for horses (EquiFACS). The results allowed some conclusions to be drawn about the effect of stress and sedation on facial expressions of pain. Below, a number of issues raised by the findings in this thesis are discussed.

6.1 Facial expressions of stress and pain

When horses were introduced to empirical stress situations, such as social isolation and transportation by road, changes in facial repertoire measurable against a neutral baseline were observed. This work was the first to demonstrate FACS-based facial muscle contractions during a stressful situation. Isolation and transportation were chosen as stress situations because descriptions in the literature indicated that these interventions induce both behavioural and physiological changes commonly linked with stress (Mal et al. 1991; Fazio & Ferlazzo 2003). Isolation and transportation are also managerial situations which are common within the equestrian community and which may be present in contexts where pain evaluations are performed. Both are also unpredictable (horses expected to follow each other outside in these papers) and uncontrollable (since horses cannot adjust during these management procedures and are not allowed to correct their situation due to fixation) in nature, factors which are considered necessary in order to meet the definition of stress (Koolhaas et al. 2011). Therefore, the observed increases in heart rate (Papers I and II) and serum cortisol (Paper II), were most likely caused by a stress response. While it is known that stress may induce an emotional response, the scope and study design did not allow for any inference about valence. Thus, as mentioned in Chapter 1, no conclusions regarding the emotional state of the horse were drawn in this thesis. However, it was possible to anticipate the presence of a physiological response in Paper I and a physiological response could be enough to produce changes in facial expressions, based on current findings in humans (Lerner et al. 2007). Paper I mainly sought to test the hypothesis that facial expressions can be quantified during stressful situations. Although Paper II had different aims, that study also measured facial expressions during social isolation, allowing presence of certain AUs during stress to be identified mainly *blinking* (AU145), *nostril dilator* (AD38) and *ear movements* (EAD101+104), *i.e.* fewer AUs than identified in Paper I. The differences in AUs selected by dimensionality reduction methods between Papers I and II, despite similarities in study design, were most likely due to differences in handling of the data, statistical methods and type of interventions studied (Paper I included transportation as well).

Since Paper I studied both transportation and social isolation, and Paper II only social isolation, it is possible that the reason for more facial expressions being observed in the first study was solely that transportation was a greater stressor and was associated with other factors affecting this intervention. The social isolation intervention took place in the same environment as the baseline (the horses' own boxes), thus limiting the number of visual and auditory inputs the horse experienced. Simple reaction to the change in environmental factors during transport thus produced a different number of AUs than the stressful experience alone. Furthermore, Papers I and II employed two different approaches for selection of relevant time sections to annotate (I: highest probability of the side of the horse's face being present, II: highest probability of the side of the horse's face being present and as close to 10 minutes into the intervention as possible) and different dimensionality reduction methods (I: Co-occurrence and HFI, II: PLS-DA), which could potentially have affected which AUs that were selected for analysis. The AUs that were similar between both studies were *blinking* (AU145), *ear movements* (EAD101+104) and *nostril dilator* (AD38), which have all been described previously as changes relevant to acute stress (Young et al. 2012), although not analysed using EquiFACS. Blinking frequency in itself has been proposed as a tool for measuring acute stress in the horse (Mott et al. 2020). *Upper lid raiser* (AU5), selected by the co-occurrence method in Paper I, indicates that the horse may be experiencing increased awareness. In humans, *upper lid raiser* is reported to be present in fearful or surprised individuals (Ekman et al. 1980). Due to the empirical evidence obtained, in conjunction with the confirmation of physiological parameters, the interventions in the Papers I and II may therefore be considered valid for induction of stress, and changes in facial expressions are likely a consequence of this stress.

Interestingly, most of the AUs selected during stress were also selected for horses in pain using the same statistical methods, except for *chin raiser* (AU17) (Rashid et al. 2020). This may indicate that these AUs are common in multiple situations and not specific for pain, or that pain induces a stress response resulting in facial expressions of both stress and pain. The main differences between facial expressions of stress and pain were in codes referring to the lower part of the face (muzzle and lips). In Paper I, the frequency of *lips part* (AU25)*, inner brow raiser* (AU101) and *tongue show* (AD19) increased in stressed horses. Rashid *et al*. (2020) did not find this in horses in pain, but instead found an increase in *chin raiser* (AU17). Both of these findings coincide with the current concept of facial expressions of pain, describing tense facial muscles around the muzzle, as both indicate less movement around the muzzle. However, the presence of *inner brow raiser* (AU101) in stress is somewhat inconsistent but may be a result of *this AU* being common in both painful and stress states or due to presence of stress in the pain inductions. Further, Paper II found less facial activity in pain compared with earlier studies overall (only blinking were statistically significant). The fact that more AUs were found by Rashid *et al.* (2020) in horses in pain, than in Paper II could be due to differences in study design. The horses in the study by Rashid *et al.* (2020) were based on data collected by Gleerup *et al.* (2015) of experimentally induced ischemic pain and clinical data from spontaneous pain. In both datasets, but particularly in the clinical data, there is a possibility that stress was present, resulting in facial expressions similar to those found in Paper I. Presence of *nostril dilator* (AD38) may have been due to purely physiological reasons, *i.e.* an increase in respiratory rate as part of preparation for a flight response in the horse, a flight animal (McGreevy 2004).

6.2 Discrimination between stress, pain and sedation

The above speculations about two unrelated studies are insufficient to draw conclusions regarding similarities between facial expressions of pain and stress while they may generate hypotheses. Differences in study design and study population may have affected the results, since the effects of stress on facial expressions are known to vary between subjects (Mayo & Heilig 2019). Therefore, the aim in Paper II was to compare these states within the same set of subjects, strictly controlling for individual variation and environmental factors by allowing the horses to act as their own control, enabling statistical modelling of individual differences. With the PLS-DA method, it was possible to use the frequency of AUs displayed during these interventions to discriminate between stress, sedation, pain and baseline by applying weights to the AUs based on partial least squares. More importantly, the AUs that were observed to be associated with stress, pain and sedation coincided with empirical data regarding descriptions of the facial expressions during these states. As discussed in section 6.1, blinking and ear movements were found during stress and the findings corroborated those in Paper I.

Sedation in horses is often described using the level of the head as an indicator of level of sedation (Kamerling et al. 1988). In this thesis, level of the head was found to be more indicative of painful states, which could be due to the way that the FACS data are presented and interpreted. More specifically, a sedated horse with the head constantly held down would have a code frequency of 1, while a horse moving the head up and down multiple times would have a higher frequency of the different codes. Facial expressions that have been reported in the equine face during sedation include closed eyes, drooping of ears and pronounced relaxation of the lips and muzzle (Oliveira et al. 2021). Since FACS is based mainly on activation and contraction of muscles, not all of these features could be observed in Papers I and II. *Eye closure* (AU143) coincides well with the description of closed eyes, whereas *lower lip depressor* (AU16) is somewhat inconsistent with relaxation of the lips, since it describes an active muscle contraction rather than relaxation. Instead, *lower lip relax* (AD160) would be expected during sedation. The reason for this AD not being more frequently detected for sedation could be due to the camera angle making it more difficult to see that particular part of the face or due to the fact that it is relaxed constantly, thus the frequency is low. Interestingly, *chin raiser* (AU17), which was found to be present in painful states in earlier studies (Rashid et al. 2020), was associated with sedation, probably due to the fact that *lower lip depressor* (AU16) and *chin raiser* (AU17) are often seen together due to their similar muscular basis (Wathan et al. 2015). The PLS-DA model was able to discriminate sedation from other states, but no single AU was statistically significant. As in the case of head position, this could be due to the way of recording activity in the FACS system, since frequency is favoured over static descriptions.

Since the PLS-DA was run on frequency data, it is not surprising that the AUs most commonly present in the dataset showed the most discriminatory power. However, only the component representing the horizontal weights in Figure 5 (component t1) was significant ($Q^2>0.05$), indicating that this had a large impact. Nevertheless, the method showed promise in discriminating between several states, allowing more insight into the context in which facial expressions can be analysed. This supports the notion that multiple facial expressions in combination need to be detected to discriminate between states and to differ stress from pain, since only *blinking* (AU145) was significantly different when evaluating the facial expressions one by one.

6.3 Compound states

While it is of great importance to be able to discriminate between discrete experimentally induced states, there is seldom a situation in the field or clinic where these situations can be carefully controlled. In reality, stress is affected by a multitude of factors, including the presence of pain. In clinic, horses are often isolated from conspecifics, have been transported and are subjected to a multitude of examinations that may be stressful in themselves (Watson & McDonnell 2018). Likewise, sedation has an effect on facial expressions and may thus influence the ability to correctly assess pain from facial expressions as shown in this thesis. The compound intervention isolation-nociception, which was intended to represent the reality of a horse experiencing low-degree pain while also stressed, could not be differentiated from stress, and thus the AUs recorded in this thesis could not determine whether pain was present in a stressed horse. While it can be assumed that stress-induced analgesia is present for low-degree pain, it could also be concluded that facial activity in itself is not specific enough to discriminate these two states from each other. Stressful interventions tended to increase the frequency of AUs in total and may thus have overshadowed facial activity during a low degree of pain. If one could be certain that there is no presence of stress when evaluating the horse, AUs could be discriminated from baseline and thus provide a surer ‘diagnosis’ of pain. Therefore, at the very least, stress needs to be acknowledged and accounted for in the assessment.

While both sedation and nociception could be discriminated from the compound intervention, nociception-sedation in combination could not be well discriminated from baseline (control). This indicates that pain in a sedated horse can be missed when only examining facial activity. In contrast to the high-arousal interventions, however, a single AU may instead be of importance. As shown in Table 5, *blinking* (AU145) was statistically significant as a simple yet distinctive variable to help detect pain in the sedated patient. However, this finding must be validated based on clinical data and in other types of pain, since it may be a response to the ischemic pain induction model used in this thesis. Overall, it proved to be more difficult to differentiate between compound states using PLS-DA, meaning that careful consideration of other states is needed when evaluating pain using facial expressions in field situations.

6.4 Current methods for analysing FACS

Dimensionality reduction methods and test accuracy of these were evaluated and compared in this thesis. The co-occurrence method showed promise, with high sensitivity for the lower threshold tested (α=0.25) and high specificity for the higher threshold (α=0.5), and outperformed methods which relied on presence and absence of AU in some instances (Figure 7). Network analysis (NetFACS) also seemed to perform rather well. Further, it appeared that the threshold selected had a large impact on the accuracy parameters. This makes sense, since the main output from all methods, including that of the temporal methods, was prototypical AUs selected for that type of intervention.

The view of prototypical facial expression for certain states mainly stems from work by Ekman et al. (1980) and means that methods developed to analyse data generally involve annotation of only a few AUs over a short period. However, the theory of a prototypical face is constantly under debate within the scientific community and there is emerging evidence of a more intricate structure of facial expressions. For instance, the temporal order of facial activity in primates has been shown to be of great significance (Parr et al. 2005), but is not included in any of the methods compared in this thesis. FACS is unique in recording this type of data, and methods considering such data could prove beneficial for the understanding of facial expressions. Thus, the concept of specifying an internal state based on a few specific parameters may not be the way forward (Kappas 2003). Although it has been shown that reduction and weighting of pain scale items can successfully increase the performance of models in comparing baseline and pain (Ask et al. 2020; Trindade et al. 2023), the results in this thesis showed that this could decrease the specificity when taking stress and sedation into account, at least when evaluating facial expressions of pain.

The results obtained in this thesis instead suggest that benefits can be gained by incorporating further measures, such as physiological measures or behaviours, to validate states detected using facial expressions. However, reduced and weighted methods may be applicable in well-defined screening situations. Setting detection thresholds for probable pain can alleviate the extensive workload of analysis by pre-screening candidates that need further evaluation. As an analogy, gross pain behaviour such as a horse scraping the ground may be a sign of pain, but in itself does not confirm a clinical diagnosis of pain. Instead, there is signal value in that behaviour to be evaluated further. Similar signal value may be present in the facial display of horses. Combinations or sole frequency of certain AUs may provide this and may be selected based on the premises in Figure 6. Some AUs may provide great signal value in contexts where specificity does not necessarily need to be perfect but where the sensitivity is crucial, for example in a clinical situation where presence of pain is suspected to be high.

6.5 Main limitations

The greatest limiting factor in Papers I-III was the low number of animals used. There is a high degree of individual variation in how horses response to pain, depending on their personality (Ijichi et al. 2014) or earlier experiences (Reicherts et al. 2016) and the same is likely true for other states. Even though this variation was controlled for in this thesis using mixed models (Paper II) and the horses were used as their own control (Papers I-III), a few animals may not represent the entire population of horses, which is important in interpretation of the results. Thus, while study population and study design allowed for good construct validity, the external validity of these studies is limited.

Inducing experimental situations allowed better control over conditions to which the horses were subjected, but not what they experienced. However, for the purpose of this study design, no other methods were realistically available if relevant conclusions were to be drawn. Induced experimental pain or stress may not represent the situation which the animal would experience from pathology or a stressed situation, which is more frightening. While essential steps were taken in order to control the interventions as much as possible, there is always the risk that the interventions were not executed in the way intended. For example, the nociception induction in Paper II was standardised in its execution, but pain tolerance and earlier experiences may have affected the outcome. Since the aim was a low degree of pain, despite the method being validated (Graven-Nielsen & Mense 2001) there could be a risk that the pain induced was too low for some individuals. The same discussion may apply to the stress interventions. Although there is both empirical and scientific evidence of physiological changes during these interventions, there is no way of knowing what the horses were experiencing. Proxies for stress were measured, in the form of physiological indicators, but these inductions also aimed towards a low degree of stress. Therefore when drawing conclusions it must be borne in mind that a higher degree of stress may induce different results, as partly shown when comparing transportation and social isolation. Furthermore, there is no guarantee that the horses did not experience some additional stressors during the interventions.

A methodological limitation with the work in this thesis was in selection of AUs, where many arbitrary thresholds were selected. For the HFI and co-occurrence methods, but also in tests of accuracy, thresholds were selected arbitrarily, which may have had an impact on the results. Another limiting factor was in handling of zero-distributed datasets. While the statistical methods applied were constructed to handle these distributions, there may be methodological factors influencing the distribution in themselves. In behavioural studies, excess of zeroes may stem from type II-errors in the annotation process. As discussed earlier, Torcivia & McDonnell (2021) found that pain behaviours manifest over a long period and that they fluctuate. The same might be true for facial expressions, and thus excess of zeroes may be the result of a too short observation window.

7. Conclusions

This thesis showed that stress and sedation produce facial expressions in horses similar to those produced during pain and that methodologies need to be improved in order to increase the accuracy and diagnostic value of EquiFACS. More specifically:

- Stressful situations can induce an increase in the Action Units *ear movements* ($p<0.001$), *blinking* ($p<0.001$), *eye white increase* ($p<0.001$), *nostril dilator* ($p<0.001$), *upper eyelid raiser* ($p<0.001$), *inner brow raiser* ($p=0.042$) and *tongue show* ($p<0.001$). However, some of these Action Units have also been observed in studies that utilized the same methods to assess facial expressions of horses experiencing pain. This suggests that there are similarities between facial expressions associated with stress and pain in horses.

- Actions Units from EquiFACS are discriminative for low-degree pain, sedation and stress by social isolation, in distinct experimental settings when applying weighting to variables (PLS-DA). It was not possible to employ the same method to determine whether a stressed horse experienced low-degree pain or to measure specific facial expressions of pain in a sedated horse. Therefore, there is a large probability that these states influence pain evaluations using facial expressions in horses. However, blinking frequency needs to be investigated as a stand-alone measurement to determine whether a sedated horse is experiencing pain.

- Dimensionality reduction, and thus selection of a few important AUs for analysis, may limit analysis of the complexity of facial expressions in different situations. Instead, methods may benefit to incorporate intensity, timing and order of facial expressions, which have been proven to be important in other studies. Longer observation periods could provide additional information. Methods based on presence or absence only may be valuable in recognizing a probable negative state using automated recognition of AUs, but have low accuracy in general.

Overall, this thesis demonstrated the importance of incorporating contextual information such as the presence of stressors and pharmacological effects of treatments with sedative (common in clinical situations) before drawing conclusions about pain evaluations using facial expressions.

8. Future considerations

This thesis revealed the complexity in evaluating the states of pain, stress and sedation using facial expressions. While the field of pain recognition in animals has gained significant ground in recent years, doubtless to the greater benefit of domestic animals, many questions still remain. Employing a standardised tool, such as EquiFACS, provides researchers with comparable results and helps build on multiple studies to draw conclusions. Translational medicine, drawing conclusions across species borders, may also benefit from the use of a standardised system. However, much remains to be discovered from a biological standpoint when using facial expressions and some technological challenges still exist.

Use of machine learning technology in this field may provide additional benefits. The suggestion in this thesis on studying facial expressions over a longer period of time would result in manual annotation or observation being simply too extensive a task to be handled manually. Automation of certain processes could produce the large datasets needed for further advances.

Technical solutions may also be of benefit in enabling biological findings. There is still a lack of knowledge regarding how the neurobiological processes underlying facial expressions operate and their meaning for the animal (questions still unanswered in humans), but technical solutions in combination with conventional behavioural methods could provide further insights. Such technical solutions could include 3D-scanning of faces, electromyography or high-resolution video recording, in combination with interventions with controllable timing, which could provide insights about temporal and intensity aspects of facial expressions.

References

Andersen, P.H., Broomé, S., Rashid, M., Lundblad, J., Ask, K., Li, Z., Hernlund, E., Rhodin, M. & Kjellström, H. (2021). Towards Machine Recognition of Facial Expressions of Pain in Horses. *Animals*, 11 (6), 1643.

Ask, K., Rhodin, M., Rashid-Engström, M., Hernlund, E. & Andersen, P.H. (2024). Changes in the equine facial repertoire during different orthopedic pain intensities. *Scientific Reports*, 14 (1), 129.

Ask, K., Rhodin, M., Tamminen, L.-M., Hernlund, E. & Haubro Andersen, P. (2020). Identification of Body Behaviors and Facial Expressions Associated with Induced Orthopedic Pain in Four Equine Pain Scales. *Animals*, 10 (11), 2155.

Augustine, G.J., Groh, J.M., Huettel, S.A., LaMantia, A.-S. & White, L.E. (2024). *Neuroscience*. 7th edition. New York, USA: Oxford University Press.

Ayala, I., Martos, N.F., Silvan, G., Gutierrez-Panizo, C., Clavel, J.G. & Illera, J.C. (2012). Cortisol, adrenocorticotropic hormone, serotonin, adrenaline and noradrenaline serum concentrations in relation to disease and stress in the horse. *Research in Veterinary Science*, 93 (1), 103–107.

Baker, C.M. & Wong, D.L. (1987). Q.U.E.S.T.: a process of pain assessment in children (continuing education credit). *Orthopedic Nursing*, 6 (1), 11–21.

Basbaum, A.I., Bautista, D.M., Scherrer, G. & Julius, D. (2009). Cellular and Molecular Mechanisms of Pain. *Cell*, 139 (2), 267–284.

Benedetti, F., Durando, J. & Vighetti, S. (2014). Nocebo and placebo modulation of hypobaric hypoxia headache involves the cyclooxygenase-prostaglandins pathway. *PAIN*, 155 (5), 921.

Boneh-Shitrit, T., Feighelstein, M., Bremhorst, A., Amir, S., Distelfeld, T., Dassa, Y., Yaroshetsky, S., Riemer, S., Shimshoni, I., Mills, D.S. & Zamansky, A. (2022). Explainable automated recognition of emotional states from canine facial expressions: the case of positive anticipation and frustration. *Scientific Reports*, 12 (1), 22611.

Bourin, M., Petit-Demoulière, B., Nic Dhonnchadha, B. & Hascöet, M. (2007). Animal models of anxiety in mice. *Fundamental & Clinical Pharmacology*, 21 (6), 567–574.

Brooks, M.E., Kristensen, K., Benthem, K.J. van, Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M. & Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9 (2), 378–400.

Buchanan, K.L. (2000). Stress and the evolution of condition-dependent signals. *Trends in Ecology & Evolution*, 15 (4), 156–160.

Butler, R.K. & Finn, D.P. (2009). Stress-induced analgesia. *Progress in Neurobiology*, 88 (3), 184–202.

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, 60 (2), 69–83.

Carey, M.P. & Fry, J.P. (1995). Evaluation of animal welfare by the self-expression of an anxiety state. *Laboratory Animals*, 29 (4), 370–379.

Castaneda, A.R., Lamberti, J., Sade, R.M., Williams, R.G. & Nadas, A.S. (1974). Open-heart surgery during the first three months of life. *The Journal of Thoracic and Cardiovascular Surgery*, 68 (5), 719–731.

Chambers, C.T. & Mogil, J.S. (2015). Ontogeny and phylogeny of facial expression of pain. *Pain*, 156 (5), 798–799.

Chayadi, E. & McConnell, B.L. (2019). Gaining insights on the influence of attention, anxiety, and anticipation on pain perception. *Journal of Pain Research*, 12, 851–864.

Coles, B., Birgitsdottir, L. & Haubro Andersen, P. (2018). Out of Sight but Not out of Clinician's Mind: Using Remote Video Surveillance to Disclose Concealed Pain Behavior in Hospitalized Horses. In: Proceedings of the International Association for the Study of Pain 17th World Congress, Boston, MA, USA. 15–18 September 2018. 471121.

Coles, B. A. (2016). No Pain, More Gain? Evaluating Pain Alleviation Post Equine Orthopedic Surgery Using Subjective and Objective Measurements. (Master's thesis 2016:41) Swedish University of Agricultural Sciences. Department of Clinical Sciences. Uppsala.

Constantinescu, G.M. (2018). *Illustrated Veterinary Anatomical Nomenclature*. 4th edition. Stuttgart, Germany: Georg Thieme Verlag.

Dalla Costa, E., Minero, M., Lebelt, D., Stucke, D., Canali, E. & Leach, M.C. (2014). Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. *PLoS ONE*, 9 (3), 1–10.

Dalla Costa, E., Bracci, D., Dai, F., Lebelt, D., Minero, M. (2017). Do Different Emotional States Affect the Horse Grimace Scale Score? A Pilot Study. *Journal of Equine Veterinary Science.* 54, 114-117.

Dantzer, R. (2001). Stress, emotions and health: where do we stand? *Social Science Information*, 40 (1), 61–78.

Descovich, K.A., Wathan, J., Leach, M.C., Buchanan-Smith, H.M., Flecknell, P., Farningham, D. & Vick, S.J. (2017). Facial expression: An under-utilized tool for the assessment of welfare in mammals. *Altex*, 34 (3), 409–429.

Driessen, B., Zarucco, L., Kalir, B. & Bertolotti, L. (2011). Contemporary use of acepromazine in the anaesthetic management of male horses and ponies: A retrospective study and opinion poll. *Equine Veterinary Journal*, 43 (1), 88–98.

Duchenne, G.-B. (1876). *Mécanisme de la physionomie humaine: où, Analyse électro-physiologique de l'expression des passions*. J.-B. Baillière.

Dyson, S. & Ellis, A.D. (2022). Application of a Ridden Horse Pain Ethogram to horses competing at 5-star three-day-events: Comparison with performance. *Equine Veterinary Education*, 34 (6), 306–315.

Ekman, P., Freisen, W.V. & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39 (6), 1125–1134.

Ekman, P., Friesen, W.V. & Hager, J.C. (2002). *Facial Action Coding System Investigator's Guide*. *Psychologist*. Research Nexus.

Fazio, E. & Ferlazzo, A. (2003). Evaluation of Stress During Transport. *Veterinary Research Communications*, 27 (1), 519–524.

Feighelstein, M., Shimshoni, I., Finka, L.R., Luna, S.P.L., Mills, D.S. & Zamansky, A. (2022). Automated recognition of pain in cats. *Scientific Reports*, 12 (1), 9575.

Fillingim, R.B. (2017). Individual differences in pain: understanding the mosaic that makes pain personal. *Pain*, 158 Suppl. 1, S11–S18.

Gaynor, J.S. & Muir, W.W. (2014). *Handbook of Veterinary Pain Management*. St. Louis, USA: Mosby Inc.

Gleerup, K.B., Forkman, B., Lindegaard, C. & Andersen, P.H. (2015). An equine pain face. *Veterinary Anaesthesia and Analgesia*, 42 (1), 103–114.

Gleerup, K.B. & Lindegaard, C. (2016). Recognition and quantification of pain in horses: A tutorial review. *Equine Veterinary Education*, 28 (1), 47–57.

de Grauw, J.C. & van Loon, J.P. a. M. (2016). Systematic pain assessment in horses. *Veterinary Journal (London, England: 1997)*, 209, 14–22.

Graven-Nielsen, T. & Mense, S. (2001). The Peripheral Apparatus of Muscle Pain: Evidence From Animal and Human Studies. *The Clinical Journal of Pain*, 17 (1), 2

Hardy, J.D., Wolff, H.G. & Goodell, H. (1947). Studies on pain: Discrimination of differences in intensity of a pain stimulus as a basis of a scale of pain intensity. *The Journal of Clinical Investigation*, 26 (6), 1152–1158.

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.

Hoffis, G.F., Murdick, P.W., Tharp, V.L. & Ault, K. (1970). Plasma concentrations of cortisol and corticosterone in the normal horse. *American Journal of Veterinary Research*, 31, 1379–1387

Hurcombe, S.D.A. (2011). Hypothalamic-Pituitary Gland Axis Function and Dysfunction in Horses. *Veterinary Clinics: Equine Practice*, 27 (1), 1–17.

Hörnicke, H., Engelhardt, W. v. & Ehrlein, H.J. (1977). Effect of exercise on systemic blood pressure and heart rate in horses. *Pflügers Archiv European Journal of Physiology*, 372 (1), 95–99.

Ijichi, C., Collins, L.M. & Elwood, R.W. (2014). Pain expression is linked to personality in horses. *Applied Animal Behaviour Science*, 152, 38–43.

Jennings, E.M., Okine, B.N., Roche, M. & Finn, D.P. (2014). Stress-induced hyperalgesia. *Progress in Neurobiology*, 121, 1–18.

Julius, D. & Basbaum, A.I. (2001). Molecular mechanisms of nociception. *Nature*, 413 (6852), 203–210.

Kamerling, S.G., Cravens, W.M.T. & Bagwell, C.A. (1988). Objective assessment of detomidine-induced analgesia and sedation in the horse. *European Journal of Pharmacology*, 151 (1), 1–8.

Kappas, A. (2003). What Facial Activity Can and Cannot Tell us About Emotions. I: Katsikitis, M. (red.) *The Human Face: Measurement and Meaning*. Springer US. 215–234.

Knych, H.K., Seminoff, K., McKemie, D.S. & Kass, P.H. (2018). Pharmacokinetics, pharmacodynamics, and metabolism of acepromazine following intravenous, oral, and sublingual administration to exercised Thoroughbred horses. *Journal of Veterinary Pharmacology and Therapeutics*, 41 (4), 522–535.

Koo, T.K. & Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15 (2), 155–163.

Koolhaas, J.M., Bartolomucci, A., Buwalda, B., de Boer, S.F., Flügge, G., Korte, S.M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M. & Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience and Biobehavioral Reviews*, 35 (5), 1291–1301.

Kunz, M., Meixner, D. & Lautenbacher, S. (2019). Facial muscle movements encoding pain - A systematic review. *Pain*, 160 (3), 535–549.

König, H.E. & Liebich, H.-G. (2020). *Veterinary anatomy of domestic mammals: textbook and colour atlas*. Seventh Edition. Stuttgart, Germany: Thieme.

Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond, T.E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., LaCroix-Fralish, M.L., Matsumiya, L., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wong, D., van den Maagdenberg, A.M.J.M., Ferrari, M.D., Craig, K.D. & Mogil, J.S. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods*, 7 (6), 447–449.

Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41 (3), 841–849.

Lazarus, R.S. (1991). *Emotion and Adaptation*. Oxford University Press.

Lenth, R.V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Lerner, J.S., Dahl, R.E., Hariri, A.R. & Taylor, S.E. (2007). Facial Expressions of Emotion Reveal Neuroendocrine and Cardiovascular Stress Responses. *Biological Psychiatry*, 61 (2), 253–260.

van Loon, J.P.A.M. & Van Dierendonck, M.C. (2015). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study. *Veterinary Journal*. 206(3), 356-364.

López-Sanromán, F.J., Cisneros, G.D., Arco, V.M. del, Llorente, S.I. & González, S.M. (2015). The use of low doses of acepromazine as an aid for lameness diagnosis in horses: An accelerometric evaluation. *Veterinary and Comparative Orthopaedics and Traumatology*, 28 (5), 312–317.

Mal, M.E., Friend, T.H., Lay, D.C., Vogelsang, S.G. & Jenkins, O.C. (1991). Behavioral responses of mares to short-term confinement and social isolation. *Applied Animal Behaviour Science*, 31 (1), 13–24.

Mayo, L.M. & Heilig, M. (2019). In the face of stress: Interpreting individual differences in stress-induced facial expressions. *Neurobiology of Stress*, 10, 100166.

McGreevy, P. (2004). *Equine behavior: A guide for veterinarians and equine scientist*. Philadelphia, USA: Saunders.

McLennan, K.M., Miller, A.L., Dalla Costa, E., Stucke, D., Corke, M.J., Broom, D.M. & Leach, M.C. (2019). Conceptual and methodological issues relating to pain assessment in mammals: The development and utilisation of pain facial expression scales. *Applied Animal Behaviour Science*, 217 (April), 1–15.

Mellor, D.J. & Beausoleil, N.J. (2015). Extending the ”Five Domains” model for animal welfare assessment to incorporate positive welfare states. *Animal Welfare*, 24 (3), 241–253.

Melzack, R. & Torgerson, W.S. (1971). On the Language of Pain. *Anesthesiology*, 34 (1), 50–59.

Mielke, A., Waller, B.M., Pérez, C., Rincon, A.V., Duboscq, J. & Micheletta, J. (2022). NetFACS: Using network science to understand facial communication systems. *Behavior Research Methods*, 54 (4), 1912–1927.

Mott, R.O., Hawthorne, S.J. & McBride, S.D. (2020). Blink rate as a measure of stress and attention in the domestic horse (Equus caballus). *Scientific Reports*, 10 (1), 1–8.

Muir, W.W. (2013). Stress and pain: Their relationship to health related quality of life (HRQL) for horses. *Equine Veterinary Journal*, 45 (6), 653–655.

Müri, R.M. (2016). Cortical control of facial expression. *Journal of Comparative Neurology*, 524 (8), 1578–1585.

Nguyen, L.H. & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology*, 15 (6), e1006907.

Oliveira, A.R. de, Gozalo-Marcilla, M., Ringer, S.K., Schauvliege, S., Fonseca, M.W., Trindade, P.H.E., Filho, J.N.P.P. & Luna, S.P.L. (2021). Development and validation of the facial scale (FaceSed) to evaluate sedation in horses. *PLOS ONE*, 16 (6), e0251909.

Parr, L.A., Cohen, M. & Waal, F. de (2005). Influence of Social Context on the Use of Blended and Graded Facial Displays in Chimpanzees. *International Journal of Primatology*, 26 (1), 73–103.

Pequito, M., Amory, H., Serteyn, D., Busoni, V., de Moffarts, B. & Sandersen, C. (2012). Comparison of the Sedative and Hemodynamic Effects of Acepromazine and Promethazine in the Standing Horse. *Journal of Equine Veterinary Science*, 32 (12), 799–804.

Poole, G.D. & Craig, K.D. (1992). Judgments of genuine, suppressed, and faked facial expressions of pain. *Journal of Personality and Social Psychology*, 63 (5), 797–805.

Pritchett, L.C., Ulibarri, C., Roberts, M.C., Schneider, R.K. & Sellon, D.C. (2003). Identification of potential physiological and behavioral indicators of postoperative pain in horses after exploratory celiotomy for colic. *Applied Animal Behaviour Science*, 80 (1), 31–43.

R: The R Project for Statistical Computing. https://www.r-project.org/

Rainville, P. (2002). Brain mechanisms of pain affect and pain modulation. *Current Opinion in Neurobiology*, 12 (2), 195–204.

Rainville, P., Carrier, B., Hofbauer, R.K., Bushnell, C.M. & Duncan, G.H. (1999). Dissociation of sensory and affective dimensions of pain using hypnotic modulation. *PAIN*, 82 (2), 159.

Raja, S.N., Carr, D.B., Cohen, M., Finnerup, N.B., Flor, H., Gibson, S., Keefe, F.J., Mogil, J.S., Ringkamp, M., Sluka, K.A., Song, X.-J., Stevens, B., Sullivan, M.D., Tutelman, P.R., Ushida, T. & Vader, K. (2020). The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *PAIN*, 161 (9), 1976.

Ransom, J.I. & Kaczensky, P. (2016). *Wild Equids: Ecology, Management, and Conservation*. USA: JHU Press.

Rashid, M., Broome, S., Andersen, P.H., Gleerup, K.B. & Lee, Y.J. (2018). What should I annotate? An automatic tool for finding video segments for EquiFACS annotation. A.J., S. (red.), Manchester, 2018. 6–8

Rashid, M., Silventoinen, A., Gleerup, K.B. & Andersen, P.H. (2020). Equine Facial Action Coding System for determination of pain-related facial responses in videos of horses. *PLOS ONE*, 15 (11), e0231608.

Reicherts, P., Gerdes, A.B.M., Pauli, P. & Wieser, M.J. (2016). Psychological Placebo and Nocebo Effects on Pain Rely on Expectation and Previous Experience. *The Journal of Pain*, 17 (2), 203–214.

Reid, K., Rogers, C.W., Gronqvist, G., Gee, E.K. & Bolwell, C.F. (2017). Anxiety and pain in horses measured by heart rate variability and behavior. *Journal of Veterinary Behavior: Clinical Applications and Research*, 22, 1–6.

Ricci-Bonot, C. & Mills, D.S. (2023). Recognising the facial expression of frustration in the horse during feeding period. *Applied Animal Behaviour Science*, 265, 105966.

Rietmann, T.R., Stauffacher, M., Bernasconi, P., Auer, J.A. & Weishaupt, M.A. (2004). The association between heart rate, heart rate variability, endocrine and behavioural pain measures in horses suffering from laminitis. *Journal of Veterinary Medicine Series A: Physiology Pathology Clinical Medicine*, 51 (5), 218–225.

Salomons, T.V., Johnstone, T., Backonja, M.-M., Shackman, A.J. & Davidson, R.J. (2007). Individual Differences in the Effects of Perceived Controllability on Pain Perception: Critical Role of the Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 19 (6), 993–1003.

Schneiders, F.I., Noble, G.K., Boston, R.C., Dunstan, A.J., Sillence, M.N. & McKinney, A.R. (2012). Acepromazine pharmacokinetics: A forensic perspective. *The Veterinary Journal*, 194 (1), 48–54.

Sellon, D.C. (2015). Chapter 2 - Pain Management in the Trauma Patient. I: Sprayberry, K.A. & Robinson, N.E. (red.) *Robinson's Current Therapy in Equine Medicine (Seventh Edition)*. W.B. Saunders. 6–10.

Shini, S. (2000). A review of diazepam and its use in the horse. *Journal of Equine Veterinary Science*, 20 (7), 443–449.

Smith, B.L., Jones, J.H., Hornof, W.J., Miles, J.A., Longworth, K.E. & Willits, N.H. (1996). Effects of road transport on indices of stress in horses. *Equine Veterinary Journal*, 28 (6), 446–454.

Smith, E.St.J. & Lewin, G.R. (2009). Nociceptors: a phylogenetic view. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 195 (12), 1089–1106.

Stevenson, D.E. (1963). The Evolution of Veterinary Anaesthesia. *British Veterinary Journal*, 119 (10), 477–483.

Straulino, E., Scarpazza, C. & Sartori, L. (2023). What is missing in the study of emotion expression? *Frontiers in Psychology*, 14:1158136

Tomberg, C., Petagna, M. & de Selliers de Moranville, L.-A. (2023). Horses (Equus caballus) facial micro-expressions: insight into discreet social information. *Scientific Reports*, 13 (1), 8625.

Torcivia, C. & McDonnell, S. (2020). In-Person Caretaker Visits Disrupt Ongoing Discomfort Behavior in Hospitalized Equine Orthopedic Surgical Patients. *Animals*, 10 (210).

Torcivia, C. & McDonnell, S. (2021). Equine discomfort ethogram. *Animals*, 11 (2), 1–19.

Trindade, P.H.E., Barreto da Rocha, P., Driessen, B., McDonnell, S.M., Hopster, K., Zarucco, L., Gozalo-Marcilla, M., Hopster-Iversen, C., Rocha, T.K.G. da, Taffarel, M.O., Alonso, B.B., Schauvliege, S., Mello, J.F.S.R. de & Luna, S.P.L. (2023). Predictive modeling toward refinement of behavior-based pain assessment in horses. *Applied Animal Behaviour Science*, 267, 106059.

Trosch, R.M., Sze, G., Brass, L.M. & Waxman, S.G. (1990). Emotional facial paresis with striatocapsular infarction. *Journal of the Neurological Sciences*, 98 (2), 195–201.

van Dierendonck, Machteld and Goodwin, Debbie (2005) Social contact in horses: implications for human-horse interactions. In, de Jonge, F.H. and van den Bos, R. (eds.) *The Human-Animal Relationship.* (Animals in Philosophy and Science, 4) Van Gorcum, pp. 65-81.

Visser, E.K., van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A. & Blokhuis, H.J. (2001). Quantifying aspects of young horses' temperament: Consistency of behavioural variables. *Applied Animal Behaviour Science*, 74 (4), 241–258.

Waller, B.M., Julle-Daniere, E. & Micheletta, J. (2020). Measuring the evolution of facial 'expression' using multi-species FACS. *Neuroscience and Biobehavioral Reviews*, 113 (February), 1–11.

Wang, M. & Saudino, K.J. (2011). Emotion Regulation and Stress. *Journal of Adult Development*, 18 (2), 95–103.

Wathan, J., Burrows, A.M., Waller, B.M. & McComb, K. (2015). EquiFACS: The equine facial action coding system. *PLoS ONE*, 10 (8), 1–35.

Wathan, J. & McComb, K. (2014). The eyes and ears are visual indicators of attention in domestic horses. *Current biology : CB*, 24 (15), R677–R679.

Watson, J.C. & McDonnell, S.M. (2018). Effects of three non-confrontational handling techniques on the behavior of horses during a simulated mildly aversive veterinary procedure. *Applied Animal Behaviour Science*, 203, 19–23.

William Revelle (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych

Wong, D. & Baker, C. (1988). Pain in children: comparison of assessment scales. *Pediatric Nursing*, (14(1)), 9–17

Young, T., Creighton, E., Smith, T. & Hosie, C. (2012). A novel scale of behavioural indicators of stress for use with domestic horses. *Applied Animal Behaviour Science*, 140 (1–2), 33–43.

Zolovick, A., Upson, D.W. & Eleftheriou, B.E. (1966). Diurnal Variation in Plasma Glukocorticosteroid Levels in the Horse (Equus Caballus). *Journal of Endocrinology*, 35 (3), 249–253.

Öhman, A. (2008). Fear and anxiety. In *Handbook of emotions*. New York, USA: Guilford Press, p. 709–729.

Popular science summary

It is important to recognise and prevent presence of pain or stress in animals. However, recognition is difficult, since pain contains an emotional experience in response to bodily harm, while stress is a protective system responding to perceived threats. Animals cannot communicate the experience of these states verbally, and instead behaviours or physiological indicators can be used. Some human patients, such as infants, also cannot report their own experience and instead evaluation tools based on facial expressions are used to determine the presence and level of pain in these patients. Because the brain and the nervous system is comparable in all mammals, similar tools utilising facial expressions have been developed for domestic animals, including horses. However, at a physiological level the response to stress is very similar to the response to pain, so the presence of stress during pain evaluations may affect the results. Sedating drugs are also often used on animals in the same situations where pain evaluation is important, for example in hospitals, and may affect the results.

This thesis evaluated whether stress and sedative drugs change facial expressions in horses and whether they interfere with the facial expressions shown during pain. An anatomical tool was applied to analyse video recordings of horse faces, which relies on human observation of muscle contractions in the video material and record all possible muscle contractions capable of producing facial expressions. Whether or not stress produces any changes in facial expressions in horses were investigated by recording the facial expressions during two stressful situations, road transportation and isolation from a companion horse. In a second study, facial expressions were measured during experimental situations where horses experienced pain and stress and effect from sedative drugs. These situations were analysed separately from each other, and facial expressions were recorded in terms of presence and frequency. Situations where the horse experienced pain in combination with stress, and pain in combination with sedation, were also analysed, to determine whether it is possible to detect pain in a stressed or sedated horse. In a third study, different methods for handling data obtained using the standardised system were evaluated and compared, and their accuracy was tested to determine how well they could detect pain, stress and sedation.

The results showed that certain facial expressions were consistently displayed during stressful situations. Raised eyelids, widened nostrils, raised inner eyebrow, exposing the tongue and increased blinking and movements of the ears were all features present in a stressed horse. This supported earlier findings for horses and showed some similarities to features of the human face in fear or surprise. Using a statistical method that combined several facial expressions, differences between separate interventions (pain, stress and sedation) were identified. This means that facial expressions can be used to describe these states if they occur alone. However, it was much more difficult to determine whether a horse was experiencing pain when it was also stressed, since the facial expressions during the stressful situation overshadowed the facial expressions indicating pain. Likewise, the facial expressions of a sedated horse in pain were very similar to those of a neutral horse, indicating a risk of pain in a sedated horse being undetected. The frequency of blinking however, increased during pain despite that the horse was sedated which needs to be investigated further.

Further analysis revealed some difficulties in the methods used to examine facial expressions in horses. Some specific methods were better at determining whether a horse was in pain or not, but tended to overestimate the number of horses in pain or stressed. Other methods risked missing horses in pain, but were specific in terms of negative results. Performance could be improved if the methods were extended to include the timing, order and intensity of facial expressions. The manual tool used showed excellent performance of different evaluators in terms of consistency and agreement.

In conclusion, when evaluating pain using facial expressions in horses, one should consider whether the horse is stressed or under the effect of sedative drugs, since this may affect the results. Inclusion of other measurements, such as behaviours or physiological factors, could improve the accuracy of the analysis. In their present form, the methods are best used to warn of negative welfare states.

Populärvetenskaplig sammanfattning

Från ett välfärdsperspektiv är det viktigt att känna igen och förebygga tillstånd såsom smärta eller stress hos djur. Att känna igen eller upptäcka dessa tillstånd är dock svårt. Smärta är en emotionell upplevelse som svar på kroppslig skada, medan stress är ett skyddssystem som svar på upplevda hot. Eftersom hästar inte kan kommunicera dessa tillstånd verbalt kan beteenden eller fysiologiska indikatorer användas istället. Hos människor finns det patienter som, likt djur, inte kan meddela sin upplevelse verbalt, exempelvis spädbarn. Hos dessa patienter har ansiktsuttryck utvärderats för att avgöra hur mycket smärta de lider av och verktyg som använder ansiktsuttryck har även skapats för djur. Dessa verktyg har fått betydelse för användning på hästkliniker och förslag om dess användning som verktyg för att bedöma djurs välfärd har lagts fram. Emellertid är stress på en fysiologisk nivå mycket likt smärta och om individen är stressad vid bedömning av smärta kan resultaten påverkas. Dessutom används lugnande medel ofta på hästar i situationer där korrekt utvärdering av smärta är viktig och kan därmed påverka resultaten.

Målet med denna avhandling var att utvärdera om stress och lugnande läkemedel producerar ansiktsuttryck hos hästar och om de i så fall påverkade uttryck av smärta. Ett verktyg för ansiktsuttryck användes, baserat på videoupptagning av hästars ansikten. Verktyget bygger på manuell detektion av muskelkontraktioner i videomaterialet och registrerade alla möjliga muskelkontraktioner som kan producera ansiktsuttryck på ett standardiserat sätt. Forskningsfrågan delades upp i mindre delar. Dels undersökte avhandlingen om stress över huvud taget kunde producera ansiktsuttryck genom att registrera ansiktsuttrycken under två situationer som vanligen orsakar stress, vägtransport och isolering från en artfrände. Ansiktsuttryck mättes även under experimentella situationer där hästar upplevde smärta, stress och påverkan av lugnande medel. Dessa situationer analyserades separat från varandra och ansiktsuttrycken registrerades. Dessutom observerades situationer där hästen upplevde smärta i kombination med stress, samt smärta hos en sederad häst. Detta för att avgöra om det var möjligt att upptäcka smärta hos en stressad eller sederad häst. Slutligen utvärderades metoder som hanterar data från det standardiserade systemet och dessa jämfördes med varandra. Noggrannheten för metoderna testades för att utvärdera hur väl de kunde upptäcka smärta, stress och sedering.

Resultaten visade att ansiktsuttryck kunde registreras under stressiga situationer. Höjda ögonlock, vidgade näsborrar, höjt inre ögonbryn, framträdande tunga, ökad blinkfrekvens och ögonrörelser var de ansiktsdrag som förekom hos en stressad häst. Dessa resultat visade likheter med tidigare fynd hos hästar och visade även likheter med drag i människans ansikte vid rädsla eller förvåning. Genom att använda en statistisk metod som kombinerade flera ansiktsuttryck kunde skillnader mellan interventioner av smärta, stress och sedering hittas, vilket innebär att ansiktsuttryck kan användas för att beskriva dessa tillstånd om de förekommer ensamma. Det var emellertid mycket svårare att se om en häst upplevde smärta när den också var stressad, eftersom ansiktsuttrycken under den stressiga situationen överskuggade de ansiktsuttryck som var närvarande vid smärta. På samma sätt finns det en risk att smärta hos en sederad häst missas eftersom ansiktsuttryck hos en sederad häst i smärta var mycket likt de hos en neutral häst. Denna avhandling belyser också vissa svårigheter med de metoder som används för att undersöka ansiktsuttryck hos hästar. Vissa metoder var bättre på att avgöra om en häst hade ont eller inte, men tenderade att överskatta antalet hästar som hade ont eller var stressade. Andra metoder riskerade att missa hästar med smärta, men var specifika när det gällde negativa resultat. Metodernas prestanda skulle kunna bli bättre om tidpunkt, ordning och intensitet av ansiktsuttryck tas med i bedömningen. Det manuella verktyget hade utmärkt prestanda när det gäller hur konsekventa utvärderarna var, och utvärderarna hade god överensstämmelse med varandra.

Sammanfattningsvis, när smärta utvärderas med hjälp av ansiktsuttryck hos hästar måste hänsyn tas till om hästen är stressad eller under påverkan av lugnande medel. Inkludering av andra parametrar såsom beteenden eller fysiologiska mått kan förbättra noggrannheten. Metoder bör utökas för att inkludera noggrannare analyser, men kan användas som de är för att fungera som ett varningssystem för välfärds- och kliniska bedömningar.

Acknowledgements

This PhD project was funded by Swedish Research Council **FORMAS** (grant number 2020-01840). My thanks go out to the faculty, department, co-workers, and current and previous head of department.

This PhD project was made possible by three exceptional researchers. I could not have wished for better supervisors. **Marie Rhodin**, thank you for being exceptionally good at what you do, for coming with great advice, for being kind and supportive and for making this thesis possible and letting me aboard your amazing group. **Pia Haubro Andersen**, thank you for guiding this research area with enthusiasm and knowledge, for always knowing what to do, and for making me grow as a researcher and person. **Elin Hernlund,** thank you for always making time for positive talks, for your excellent skills you share and for your wonderful supervision style.

Warm thanks to current and former members of Team Biomechanics (and Pain) at SLU, with **Emma Persson-Sjödin, Katrina Ask, Britt Coles, Camilla Frisk, Anna Leclercq, Ebba Zetterlund, Ellen Law, Ylva Mellbin, Maja Söderlind, Sara Weidenblad, Marie Hammarberg, Linda Wright, Zala Zgank, Anna Byström** and **Hedvig Kjellström.**

Thanks for their contributions as co-authors on the papers to **Maheen Rashid-Engström, Sara Hidén Rudander** and **Hanna Bjarnestig**.

Thanks to **Erika Roman, Claudia von Brömssen, Jesper Rydén** and **Mary McAfee** for exceptional contributions to this thesis.

Tack till **vänner, familj, mamma, pappa, syster**, som stöttat genom hela resan. Till djuren i mitt liv **Wheatley, Aslan, Amber, Eevee.**

Linn, tack för att du finns!

Lastly, if you are reading this page, you probably helped me through my PhD in some way, be it with a board game night (or weekend), a phone call, or professional help. I'm grateful for the support!

PLOS ONE

RESEARCH ARTICLE

Effect of transportation and social isolation on facial expressions of healthy horses

Johan Lundblad[1]*, **Maheen Rashid**[2], **Marie Rhodin**[1], **Pia Haubro Andersen**[1]

1 Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden, **2** Department of Computer Science, University of California, Davis, California, United States of America

* Johan.Lundblad@slu.se

Abstract

Horses have the ability to generate a remarkable repertoire of facial expressions, some of which have been linked to the affective component of pain. This study describes the facial expressions in healthy horses free of pain before and during transportation and social isolation, which are putatively stressful but ordinary management procedures. Transportation was performed in 28 horses by subjecting them to short-term road transport in a horse trailer. A subgroup (n = 10) of these horses was also subjected to short-term social isolation. During all procedures, a body-mounted, remote-controlled heart rate monitor provided continuous heart rate measurements. The horses' heads were video-recorded during the interventions. An exhaustive dataset was generated from the selected video clips of all possible facial action units and action descriptors, time of emergency, duration, and frequency according to the Equine Facial Action Coding System (EquiFACS). Heart rate increased during both interventions (p<0.01), confirming that they caused disruption in sympato-vagal balance. Using the current method for ascribing certain action units (AUs) to specific emotional states in humans and a novel data-driven co-occurrence method, the following facial traits were observed during both interventions: *eye white increase* (p<0.001), *nostril dilator* (p<0.001), *upper eyelid raiser* (p<0.001), *inner brow raiser* (p = 0.042), *tongue show* (p<0.001). Increases in 'ear flicker' (p<0.001) and blink frequency (p<0.001) were also seen. These facial actions were used to train a machine-learning classifier to discriminate between the high-arousal interventions and calm horses, which achieved at most 79% accuracy. Most facial features identified correspond well with previous findings on behaviors of stressed horses, for example flared nostrils, repetitive mouth behaviors, increased eye white, tongue show, and ear movements. Several features identified in this study of pain-free horses, such as dilated nostrils, eye white increase, and inner brow raiser, are used as indicators of pain in some face-based pain assessment tools. In order to increase performance parameters in pain assessment tools, the relations between facial expressions of stress and pain should be studied further.

OPEN ACCESS

Citation: Lundblad J, Rashid M, Rhodin M, Haubro Andersen P (2021) Effect of transportation and social isolation on facial expressions of healthy horses. PLoS ONE 16(6): e0241532. https://doi.org/10.1371/journal.pone.0241532

Editor: Elisabetta Palagi, Universita degli Studi di Pisa, ITALY

Received: October 16, 2020

Accepted: May 19, 2021

Published: June 4, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files.

Funding: This study was partly financed by the platform SLU Future One Health awarded to P.H.A and partly financed by the Swedish Research Council Formas, grant no 2020-01840 (http://www.formas.se/) awarded to P.H.A. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In horses, which are prey animals [1], a multitude of emotional and physical challenges may be present during both ordinary and extraordinary management situations. These situations may include competitions, transportation by road, separation from the herd, social isolation during transportation, introduction to a new environment, and confinement during veterinary diagnostic procedures and treatment. Most of these experiences are putatively stressful and have been shown to induce a physiological stress response [2–4].

Stress is defined as the animal's non-specific reaction to challenges that require the individual to cope with environmental conditions or psychological challenges [5]. The stress response is a result of the impact from either the environment (external stressors) or the horse itself (internal stressors). An affective component of internal stressors is associated with a stressful experience, generally characterized by a high level of arousal with negative valence [6], thought to be caused by uncontrollability or unpredictability of the animal's situation [7]. However, internal stressors such as pain may also be relevant. These stress responses are associated with a number of physiological and behavioral changes [8]. In mammals, the response involves activation of two systems, the sympathetic-adrenal medulla axis and the hypothalamic-pituitary-adrenal cortex axis [9]. It may manifest as elevated heart and respiratory rate, blood pressure, and temperature [8]. It may even induce some degree of analgesia [10] or hyperalgesia [11], at least experimentally.

However, many stress-related physiological changes are not specific to stress. Cortisol release shows a diurnal variation [12] and may be affected by pathologies or pain [13]. Heart rate and blood pressure may be elevated in response to purely high-arousal activities, such as exercise [14], or during experience of another affective state, such as pain [13]. This renders physiological markers suitable for measurements of stress in controlled settings, but not in the field, where discrimination between stress and other experiences is important in decision making for both clinical and welfare applications.

Bodily behavioral changes are associated with the fight-flight nature of the horse, while facial behaviors are thought to convey communication to conspecifics [15]. Facial activity can generate a wide array of different observable expressions [16], and has been suggested as a tool for assessment of welfare in mammals [17]. Horses have the ability to generate a remarkable repertoire of facial expressions, which can be described by 17 action units [18]. This is a smaller repertoire than that of humans [16], but larger than that of e.g., chimpanzees or dogs [19,20]. Interestingly, the facial expressions of pain are conserved across mammal species, including humans [21]. It is known that the affective component of pain is expressed by prototypical facial expressions [17]. Recently, it has been shown that horses can display facial changes which are specific to pain [22–24]. However, studies on facial expressions originating from other experiences in horses are very sparse, limiting the use of facial cues for pain assessment in horses since the specificity in relation to other common affective states, e.g., stress [25], is not known. In humans, facial expressions remain a valuable tool for assessing emotional states [16] and furthermore stress induces typical facial expressions [26]. Only a few studies of facial expressions during potentially stressful management situations have been performed in horses, most focusing on features around the eye [27] or blinking frequency [28,29]. In order to address other facial features during these interventions, a tool called the Equine Facial Action Coding System (EquiFACS) [18] can be used. EquiFACS records facial expressions by observing onset and offset of anatomically based action units (AUs) and action descriptors (ADs) over time. The method does not infer anything about the meaning of facial movements observed, leaving less space for subjective judgment. The resulting dataset contains spatio-temporal data on the occurrence of different AUs, time of onset, offset, and duration, and their temporal overlap with other active AUs.

These datasets tend to be large, even with relatively small sample sizes, and thus they are difficult to classify without manual interference. To determine AUs that are typical for pain in humans, methods based on frequencies of AUs have been proposed [30]. However, statistical methods for analyzing FACS data on animals are not yet well-developed. To address this problem, use of data-driven machine learning principles has been applied for the analysis of facial expressions [31]. Such methods have been proven feasible when analyzing other large and variable datasets in the biological sciences, for example behavioral studies [32]. In a recent study, EquiFACS data were used for determination of facial expressions of pain [33]. Using a machine learning method utilizing the temporal overlap of AUs in observation windows of different lengths, co-occurring facial expressions which discriminated between painful and non-painful horses were determined in a very small dataset in that study (N = 6) [33]. The results of this method largely agreed with those of the frequency-based method, but it was also able to identify less frequent, but distinct, AUs of relevance for pain [33]. To our knowledge, exploration of facial expressions originating from affective states using this or other methods has yet to be explored in horses, and it is not clear whether facial expressions of pain in horses can be affected by other cognitive states.

The aim of this study was therefore to describe facial expressions during two common horse management events which putatively induce a physiological stress response in healthy individuals. Based on clinical and ethological descriptions during similar events, we expected to identify facial action patterns, with the most prominent being changes in repetitive mouth behaviors, flared nostrils, flattened ears [34], and the ADs yawning and tongue show [35]. We also expected an increased number of AUs in response to visual or auditory inputs, displayed as increased frequencies of ear movement and eye blinks [28,29,36]. To our knowledge, a complete set of EquiFACS facial expressions in these situations has not been described previously.

We further hypothesized that the frequency methods applied in human research can identify important AUs and ADs in horses, but that methods using temporal distribution (co-occurrence of facial expressions) are important, especially when frequency and duration of distinct facial traits are low or environmental input is high. Finally, we explored whether facial expressions during the interventions can be classified using a Linear Support Vector Machine, to support the construct validity of the facial expressions selected by the two methods.

Materials and methods

Ethical statement

This study was approved by the Ethics Committee for Animal Experiments in Uppsala, Sweden (Approval no. 5.8.18-10767/2019). Owner consent for the use of privately owned horses was obtained before experimentation.

Study design

For this study, consisting of one observational part and one experimental part, two standard horse management practices were used: short-term transportation and short-term isolation. These interventions are generally considered to be linked to psychologically induced stress. Video footage was recorded during the events, and during the horses' normal living conditions before or after the intervention. A body-mounted, remote-controlled heart rate monitor provided continuous heart rate measurements in all three situations.

Study groups

A total of 28 horses were used in the study. A heterogeneous study group, consisting of 18 privately owned horses (PRI), was included. They comprised 10 geldings, seven mares and one stallion, of the breeds Thoroughbreds (n = 5), mixed-breed ponies (n = 4), Standardbred trotters (n = 3), and Swedish warmblood/riding breeds (n = 6), with body weight ranging between approximately 400 and 600 kg. The median age of horses in this group was 10 years (range 3–24 years). They were considered healthy by their caretakers, had not been subjected to veterinary treatment for the previous two months, and had not been treated with analgesics during that period. The horses were managed at home, by the horse owner, in the routines to which they were accustomed. Most of these horses had previously been introduced to transportation. All were kept in stables except for the thoroughbreds, which were kept in a free-range system. Three of the PRI horses were kept at the university but were treated as though they were privately owned.

A more homogeneous study group was included from the university herd (UNI), consisting of nine Standardbred trotters (seven mares and two geldings) and one warmblood mare. They were considered healthy at routine examinations during the previous four months, were of median age 12 years (range 8–19 years) and had roughly similar body weight. They were kept in an authorized research facility at the Swedish University of Agricultural Sciences. These horses were fed hay four times a day, and oats once a day according to a nutritional plan that supported normal condition. All horses were allowed out on pasture for 6 hours a day and otherwise kept in individual 3 m x 4 m boxes. The horses were transported once a year to summer pasture, but other than that not regularly accustomed to transportation. During the experimental part, horses were moved to other boxes in the same facility and acclimatized for at least 16 hours. Horses were moved together in pairs, stabled besides each other, and kept in their regular stable herd (together for at least the previous six months). Each pair of horses had the same feeding and housing routine and had the same caretakers in all stables.

Horses in the two study groups, PRI (N = 18) and UNI (N = 10), all underwent the transportation intervention. The PRI horses were studied in their own stable and were transported in their own trailer. The UNI horses were transported in a standard horse trailer, which was novel for the horses, for 20 minutes. All UNI horses showed reluctance to enter the trailer and some loading procedures took up to 30 minutes before the horses entered the transport. All horses from UNI (N = 10) were used to create a subgroup, which in addition was subjected to social isolation on a subsequent occasion. Social isolation was performed by taking out the herd mate, leaving the horse alone in the stable for at least 15 and at most 30 minutes. The horses were kept in the same box as during the control intervention, making the environmental factors the same.

Video-recording

Video-recordings of the horses were made during the two interventions and during baseline without the presence of an observer. During the transport intervention, video-recordings were made in the box and inside the horse trailer, using GoPro Hero 3+ Silver Edition and GoPro Hero 7 Black cameras (Gopro Inc., San Mateo, California, USA). Resolution was set to 1080p at 30 fps and videos were exported to mp4-format. The cameras were mounted depending on the layout of the box, so that the entire horse and its box could be seen in the footage. If the stable had no regular box, the horses were filmed in their grooming spot. In the trailer, the halter of the horse was tied to a front bar in a standard manner, and the camera was mounted in line with the horse's head height and angled approximately 45–60 degrees from the horse's medial plane. The cameras recorded for 10 to 20 minutes during transportation, and for at least 30 minutes during baseline.

During the experimental social isolation intervention and during the baseline for the UNI subgroup, the horses were filmed in their own boxes. These video-recordings were made using two wall-mounted standard surveillance cameras with night vision (WDR EXIR Turret Network Camera, HIKVISION, Hangzhou, China). Extra light was provided with nine standard fluorescent lights mounted in the ceiling, programmed to provide light during daytime hours. The cameras were mounted in each corner in the front of the box so only the horse and its box could be seen in the footage, in order to ensure blinding. Resolution was set to maximum and images were exported to mp4-format. The cameras recorded all baseline sessions for a minimum of 30 minutes and social isolation sessions for a minimum of 15 minutes.

Heart rate monitoring

A remotely controlled heart rate monitor (Polar Wearlink, Polar Electro OY, Kempele, Finland), made for equine use, was used to obtain continuous heart rate measurements without the interference of an observer. The Wearlink device was fastened using a girth, which was soaked in water before attachment. Heart rate measurements started well before the interventions and the horses were allowed to adjust to the transmitter for at least 10 minutes before filming began [37]. The heart rate monitor was time synchronized with the videos, using a gesture in the video when the transmitter was started or using the time-stamped files produced by the cameras and heart rate transmitter. Files containing R-R intervals were exported through Polar ProTrainer Equine Edition (Polar Electro OY, Kempele, Finland). Anomalies in the heart rate measurements were removed using the program's own algorithm with the medium filter and minimum protection zone of six beats per minute. Heart rate measurements were extracted as a mean during five minutes, with onset two minutes and 15 seconds before the 30 second annotation clip and offset two minutes and 15 seconds after the clip ended. A Wilcoxon signed rank test was used to calculate significance in the PRI group. In the UNI group, a Wilcoxon signed rank test was used to test for the specific rise in heart rate between the baseline and the respective intervention. In the latter, the p-values were corrected according to the Bonferroni-Holm method.

Video processing and annotation

The identity of the video-recordings of the transportation group could only be blinded for horse, and not for intervention, since the location in the trailer and its movements could not be hidden. Selection of clips was made by manual inspection and 30-second clips of suitable footage were cut from the videos. If the face was visible and scorable for more than 30 seconds, a random number generator was used for video selection.

The identity of the video-recordings from the experimental social isolation intervention was blinded in relation to horse and intervention before annotation. Selection of videos for the social isolation group was performed using an automated horse face detection software [38], where sequences were selected if the head position of the horse was visible and suited for annotation. Thirty-second sequences of video with a side- or front-view confidence of at least 60% were selected. If several selections were available, a random number generator was used to select one clip. The selected clips were manually inspected to ensure that the software had successfully identified a face. If not, a new clip was randomly selected.

All films were annotated in a blinded manner by two EquiFACS-certified state-approved veterinarians with a minimum of 70% correct annotations compared with expert raters. All transportation and baseline films were also annotated by one of the authors (JL), who is also certified in EquiFACS. Annotation was performed using a template consisting of all codes in EquiFACS, including supplemental codes and the visibility code VC74 (code for unscorable), but without head movements (AD51-AD55). Annotation was performed with the open-source program ELAN [39]. The annotators coded the onset and offset of the facial AUs, allowing calculation of frequency and duration, i.e., how frequently an AU or AD occurred and how long it remained active. The annotators set the onset of the AU to when the muscle started contraction and the offset to when it was fully back to neutral again. Inter-rater agreement between the coders was calculated using the Wexler ratio as described by Ekman et al. [16], using all 30-second clips. Inter-rater agreement was found to be on average 0.75 (coder 1–2: 0.76; coder 2–3: 0.76; coder 1–3: 0.71), indicating good agreement between raters.

Selection of EquiFACS codes

Since inter-rater agreement was good, one set of annotations was randomly selected and used for each video. For each selected AU or AD, frequency and duration were observed. It was anecdotally noted that the frequency of ear-related ADs had a high presence in the dataset. In order to determine whether these movements were due to the ear moving back and forth or the ears focusing on a certain point, a facial movement index (FMI) was created. To describe *ears forward* (EAD101) and *ear rotator* (EAD104) occurring together within a one-second interval the term "ear flicker" was used. The FMI was created prior to any hypothesis testing where the EquiFACS codes were selected. It is important to note that this is not an AD, but an index describing a series of specific facial movements (ADs) that occur in succession to constitute the "ear flicker".

EquiFACS codes and the "ear flicker" were analyzed using the method described by Kunz et al. [30], here called the Human FACS Investigation (HFI) method. Action units that accounted for more than 5% of total AU occurrences in stress videos were selected. From this subset, AUs detected at higher frequency in the intervention videos than in control videos were selected as the final set of intervention AUs. While the HFI AU selection method ensures that selected codes are frequent and distinct, they may have only a slightly stronger correlation with the experienced state and can exclude less frequent, but highly discriminative, AUs. Therefore, the relative temporal distribution of AUs was also considered. In order to do this, the method of Rashid et al. [33], here referred to as the Co-occurrence method, was used to calculate the co-occurrence of AUs. This method selected EquiFACS codes that occurred together with other EquiFACS codes more frequently in stress than in no-stress states. Since onset and offset of EquiFACS codes were recorded in ELAN, codes which appeared simultaneously or in close relation to each other could be further studied. EquiFACS codes that occurred within a predetermined period (observation window size, OWS) were recorded as co-occurring. Action units that exhibited the largest difference in co-occurrence patterns between intervention and control were selected. The method uses directed graphs to record and calculate differences in co-occurrence patterns. Furthermore, a paired t-test for mean values was used to test significance, with $p<0.05$ considered significant.

For both the HFI and Co-occurrence methods, occurrences of *ears forward* (EAD101) and *ear rotator* (EAD104), which were included in the "ear flicker" category, were not double-counted for EAD101 and EAD104 separately. As a result, occurrence counts of EAD101 and EAD104 did not occur within a one-second interval of one another.

Classification of facial expressions during the interventions

The EquiFACS codes selected by the HFI and Co-occurrence methods were used to train a machine learning classifier, Linear Support Vector Machine (LSVM), for intervention versus control classification. Twenty-five control videos and 35 intervention videos (10 from social isolation, 25 from transportation) were used. The frequency and duration features in the clips were used to represent each video sequence, in order to train the LSVM for the classification. This was done without the "ear flicker" adjustment, in order to have pure data. Using five-fold cross-validation, the optimum regularization parameter *C* and balanced class weights were selected. The Python Scikit-Learn library [40] and the Leave-One-Out (LOO) protocol were used to train and test the models, meaning that the features of all videos except one were used to train an LSVM, which then used the same features on the remaining video to determine whether it showed a stressful intervention. The LSVM predictions were collated across the entire dataset, and precision and recall were calculated. Precision was reported as the proportion of true positives in the total number of predictions and recall as the proportion of true positives which could be identified by the model. Overall accuracy, the number of correct predictions in the number of total predictions, was also calculated. The performance of the LSVM models indicated how well the selected EquiFACS codes captured the facial expressions during transportation and social isolation, thus acted as a type of construct validity to classify the interventions.

Results

Heart rate during interventions

The means of the five-minute heart rate periods during interventions are shown in Fig 1. For the PRI group, heart rate increased from a pooled mean of 54 bpm (SD 25.5) during baseline to 77 bpm (SD 32.3) during transportation (p = 0.008). For the UNI group, heart rate increased from 35 bpm (SD 4.4) to 65 bpm (SD 30.5) during social isolation (p = 0.008) and to 88 bpm (SD 31.5) during transportation (p = 0.004), meaning that all interventions caused a rise in heart rate compared with the control situation.

Selected annotations

A full set consisting of 7900 annotations was created from the films. The horses displayed higher frequencies of facial movements during the interventions than during baseline, with the horses having an average of 38 annotations during social isolation and 57 annotations during transportation, compared with 26 annotations during baseline.

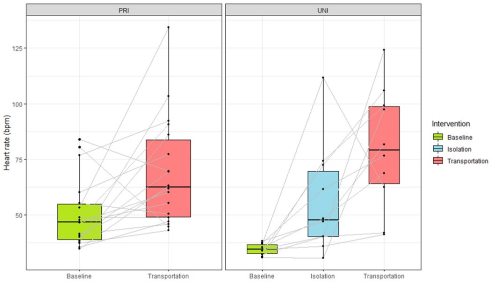


Fig 1. Heart rate during interventions. Boxplots showing the heart rate of (left) privately owned horses (PRI) and (right) university horses (UNI) during baseline, social isolation, and transportation.

https://doi.org/10.1371/journal.pone.0241532.g001

Table 1. Facial expressions during the interventions as defined by the HFI method.

	Eye white increase (AD1)	*Nostril dilator* (AD38)	*Inner brow raiser* (AU101)	*Blink* (AU145)	*Half blink* (AU47)	*Upper lid raiser* (AU5)	"Ear flicker"	*Ear rotator* (EAD104)
				Transportation				
Percentage of AUs during intervention / control	8.2% / 4.8%	13.1% / 8.4%	5.3% / 8.1%	12.7% / 19.6%	7.7% / 11.2%	8.0% / 5.7%	18.9 / 17.7%	Not selected
Difference in frequency	113.7%	106.2%	31.4%	30.1%	35.8%	98.6%	76.2%	Not selected
				Social isolation				
Percentage of AUs during intervention / control	7.8% / 3.9%	15.0% / 7.0%	15.0% / 12.1%	18.1% / 19.8%	Not selected	Not selected	16.6% / 20.2%	5.3% / 6.2%
Difference in frequency	85.7%	90.9%	43.0%	12.8%	Not selected	Not selected	1.9%	6.1%
				Combined				
Percentage of AUs during intervention / control	8.2% / 4.8%	13.4% / 8.4%	7.2% / 8.1%	13.8% / 19.6%	8.0% / 11.2%	7.0% / 5.7%	18.4% /17.7%	Not selected
Difference in frequency	106.8%	101.8%	52.1%	29.3%	30.5%	80.7%	66.4%	Not selected

Action units (AUs) and action descriptors (ADs) selected using the Human FACS Investigation (HFI) method to represent stressful interventions in horses in the transportation and social isolation groups and together as a combined group.

https://doi.org/10.1371/journal.pone.0241532.t001

HFI method. Action units which comprised at least 5% of AUs recorded during the interventions, and the percentage difference between interventions and control, are presented in Table 1. The results for the transportation and social isolation interventions are presented both separately and combined, to display differences and similarities between the groups. Generally, similar codes were selected in both groups with the exception of *Upper lid raiser* (AU5) and *Half blink* (AU47), which were not selected during social isolation, and *Ear rotator* (EAD104), which was selected only during social isolation. Blink AUs (AU145 and AU47) and *inner brow raiser* (AU101) had the most similar rate of occurrence between intervention and control, while *eye white increase* (AD1), *nostril dilator* (AD38) exhibited the largest difference in frequency between intervention and control recordings. However, an AU not selected during social isolation, *upper lid raising* (AU5), exhibited the largest difference in frequency between transportation and control recordings. The movement index "ear flicker" was also more frequent and more pronounced in transportation than in social isolation.

When combining both groups, all AUs that comprised at least 5% of stress AU occurrences were also more frequent during intervention videos than control videos. Other than the exceptions mentioned earlier, the chosen AUs for social isolation were identical to those selected for transportation stress, but the percentage difference between control and intervention frequency counts was noticeably larger for *inner brow raiser* (AU101).

Co-occurrence method. Action units and ADs selected using the Co-occurrence method are presented in Table 2. Of the selected codes, *nostril dilator* (AD38), *tongue show* (AD19), *mouth open* (AU25), *upper lid raiser* (AU5), *eye white show* (AD1), and "ear flicker" showed significance in all OWS. *Inner brow raiser* (AU101) was selected by the HFI method and significant (up to a 5-second OWS) using this method.

Frequency and duration patterns

In order to study each facial expression in detail, average frequency and maximum duration patterns for the above selected ADs are further presented in Figs 2 and 3, respectively. Action unit frequency increased for eight of the selected codes, mainly during transportation. With just 10 horses in the group, AU frequency was rarely significant for isolation stress. Only *Nostril dilator* (AD38) increased in frequency during social isolation. *Inner brow raiser* (AU101),

Table 2. Facial expressions during the interventions (combined) as defined by the Co-occurrence method.

OWS	*Inner brow raiser* (AU101)	*Lips part* (AU25)	*Tongue show* (AD19)	*Nostril Dilator* (AD38)	"Ear flicker"	*Blink* (AU145)	*Eye white increase* (AD1)	*Nostril lift* (AUH13)	*Upper lid raiser* (AU5)	*Half blink* (AU47)	*Ears forward* (EAD101)	*Ear rotator* (EAD104)
2	✓	✓	✓	✓	✓	✓	✓	✓	✓			
	(p = 0.042)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p = 0.064)	(p<0.001)			
5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	(p = 0.024)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p = 1.000)	(p<0.001)	(p = 0.107)		
10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	(p = 0.051)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p = 0.013)	(p<0.001)	(p = 0.450)	(p<0.001)	(p = 0.185)		
15	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	(p = 0.052)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p = 0.037)	(p<0.001)	(p = 0.576)	(p = 0.001)	(p = 0.373)	(p = 0.173)	
20	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	(p = 0.090)	(p = 0.001)	(p = 0.001)	(p<0.001)	(p<0.001)	(p = 0.053)	(p<0.001)	(p = 0.383)	(p = 0.003)	(p = 0.238)	(p = 0.217)	
30	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	(p = 0.179)	(p = 0.018)	(p = 0.017)	(p<0.001)	(p = 0.001)	(p = 0.079)	(p<0.001)	(p = 0.450)	(p = 0.018)	(p = 0.210)	(p = 0.252)	(p = 0.641)

Action units (AUs) and action descriptors (ADs) selected using the Co-occurrence method to represent the interventions in horses using different observation window sizes (OWS).

https://doi.org/10.1371/journal.pone.0241532.t002

despite its high frequency in the social isolation intervention, was not statistically significant. All AUs selected by the HFI method had $p<0.01$ for at least one representation and intervention. Additionally, *tongue show* (AD19) and *lips part* (AU25), which were only selected by the Co-occurrence method, showed $p<0.01$ across all groups and representations, for either frequency or maximum duration, when tested separately.

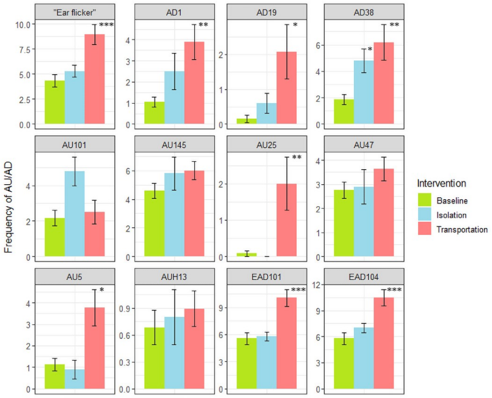


Fig 2. Frequency of EquiFACS codes. Changes in action unit (AU) and action descriptor (AD) frequency patterns between interventions and control. Asterisk marks significant difference from control (*p<0.05, **p<0.01; ***p<0.001).

https://doi.org/10.1371/journal.pone.0241532.g002

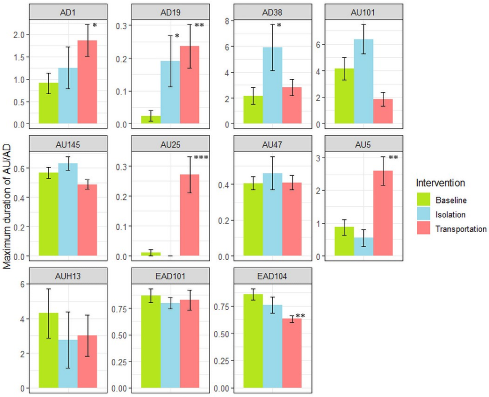


Fig 3. Maximum duration of EquiFACS codes. The interventions affected the duration (s) of activity for an action unit (AU). Asterisk marks significant difference from control (*p<0.05, **p<0.01; ***p<0.001).

https://doi.org/10.1371/journal.pone.0241532.g003

Leave-One-Out classification

The selected AUs in Tables 1 and 2 were used to train the LSVM for stress or no-stress classification, in order to check the validity of the selected AUs. The results of the LOO classification are presented in Table 3, which shows both the precision (proportion of positive predictions that were correct), and recall (proportion of positive instances that were correctly predicted). Accuracy (proportion of test instances correctly predicted, both positive and negative), is also shown. The best classification was obtained using both frequency and maximum duration, reaching an impressive 95% recall rate for the AUs selected by the HFI method and 78% precision rate for the AUs selected by the Co-occurrence method. Interestingly, the model was able to classify between the interventions and control almost as accurately even without selection of important AUs, although the precision and recall were better when focusing on a few AUs.

Table 3. Results of Leave-One-Out classification for action units with and without pre-selection.

	Frequency	Max duration	Both
HFI method			
Precision	75.56%	66.67%	75.00%
Recall	89.47%	68.42%	94.74%
Accuracy	77.27%	62.12%	78.79%
Co-occurrence method			
Precision	73.68%	66.67%	73.91%
Recall	73.68%	68.42%	89.47%
Accuracy	69.70%	62.12%	75.76%
Without selection of AUs			
Precision	71.79%	62.86%	76.74%
Recall	73.68%	57.89%	86.84%
Accuracy	68.18%	56.06%	77.27%

https://doi.org/10.1371/journal.pone.0241532.t003

Discussion

This study investigated whether common putatively stressful management procedures can induce facial expressions in horses, and whether these expressions can be recorded and identified using an objective facial coding system that exhaustively codes all facial activity, and not only predetermined actions. Transportation and social isolation were selected as interventions, since both are well described in the literature and practice, increasing the relevance of the study for horse management [41,42]. However, the literature mainly concentrates on the physiological or clinical component of the stress response or welfare issues during transportation and isolation [42], while the emotional characteristics are less well described. According to the dimensional approach, the dimensions of arousal and valence should be considered, but this is a particular difficulty in non-verbal species [43,44]. Empirical evidence of negative valence of both procedures in the majority of horses is very high, i.e., most horses avoid entering a trailer unless encouraged to, and horses have evolved to live in social groups and continue to avoid isolation from conspecifics [41]. The dimension of arousal is empirically less obvious, as some horses seem calmer than others and this may depend on many factors, including earlier life experiences and temperament [45].

Physical characteristics, such as age and sex, have been shown to have little effect on the physiological stress response in some instances [45] and even horses accustomed to transportation can show physiological changes characteristic of HPA activation [46]. However, sex and age could have a large impact on results when analyzing facial expressions. For example, the previous experience of the stressor, a factor strongly influenced by age, seems to decrease the response to some extent [45]. Irrespective of the horse's previous experience, assessment of these interventions as emotionally stressful remains subjective. A rise in heart rate was observed during both the social isolation and transportation interventions in this study, which might indicate an increase in arousal/alertness or at least some form of physiological response, e.g. due to physical activity. It is also important to note that the increase is still within the limit of vagal variation within the horse, which makes conclusions strenuous. In the PRI group, earlier experience of transportation differed and some of the horses were even accustomed to travel by road on a weekly or monthly basis. Whether or not these horses experienced positive or negative valence to the transportation intervention is not known. This was one of the reasons for including the UNI subgroup, where all horses were unaccustomed to travel by road transport and all horses showed avoidance behaviors when being loaded. However, the results in this study should not be interpreted as a true measurement of the horses' emotions, but rather as proof of changes in facial expressions due to high-arousal interventions. It is possible that other emotional states, such as excitement or fear, could be the source of the change in facial expressions and rise in heart rate observed during the interventions, meaning that the true emotional experience cannot be determined from the results in this study. The fact that horses react individually to transportation and social isolation is clearly illustrated in the results, with some horses showing little to no physiological changes during some interventions. The variance was generally higher in the PRI group, which probably reflects the heterogeneity of this population, both physically and mentally.

Despite the large variation in our experimental horses, significant changes in facial activities were recorded after both transportation and social isolation (Fig 4). According to the HFI method, there was increased frequency of the AUs *upper lid raiser* (AU5) and *inner brow raiser* (AU101), as well as *blink* (AU145) and "ear flicker". The frequency of the ADs *nostril dilator* (AD38) and *eye white increase* (AD1), not describing certain muscle-induced movements but rather the effects of two or more muscle movements, was also significantly increased.

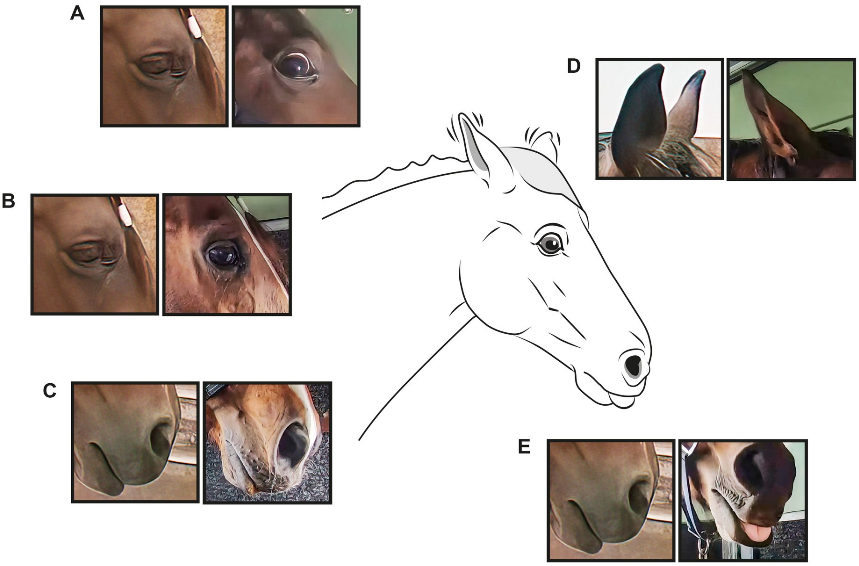


Fig 4. Illustration of facial expressions during the interventions. Action units (AU)/action descriptors (AD) relevant for the interventions. A: *Upper lid raiser* (AU5) and *Eye white increase* (AD1). B: *Inner brow raiser* (AU101). C: *Nostril dilator* (AD38). D: "Ear flicker"/ *Ear rotator* (EAD104). E: *Tongue show* (AD19). Action codes (right) are compared with a "neutral" horse (left). Illustration by Anders Rådén/ARDI.

https://doi.org/10.1371/journal.pone.0241532.g004

According to the Co-occurrence method, *tongue show* (AD19) and *mouth open* (AU25) were also important.

With this in mind, it is also possible that a number of external physical inputs inevitably associated with transportation, e.g., exposure to new environment, wind, confined space, ventilation or movement restriction, had some effect on the results presented in this study. In order to address this, the more homogeneous UNI subgroup underwent the social isolation intervention. Social isolation was associated with the same external inputs as the control, because the horses stayed in the same environment during the intervention. Changes in the facial expressions in this group should therefore be due to the experience of the horse being left alone, and not environmental factors. During social isolation, *Upper lid raiser* (AD5) and *Half blink* (AU47) were not selected. An explanation for this could be the environmental factors (wind, sounds, smells, movements) experienced during transportation. A distinct increase in the movement index "ear flicker" was apparent during both transportation and social isolation stress. Ear movements are very communicative [36], but during transportation ear movements due to sound might be a more likely cause of the high "ear flicker" frequency. During social isolation, a likely cause of ear movements is increased awareness of the surroundings due to arousal.

A reason for *upper lid raiser* (AU5) being more prominently seen during transportation stress, but not selected when analyzing isolation stress, could be that tension in *m. levator palpebrae superioris* (proposed basis for AU5) would hide tension in *m. levator anguli occuli medialis* (proposed basis for AU101) due to environmental factors but needs to be studied further.

The frequency of *blink* (AU145) increased during both transportation and social isolation. An earlier study also reported an increase in blinks during stressful situations [47]. However, Merkies et al. [28] found that full blink diminished during stress. In the present study, the increase was only statistically significant for the Co-occurrence method during transportation stress. This may be a result of the greater number of horses in the transportation group. Differences in frequency of full blinks were not significant between baseline and the interventions (Fig 2). The only AU selected as indicative of a stressful intervention for the lower face was *lips part* (AU25). Concurrently, increased frequency of *tongue show* (AD19) was noted. This coincides well with earlier findings on behaviors of the tongue and repetitive mouth and licking behaviors during stress [34,48]. *Tongue show* (AD19) may be interpreted as a coping mechanism in horses subjected to stress, which is supported by the fact that oral stereotypies are often reported as a long-term consequence of inability to perform natural behavior in horses (e.g., cribbing).

When comparing the HFI method with the Co-occurrence method for two-second OWS, these two codes related to mouth movements were the only added codes. Since HFI is a frequency-based method, less frequent AUs such as *tongue show* (AD19) are not picked up using the HFI method, but were still sufficiently distinct to differentiate between stress and no-stress states. The logical interpretation of this pattern is that *tongue show* (AD19) and *lips part* (AU25) are sufficiently distinct to discriminate between stress and neutral states, but absence of the codes cannot exclude stress. This indicates the importance of the Co-occurrence method for selecting distinct and useful EquiFACS codes.

When comparing the facial expressions recorded in this study to facial activities previously described during stress in horses, similarities and differences were detected. Flared nostrils, repetitive mouth behaviors, increased eye white, and an increase in eye movements are features previously described during stressful interventions [23,24,30], However, increased activity of the *inner brow raiser* (AU101) is associated with pain in some pain assessment tools [22–24], and was not expected to be displayed during these management procedures where pain was not present. It is therefore relevant to discuss the possible presence of other states during the interventions. We recruited horses that were perceived as healthy and free from pain, and horses were used as their own control, so the risk of presence of pain in the majority of horses can be considered low, although not completely eliminated, since there is no 'gold standard' for evaluating pain.

The specificity of facial expressions across emotional states is of interest for their use as an emotional indicator [44]. To our knowledge, facial expressions during pain are the only experience to be analyzed to date using EquiFACS. Since pain is an internal stressor, while stress is not painful, comparison of facial expressions of pain and stress is needed. Rashid *et al.* [33] found that *nostril dilator* (AD38) and *chin raiser* (AU17) were indicative of pain when using both the HFI and Co-occurrence methods. The fact that *nostril dilator* (AD38) is also present during stressful management conditions could indicate that this AD is common during simple management interventions and less significant for determining pain. During both stress and pain, respiratory rate of the horse tends to increase, which may be a reason for *nostril dilator* (AD38) being common during both interventions.

As mentioned above, face-based pain scoring tools include facial expressions that were also present during the pain-free management interventions in this study. For example, the horse grimace scale [22] includes *ear flattener* (EAD103) and *ear rotator* (EAD104) as elements of the pain scale, while the FAP scale [24] uses eye white increase as an element. The "equine pain face" shows the features "tension of the lower face, rotated ears, dilated nostril and tension above the eye" [23]. All but "tension of the lower face" was seen in the pain-free stressed horses in this study. When discussing both physiological and behavioral aspects of pain assessment, stress is often described as a complicating factor [25]. This, together with our results, suggests a need for caution when using facial expressions for assessments during potentially stressful situations, since simple management procedures could induce similar facial expressions. Further, facial expressions during different emotional states should be studied in more controlled experiments in order to increase the validity of pain evaluations.

Despite great variation in the study group, the overall impressive recall and precision rates of the LOO classification indicate that the AUs/ADs selected by both the HFI method and the Co-occurrence method are indeed different from the baseline, and can successfully differentiate a high arousal state in horses from a control state. Interestingly, training the LOO classification on the videos with all AUs included generated almost as good results as only including relevant EquiFACS-codes selected by the HFI or Co-occurrence methods. This is probably because high-arousal states produced more facial activity, and therefore higher frequencies of EquiFACS codes, than the resting horse in its regular environment. However, on comparing two high-arousal states with many AUs, where specific features are more important to differentiate between states, the results would probably differ significantly more. Further studies should focus on using this method for comparing horses during high-arousal states during pain. Since these states both produce many AUs, this method could show promise in differentiating different affective states in the horse.

Conclusions

It proved possible to induce and objectively record the presence of facial expressions in healthy horses under field conditions, using simple equipment and ordinary management practices. Applying two different frequency and duration-based methods revealed that two types of common management procedures (social isolation and transportation) induced increased frequencies of several facial movements. *Eye white increase* (AD1), *nostril dilator* (AD38), *inner brow raiser* (AU101), *upper lid raiser* (AU5), *tongue show* (AD19) and the facial movement index "ear flicker" were recorded when the horses underwent transportation and social isolation. These results partly corroborate earlier findings on behavioral aspects during stress. However, some of the facial activities (dilation of the nostril, contraction of m. *occulus levator angulii*) observed during pain-free transportation and social isolation are also commonly used in face-based pain assessment tools.

Supporting information

S1 Dataset. Data used for analysis.
(XLSX)

S1 File. HRM files (compressed R-R intervals for heart rate analysis).
(ZIP)

S1 Video. Sample clip of a baseline video.
(MP4)

S2 Video. Sample clip of an isolation video.
(MP4)

S3 Video. Sample clip of a transportation video.
(MP4)

Acknowledgments

DVM Camilla Frisk and DVM Alina Silventoinen are thanked for annotating the films. Horses and owners who participated in this study are warmly thanked for their contributions.

Author Contributions

Conceptualization: Johan Lundblad, Marie Rhodin, Pia Haubro Andersen.

Data curation: Johan Lundblad.

Formal analysis: Johan Lundblad, Maheen Rashid.

Funding acquisition: Marie Rhodin, Pia Haubro Andersen.

Investigation: Johan Lundblad.

Methodology: Johan Lundblad, Maheen Rashid, Pia Haubro Andersen.

Project administration: Johan Lundblad, Marie Rhodin, Pia Haubro Andersen.

Resources: Johan Lundblad, Pia Haubro Andersen.

Software: Maheen Rashid.

Supervision: Marie Rhodin, Pia Haubro Andersen.

Validation: Johan Lundblad, Maheen Rashid.

Visualization: Johan Lundblad, Maheen Rashid.

Writing – original draft: Johan Lundblad, Pia Haubro Andersen.

Writing – review & editing: Johan Lundblad, Maheen Rashid, Marie Rhodin, Pia Haubro Andersen.

References

1. McGreevy P. Equine behavior: A guide for veterinarians and equine scientist. 2nd ed. Philadelphia: Saunders; 2004.
2. Budzyńska M. Stress reactivity and coping in horse adaptation to environment. J Equine Vet Sci. 2014; 34: 935–941. https://doi.org/10.1016/j.jevs.2014.05.010
3. Schmidt A, Hödl S, Möstl E, Aurich J, Müller J, Aurich C. Cortisol release, heart rate, and heart rate variability in transport-naive horses during repeated road transport. Domest Anim Endocrinol. 2010; 39: 205–213. https://doi.org/10.1016/j.domaniend.2010.06.002 PMID: 20688463
4. Mal ME, Friend TH, Lay DC, Vogelsang SG, Jenkins OC. Behavioral responses of mares to short-term confinement and social isolation. Appl Anim Behav Sci. 1991; 31: 13–24. https://doi.org/10.1016/0168-1591(91)90149-R
5. König U, Visser EK, Hall C. Indicators of stress in equitation. Appl Anim Behav Sci. 2017; 190: 43–56. https://doi.org/10.1016/j.applanim.2017.02.018
6. Liapis A, Katsanos C, Sotiropoulos DG, Karousos N, Xenos M. Stress in interactive applications: analysis of the valence-arousal space based on physiological signals and self-reported data. Multimed Tools Appl. 2017; 76: 5051–5071. https://doi.org/10.1007/s11042-016-3637-2
7. Koolhaas JM, Bartolomucci A, Buwalda B, de Boer SF, Flügge G, Korte SM, et al. Stress revisited: A critical evaluation of the stress concept. Neurosci Biobehav Rev. 2011; 35: 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003 PMID: 21316391
8. Buchanan KL. Stress and the evolution of condition-dependent signals. Trends Ecol Evol. 2000; 15: 156–160. https://doi.org/10.1016/s0169-5347(99)01812-1 PMID: 10717685
9. Minton JE. Function of the hypothalamic-pituitary-adrenal axis and the sympathetic nervous system in models of acute stress in domestic farm animals. J Anim Sci. 1994; 72: 1891–1898. https://doi.org/10.2527/1994.7271891x PMID: 7928769

10. Butler RK, Finn DP. Stress-induced analgesia. Prog Neurobiol. 2009; 88: 184–202. https://doi.org/10.1016/j.pneurobio.2009.04.003 PMID: 19393288

11. Jennings EM, Okine BN, Roche M, Finn DP. Stress-induced hyperalgesia. Prog Neurobiol. 2014; 121: 1–18. https://doi.org/10.1016/j.pneurobio.2014.06.003 PMID: 25010858

12. Zolovick A, Upson DW, Eleftheriou BE. Diurnal Variation in Plasma Glukocorticosteroid Levels in the Horse (Equus Caballus). J Endocrinol. 1966; 35: 249–253. https://doi.org/10.1677/joe.0.0350249 PMID: 5926080

13. Rietmann TR, Stauffacher M, Bernasconi P, Auer JA, Weishaupt MA. The Association between Heart Rate, Heart Rate Variability, Endocrine and Behavioural Pain Measures in Horses Suffering from Laminitis. J Vet Med Ser A. 2004; 51: 218–225. https://doi.org/10.1111/j.1439-0442.2004.00627.x PMID: 15315700

14. Hörnicke H, Engelhardt WV., Ehrlein H-J. Effect of exercise on systemic blood pressure and heart rate in horses. Pflügers Arch. 1977; 372: 95–99. https://doi.org/10.1007/BF00582212 PMID: 563589

15. Williams AC. Facial expression of pain: An evolutionary account. Behav Brain Sci. 2002; 25: 439–455. https://doi.org/10.1017/s0140525x02000080 PMID: 12879700

16. Ekman P, Friesen WV, Hager JC. Facial Action Coding System. Salt Lake City: Research Nexus; 2002.

17. Descovich KA, Wathan J, Leach MC, Buchanan-Smith HM, Flecknell P, Farningham D, et al. Facial expression: An under-utilized tool for the assessment of welfare in mammals. ALTEX. 2017; 34: 409–429. https://doi.org/10.14573/altex.1607161 PMID: 28214916

18. Wathan J, Burrows AM, Waller BM, McComb K. EquiFACS: The Equine Facial Action Coding System. PLoS ONE. 2015; 10(8): e0131738. https://doi.org/10.1371/journal.pone.0131738 PMID: 26244573

19. Vick SJ, Waller BM, Parr LA, Pasqualini MCS, Bard KA. A cross-species comparison of facial morphology and movement in humans and chimpanzees using the Facial Action Coding System (FACS). J Nonverbal Behav. 2007; 31: 1–20. https://doi.org/10.1007/s10919-006-0017-z PMID: 21188285

20. Waller BM, Peirce K, Caeiro CC, Scheider L, Burrows AM, et al. Paedomorphic Facial Expressions Give Dogs a Selective Advantage. PLoS ONE. 2013; 8(12): e82686. https://doi.org/10.1371/journal.pone.0082686 PMID: 24386109

21. Chambers CT, Mogil JS. Ontogeny and phylogeny of facial expression of pain. Pain. 2015; 156: 798–799. https://doi.org/10.1097/j.pain.0000000000000133 PMID: 25887392

22. Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach C. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. PLoS ONE. 2014; 9 (3): e92281. https://doi.org/10.1371/journal.pone.0092281 PMID: 24647606

23. Gleerup KB, Forkman B, Lindegaard C, Andersen PH. An equine pain face. Vet Anaesth Analg. 2015; 42: 103–114. https://doi.org/10.1111/vaa.12212 PMID: 25082060

24. van Loon JPAM Van Dierendonck MC. Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study. Vet J. 2015; 206: 356–364. https://doi.org/10.1016/j.tvjl.2015.08.023 PMID: 26526526

25. McLennan KM, Miller AL, Dalla Costa E, Stucke D, Corke MJ, Broom DM, et al. Conceptual and methodological issues relating to pain assessment in mammals: The development and utilisation of pain facial expression scales. Appl Anim Behav Sci. 2019; 217: 1–15. https://doi.org/10.1016/j.applanim.2019.06.001

26. Mayo LM, Heilig M. In the face of stress: Interpreting individual differences in stress-induced facial expressions. Neurobiol Stress. 2019; 10: 100166. https://doi.org/10.1016/j.ynstr.2019.100166 PMID: 31193535

27. Hintze S, Smith S, Patt A, Bachmann I, Wurbel H. Are Eyes a Mirror of the Soul? What Eye Wrinkles Reveal about a Horse's Emotional State. PLoS ONE. 2016; 11(10): e0164017. https://doi.org/10.1371/journal.pone.0164017 PMID: 27732647

28. Merkies K, Ready C, Farkas L, Hodder A. Eye Blink Rates and Eyelid Twitches as a Non-Invasive Measure of Stress in the Domestic Horse. Animals. 2019; 9: 562. https://doi.org/10.3390/ani9080562 PMID: 31443315

29. Mott RO, Hawthorne SJ, McBride SD. Blink rate as a measure of stress and attention in the domestic horse (Equus caballus). Sci Rep. 2020; 10: 1–8.

30. Kunz M, Meixner D, Lautenbacher S. Facial muscle movements encoding pain—A systematic review. Pain. 2019; 160: 535–549. https://doi.org/10.1097/j.pain.0000000000001424 PMID: 30335682

31. Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J. Recognizing facial expression: machine learning and application to spontaneous behavior. 2005 IEEE Computer Society Conference

on Computer Vision and Pattern Recognition (CVPR'05); 2005 June 20–25; San Diego, USA. IEEE Xplore. 2005. p. 568–573 vol. 2.

32. Pons P, Jaen J, Catala A. Assessing machine learning classifiers for the detection of animals' behavior using depth-based tracking. Expert Syst Appl. 2017; 86: 235–246. https://doi.org/10.1016/j.eswa.2017.05.063
33. Rashid M, Silventoinen A, Gleerup KB, Andersen PH. Equine Facial Action Coding System for determination of pain-related facial responses in videos of horses. PLoS ONE. 2020; 15(11): e0231608. https://doi.org/10.1371/journal.pone.0231608 PMID: 33141852
34. Young T, Creighton E, Smith T, Hosie C. A novel scale of behavioural indicators of stress for use with domestic horses. Appl Anim Behav Sci. 2012; 140: 33–43. https://doi.org/10.1016/j.applanim.2012.05.008
35. Tateo A, Padalino B, Boccaccio M, Maggiolino A, Centoducati P. Transport stress in horses: Effects of two different distances. J Vet Behav Clin Appl Res. 2012; 7: 33–42. https://doi.org/10.1016/j.jveb.2011.04.007
36. Wathan J, McComb K. The eyes and ears are visual indicators of attention in domestic horses. Curr Biol. 2014; 24: 677–679.
37. Rietmann TR, Stuart AEA, Bernasconi P, Stauffacher M, Auer JA, Weishaupt MA. Assessment of mental stress in warmblood horses: Heart rate variability in comparison to heart rate and selected behavioural parameters. Appl Anim Behav Sci. 2004; 88: 121–136. https://doi.org/10.1016/j.applanim.2004.02.016
38. Rashid M, Broome S, Andersen PH, Gleerup KB, Lee YJ. What should I annotate? An automatic tool for finding video segments for EquiFACS annotation. In: Grant R, Allen T, Spink A, Sullivan M, editors. Measuring Beaviour Conference Proceedings; 2018 June 5–8; Manchester, UK. p. 165–166.
39. ELAN (Version 5.4). Nijmegen: Max Planck Institute for Psycholinguistics; https://tla.mpi.nl/tools/tla-tools/elan/.
40. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn: Macine Learning in Python. GetMobile Mob Comput Commun. 2015; 19: 29–33. https://doi.org/10.1145/2786984.2786995
41. Hartmann E, Søndergaard E, Keeling LJ. Keeping horses in groups: A review. Appl Anim Behav Sci. 2012; 136: 77–87. https://doi.org/10.1016/j.applanim.2011.10.004
42. Friend TH. A review of recent research on the transportation of horses. J Anim Sci. 2001; 79: E32–E40. https://doi.org/10.2527/jas2001.79E-SupplE32x
43. Russell JA, Lewicka M, Niit T. A Cross-Cultural Study of a Circumplex Model of Affect. J Pers Soc Psychol. 1989; 57: 848–856. https://doi.org/10.1037/0022-3514.57.5.848
44. Lansade L, Nowak R, Lainé AL, Leterrier C, Bonneau C, Parias C, et al. Facial expression and oxytocin as possible markers of positive emotions in horses. Sci Rep. 2018; 8: 1–11.
45. Waran NK, Cuddeford D. Effects of loading and transport on the heart rate and behaviour of horses. Appl Anim Behav Sci. 1995; 43: 71–81. https://doi.org/10.1016/0168-1591(95)00555-7
46. White A, Reyes A, Godoy A, Martínez R. Effects of transport and racing on ionic changes in thoroughbred race horses. Comp Biochem Physiol A Physiol. 1991; 99: 343–346. https://doi.org/10.1016/0300-9629(91)90011-Z
47. Roberts K, Hemmings AJ, Moore-Colyer M, Parker MO, McBride SD. Neural modulators of temperament: A multivariate approach to personality trait identification in the horse. Physiol Behav. 2016; 167: 125–131. https://doi.org/10.1016/j.physbeh.2016.08.029 PMID: 27597134
48. Munsters CCBM, de Gooijer J-W, van den Broek J, van Oldruitenborgh-Oosterbaan MMS. Heart rate, heart rate variability and behaviour of horses during air transport. Vet Rec. 2013; 172: 15–15. https://doi.org/10.1136/vr.100952 PMID: 23143989

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:38

This thesis investigates how stress and pharmacological sedation affects facial expressions and impacts their capacity to evaluate pain. Facial activity was high during stressful managerial situations but low during sedation. Furthermore, simultaneous presence of pain in combination with these states made discrimination harder. Methods currently available for analysis of facial expressions perform well in certain instances but are in need of expanding upon in order to achieve better performance.

Johan Lundblad received his postgraduate education at the Department of Animal Biosciences (formerly Anatomy, Physiology and Biochemistry), Swedish University of Agricultural Sciences (SLU). He obtained his veterinary degree at SLU.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-340-9
ISBN (electronic version) 978-91-8046-341-6