Title: Coding and quantification of a facial expression for pain in lambs.

Authors: M.J. Guesgen[a], N.J. Beausoleil[a], M. Leach[b], E.O. Minot[c], M. Stewart[d], K.J. Stafford[a]

[a] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

[b] School of Agriculture, Food & Rural Development, Newcastle University, Newcastle upon Tyne, United Kingdom

[c] Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand

[d] InterAg, Ruakura Research Centre, Hamilton, New Zealand.

Correspondence address:

Mirjam Guesgen, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand 4442.

mguesgen@gmail.com

Abstract

Facial expressions are routinely used to assess pain in humans, particularly those who are non-verbal. Recently, there has been an interest in developing coding systems for facial grimacing in non-human animals, such as rodents, rabbits, horses and sheep. The aims of this preliminary study were to: 1. Qualitatively i~~I~~dentify facial feature changes in ~~docked~~ lambs experiencing pain as a result of tail-docking and compile these changes to create a Lamb Grimace Scale (LGS); 2. Determine whether human observers can use the LGS to differentiate tail-docked lambs from control lambs and differentiate lambs before and after docking; 3. Determine whether changes in facial action units of the LGS can be objectively quantified in lambs before and after docking; 4. Evaluate effects of restraint of lambs on observers' perceptions of pain using the LGS and on quantitative measures of facial action units. By comparing images of lambs before (no pain) and after (pain) tail-docking, the LGS was devised in consultation with scientists experienced in assessing facial expression in other species. The LGS consists of five facial action units: Orbital Tightening, Mouth Features, Nose Features, Cheek Flattening and Ear Posture. The aims of the study were ~~evaluated using~~addressed in two experiments. In Experiment I, still images of the faces of ~~seven~~ restrained lambs were taken from video footage before and after tail-docking (n = 4) or sham tail~~-~~docking (n = 3). These images were scored by ~~two~~a ~~different~~ group~~s~~ of five naïve human observers using the LGS. Because lambs were restrained for the duration of the experiment, Ear Posture was not scored. The scores for the images were averaged to provide one value per feature per period and then scores for ~~the four~~the four LGS action units ~~assessed~~ were averaged to give one LGS score per lamb per period. In Experiment II, still images of the faces ~~of unrestrained, as well as restrained,~~nine lambs ~~(n = 9)~~ were taken before and after tail-docking. Stills were taken when lambs were restrained and unrestrained in each period. A different group of five human observers scored the images from Experiment II. Changes in facial action units were~~as~~ also quantified objectively by a researcher using image measurement software. In both experiments LGS scores were~~as~~ analyzed using a linear MIXED model to evaluate the effects of tail docking on observers' perception of facial

expression changes. Kendall's Index of Concordance was used to measure reliability among~~between~~ observers. In Experiment I, ~~H~~human observers were, ~~to some extent,~~ able to use the LGS to differentiate docked lambs from control lambs. LGS scores significantly increased from before to after treatment in docked lambs but not control lambs~~ in Experiment I~~. In Experiment II there was a significant increase in LGS scores after docking. This was coupled with changes in other validated indicators of pain after docking in the form of pain-related behaviour. Only two components, Mouth Features and Orbital Tightening, showed significant quantitative changes after docking. The direction of these~~se~~ changes agree with the description of these facial action units in the LGS. Restraint affected people's perceptions of pain ~~and~~ as well as quantitative measures of LGS components.~~ with f~~ Freely moving lambs ~~being~~ were scored lower using the LGS over both periods and had a significantly smaller eye aperture and smaller nose and ear angles than when they were held. Agreement among~~between~~ observers for LGS scores were fair overall (Experiment I: W = 0.60; Experiment II: W = 0.66). This preliminary study ~~is the first to~~ demonstrates changes in lamb facial expression associated with pain. The results of these experiments should be interpreted with caution due to low lamb numbers. ~~We encourage other researchers to investigate lamb, and sheep, facial expression further.~~

Keywords: Facial expression; Grimace; Lamb; Pain

1. Introduction

Most mammals can change their facial expression in response to a range of stimuli or experiences (Diogo et al. 2009). Mammalian facial expressions may serve an adaptive function, whereby information about emotion, intent or the environment can be sent to a nearby observer. In humans, describing facial expression is nearly synonymous with describing emotion (Waller and Micheletta 2013). Several studies identify a specific facial expression, or grimace, for pain in humans. This can be described in terms of brow lowering, cheek raise, eyelid tightening, nose wrinkle and eye closing (Craig, Prkachin, and Grunau 1992; Prkachin 1992). Identifying human pain via facial grimace is useful, as it allows clinicians to assess pain in non-verbal patients (Hicks et al. 2001).

Recently, there has been interest in developing coding systems for grimacing in non-human mammals. A Mouse Grimace Scale (MGS) was developed using the same method ~~as~~ used to identify pain related changes in human facial expressions. Photographs of mice from before and after a range of routine nociceptive tests (for example administration of an irritating substance into the abdominal cavity) were compared to identify which facial features changed in response to these painful procedures (Langford et al. 2010). The MGS consists of five facial action units: orbital tightening, nose bulge, cheek bulge, ear position and whisker change. Rat and Rabbit Grimace Scales were subsequently developed in a similar manner (Sotocinal et al. 2011; Keating et al. 2012), with the majority of action units being broadly similar to those of the MGS with the exception of cheek flattening in rats and rabbits as opposed to bulging in mice. The Horse Grimace Scale has been developed using animals undergoing routine castration. This study identified six facial action units: stiffly backward ears, orbital tightening, tension above the eye area, prominent strained chewing muscles, mouth strained and pronounced chin, and strained nostrils and flattening of the profile (Dalla Costa et al. 2014). Similar changes were noted in horses when a tourniquet was applied on the antebrachium or with the application of an irritant, but were described as low ears, angled eye, withdrawn and tense stare, square-like nostrils, tension of the muzzle, and tension of the mimic muscles (Gleerup et al. 2015).

Facial expressions of pain in these species are consistently recognized by human observers as all scales demonstrated high inter-observer reliability and accuracy. Further studies have been conducted to validate these scales by using them to assess post-procedural pain and efficacy of routinely used analgesics as well as by comparing the findings to behavioural and physiological indicators of pain (Keating et al. 2012; Leach et al. 2012; Dalla Costa et al. 2014; Gleerup et al. 2015; Matsumiya et al. 2012).

Recently, a facial expression scale was developed for sheep with naturally-occurring pain due to foot-rot. Consistent with scales for other mammals, the Sheep Pain Facial Expression Scale describes changes in orbital tightness, cheek tightness, ear position, lip and jaw profile, and nostril and philtrum position (McLennan et al. 2016). Lambs may also show a noticeable change in facial expression due to acute pain. Domestic lambs experience pain routinely as they undergo painful husbandry procedures including ear tagging, tail docking and castration (Mellor and Stafford 2000). ~~There is evidence that sheep pay attention to, and remember faces of conspecifics (Kendrick 2008; Kendrick et al. 1995; Kendrick et al. 1996; Kendrick et al. 2007). Sheep are also a social species with a strong tendency to form groups (Esztevez, Andersen, and Naevdal 2007) and are diurnal, meaning that any change in facial expression may be recognized and responded to by other members of the group. The last two points taken together may indicate a communicative function of facial expression in a social species such as sheep (Williams 2002).~~

The aims of this preliminary study were to: 1. Qualitatively identify facial feature changes in lambs experiencing pain as a result of tail-docking and compile these changes to create a Lamb Grimace Scale (LGS); 2. Determine whether human observers can use the LGS to differentiate tail-docked lambs from control lambs and differentiate lambs before and after docking; 3. Determine whether changes in facial action units of the LGS can be objectively quantified in lambs before and after docking; 4. Evaluate effects of restraint of lambs on observers' perceptions of pain using the LGS and on quantitative measures of facial action units.

2. Methods

The Massey University Animal Ethics Committee approved all procedures for both experiments (Protocol 12/104). Tail-docking occurred as part of routine husbandry practices in New Zealand and in accordance with the codes of practice outlined in the Painful Husbandry Procedures Code of Practice (Anonymous 2005). Both experiments were undertaken at the Massey University Keebles Farm in Palmerston North, New Zealand.

2.1 Experiment I

2.1.1 Animals

Nine 5 to 6 week old Romney cross lambs were used in this study (four females and five males). Lambs were randomly selected from a flock of 40 lambs and their dams. Prior to testing, the ewes and lambs were kept on pasture according to normal husbandry practice.

2.1.2 Experimental Procedure

Testing was undertaken in an outdoor yard with concrete floors. On the day of testing, the dams and lambs were brought in from the paddock as a flock and kept in a holding yard. One lamb at a time was randomly selected for testing. The same experimenter picked the lambs up and held them in a seated position for the duration of tail docking and subsequent observation. All lambs were tested over one day.

Lambs were alternately allocated to one of two treatments: they were either tail-docked using a rubber ring or sham-docked (control). The treatments were applied by the farm manager and the lambs were restrained for the duration of observation. The rubber ring was applied using an elastrator between two tail vertebrae at a point allowing sufficient tail proximal to the ring to cover the anus (and vulva for female lambs). During sham docking the farm manager handled the tail area for 15 s, to replicate docking without the ring being applied. The lamb's face was recorded for 1 minute before (pre) and 10 minutes after (post) treatment using a high definition video camera from the front angle (Sony Handycam DCR-SR20, Sony Electronics Asia Pacific Pte Ltd., Tokyo, Japan). Ten minutes after the treatment was applied, the lambs were released back to the paddock.

Two lambs were excluded from the analysis. One received an ear notch prior to filming, which may have altered its response to tail docking. The second had been allocated to the control (sham-docking) treatment group and fell asleep during the handling procedure and hence its facial expression may have been incorrectly interpreted (Langford et al. 2010; Sotocinal et al. 2011). Therefore, data from seven lambs were analysed (3 control (1 female), 4 docked (2 female)).

2.1.3 Frame Capture

For each lamb, four still images were extracted manually from the video recordings for each period (pre- and post-treatment) to produce 8 images per lamb. Stills were selected from across the 1 min pre-docking period and in the last 5 min of the post-treatment period, as freely behaving lambs show a high frequency of other pain-related behaviours around this time after docking (Molony and Kent 1997). Images were selected every 15 s across the 1 min pre-period and every 75 seconds in the post-period. When a lamb was moving at that particular moment the image was taken immediately after or before the selection time point. We attempted to blind the person selecting frames to the treatment group by randomly numbering the videos used for selection, however due to the lambs being filmed for different time periods before and after treatment, as well as docked lambs displaying overt pain-related behaviours post-docking, this was not completely possible. Individual frames were "grabbed" using screen capture and cropped using Preview (Apple Inc., California, USA) so that the body and most of the background was no longer visible. All 56 images were used for development of the grimace scale and subsequently for scoring by human observers.

2.1.4 LGS Development

The Lamb Grimace Scale (LGS) was developed using methods similar to those used to develop the other grimace scales (Keating et al. 2012; Langford et al. 2010; Sotocinal et al. 2011). By comparing multiple images of the four docked lambs before (no pain) and after (pain) docking (that is, within-subjects design), the LGS was devised in consultation with scientists experienced in assessing facial expression in other species. Facial Action Units were selected on the basis of their presence and consistency across lambs in pain. The LGS consists of five facial action units (Table 1): Orbital Tightening, Nose Features, Mouth Features, Cheek Flattening, and Ear Posture. As with the rat, mouse, and rabbit scales, all facial action units were unweighted.

2.1.5 Lamb Grimace Scale Scoring

The 56 image files were numbered and their order randomized using a random number generation system (Random.org). Each image was then copied into Excel (Microsoft Corporation, Redmond, Washington) with one image per tab.

The group of human observers consisted of three animal science postgraduate students and two animal welfare scientists. All had experience observing animals including sheep, however only one animal welfare scientist was familiar with grimace scale scoring.

Each observer was given two files: A detailed instruction sheet, describing the facial action units and providing visual examples of each grade of each action unit; and the Excel scoring spreadsheet containing the 56 images. Observers were blinded to treatment and period. For each image, observers were asked to score the presence of each action unit on a three-point scale 0, 1, 2 (Table 1). If they were unable to see or score a particular action unit, they were asked to score it as a 9. All observers received the images in the same order. Ear Posture was not scored in Experiment I as the restraint applied to the lamb may have confounded ear posture scores. The remaining four action units (Orbital Tightening, Nose Features, Mouth Features and Cheek Flattening) were scored as follows. A score of zero indicated confidence by the scorer that the action unit was absent. 'One' indicated confidence by the scorer that the action unit was present to a moderate degree and 'two' indicated confidence by the scorer that the action unit was present to an obvious degree. Nine indicated the scorer didn't know, or did not feel confident assigning a degree of pain score to a particular action unit for this image.

2.1.6 Statistical Analysis

Analysis was performed using SAS Version 9.2 (SAS Institute Inc., North Carolina, USA). When one or two images (out of a possible four for a lamb in each period) were scored as 'don't know' (9) for a particular facial action unit, a value was imputed from an average of the remaining three, or two, images. For each lamb in each period, the scores for the four images were averaged to provide one value per FAU. Then scores for the four LGS action units were averaged to give one LGS score per lamb per period. Cases where there were more than two images scored as 'don't know' were treated as missing values. The maximum possible LGS score after 'don't know' scores were imputed was 2 and the minimum was 0.

Residuals were generated and tested for normality of distribution. The data were also tested for homogeneity of variance between periods. If the assumptions for robust parametric statistical methods were met, analyses were performed on raw data. If not, data were transformed using Blom's normalized ranks before analysis.

LGS score was analyzed using a linear MIXED model to evaluate the effects of tail docking on observers' perception of facial expression changes with period (pre-, post-docking), treatment (docked, control) and observer (1-5) as fixed effects, and lamb (1-7) as the repeated measures subject and treatment as the group. Sex could not be included in the analysis because of low lamb numbers.

Kendall's Index of Concordance was used to measure reliability among observers. Tests were conducted on ranked LGS scores. Lambs with one or more data point missing (that is, a value for a still image could not be imputed because of too many 'I don't know' responses) were excluded from the concordance analysis.

2.2 Experiment II

2.2.1 Animals

Nine 5 to 6 week old Romney cross lambs were used in this study ((4 female, 5 male) different from those used in Experiment I). Prior to testing, the sheep were kept on pasture. Testing was undertaken in a semi-covered outdoor yard with concrete floors.

2.2.2 Preparation Procedure

Lambs served as their own controls in this experiment, so all lambs were tail-docked. Dams and their lambs were brought in from the paddock as a flock and kept in a holding yard. The lambs to be tested were randomly selected from the flock and placed in a pre-testing pen approximately 3 m^2. The flock, including the test lambs' dams, was then returned to the paddock approximately 100 m away from the testing pen. Each test lamb was caught and restrained while the researcher placed twelve 1 cm^2 sticker markers at specific locations on its face as well as applying a 30 mm by 36 mm rectangular calibration sticker, secured to the centre of its forehead (Figure 1). The calibration sticker identified each lamb by a number and was later used for calibrating images for analysis (relating the number of pixels in an image to an actual length in mm. All markers were made from Tiki-Tape™ multi-purpose cloth tape (Tiki-Tape N.Z. Ltd., Wellington).

2.2.3 Testing Procedure

Testing began 5 min after all the lambs had undergone the preparation procedure, to allow the lambs to settle into the novel environment. Lambs were video recorded in 2 min periods: Pre-docking pen (PrP); pre-docking held (PrH); post-docking held (PH); post-docking pen (PP) (Figure 2).

The experimenters filmed from outside the pens to minimize disturbance to the lambs. The docking procedure was carried out in the same way as in Experiment I. After docking, the lamb was put in a post-testing pen, adjacent to the pre-testing pen and of similar size. Lambs were filmed in the same order in which they were docked, so that the time between docking and filming was consistent for all lambs.

All videos were recorded from a front-on angle, providing a close-up of the lamb's face. When lambs were freely behaving in the pens, the experimenter followed the lamb from outside the pen with the aim of maintaining the front-on angle for as much of the time as possible.

2.2.4 Frame Capture

For each of the nine lambs, three still images were extracted manually from each of the four videos (PrP, PrH, PH, PP) to produce 12 images per lamb (total n = 108). Individual frames were "grabbed" using screen capture and cropped using Preview (Apple Inc., California, USA) so that the body and most of the background was no longer visible. Stills were selected across the 2 min videos for the PrP, PrH and PH periods. Images were selected when the head angle to the camera was appropriate, that is front-on. For the PP period, stills were taken when the lamb was displaying body behaviour indicative of pain, as has been previously validated by Molony and Kent (1997). These behaviours were either: neck arch, lateral lying or stamp. This was not done for the PH period, as the movement of the lamb was restricted. Two lambs did not appear to show any of these body behaviours in which case stills were taken at regular intervals across the 2-min PP period.

2.2.5 Lamb Grimace Scale Scoring

Five agriculture or animal science postgraduate students (different from those in Experiment I) scored the 108 images according to the procedure described in the section 'Lamb Grimace Scale Scoring'. In this experiment, all five FAU were scored and included in the analysis. All observers had experience observing animals including sheep, however none were familiar with grimace scale scoring.

In addition, each image file was also scored quantitatively by MG using the sticker markers with ImageJ software (National Institutes of Health, http://rsbweb.nih.gov/ij/index.html). Each image was first calibrated by drawing a line that was the length of the calibration sticker and assigning the known horizontal length of 36 mm to the length of that line in pixels. This allowed measurements taken across different images to be compared directly. Six facial features were measured according to Table 2. An example image is provided (Figure 1). In some cases, the lamb was positioned in such a way that it was not possible to measure a particular feature, for example ear angle when the lamb was leaning against a pen wall. In these cases, a missing value was recorded.

2.2.6 Statistical Analysis

Analysis was performed using SAS Version 9.2 (SAS Institute Inc., North Carolina, USA). Data were tested for appropriateness of parametric analysis as described in Experiment I.

For images where a 'don't know' score was given, a value was imputed from an average of the other images for that period. For each lamb in each period, the scores for each FAU for the three images were averaged. Then scores for the five LGS action units were averaged to give one LGS score per lamb per period. Cases where there was more than one image (out of a possible three) scored as 'don't know' were treated as missing values.

LGS score and quantitative measures of the six facial features were each analyzed using a linear MIXED model to evaluate the effects of tail docking and restraint on facial expression changes, with period (pre-, post-docking), restraint (held, pen) and observer (1-5) as fixed effects, and lamb (1-9) as the repeated measures subject.

Kendall's Index of Concordance was used to measure reliability among observers. Tests were conducted on ranked LGS scores. Lambs with one or more data point missing (that is, a value for a still image could not be imputed because of too many 'I don't know' responses) were excluded from the concordance analysis.

3. Results

3.1 Experiment I

Period and treatment had an interactive effect on observers' LGS scores ($F_{(1,5)} = 11.23$, $P = 0.02$). Observers' LGS scores were significantly higher after docking than before for docked lambs but there was no significant change for control lambs (Raw LGS scores ± SE, docked: pre 0.71 ± 0.14, post 1.19 ± 0.14, $t = 10.03$, $P < 0.001$; control: pre 0.88 ± 0.06, post: 1.04 ± 0.06).

Observer had a significant effect on LGS scores ($F_{(4,24)} = 5.06$, $P = 0.004$). Observer one provided lower LGS scores than observers three and four and observer two provided lower scores than observer four (Raw LGS scores ± SE, One 0.70 ± 0.10, Two 0.80 ± 0.10, Three 1.07 ± 0.10, Four 1.15 ± 0.10, Five 1.02 ± 0.10, $P < 0.05$).

Observers were moderately consistent in their scoring of lamb faces overall, and agreement among observers was not due to chance. Observers agreed to a greater degree when scoring Orbital Tightening but were not very consistent when scoring Cheek Flattening (Table 3a).

3.2 Experiment II

3.2.1 Observer scores

Tail docking resulted in a significant increase in observers' LGS scores (Period effect, $F_{(1,8)} = 40.48$, $P < 0.001$: Raw mean ± SE, LGS scores: pre 0.34 ± 0.11, post 1.06 ± 0.11).

Restraint had a significant effect on observers' LGS scores ($F_{(1,8)} = 33.47$, $P < 0.001$). When lambs were in the pen observers scored LGS lower than when lambs were held regardless of period (Raw mean ± SE, LGS scores: pen 0.49 ± 0.10, held 0.92 ± 0.10). There was no significant interactive effect of restraint and period.

Observer had a significant effect on LGS scores ($F_{(4,32)} = 21.81$, $P < 0.0001$). Observer three provided significantly lower LGS scores than the other observers and observer two scored higher than all other observers (Raw LGS scores ± SE, One 0.65 ± 0.12, Two 0.95 ± 0.12, Three 0.45 ± 0.12, Four 0.74 ± 0.12, 0.75 ± 0.12, $P < 0.05$).

Observers were reasonably consistent in their scoring of lamb faces overall and agreement between observers was not due to chance. There was strong agreement among observers when scoring Ear Posture, but low agreement when scoring Mouth Changes and Cheek Flattening (Table 3b).

3.2.2 Quantitative scores

Tail docking resulted in a significant increase in mouth angle and a significant decrease in aperture of the right eye (Period effect, Mouth Angle: $F_{(1,8)} = 8.58$, *P = 0.019*; Mean of ranks ± SE: pre -0.41 ± 0.23, post 0.41 ± 0.23. Right Eye Aperture: $F_{(1,8)} = 6.14$, *P = 0.038*; Raw mean ± SE: pre 12.15 mm ± 0.72, post 10.80 mm ± 0.72).

Restraint significantly affected the aperture of the right eye, nose angle and the ear angle for both ears. Lambs in the pen had a significantly smaller eye aperture and smaller nose and ear angles than when they were held (Right Eye Aperture: $F_{(1,8)} = 5.30$, *P = 0.05*; Raw mean ± SE: pen 11.09 mm ± 0.68, held 11.85 mm ± 0.68; Nose Angle: $F_{(1,8)} = 10.98$, *P = 0.01*; Mean of ranks ± SE: pen -0.40 ± 0.21, held 0.40 ± 0.21; Right Ear Angle: $F_{(1,8)} = 17.13$, *P = 0.003*; Raw mean ± SE: pen 53.23° ± 8.62, held 86.91° ± 8.62; Left Ear Angle: $F_{(1,7)} = 7.68$, *P = 0.028*; Mean of ranks ± SE: pen -0.29 ± 0.26, held 0.29 ± 0.26).

4. Discussion

The aims of this preliminary study were to: 1. Identify facial feature change in docked lambs to create a Lamb Grimace Scale (LGS); 2. Determine whether observers can use the LGS to differentiate lambs in pain from those not; 3. Determine whether changes in facial action units of the LGS can be quantified in lambs before and after docking; 4. Evaluate effects of restraint of lambs on observers' perceptions of pain using the LGS and on quantitative measures of facial action units.

We identified five facial action units that were consistently altered by the pain of tail-docking. Unweighted scores for those FAU were averaged to produce the LGS, scores that reflect the degree to which facial expression changed. The Sheep Pain Facial Expression Scale, recently developed independently, identified similar FAU changes in sheep experiencing pain from foot-rot and mastitis (McLennan et al. 2016), supporting the notion that this expression reflects pain more generally in this species. In addition, there was considerable overlap between the facial features identified in the LGS with those of the rat, rabbit, mouse and horse grimace scales. All scales identified orbital tightening or tension in the eye area and ears held back as features that changed when the animal was in pain. Similarly, a pointed nose and flattened cheeks are features that both pained lambs and rabbits show.

The similarities among sheep, mouse, rat, rabbit and horse grimace scales lend evidence to the prediction of similar facial expressions for emotions across mammalian species (Darwin 1872; Williams 2002; Dalla Costa et al. 2014). However, it is worthwhile noting that the same people were involved in developing all grimace scales, therefore some overlap may be due to *a priori* knowledge of facial feature changes. Any differences in FAUs among grimace scales may be due to the disparity in size, composition and musculature between lamb, mouse, rat and rabbit faces. It may also be the case that different types of pain (thermal, chemical, mechanical) induce slightly different facial expressions, although this remains to be investigated.

Human observers instructed in its use were able to apply the LGS to differentiate docked lambs from control lambs. LGS scores significantly increased from before to after treatment in docked lambs but not control lambs in Experiment I. Likewise, in Experiment II there was a significant increase in LGS scores after docking. These findings suggest that the LGS reflects facial expression of pain in lambs that is recognizable by human observers.

Agreement among observers for LGS scores was moderate overall ($W = 0.60$ for Experiment I and 0.66 for Experiment II). However, there was greater agreement among observers when scoring the facial action units of Ear Posture and, to a lesser extent, Nose Changes and Orbital Tightening. Ears are a prominent feature of the lamb face and therefore changes in ear posture are likely to be more recognizable than changes in other facial features.

There were systematic differences in LGS scores among observers in both experiments. In Experiment I over both periods for both docked and control lambs, observer one provided lower LGS scores than most other observers and observer two provided lower LGS scores than observer four. Observer one was experienced with scoring facial action units in laboratory animals and observer two had spent the most time researching sheep and being in the presence of lambs. These factors may account for observational differences as observers with such experience may be more stringent or cautious in attributing pain to lambs. In Experiment II over all periods observer three provided significantly lower LGS scores than the other observers and observer two scored higher than all other observers. Observer three had the most experience with sheep out of the five observers in Experiment II, which again may lead to more cautious scoring. In contrast, observer two was the youngest observer and the newest to the study of animal science and behaviour and therefore may have been the most susceptible to being primed to look for a facial grimace. The other three observers had fairly similar experience with sheep and none had experience scoring facial features. This indicates that individual differences, possibly personal pain tolerance or ability to empathize with the animals, may influence how observers use the LGS (Noring et al. 2014; Furnham, McManus, and Scott 2003). In addition, we do not know how well observers learnt to use the scale before scoring, which may have also resulted in differences between observers. The statistical effect of observer only indicates differences between, and not within, observers so we are unable to determine how reliably observers scored different images of the same lamb.

We also investigated whether changes in the components of the LGS could be objectively quantified using software that compares distances between facial markers in lambs before and after docking. Only two components, Mouth Features and Orbital Tightening, showed significant quantitative changes after docking. The direction of these changes agree with the description of these facial action units in the LGS i.e. docked lambs had a larger mouth angle, so their mouth looked more like a horizontal line from the front angle. In addition, the aperture of the right eye was smaller, indicating squinting of the eye.

There were no significant quantitative changes in any of the other facial action units. This may be because of asymmetry in lamb facial expression, varying camera angles and difficulties accounting for depth. It is likely that lamb facial expression is asymmetrical, meaning that one eye or ear may not show significant change after docking. This may account for the fact that we only found a period effect for the right eye aperture and not the left. Lateralization of facial expression, particularly eye and ear movements, has been demonstrated in humans, other primates and dogs (Nagasawa et al. 2013; Rogers and Andrew 2002). Second, quantitative measures are more sensitive to slight changes in the camera angle and are affected by depth perception more so than qualitative ones. Observers may be able to intuitively correct for slight variations in camera angle or perceive depth within a photograph while the software cannot. Thus they may tell whether the ears, for example, are pointed backwards or are horizontal. In contrast, when an angle or length measurement is superimposed on top of a two-dimensional photograph, this information is lost.

Informal observations made in Experiment I suggested that facial expression may be affected by restraint. We therefore set out to test this idea in Experiment II. Freely moving lambs were scored lower using the LGS over both periods and had a significantly smaller eye aperture and smaller nose and ear angles than when they were held. Orbital tightening and smaller nose angle contradict the lower LGS reported for penned lambs, while the smaller ear angle is consistent and reflects a more 'ears forward' position. Lambs in the pen may have had their ears forward to gather information about their conspecifics (Guesgen et al. 2016) and, as noted, this FAU may have had the strongest effect on LGS.

These observations suggest that the changes observed due to restraint are not coherent with regard to the lamb pain expression, that is, they are a mixture of changes that do not match the specific suite of changes seen in pained lambs. Nonetheless, restraint had some effect on facial expression and influenced both qualitative and quantitative measures of that expression. In this design, restraint was also confounded with time, i.e. relative to the start of yarding (pre) and to the application of the ring (post). Therefore the facial feature changes we observed were likely the result of an interaction of effects associated with the experimental procedures and restraint and the pain experience associated with tail-docking. The results also suggest that not all facial features may be affected equally by stress or some other aspect of restraint. An alternative explanation for the restraint effect is that video recording methods in the pen versus when the lamb was restrained altered people's perceptions of lamb pain. It was easier to maintain a constant camera angle when the lamb was held as opposed to free in the pen. Camera angle may have a particularly pronounced effect on quantitative measures of FAU changes. This point is discussed in more detail below.

4.1 Study limitations

This study brought to light some issues that should be considered when investigating lamb facial expression including the selection of images used to develop the LGS in Experiment I, the order of images being the same for each human observer, some lambs not showing body behavioural signs of pain at the time points the stills were selected, potential inconsistencies in observer training, and differences in camera angle which are discussed in the following paragraphs.

Two out of nine docked lambs in Experiment II did not show other behavioural signs that they were in pain. However, it has been observed previously that not all individuals display overt or active behavioural signs of pain after docking (Petrie et al. 1995) and so this finding is perhaps not surprising. These particular lambs may have been displaying a reactive coping style to pain, which is characterized by immobility and passivity (Koolhaas et al. 1999). Their LGS scores did not fall outside two standard deviations of the group means so they were not deemed outliers in this regard. Therefore, even if these animals were reactive copers, this did not appear to influence their facial expressions exhibited following docking, unlike other behavioural expressions.

In previous studies examining facial expression in non-human animals, accuracy of detecting pain using facial grimace scales was assessed by comparing an overall pain score to the *a priori* knowledge of whether the animal was in pain. By doing so, it was possible to calculate a percentage of correct identifications of pain/not pain which ranged from 73.3% (Dalla Costa et al. 2014) to 83.6% (Keating et al. 2012). However, this measure of accuracy is based on the assumption that all animals were showing facial feature changes. Therefore, any such accuracy measures would have to have been based on a decision by the researchers as to whether a lamb was displaying a grimace.

In Experiment II, it was difficult to maintain the same camera angle as the lamb was moving in the pen, but easier to do so when the lamb was restrained. This may have affected observers' interpretation of facial expression or altered the quantitative measurements of those features. In part, this may explain some of the restraint effect we observed. Ideally, we would use some kind of head-mounted system to capture sheep facial expressions so that the camera angle was consistent throughout. It may also be possible to use a motion capture system, similar to those used in the film industry, to more accurately track facial feature changes over time. The feasibility of such technology for an animal science application is an avenue for further investigation.

Finally, we are unable to know how well observers learnt to use the scale before scoring. Because of this, we would recommend getting a group of observers with similar experience and training all observers together on the LGS before commencing scoring. Furthermore, while the five facial action units were appropriate based on expert identification, the comment was made by some observers that cheek flattening was a difficult feature to assess due to differences in camera angle or lighting. This feature also had the lowest concordance among observers, suggesting that this action unit contributed little to the LGS and therefore could be excluded. In this study, as in previous studies of animal facial expression, facial action units were unweighted. However, it may be interesting in future to assess how much each particular feature is contributing to observers' impressions of lamb pain.

Conclusion

This preliminary study demonstrates changes in lamb facial expression associated with docking pain. Human observers were able to use the LGS to distinguish between lambs in pain and those not. Furthermore, we found significant quantitative changes in some facial action units. However, the changes in observer scores for Experiment II did not correlate well with the features that changed quantitatively. The results of these experiments should be interpreted with caution due to low lamb numbers. We encourage other researchers to investigate lamb, and sheep, facial expression to further develop this method of pain assessment in animals.

Acknowledgements

We would like to thank technician Dean Burnham, and Lorelle Barrett for practical assistance.

References

Anonymous. 2005. Animal Welfare (Painful Husbandry Procedures) Code of Welfare, Ministry of Agriculture and Forestry (MAF), Wellington, New Zealand.

Craig, K.D., Prkachin, K.M. and Grunau, R.V.E. 1992. The facial expression of pain. In: D.C. Turk and R. Melzack (Editors), Handbook of pain assessment, Guilford Press, New York.

Dalla Costa, E., Minero, M., Lebelt, D., Stucke, D., Canali, E. and Leach, M. 2014. Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. PLoS ONE, 9: e92281.

Darwin, C. 1872. *The Expression of the Emotions in Man and Animals*. Albemarie, London.

Diogo, R., Wood, B.A., Aziz, M.A. and Burrow, A. 2009. On the origin, homologies and evolution of primate facial muscles, with a particular focus on hominoids and a suggested unifying nomenclature for the facial muslces of the Mammalia. Journal of Anatomy, 215: 300-319.

Esztevez, I., Andersen, I.L. and Naevdal, E. 2007. Group size, density and social dynamics in farm animals. Applied Animal Behaviour Science, 103: 185-204.

Furnham, A., McManus, C. and Scott, D. 2003. Personality, empathy and attitudes to animal welfare. Anthrozoos: A Multidisciplinary Journal of the Interactions of People and Animals, 16: 135-146.

Gleerup, K.B., Forkman, B., Lindegaard, C. and Andersen, P.H. 2015. An equine pain face. Veterinary Anaesthesia and Analgesia, 42: 103-114.

Guesgen, M., Beausoleil, N., Minot, E., Stewart, M., Stafford, K. and Morel, P. 2016. Lambs show changes in ear posture when experiencing pain. Animal Welfare, 25: 171-177.

Hicks, C.L., von Baeyer, C.L., Spafford, P.A., van Korlaar, I. and Goodenough, B. 2001. The Faces Pain Scale - Revised: toward a common metric in pediatric pain measurement. Pain, 93: 173-183.

Keating, S.C.J., Thomas, A.A., Flecknell, P.A. and Leach, M.C. 2012. Evaluation of EMLA cream for preventing pain during tattooing of rabbits: Changes in physiological, behavioural and facial expression responses. PLoS ONE, 7: e44437.

Kendrick, K.M. 2008. Sheep Senses, Social Cognition and Capacity for Consciousness. In: C.M. Dwyer (Editor), The Welfare of Sheep, Springer Science & Business Media B.V.

Kendrick, K.M., Atkins, K., Hinton, M.R., Broad, K.D., Fabre-Nys, C. and Keverne, B. 1995. Facial and vocal discrimination in sheep. Animal Behaviour, 49: 1665-1676.

Kendrick, K.M., Atkins, K., Hinton, M.R., Heavens, P. and Keverne, B. 1996. Are faces special for sheep? Evidence from facial and object discrimination learning tests showing effects of inversion and social familiarity. Behavioural Processes, 38: 19-35.

Kendrick, K.M., da Costa, A.P., Leigh, A.E., Hinton, M.R. and Peirce, J.W. 2007. Sheep don't forget a face. Nature, 447: 346-346.

Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A.W. and Blokhuis, H.J. 1999. Coping styles in animals:

current status in behavior and stress-physiology. Neuroscience and Biobehavioral Reviews, 23: 925-935.

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., Ferrari, M.D., Craig, K.D. and Mogil, J.S. 2010. Coding of facial expressions of pain in the laboratory mouse. Nature Methods, 7: 447-449.

Leach, M.C., Klaus, K., Miller, A.L., Scotto di Perrotolo, M., Sotocinal, S.G. and Flecknell, P.A. 2012. The assessment of post-vasectomy pain in mice using behaviour and the Mouse Grimace Scale. PLoS ONE, 7: e35656.

Matsumiya, L.C., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wieskopf, J.S., Zaloum, A., King, O.D. and Mogil, J.S. 2012. Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. Journal of the American Association for Laboratory Animal Science, 51: 42-49.

McLennan, K.M., Rebelo, C.J.B., Corke, M.J., Holmes, M.A., Leach, M.C. and Constantino-Casas, F. 2016. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. Applied Animal Behaviour Science, 176: 19-26.

Mellor, D.J. and Stafford, K.J. 2000. Acute castration and/or tailing distress and its alleviation in lambs. New Zealand Veterinary Journal, 48: 33-43.

Molony, V. and Kent, J.E. 1997. Assessment of acute pain in farm animals using behavioral and physiological measurements. Journal of Animal Science, 75: 266-272.

Nagasawa, M., Kawai, E., Mogi, K. and Kikusui, T. 2013. Dogs show left facial lateralization upon reunion with their owners. Behavioural Processes, 98: 112-116.

Noring, M., Wikman, I., Hokkanen, A.H., Kujala, M.V. and Hänninen, L. 2014. Empathic veterinarians score cattle pain higher. The Veterinary Journal, 200: 186-190.

Petrie, N.J., Stafford, K.J., Mellor, D.J., Bruce, R.A. and Ward, R.N. 1995. The behaviour of calves tail docked with a rubber ring used with or without local anaesthesia. Proceedings of the New Zealand Society of Animal Production, 55: 58-60.

Prkachin, K.M. 1992. The consistency of facial expressions of pain: a comparison across modalities. Pain, 51: 297-306.

Rogers, L.J. and Andrew, R.J. 2002. *Comparative Vertebrate Lateralization*. Cambridge University Press, Cambridge.

Sotocinal, S.G., Sorge, R.E., Zaloum, A., Tuttle, A.H., Martin, L.J., Wieskopf, J.S., Mapplebeck, J.C.S., Wei, P., Zhan, S., Zhang, S., McDougall, J.J., King, O.D. and Mogil, J.S. 2011. The Rat Grimace Scale: a partially automated method for quantifying pain in the laboratory rat via facial expressions. Molecular Pain, 7: 1-10.

Waller, B.M. and Micheletta, J. 2013. Facial expression in nonhuman animals. Emotion Review, 5: 54-59.

Williams, A.C.D. 2002. Facial expression of pain: an evolutionary account. Behavioral and Brain Sciences, 25: 439- 488.

Tables

Table 1. Description of the Lamb Grimace Scale action units. Note that Ear Changes were not scored in Experiment I.

Action Unit	Description
Orbital Tightening	Lambs in pain show “squeezing” of the eye or closing of the eye, described as orbital tightening. This may only occur, or occur more strongly, in one eye. If the eye closure reduces the visibility of the eye by more than half, it would be scored as obvious (2).
Nose Features	The nose of lambs in pain appears tightened with a decrease in nostril size. Tightening may be depicted through flattening or ‘pointing’. Flattening makes the nose appear more like a horizontal line in frontal headshots, whereas pointing makes the nose appear more ‘V’ rather than ‘U’ shaped in frontal headshots.

Mouth Features	The lips of a lamb in pain are flattened and tightened. The lips appear more like a horizontal line in frontal headshots. There is lack of the ‘upwards curl’ at the edge of the lips that gives lambs their ‘smiling’ appearance when not in pain.
Cheek Flattening	Lambs in pain show less bulging of the nose and cheek area. In obvious cases, the cheek has a ‘hollowed’ appearance. When not in pain, the cheeks appear rounded in frontal headshots.
Ear Posture	Lambs in pain have ears that are tense and point backwards or downwards so that the inner part of the ear is not visible. As a result, ears may appear narrower, and flattened dorsally. When lambs are not in pain, their ears are relaxed and horizontal, or slightly forward of the head and the inner ear is visible. Note, discretion should be used when scoring lambs where the ear posture is obscured by leaning against objects, e.g. pen wall.

Table 2. Descriptions of how measurements of the facial features were made in ImageJ for Experiment II.

Feature Name	Description	Letter on example Image
Left Eye Aperture	Height of the lamb’s left eye, taken from the center of the eye vertically.	A
Right Eye Aperture	Height of the lamb’s right eye, taken from the center of the eye vertically.	(not shown)
Nose Angle	The angle formed when the three sticker markers on the tips of the nose and the center of the nose are joined by a line. The inside corners of the stickers were used as a guide.	B
Mouth Angle	The angle formed when the three sticker markers on the outside corners of the mouth and the sticker on the chin are joined by a line. The inside corners of the stickers were used as a guide.	C
Left Ear Angle	The angle from a perpendicular line to the lamb’s head to the tip of the	D

	lamb's left ear. The inside corner of the sticker on the ear tip was used as a guide.	
Right Ear Angle	The angle from a perpendicular line to the lamb's head to the tip of the lamb's right ear. The inside corner of the sticker on the ear tip was used as a guide.	(not shown)

Table 3. Kendall's Index of Concordance (W) among observers for Lamb Grimace Scale (LGS) scores as well as for each facial action unit for Experiments I (a) and II (b). $P < 0.05$ indicates that the concordance among observers is not due to chance (that is, significantly different from 0.5).

a.

Facial Feature	W	ChiSq	*P*
LGS Score	0.60	146.85	< ***0.0001***
Orbital Tightening	0.55	133.43	< ***0.0001***
Nose Changes	0.69	169.13	< ***0.0001***
Mouth Changes	0.49	120.21	< ***0.0001***
Cheek Flattening	0.34	93.05	***0.0002***

df = 49

b.

Facial Feature	W	ChiSq	*P*
LGS Score	0.66	311.31	< ***0.0001***
Orbital Tightening	0.56	196.65	< ***0.0001***
Nose Changes	0.56	222.68	< ***0.0001***
Mouth Changes	0.49	192.72	< ***0.0001***
Cheek Flattening	0.46	180.80	< ***0.0001***
Ear Posture	0.79	309.93	< ***0.0001***

df = 79

Figure Captions

Figure 1. Timeline of experimental procedure for Experiment II.

Figure 2. Example image outlining how the facial features of each lamb were measured quantitatively including placement of sticker markers on the test lambs. The larger rectangle denotes a calibration sticker of a known size for later analysis.

Supplement 1. Visual examples of the Lamb Grimace Scale action units and how they are scored.

Supplement 2. Visual examples of lamb facial expression when in pain or not in pain in Experiments I and II.