

Identification and development of measures suitable as potential breeding traits regarding dairy cows' reactivity towards humans

Asja Ebinghaus, Silvia Ivemeyer, Julia Rupp, Ute Knierim

Farm Animal Behaviour and Husbandry Section, Faculty of Organic Agricultural Sciences, University of Kassel, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany, phone: +49 (0)561 804-1748, fax +49 (0)561 804-1646, email ebinghaus@uni-kassel.de

Abstract

Behavioural indicators of the human-animal relationship (HAR) are predominantly used in animal welfare science. However, the reactivity of dairy cows – as part of the HAR - is also of interest in the context of dairy breeding, due to its estimated moderate heritability. The avoidance distance (AD) towards an unfamiliar experimenter in a standardized test is regarded an established behavioural indicator of the HAR, but for breeding purposes more feasible measures would be advantageous.

The aim of the present pilot study was to identify and develop potential measures of the cow's reactivity towards humans, which are promising as breeding traits with regard to feasibility, reliability and criterion validity.

On three German dairy farms with loose housing and herd sizes of 45 to 195 cows of the Holstein-Friesian and German Black Pied Cattle breed, AD at the feeding place and AD in the barn as well as four alternative HAR measures were recorded and tested for inter-observer reliability (IOR) and inter-test associations for the assessment of criterion validity. Alternative measures were (1) tolerance to standardised tactile interaction (TTI), (2) release behaviour after restraint (RB), (3) qualitative behaviour assessment (QBA) of the cow's during the TTI and RB Test, and (4) facial hair whorl position and form (HW). TTI, RB and QBA were additionally tested for intra-observer reliability using video recordings of 31 cows.

IOR was assessed based on Spearman rank or Kendall W correlation coefficients (in case of QBA with three observers) for metric and ordinal data and based on PABAK coefficients in case of nominal data (HW). Intra-observer reliability was assessed based on Spearman rank correlation coefficients. Inter-test associations between AD at the feeding place and HW were analysed using a General Linear Model and between all other measures using Spearman rank correlation.

IOR was good to very good for all measures: AD feeding place $r_s = 0.79$ ($n = 84$, $p < 0.01$); AD barn $r_s = 0.83$ ($n = 36$, $p < 0.01$); TTI $r_s = 0.93$ ($n = 55$, $p < 0.01$); RB $r_s = 0.90$ ($n = 54$, $p < 0.01$); QBA $W = 0.95$ ($n = 32$, $N = 3$, $p < 0.01$); HW PABAK = 0.77 – 0.83 ($n = 58$). Intra-observer reliability of the alternative behavioural measures was also very good: TTI $r_s = 0.94$ ($n = 31$, $p < 0.01$), RB $r_s = 0.89$ ($n = 31$, $p < 0.01$); QBA $r_s = 0.93$ ($n = 31$, $p < 0.01$).

High inter-test correlations were found between AD feeding place and AD barn ($r_s = 0.77$, $n = 44$, $p < 0.01$), between TTI and RB ($r_s = 0.78$, $n = 52$, $p < 0.01$) as well as between QBA and RB ($r_s = 0.76$, $n = 18$, $p < 0.01$). Moderately correlated were QBA and TTI ($r_s = 0.68$, $n = 18$, $p < 0.01$), AD feeding place and TTI ($r_s = 0.50$, $n = 44$, $p < 0.01$), and AD feeding place and RB ($r_s = 0.45$, $n = 43$, $p < 0.01$). No significant associations were found between HW and AD.

The present results suggest that TTI, RB and QBA alongside the established AD measures are suitable reactivity measures. They partly reflect similar and partly different aspects of the HAR, with an apparent clustering into distance and handling measures.

Keywords: dairy cattle, human-animal relationship, inter-observer reliability, inter-test associations

1 Introduction

Close interactions between stockpeople and cows are a regular part of dairy farming. In this situation, the quality of the human-animal relationship (HAR) is especially important. Effects of the HAR on work safety and quality for the human as well as on the welfare and productivity of the animals were found in numerous studies (reviewed by Hemsworth, 2003; Rushen and de Passillé, 2010; Waiblinger et al. 2006). Cows that are easily irritated or fearful towards humans are more difficult to handle during the milking routine (e.g. Breuer et al., 2000). Further, the quality of the HAR can influence milk ejection and milk yield (e.g. Waiblinger et al., 2002) as well as aspects of udder health (Ivemeyer et al., 2011).

Behavioural measures to assess the quality of the HAR, or more precisely the reaction of dairy cows towards humans, have predominantly been used in the context of animal welfare science, and are regarded to reflect the animals' level of fear or confidence in humans (Waiblinger et al., 2006). In dairy cows, mainly distance measures are used. The 'avoidance distance' (AD), for instance, measures the distance an animal allows a moving person to approach. AD can be analysed on individual or on farm level, and has been applied in various studies at the feeding place or in the barn (e.g. Ivemeyer et al., 2011; Waiblinger et al., 2003; Windschnurer et al., 2009). According to Waiblinger et al. (2006), AD reflects the HAR well and is suitable as a basis for on-farm assessments. Furthermore, the Welfare Quality[®] protocol for cattle includes the AD to assess the HAR on farm level (Welfare Quality[®], 2009). However, other measures might be more easily integrated in the assessment of breeding values, might be less time-consuming to apply or easier to be trained.

Besides AD and further quantitative HAR measures (reviewed in Waiblinger et al. 2006) also qualitative approaches have been developed in recent years. Based on the qualitative behaviour assessment (QBA) developed by Wemelsfelder et al. (2000, 2001), the animals' body language has been assessed in different handling situations. These studies related to calves' responses in a handling situation (Ellingsen et al., 2014), Nellore cattle temperament after exiting the crush (Sant'Anna and da Costa, 2013), pre-slaughter behaviour in Angus

steers (Stockman et al., 2012), and stress during transport (Stockman et al., 2011). To our knowledge, the QBA has not been applied explicitly in a situation of human-animal interaction in dairy cows yet.

The reactivity of cows towards humans is not only of interest in the context of animal welfare research but also for dairy cattle breeding. Estimated heritability of traits relating to the HAR in cattle, e.g. milking temperament or responses in a chute test, ranged from $h^2 = 0.07$ to 0.53 (reviewed by Adamczyk et al., 2013; Haskell et al., 2014; Schutz and Pajor, 2001). Breeding related research focuses on the animals' temperament that also contributes to the HAR. Breeding associations often routinely use the trait "milking temperament", i.e. a subjective evaluation by the animal owner of the cow's behaviour during milking (Adamczyk et al., 2013). Since milking temperament is genetically correlated with milking speed, measured as average milk flow per minute ($r_g = 0.247 \pm 0.075$; Sewalem et al., 2011), in some countries both traits are used in combination to select for enhanced milkability and manageability (e.g. Interbull, 2009; VIT, 2015). The evaluation of breeding values is done, as a rule, once in a cow's lifetime during the first lactation and includes production traits, exterior and functional traits (linear assessment). The linear assessment is conducted by a trained assessor, who records these traits on the individual primiparous cow on-farm. Milking temperament and milk flow are assigned to the group of functional traits. However, information on the reliability of these breeding traits and their validity regarding the HAR is insufficient (Haskell et al., 2014). Milk flow is likely stronger affected by milking management and physiological factors (Bruckmaier et al. 1996, Sandrucci et al. 2007) than by HAR. The farmers' evaluation of milking temperament may not be reliable enough, particularly in large herds, with changing employees or on farms using automatic milking systems.

Behavioural measures, which capture aspects of the HAR for breeding purposes, were mainly evaluated in beef cattle, so far. Hoppe et al. (2010), for instance, suggested that the categorical assessment of beef calves' behaviour on a 5-point scale in the 'chute test' and the flight-speed test' after release from restraint are suitable to improve temperament traits.

However, these or similar temperament measures have barely been tested in dairy cattle so far.

In addition, position and form of facial hair whorls were found to be related to responses of extensively managed beef cattle to the chute and flight speed test (Grandin et al., 1995). Cattle with a round hair whorl located above the eyes were significantly more agitated while restrained and handled in a chute and while exiting after release from restraint. This relationship may possibly be explained by simultaneous development of hair patterns and brain in the fetus (Smith and Gong, 1974). The findings by Grandin et al. (1995) and confirming results by Randle (1998) and Lanier et al. (2001) suggest that facial hair whorl position and form might be a potential trait in selection to improve cattle behaviour towards human handling.

Thus, the aim of the present pilot study was to identify or develop measures that are appropriate for the application as dairy cattle breeding traits, i.e. that (1) allow valid conclusions, and (2) can reliably be recorded on individual animal level by different observers (3) without requiring major training, time and other resources.

2 Animals, material and methods

2.1 Selection and adaptation of measures to be tested

Based on a literature search, measures of the cows' reactivity towards humans were selected which were applicable on-farm (e.g. no necessity of an experimental arena or complex technical equipment) on individual animal level (Table 1).

Table 1 near here

The QBA according to Wemelsfelder et al. (2009) has not been used to specifically assess the HAR in dairy cows before. The chute and flight speed tests have been applied while or after restraint of the animals in a squeeze chute, which is not feasible for on-farm assessment in dairy cattle. Therefore the chute and flight speed test as well as the QBA were

adapted for routine assessment on dairy farms and tested regarding practicability at the research farm of the University of Kassel, Germany.

Data collection for the testing of inter-observer reliability was carried out by two and in the case of QBA by three trained observers (observer A, B, and C) They were all females, wearing similar green overalls and gumboots and all having a minimum experience in working with cows and in behavioural observation, but on different levels. In addition, for testing the intra-observer reliability, observer A assessed 31 cows based on video recordings twice with a time interval of seven months.

2.2 Farms and animals

Data were collected from May to July 2014 on three dairy farms located in Northern Hesse, Germany. All farms had loose housing systems and at least in part self-locking feeding gates. While one farm with 100 horned German Black Pied cows was managed organically and offered pasture, the other two with 45 and 195 dehorned Holstein Friesian cows, respectively, were managed conventionally with zero grazing.

Measures were applied on a convenience sample of lactating cows with various parities. Dry cows and pregnant heifers were not assessed. For organisational reasons, the different measures could not always be applied on the same cows. Thus, sample sizes for the analyses were not identical and varied between $n = 32$ to 84 for IOR and between $n = 18$ and 88 for the inter-test associations.

2.3 HAR measures

The AD at the feeding place (AD_{feed}) was applied according to the Welfare Quality[®] protocol for cattle (Welfare Quality[®], 2009). Data collection was carried out after milking, when most cows were restrained in the feeding gate. The observer approached the test cow in a standardised way and recorded the distance at signs of withdrawal by the cow. These were, when the cow “moves back, turns the head to the side, or pulls back the head trying to get out of the feeding gate; head shaking can also be found” (Welfare Quality[®] 2009). For

practical reasons observer A assessed the animals always at first. Observer B followed about 10 minutes after observer A had finished the assessment. The next measurement was the AD in the barn (ADbarn), carried out after the main feeding period, when most cows moved freely in the barn. It was applied as described by Waiblinger et al. (2002, 2003). When the cow could be touched at the muzzle, the observer recorded 0.1 cm. When the cow could also be touched at the side of the head without indications of avoidance, 0.0 cm was recorded. Observer A and B collected data at the same time in different sections of the barn, but altogether on the same animals. If practicably possible, ADfeed and ADbarn measurements were repeated once and the arithmetic mean of the two values was used for data analysis.

Based on the principles of temperament tests on beef cattle (e.g. Grandin, 1993; Hoppe et al., 2010), the measurements 'tolerance to tactile interaction' (TTI) and 'behaviour during release from restraint' (RB) were developed. The cows were restrained in the feeding gate for both tests. Observer A looked at the test cow from about 2 m distance from behind, left hand and right hand side for about 30 seconds altogether to mimic the procedure of linear assessment of breeding value. Subsequently, the observer approached the cow from the left hand side and stroked slowly with gentle pressure three times in an arch from slightly below and behind the withers down to the fold of flank. The cows' reaction was rated on a 5-point scale (Table 2).

RB was assessed directly after the TTI test, during and after the feeding gate was opened by observer A as gently as possible. The cows' manner of leaving the feeding gate was also rated on a 5-point scale (Table 2). Observer B assessed the TTI and RB concurrently with observer A from the feeding table.

ADfeed, ADbarn, TTI and RB were recorded using printed record sheets.

Table 2 near here

The QBA was developed based on Wemelsfelder et al. (2001, 2002), but using a fixed list of descriptors based on Wemelsfelder et al. (2009). To assess individual cows specifically in a human-animal interaction, an appropriate list of descriptors was developed at first. 15 video clips á 1 to 2 minutes each showing a real or staged procedure of linear assessment of a dairy cow were presented to a team of 16 persons – farm animal scientists, technicians and agricultural students at the University of Kassel. After each clip, the participants were asked to write down spontaneously adjectives to describe the body language of the cow shown in the video. The total list of all observers and all video clips contained 116 different adjectives. This collection was systematically sorted according to synonyms and frequency of usage and shortened to a final list of 20 descriptors (attentive, hectic, relaxed, fearful, confident, insecure, contact-seeking, nervous, indifferent, suspicious, patient, evading, panicking, reluctant, aggressive, trustful, uneasy, friendly, tense, distressed).

Prior to data collection, three observers trained using the descriptors and scoring them on visual analogue scales from 0 to 125 mm. On these scales the degree of the respective expression was to be marked, with 0 mm being the minimum and 125 mm being the maximum. For the training, the same video clips were used as for developing the list of descriptors. No definitions of the descriptors were given, but the observers were encouraged to discuss and agree on their meaning.

On-farm the QBA was applied in connection to the staged linear assessment, the TTI and RB test on 32 cows. Observer A performed the human-animal interaction and finished rating the TTI and RB first, before applying the QBA. Simultaneously to observer A, observer B and C assessed the respective cow from the feeding table, using a printed record sheet containing the fixed descriptors with visual analogue scales. Values for each descriptor were determined by measuring the position of marks with a ruler.

Separate Principal Component Analyses (PCA, correlation matrix, no rotation, SPSS 22) for each of the three observers were carried out using the data of the assessed cows. The first principal component (PC1) was characterised by descriptors relating to

relaxation/attraction/trust on the negative and descriptors relating to fear/stress/aversiveness on the positive end (Figure 1). The cows' individual scores of the PC1 were used.

Figure 1 near here

The facial hair whorl (HW) position and form was categorised based on the method of Grandin et al. (1995) and Lanier et al. (2001). The centre of the HW was used as the reference point to determine its vertical and lateral position. The vertical position was categorised high (above the eyes), middle (between the eyes) or low (below the eyes); the lateral position was categorised as middle (on a 2.5 cm facial centre line) or as next to the centreline (Figure 2). The form was defined "normal" when the HW had a circular centre that created a swirl pattern in the hair. Deviations were considered "abnormal". Animals with no HW on the forehead were also recorded. HW position and form was judged from 90 photographs taken from each individual animal frontally at the feeding place from about 100 cm distance.

Figure 2 near here

2.4 Statistical data analysis

Inter-observer reliability concerning ADfeed, ADbarn, TTI, and RB, as well as intra-observer reliability concerning TTI, RB and QBA, were assessed on the basis of Spearman rank correlation coefficients, concerning QBA, of Kendall coefficients of concordance (Kendall W) using SPSS statistics 22 (IBM, 2013). In addition, scatterplots of the data were visually checked for possible systematic observer bias. Agreement concerning the nominal data of the HW position and form was analysed using the Prevalence Adjusted Bias Adjusted Kappa (PABAK)-statistic (Brenninkmeyer et al., 2007; Byrt, 1993), calculated in Excel (for Mac 2011, Microsoft): $PABAK = [(k*p)-1]/(k-1)$, with k = number of categories, p = proportion of matchings between observers.

For descriptive statistics and for analyses of inter-test associations, data collected by observer A were used. Associations between ADfeed and ADbarn, TTI, RB and QBA were

analysed using Spearman rank correlation. To analyse associations between HW and ADfeed, a saturated General Linear Model (GLM; SPSS 22) with the factors vertical position, lateral position, form and their interactions was used. Normal distribution of the residuals was visually estimated by QQ-Plots and homogeneity of variance was tested by Levene's test. Model requirements were only fulfilled for ADfeed as dependent variable.

3 Results

Descriptive statistics for ADfeed, ADbarn, QBA, TTI, RB, and HW are shown in Table 3.

Table 3 near here

3.1 Inter- and intra-observer reliability

Correlation coefficients for ADfeed, ADbarn, TTI, RB and QBA between the different observers ranged between 0.79 and 0.94 (Figure 4 a-f).

PABAK coefficients for vertical HW position ($k = 3$) were 0.77, for lateral HW position and HW form (both $k = 2$) were both 0.83 ($n = 58$ for all three measures).

Figure 3 near here

Correlation coefficients regarding intra-observer reliability of TTI were $r_s = 0.94$, for RB $r_s = 0.89$, and for QBA $r_s = 0.93$ ($n = 31$, $p < 0.01$).

3.2 Inter-test associations

Between the behavioural HAR measures significant inter-test associations of different strengths were found ($r_s = 0.45 - 0.78$), except for ADfeed with QBA as well as for ADbarn with TTI, RB and QBA. Strong correlations were found between the two established AD measures ($r_s = 0.77$) and within the alternative measures TTI, RB and QBA ($r_s = 0.68 - 0.78$) (Table 4). Between the vertical and lateral position and the form of HW and ADfeed no significant associations were found – neither for individual factors (position, form) nor their interactions (Table 5).

Table 4 near here

Table 5 near here

4 Discussion

4.1 Inter- and intra-observer reliability

For all measures high to very high agreements between observers were achieved. Regarding the avoidance distance at the feeding place (AD_{feed}) and in the barn (AD_{barn}), this corresponds with results from Windschnurer et al. (2008). Since the ADs have not only been recorded, but also performed by different observers, the results reflect similar responses by the cows towards different persons as well. The consecutive testing by observer A and B resulted in only slightly higher AD_{feed} towards observer A, thus the potential habituation effect was so small that it did not lead to a strong systematic observer bias. However, slight over- or underestimation by one observer cannot be excluded, since for practical reasons the order of assessments of observer A and B was not randomised, which would have been preferable. Regarding the cows' tolerance to tactile interaction (TTI) and release behaviour (RB) the only slight deviations of a maximum of one score between observers might sometimes have been due to different observation perspectives. Observer A, who performed the human-animal interaction for the tests, was inside the barn and observer B was observing from the feeding table. Regarding the RB assessment it became apparent that in the case of lameness the differentiation between normal (category 2) and fast walking speed (category 3) might be a challenge. However, also the agreement between two assessments by the same observer regarding TTI and RB with a time interval of seven months was high to very high, underlining the generally high reliability potential of the measures.

The inter- and intra-observer agreement regarding the qualitative behaviour assessment (QBA) was considerably higher than in other studies with dairy cows (e.g. Wemelsfelder et al. 2009, Bokkers et al. 2012). This might be explained by the intensive training of the

observers during the development phase, by the application of QBA in a standardised situation and by the assessment of individual cows instead of animal groups.

With regard to the alternative behavioural measures of TTI, RB and QBA, there was only one person (observer A) that had been handling the cows. In future studies it should also be tested whether the cows react in the same way towards different persons.

4.2 Inter-test associations

A number of moderate to strong inter-test correlations were found in this study. Conforming to results from Windschnurer et al. (2008), AD at the feeding place and in the barn correlated strongly, possibly reflecting similar aspects of the animals' perception of humans. Still, the differences between both measures may be related to the cows' distraction by food at the feeding place or by different experiences of the cows at the feeding gate (e.g. being fed manually) and in the barn (e.g. being driven to the milking parlour).

Also the moderate positive correlations between the established measure ADfeed and the alternative measures TTI and RB likely indicate some common motivational aspects in the cows' responses. In fact, all these measures capture responses of restrained cows at the feeding place towards an approaching human. However, restrained cows might perceive an approaching human differently depending on the location of the human. Approaches from inside the barn might be associated with aversive situations such as artificial insemination or veterinary examination or treatment. In addition, tactile contact is a much closer interaction than mere approach. Actually, all alternative handling measures involving close interaction showed a relatively high correlation with each other. It was for TTI and RB as well as for QBA and RB on a comparable level to the correlation between ADfeed and ADbarn. The differences between TTI and RB nevertheless might be explainable by the intensity of interaction. While TTI involved forced physical contact at the cows' body side, RB involved only close approach at the head. Furthermore, during the TTI test the cow stayed restrained, while for RB the cow was released and could rather control the situation, i.e. decide to stay or to move away. Differences to the QBA might also be attributable to different assessment

methods: TTI and RB used observation of defined behaviours, whereas QBA used observation of the whole body language.

Correlations of the handling measures with AD in the barn were generally lower and, due to the reduced sample size, not significant. The same was true for QBA in relation to AD at the feeding place.

Altogether the presented correlations indicate that the applied measures partly reflect similar aspects of the HAR but, in line with the suggestion of Windschnurer et al. (2008), that they also partly reflect different cows' perceptions of the human in the different tests. The measures apparently cluster into distance and handling measures, relating to specific aspects of the cows' responsiveness. The measures should be further investigated under practical or experimental conditions to further specify their meaning and to test their robustness under differing testing conditions, and also over time.

Grandin et al. (1995) found temperament ratings during and after handling in the chute affected by the facial hair whorl position in infrequently handled beef cattle. Olmos and Turner (2001) confirmed these findings with regard to the chute test for frequently handled beef cattle, but not for flight speed upon exit from the chute. Randle (1998) found no association between the facial hair whorls and the response towards a familiar human in beef cattle, but towards an unfamiliar human. Individuals with mid hair whorls showed significantly stronger reactions in an AD and a voluntary approach test than individuals with low whorls (Randle, 1998). Thus, the associations found in previous studies were partly between different HW characteristics and behaviour. In the present study no significant association between ADfeed and HW positions and form were found. The absence of associations might be explained by generally less extreme reactions towards humans due to closer and more frequent human-animal interaction in dairy farming compared to beef cattle farming. However, also the studies investigating HW and reactivity towards humans and human handling in beef cattle yielded ambiguous results. Overall, the association appears to be sensitive to circumstances and the specific aspect of reactivity.

4.3 Practicability

All tested behavioural measures assess the cows' reactivity towards humans or human handling at the feeding place and are applicable on various farm types independently from bedding design and milking system.

With regard to resources needed, recording ADfeed, TTI, and RB requires a reasonable expenditure of time and no technical equipment. Yet, at least a part of the housing has to be equipped with self-locking feeding gates. The application of ADbarn is less feasible in the course of larger scale studies, because identification of individual cows is more complicated and time consuming, especially in larger herds or under confined housing conditions. However, this limitation is less relevant for the assessment of individual cows for breeding purposes.

The RB assessment may be biased by different feeding gate types or impaired functions of old or defect facilities. For instance, the cows' responses might be more vigorous just because the opening is more complicated or takes longer.

Performance of the QBA requires high concentration on the individual cow. To ensure acceptable reliability in larger herds, QBA should be applied only on a limited sample or including sufficient breaks. Based on the experience from the present study, an observer appears to be able to assess about ten to 15 individual animals a day. This limitations might also be less relevant for breeding assessments, since as a rule only primiparous cows are to be evaluated.

All behavioural measures require training of the observers (e.g. breeding assessors), and particularly for the handling measures TTI, RB and QBA as well as for ADbarn experience with dairy cows is an advantage. This does apply less for ADfeed, since the observer stays outside the herd. However, for the application as breeding trait this is less relevant, as the breeding assessor approaches the cow inside the barn anyhow to conduct the linear assessment.

Overall, with regard to practical applicability as breeding traits, the alternative measures TTI, RB and QBA, seem to be equivalent to ADfeed and ADbarn. Further investigations of reliability and validity aspects have to show, which measure is most suitable. Moreover, besides the practical and methodological suitability of measures as breeding traits, also their genetic component of variance needs to be considered. Genetic correlations and heritability of the measures are to be investigated in sufficiently large data sets in the further course of the project.

5 Conclusions

Avoidance distances (AD), tolerance to tactile interaction (TTI), release behaviour (RB) and qualitative behaviour assessment (QBA) were found to be promising measures of dairy cows' reactivity towards humans, being suitable as breeding traits in terms of practical application and reliability. Good to very good agreement between trained observers could be reached, for TTI, RB and QBA also within one observer after seven months. They therefore proved to be sufficiently feasible to apply. Significant moderate inter-test correlations between ADfeed, TTI and RB suggest that they partly reflect similar and partly different aspects of the human animal relationship. Higher correlations between the two AD measures on the one hand and between TTI, RB and QBA on the other hand indicate a distinction between distance and handling measures.

6 Acknowledgements

We are very grateful to the participating farmers for their openness and cooperation and to Françoise Wemelsfelder for her great support in developing QBA to assess the HAR.

The present study is part of the LOEWE research cluster "Animals – Humans – Society: Interdisciplinary Animal Studies" at the University of Kassel, Germany and has been funded by the Hessen State Ministry of Higher Education, Research and the Arts.

7 References

- Adamczyk, K., Pokorska, J., Makulska, J., Earley, B., Mazurek, M., 2013. Genetic analysis and evaluation of behavioural traits in cattle. *Livestock Science* 154, 1–12.
- Bokkers, E.A.M., de Vries, M., Antonissen, I.C.M.A., de Boer, I.J.M., 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Animal Welfare* 21, 307-318.
- Brenninkmeyer, C., Dippel, S., March, S., Brinkmann, J., Winckler, C., Knierim, U., 2007. Reliability of a subjective lameness scoring system for dairy cows. *Animal Welfare* 16, 127-129.
- Breuer, K., Hemsworth, P.H., Barnett, J.L., Matthews, L.R., Coleman, G.J., 2000. Behavioural response to humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science* 66, 273–288.
- Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. *Journal of clinical epidemiology* 46, 423-429.
- Curley, K.O., Paschal, J.C., Welsh, T.H., Randel, R.D., 2006, Technical note: Exit velocity as a measure of cattle temperament is repeatable and associated with serum concentration of cortisol in Brahman bulls. *Journal of Animal Science* 84, 3100-3103.
- Ellingsen, K., Coleman, G.J., Lund, V., Mejdell, C.M., 2014. Using qualitative behaviour assessment to explore the link between stockperson behaviour and dairy calf behaviour. *Applied Animal Behaviour Science* 153, 10-17.
- Grandin, T., 1993. Behavioral agitation during handling of cattle is persistent over time. *Applied Animal Behaviour Science* 36, 1-9.
- Grandin, T., Deesing, M.J., Struthers, J.J., Swinker, A.M., 1995. Cattle with hair whorl patterns above the eyes are more behaviorally agitated during restraint. *Applied Animal Behaviour Science* 46, 117-123.
- Haskell, M.J., Simm, G., Turner, S.P., 2014. Genetic selection for temperament traits in dairy and beef cattle. *Frontiers in genetics* 5, doi: 10.3389/fgene.2014.00368.

- Hemsworth, P.H., 2003. Human-animal interactions in livestock production. *Applied Animal Behaviour Science* 81, 185-198.
- Hoppe, S., Brandt, H.R., König, S., Erhardt, G., Gauly, M., 2010. Temperament traits of beef calves measured under field conditions and their relationships to performance. *Journal of Animal Science* 88, 1982-1989.
- Interbull, 2009. Genetic Evaluation for workability traits, August 2009. http://www.interbull.org/web/static/mace_evaluations_archive/Workability/wo-aug09.html (accessed 26.07.16).
- Ivemeyer, S., Waiblinger, S., Knierim, U., 2011. Effect of Human-Animal Relationship and Management on Udder Health in Swiss Dairy Herds. *Journal of Dairy Science* 94, 5890–5902.
- Lanier, J.L., Grandin, T., 2002. The relationship between Bos taurus feedlot cattle temperament and cannon bone measurements. in: *Proceedings of the Western Section of the American Society of Animal Science* 53, 97-98.
- Lanier, J.L., Grandin, T., Green, R., Avery, D., McGee, K., 2001. A note on hair whorl position and cattle temperament in the auction ring. *Applied Animal Behaviour Science* 73, 93-101.
- Olmos, G., Turner, S.P., 2008. The relationships between temperament during routine handling tasks, weight gain and facial hair whorl position in frequently handled beef cattle. *Applied Animal Behaviour Science* 115, 25-36.
- Randle, H.D., 1998. Facial hair whorl position and temperament in cattle. *Applied Animal Behaviour Science* 56, 139-147.
- Rushen, J., de Passillé, A.M., 2010. The importance of good stockmanship and its benefits for the animals. in: Grandin, T. (Ed.), *Improving animal welfare: a practical approach*. CAB International, Wallingford, 50–63.
- Rushen, J., Taylor, A.A., de Passillé, A.M., 1999. Domestic animals' fear of humans and its effect on their welfare. *Applied Animal Behaviour Science* 65, 285-303.
- Sant'Anna, A.C., da Costa, M.J.P., 2013. Validity and feasibility of qualitative behavior assessment for the evaluation of Nelore cattle temperament. *Livestock Science* 157, 254-262.

- Schutz, M.M., Pajor, E.A., 2001. Genetic Control of Dairy Cattle Behavior. *Journal of Dairy Science* 84 (E. Suppl.), E31–E38.
- Sewalem, A., Miglior, F., Kistemaker, G.J., 2011. Genetic parameters of milking temperament and milking speed in Canadian Holsteins. *Journal of Dairy Science* 94, 512-516.
- Smith, D. W., Gong, B. T., 1974. Scalp - hair patterning: Its origin and significance relative to early brain and upper facial development. *Teratology* 9, 17-34.
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naïve and habituated to road transport. *Animal Production Science* 51, 240-249.
- Stockman, C.A., McGilchrist, P., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Greenwood, P.L., Cafe, L.M., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012. Qualitative Behavioural Assessment of Angus steers during pre-slaughter handling and relationship with temperament and physiological responses. *Applied Animal Behaviour Science* 142, 125-133.
- Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K., Jones, R.B., 2006. Assessing the human-animal relationship in farmed species: A critical review. *Applied Animal Behaviour Science* 101, 185-242.
- Waiblinger, S., Menke, C., Coleman, G., 2002. The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Applied Animal Behaviour Science* 79, 195-219.
- Waiblinger, S., Menke, C., Fölsch, D.W., 2003. Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. *Applied Animal Behaviour Science* 84, 23-39.
- Welfare Quality®, 2009. Welfare Quality® assessment protocol for cattle (online)., <http://www.welfarequality.net/network/45848/7/0/40> (accessed 26.07.2016).
- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science* 67,193-215.

- Wemelsfelder, F., Hunter, T.E., Mendl, M.T., Lawrence, A.B., 2001. Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour* 62, 209-220.
- Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009. Qualitative Behaviour Assessment. in: Forkman, B., Keeling, J., (Eds.), *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*, Welfare Quality® Reports No. 11, 215-224.
- Windschnurer, I., Schmied, C., Boivin, X., Waiblinger, S., 2008. Reliability and inter-test relationship of tests for on-farm assessment of dairy cows' relationship to humans. *Applied Animal Behaviour Science* 114, 37-53.
- Windschnurer, I., Schmied, C., Boivin, X., Waiblinger, S., 2009. Assessment of human-animal relationship in dairy cows. in: Forkman, B., Keeling, J. (Eds.), *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*, Welfare Quality® Reports No. 11, 137- 152.
- VIT (2015): Beschreibung der Zuchtwertschätzung für Milchleistungsmerkmale, Zellzahl, Exterieurmerkmale, Nutzungsdauer und Zuchtleistungsmerkmale [Description of breeding evaluation for milk production traits, somatic cell counts, confirmation traits, longevity, and breeding performance traits]. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_deu.pdf (accessed 26.07.2016).

Table 1: Selected measures of the reactivity towards humans

Measure	Method	Reference
ADfeed	Distance measurement at the feeding place	e.g. Welfare Quality®, 2009; Windschnurer et al., 2008
ADbarn	Distance measurement in the barn	e.g. Ivemeyer et al., 2011 ;Waiblinger et al., 2002, 2003
Chute test ¹⁾	Categorical behaviour assessment on a scale from very quiet to very nervous, when the animal is handled in a squeeze chute	e.g. Curley et al., 2006; Grandin, 1993; Hoppe et al., 2010
Flight-speed / exit behaviour test ¹⁾	Categorical behaviour assessment on a scale from quiet to very fast/jumping, when the animals is released from restraint	e.g. Hoppe et al., 2010; Lanier and Grandin, 2002
QBA	Qualitative behaviour assessment in handling situations	e.g. Ellingsen et al., 2014; Sant'Anna and Paranhos da Costa, 2013
HW	Categorisation of facial hair whorl position and form	e.g. Grandin et al., 1995; Lanier et al., 2001, Olmos and Turner, 2008

¹⁾ Mainly used to assess temperament in beef cattle

ADfeed = avoidance distance at the feeding place, ADbarn = avoidance distance in the barn, QBA = qualitative behaviour assessment, HW = hair whorl

Table 2: Rating scales of the tolerance to tactile interaction (TTI) and release behaviour (RB) tests

Category	TTI	RB
1	Cow stays calm, no stepping* or kicking** with hindlegs	Cow stays calm, leaves the feeding place slowly or only after some time still standing in or in front of the feeding gate
2	Cow lowers the hindquarters at least at the first and second stroke or steps maximum twice, no kicking	Cow leaves the feeding place walking in intermediate speed
3	Cow steps maximum five times, no kicking	Cow leaves the feeding place walking fast
4	Cow steps more than five times or kicks at least once	Cow leaves the feeding place running or jumping
5	Cow is reacting violently, is panicking; touching the cow is (barely) not possible	Cow reacts violently, is panicking; opening the feeding gate (barely) not possible

* lifting the foot maximum 15 cm, ** lifting the foot more than 15 cm

Table 3: Descriptive statistics of different HAR measures, based on data collected by observer A

	n	Median	Min	Max	25 % Percentile	75 % Percentile
ADfeed (cm)	94	5	0	120	0	20
ADbarn (cm)	64	30	0	300	0	100
QBA (PC1 score, explained variance = 78.6 %)	32	-0.239	-1.187	2.073	-0.859	1.059
		Category 1	Category 2	Category 3	Category 4	Category 5
TTI (% of cows)	55	60.0	18.2	14.5	7.3	0.0
RB (% of cows)	54	59.3	18.5	14.8	7.4	0.0
		high	middle	low	no HW	
HW vertical position (number of cows)	90	12	55	18	5	
		middle	next to centre line	no HW		
HW lateral position (number of cows)	90	62	23	5		
		normal	abnormal	no HW		
HW form (number of cows)	90	59	26	5		

ADfeed = avoidance distance at the feeding place, ADbarn = avoidance distance in the barn, QBA = qualitative behaviour assessment (PC1 = first principal component), TTI = tolerance to tactile interaction, RB = release behaviour, HW = facial hair whorl

Table 4: Inter-test associations (Spearman rank correlation coefficients r_s) between HAR measures, i.e. avoidance distance at the feeding place (ADfeed) and in the barn (ADbarn), tolerance to tactile interaction (TTI), release behaviour (RB) and qualitative behaviour assessment (QBA), based on data collected by observer A

		AD barn	TTI	RB	QBA (PC1)
AD feeding place	r_s	0,77**	0,50**	0,45**	0,33
	n	44	44	43	25
AD barn	r_s		0,37	0,38	0,42
	n		24	24	19
TTI	r_s			0,78**	0,68**
	n			52	18
RB	r_s				0,76**
	n				18

* significant on the level 0.05 (two-tailed), ** significant on the level of 0.01 (two-tailed)

Table 5: Tests of effects of HW factors, i.e. vertical position, lateral position, and form, on avoidance distance at the feeding place (AD_{feed}) using a GLM (n = 86), based on data collected by observer A

Factors	df	F	p value
Intercept	1	11.811	0.001
Position vertical	2	0.149	0.862
Position lateral	1	0.079	0.779
Form	1	0.954	0.332
Position vertical * Position lateral	1	0.127	0.723
Position vertical * Form	2	0.421	0.658
Position lateral * Form	1	1.617	0.207
Position vertical * Position lateral * Form	1	0.503	0.480

Figure 1: Component plot of the PCA on QBA data by observer A, based on correlation matrix, without rotation (n = 32, explained variance PC1 = 78,6 %, explained variance PC2 = 6,4 %)

Figure 2: Method used to determine the vertical and lateral position of facial hair whorls (HW) in cows (following Grandin et al. 1995, Lanier et al. 2001)

Figure 3: Inter-observer reliability regarding avoidance distances at the feeding place (AD_{feed}) and in the barn (AD_{barn}), tolerance to tactile interaction (TTI), release behaviour (RB) and qualitative behaviour assessment (QBA). The scatterplots show the correlation of results from observer A and B, and in the case of QBA also from observer A and C; y = slope of the trend line forced through zero, r_s = Spearman rank correlation coefficient