



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

 ScienceDirect

APPLIED ANIMAL  
BEHAVIOUR  
SCIENCE

Applied Animal Behaviour Science 101 (2006) 40–53

[www.elsevier.com/locate/applanim](http://www.elsevier.com/locate/applanim)

# Qualitative assessment of social behaviour of dairy cows housed in loose housing systems

Tine Rousing<sup>a,\*</sup>, Francoise Wemelsfelder<sup>b</sup>

<sup>a</sup> *Danish Institute of Agricultural Sciences, Department of Animal Health, Welfare and Nutrition, Research Centre Foulum, P.O. Box 50, 8830 Tjele, Denmark*

<sup>b</sup> *Sustainable Livestock Systems, Scottish Agricultural College, Bush Estate, Penicuik EH26 0PH, United Kingdom*

Accepted 5 December 2005

Available online 31 January 2006

---

## Abstract

This study evaluates the qualitative assessment of dairy cows' social behaviour on farm with regard to its inter- and intra-observer reliability and its correlation to quantitative ethogram-based assessment. Qualitative behaviour assessment is a method based upon the integration by observers of perceived animal behaviour expression, using descriptors such as 'calm', 'aggressive', 'sociable' or 'indifferent'. Cows' behaviour at the drinker was video recorded in five commercial dairy herds with loose housing systems. Qualitative assessment of 25 video clips showing various types of cows' interaction was provided in two replicate studies by 12 experienced dairy cow observers, through the use of a methodology called free choice profiling (FCP). This method gives the observers complete freedom to choose their own descriptive terms. Furthermore, an ethogram was used to quantify the cows' social behaviour in the same 25 video clips. The ethogram included frequency and duration of social licking, head and body sniffing, pushing, head butting, fighting and behavioural response to pushing or head butting. Data of the qualitative assessment were analysed with generalised procrustes analysis (GPA), a multivariate statistical technique associated with FCP. The correlation between qualitative and the quantitative assessment of the 25 video clips was investigated by calculating Spearman rank correlation between the qualitative assessments and the calculated frequencies and proportional durations of the ethogram measures. The results indicate that observers showed significant agreement in their qualitative assessments ( $P < 0.001$ ) and could accurately repeat these assessments ( $P < 0.001$ ). The GPA found two main dimensions of assessed social behaviour expression in dairy cattle which together explain 74% of the variation observed. Dimension 1

---

\* Corresponding author. Tel.: +45 89 99 13 50; fax: +45 89 99 15 00.  
E-mail address: [tine.rousing@agrsci.dk](mailto:tine.rousing@agrsci.dk) (T. Rousing).

was characterised as ‘relaxed’/‘calm’ versus ‘aggressive’/‘bullying’ and dimension 2 as ‘passive’/‘indifferent’ versus ‘playful’/‘sociable’. The qualitative scores of individual social interactions on these dimensions were correlated significantly to the quantitative measurements of cows showing social licking, head butting and response to pushing or head butting in the respective video clips. Thus, cows showed more social licking in social interactions characterised as ‘relaxed’/‘calm’ ( $r_{\text{frequency}} = 0.68$ ;  $r_{\text{duration}} = 0.68$ ; both  $P < 0.001$ ) and ‘playful’/‘sociable’ ( $r_{\text{frequency}} = -0.58$ ;  $r_{\text{duration}} = -0.59$ ; both  $P < 0.01$ ) while in ‘aggressive’/‘bullying’ social interactions cows showed more head butting ( $r_{\text{frequency}} = -0.55$ ,  $P < 0.01$ ;  $r_{\text{duration}} = -0.62$ ,  $P < 0.001$ ) and response to pushing and head butting ( $r_{\text{frequency}} = -0.42$ ,  $P < 0.05$ ). These results suggest that qualitative behaviour assessment may be a reliable method for the assessment of on-farm social interactions in dairy cows.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Dairy cattle; Welfare; Social behaviour; Qualitative behaviour assessment; Free choice profiling

---

## 1. Introduction

Loose housing systems provide dairy cows with the possibility for locomotion and allow them to express a variety of natural behaviours. Furthermore, a well-established social environment may have a positive effect on the adjustment of individuals to the environment through social facilitation and learning, and it has been suggested that a stable social relationship within a herd may be beneficial in reducing the effect of generally stressful conditions (Bouissou et al., 2001). However, when dairy cows are housed in groups there is also a risk of aggression and social disturbance. Aggressive interactions occur in response to establishing and maintaining social order in dynamic groups. Competition for resources (food, water, resting areas, etc.) as well as inexpedient housing design are important causal factors that may induce and increase social stress and aggressive behaviour. (Bouissou et al., 2001).

Social behaviour is thus an important welfare issue in loose housed dairy cattle herds. Although different aspects of social behaviour have been thoroughly studied, the inclusion of social behaviour in on-farm welfare assessment systems is not yet widespread. Some researchers (e.g. Winckler et al., 2002; Haskell et al., 2003) have included measurements of social behaviour in their on-farm welfare assessment systems. Furthermore in a recent study Plusquelle and Bouissou (2001) used detailed social behaviour measurements to characterize temperament differences in two dairy cow breeds. What these approaches have in common is that they address social behaviour measurement quantitatively, based on the use of ethograms that consist of social behaviour elements with varying incidence. Plusquelle and Bouissou (2001) for example assessed fighting ability and dominance in cows by quantifying the latency, frequency and duration of fights, butts, threats, spontaneous withdrawals, head to rump orientation, mounting, sniffing, licking, rubbing, mock fighting, eating, and social distance in different situations including test situations. The authors summarised these data by pointing out whether one breed showed more or less of the listed behaviours than the other breed, and on this basis characterised the fighting and dominance ability of the two breeds.

Such a retrospective evaluation of quantified behavioural characteristics however may omit potentially important information. It is unlikely to take account of subtle variations in social patterns of behaviour, and is not well equipped to evaluate measures that occur with low incidence or are difficult to quantify. We suggest therefore that the use of social behaviour as an indicator of animal welfare at herd level may benefit from supplementing traditionally retrospective, quantified judgements with a qualitative approach. The qualitative assessment of behaviour consists of a process of integrating measurement and interpretation. This may include integrating subtle details of movement and posture, changes in behaviour over time, as well as aspects of the context in which the behaviour occurs, into a qualitative evaluation of the ‘animal-as-a-whole’. Such assessments describe behaviour as a dynamic, expressive process; e.g. as ‘confident’, ‘nervous’, ‘calm’, or ‘excitable’ (Stevenson-Hinde, 1983; Feaver et al., 1986; Kessler and Turner, 1999; Wemelsfelder et al., 2000, 2001).

Behavioural scientists may question such assessments, fearing it is an anthropomorphic judgements of uncertain validity. In theory however it is possible that assessments of animal behavioural expression are based on observable aspects of behavioural organisation, and are amenable to scientific analysis (Wemelsfelder, 1997, 2001). To test this hypothesis, Wemelsfelder et al. (2001) adapted a free choice profiling (FCP) methodology originally developed in food science (Arnold and Williams, 1985) for use in animal behaviour studies. It is characteristic of this methodology that it gives observers complete freedom to choose their own descriptive terms, allowing observers to interpret their own perceptions without the bias introduced by pre-fixed qualitative rating lists. Using this FCP-based approach to assess the reliability of qualitative assessments of individual pig behaviour, high levels of both inter- and intra-observer reliability were found. Observers generated similar terminologies, and showed good agreement in the way they used these terminologies as measurement frameworks for scoring the pigs’ expressive behaviour style (Wemelsfelder et al., 2000, 2001; Wemelsfelder and Lawrence, 2001). Qualitative assessment has been suggested applied to the response of dairy cows to humans (Haskell et al., 2003), however is has yet not been applied to the assessment of social behaviour in cattle or in any other animal species. Thus, the aim of the present study was to investigate the inter- and intra-observer reliability of FCP-based qualitative assessment when applied to dairy cows’ social behaviour on farm by a group of experienced observers and to investigate the correlation of these assessments to quantitative ethogram measures of the same behaviour.

## 2. Methods

### 2.1. *Animals and housing*

Five private Holstein-Friesian dairy farms located in Scotland and England were visited over a period of 2 months. Each farm housed between 40 and 105 cows. Four farms were loose housing systems consisting of cubicles with concrete and/or slatted floor, while one farm was a loose housing system with deep litter.

## 2.2. Experimental procedures

### 2.2.1. Video recordings

On each of the five farms visited the social interaction of cows around a drinker was recorded with a digital video camera (Sony digital camcorder). The camera was mounted on a pole out of cow reach to ensure undisturbed recording of social activities. Recording took place on three successive days for 2–3 h in the morning, starting approximately 1 h after the morning feeding, and 2–3 h in the afternoon before and after afternoon milking. From this video footage 25 clips of approximately 1 min duration were selected. This selection was designed to be a representative sample of the variation of social interactions observed at the drinkers on the different farms, including agonistic and non-agonistic interactions. The 25 video clips included 25 social events of a total of 66 cows: 14 video clips of two cows, eight video clips of three cows, two video clips of four cows and one video clip of six cows. All farms were equally represented in the video clips. These 25 clips were then edited on to two VHS tapes at a professional studio. Tape 1 showed the clips in varying order of agonistic and non-agonistic events, while Tape 2 showed the same 25 clips in reverse order from Tape 1.

### 2.2.2. Observers

Qualitative assessment of these video tapes was provided by 12 observers, five of whom were researchers of animal science, five were Ph.D. students of animal science, and two were stockmen familiar with daily routines in dairy herds. All observers had practical experience in handling cows and observing cow behaviour. These observers were gathered at the start of the study, and given detailed instructions about free choice profiling experimental procedures (see below). Particular attention was paid to explaining how to apply these procedures to the assessment of social interaction in groups of animals. Observers were instructed not to assess cows individually, but to assess the interaction between cows, and to interpret each cow's behavioural expression in relation to that of other cows. After completion of the instructions observers were divided into two groups, with each group seated in front of a wide screen TV monitor to watch the recorded video tapes.

### 2.2.3. Free choice profiling

The FCP procedure used in this study have been described in detail by [Wemelsfelder et al. \(2001\)](#). Generally FCP consists of two phases; phase 1 allows observers to generate their own individual descriptive terminologies, while in phase 2 observers use these personal terminologies as a quantitative measurement tool. Accordingly, in phase 1 of this study observers generated their own qualitative descriptors while watching the 25 video clips on Tape 1. After each 1 min clip observers had 2 min to write down terms that in their view best summarised the expressive qualities of the cows' social interactions. They were free to choose as few or as many terms as they wanted for each clip, and to repeat the use of terms or select new terms for each clip. To ensure the independence of FCP assessments observers were told not to discuss their terms with others during the course of the experiment. At the end of the video session personal general lists of terms describing the expressive qualities of observed social behaviour were created.

In phase 2 observers were instructed to use their own personal descriptive terminologies as quantitative rating scales. The experimenters had created individual rating forms for each observer by adding a visual analogue scale of 12.5 cm to each observer term. The visual analogue scales ranged from ‘minimum’ (interpreted as a characteristic being absent in the observed social interaction), to ‘maximum’ (interpreted as a characteristic dominating the observed social event). Observers were instructed to again watch the 25 video clips of Tape 1, and during the 2 min period at the end of each clip, to tick every scale at an appropriate point between ‘minimum’ and ‘maximum’. Thus they provided quantitative scores (measured as the distance in millimetres between the minimum point and the observer’s tick), indicating the level of each expressive characteristic for each observed social behaviour event.

A week later this procedure was repeated by showing observers the same 25 video clips in reversed order on Tape 2. Observers were informed that the clips were the same, to avoid speculation and to encourage them to get on with the task at hand.

#### 2.2.4. *Quantitative assessment*

The 25 video clips used in FCP assessment were also analysed quantitatively with the aid of a simple conventional ethogram. This ethogram was based on work by Bouissou et al. (2001) and consisted of the following behaviour elements:

- ‘social licking’, defined as licking another cow’s head, neck and/or shoulder areas;
- ‘sniffing head’, defined as head or muzzle stretched towards/maybe touching another cows head;
- ‘sniffing body’, defined as head or muzzle stretched towards/maybe touching another cows body;
- ‘gentle pushing’, defined as pressing body against body;
- ‘moderate pushing’, defined as a hard push of body against body;
- ‘head butting’, defined as a blow with the forehead directed at another cow;
- ‘fighting’ defined as head to head pushing, sometimes followed by head to neck pushing and manoeuvring for position.

In addition, the response of the ‘receiving’ cow/other cows to these behaviours was recorded; i.e. whether or not cows responded with one of the mentioned behaviour elements and/or subsequently did or did not ‘withdraw’. The latter was defined as moving away from the drinker. The frequency and proportion of time spent engaged in these elements of behaviour was calculated.

### 2.3. *Statistical methodology*

The FCP study generated two sets of 12 observer data matrices that were based on the observers’ individually generated descriptive terminologies, but that all attributed scores to the same 25 social interaction events in dairy cows. To analyse this information for inter- and intra-observer reliability, a multivariate statistical technique that does not rely on fixed variables is required. Generalised procrustes analysis (GPA) is such a technique (Gower, 1975; Arnold and Williams, 1985; Gower and Dijksterhuis, 1994). GPA basically

transforms individual observer data matrices into multidimensional configurations and determines the similarity between these configurations through a process of complex geometric transformation. Thus, it finds a ‘best fit’ of observer scoring patterns, generally referred to as the ‘consensus profile’. The significance of this consensus profile is calculated against a mean profile, obtained by re-running GPA with randomised observer data sets a hundred times. A detailed description of these GPA procedures can be found in [Wemelsfelder et al. \(2000, 2001\)](#). In our study, we generated three consensus profiles; one for each of the two repeated sessions and one for a merged data file of the two sessions.

### *2.3.1. Inter-observer reliability*

Precisely how well individual observer configurations fit the consensus profile is quantified by the procrustes statistic for the consensus profile. In addition, GPA provides a procrustes statistic for each pair of transformed observer configurations, which can be thought of as a measure of the distance between the observers’ configurations relative to the consensus profile. Principle coordinate analysis of these relative distances leads to a so-called observer plot, which maps the relative distance between observer configurations on two to three dimensions. GPA also estimates a centre of distribution of the relative distance between observer configurations together with a standard deviation, and thus determines a 95% confidence region for the consensus profile.

The calculation of the consensus profile takes place independently of the semantic information provided by the terminologies chosen by the observers. Semantic interpretation of the consensus profile takes place after its calculation. Through a principal component analysis (PCA), the number of dimensions of the consensus profile is reduced to one or more dimensions explaining the variation between the expressive social interactions in the video clips. These dimensions are subsequently interpreted by correlating them to the original individual observer data matrices. This step of the analysis produces two-dimensional individual observer interpretative word charts. In each chart, all terms of a particular observer are correlated with the two or more principal axes of the consensus profile. These observer word charts can be used for the interpretation of the main dimensions, in that the higher a term correlates with an axis the more weight it has as a descriptor for that axis. Thus, close comparison of the observer word charts is an important part of investigating the level of agreement between observers.

### *2.3.2. Intra-observer reliability*

As a final step GPA allocates each social event (=video clip) with a score on each of the main dimensions of the consensus profile. The intra-observer reliability of FCP assessments can then be calculated by determining the repeatability of those scores between the two repeat studies. However, in order to be able to correlate the scores of the repeat studies, data require to be calculated as part of the same GPA analysis. Thus, the two data matrices need to be merged and analysed as if they were one ‘merged’ data file. In the ensuing ‘merged’ consensus profile, the 25 social events (=video clips) each receive two scores on the main dimensions, one for each repeat study. These scores can be correlated using a Pearson correlation, with the level and significance of this correlation indicating how well observers were able to repeat their assessment of each individual video clip.

### 2.3.3. The relationship between qualitative and quantitative measures of social behaviour

The relationship between the qualitative FCP assessments and the quantitative, ethogram-based measurements of the 25 video clips was investigated using a Spearman rank correlation. The qualitative scores obtained from the ‘merged’ consensus profile were correlated to the frequencies and proportional durations of the quantitative behavioural categories.

## 3. Results

### 3.1. Inter-observer reliability

The procrustes statistic values of the three consensus profiles, one for each of the two repeated sessions and one for a merged version of the data, are shown in [Table 1](#). These values indicate significant inter-observer reliability, in that the consensus profiles all explain a significantly higher percentage of the variation between observer matrices than the mean of 100 randomised profiles.

The observer plots of each of the two sessions as well as of the merged data analysis are shown in [Fig. 1](#). The majority of observers fall within the 95% confidence region. Three observers in each of the two sessions are shown to be outliers.

### 3.2. Interpretation of the consensus profiles

The interpretative word charts describing the consensus profiles of the two repeat sessions are similar to those of the merged data analysis, and therefore the following presentation of results and discussion will be based on the merged analysis only.

The first dimension of the merged consensus profile explains 64.4% of the variation between the expressive qualities of observed social behaviour in the 25 video clips, while the second dimension explains 9.3% of this variation. The question is how these dimensions are to be interpreted. There is not the space to present all 12 individual observer word charts describing the consensus dimensions, for which reason word charts of observers 2 and 6 are shown as examples in [Fig. 2](#). The axes of these charts reflect the first two main dimensions of the consensus profile, and indicate which of each particular observer’s terms best correlate with those dimensions.

Thus, dimension 1 of the word chart of observer 2 ranges from ‘at ease’, ‘relaxed’, and ‘quiet’ to ‘bossy’, ‘irritated,’ and ‘pestering’, while dimension 2 ranges from ‘bored’ and

Table 1  
Procrustes statistics of the repeated sessions and merged data analysis

Procrustes statistic	Session 1	Session 2	Merged session
Consensus profile	71.4	66.44	64.43
Mean randomised profile $\pm$ S.D.	43.73 $\pm$ 0.23	40.03 $\pm$ 0.26	32.08 $\pm$ 0.08
$t_{99}$	57.7***	51.63***	110.5***

\*\*\*  $P < 0.001$ .

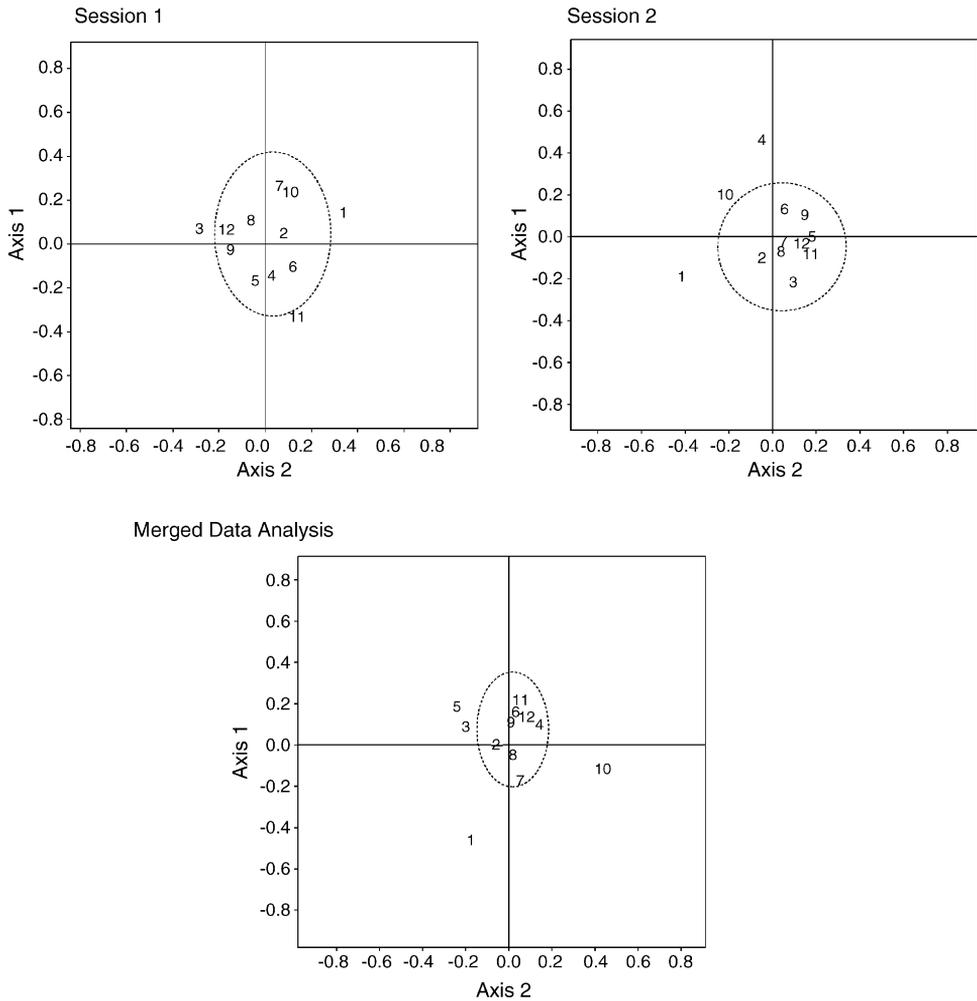


Fig. 1. The observer plots of sessions 1 and 2, and the merged data analysis. the first and second axes reflect GPA scaling values for the relative distance between the 12 observer configurations. Numbers represent observers and the dotted ellipse reflects the statistical 95% confidence region for the consensus profile. Observers positioned outside this region may be seen as outliers.

‘quiet’ to ‘interactive’, ‘matey’, and ‘interested’. Dimension 1 of the word chart of observer 6 ranges from ‘relaxed’, ‘calm’, ‘content’ and ‘peaceful’ to ‘aggressive’, ‘obtuse’ and ‘superior’, while dimension 2 ranges from ‘passive’ and ‘restful’ to ‘bonding’, ‘reaffirming’ and ‘playful’.

To give a more general overview of the observer interpretations than these two examples, [Table 2](#) lists the three terms for each of the 12 observers that correlated most strongly with the two dimensions of the consensus profile. Thus the terms used most frequently to characterise the first dimension of the consensus profile were ‘relaxed’,

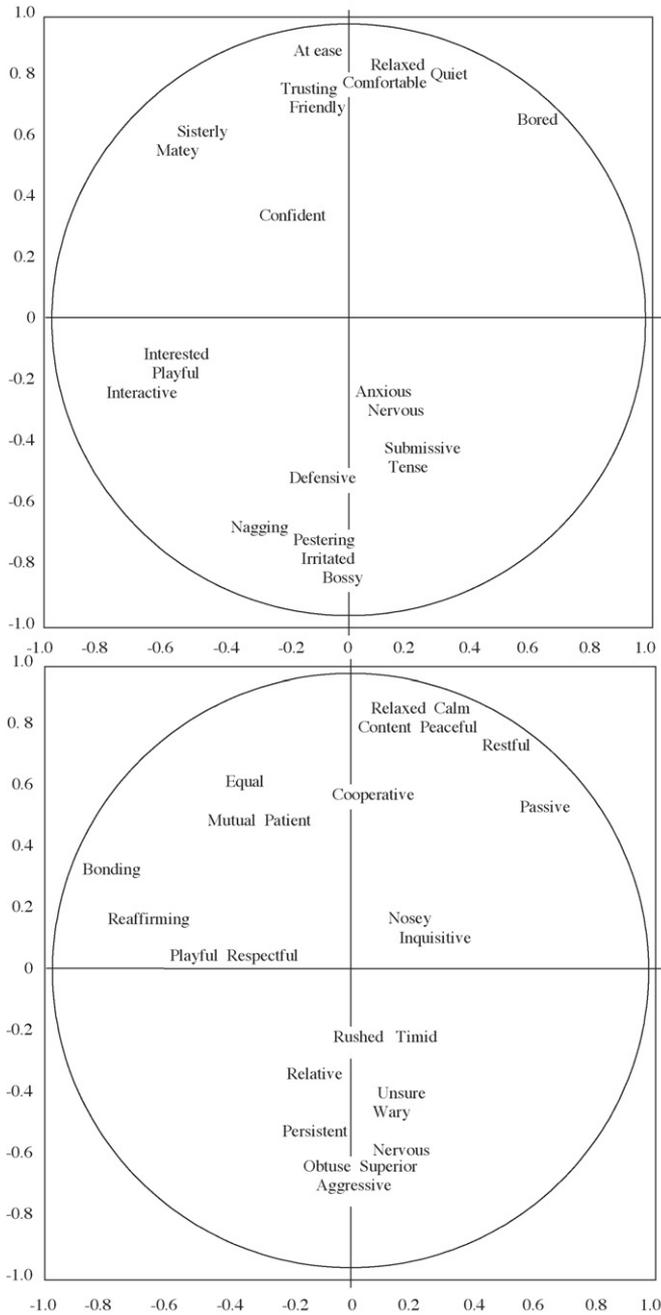


Fig. 2. Word charts of observer 2 (top) and 6 (below). Axes reflect the level of correlation at which observer's terms correlates with the first (vertical) and second (horizontal) dimension of the consensus profile.

Table 2

Terms that show the first, second and third highest positive and negative correlation with the first and second dimension of the consensus profile of the merged analysis

Positive correlation	Negative correlation
First dimension of the consensus profile	
Relaxed (5), calm (5), content (5), quiet (3), at ease/easygoing (3), friendly (2), cooperative (2), sociable (1), giving (1), accepting (1), happy (1), companionable (1), docile (1), comfortable (1), peaceful (1), passive (1), gentle (1), amiable (1)	Aggressive (7), bullying (3), dominant (2), irritated/irritable (2), forceful (2), pushy (2), challenging (2), retreating (1), bossy (1), pestering (1), assertive (1), intimidating (1), defensive (1), stubborn (1), obtuse (1), superior (1), touchy (1), threatening (1), antagonistic (1), focused (1), selfish (1), difficult (1), awkward (1)
Loadings varying from 0.48 to 0.91	Loadings varying from –0.58 to –0.94
Second dimension of the consensus profile	
Passive (5), cautious (4), indifferent (4), bored (3), submissive (2), calm (2), enquiring (1), quiet (1), worried (1), careful (1), curious (1), contented (1), frightened (1), restless (1), nosy (1), weary (1), aloof (1), weak (1), intimidated (1), neutral (1), bland (1), timid (1)	Playful (7), sociable (3), interested (3), giving (1), amused (1), interactive (1), matey (1), affectionate (1), caring (1), helpful (1), confrontational (1), ingratiating (1), cooperative (1), submissive (1), bonding (1), reaffirming (1), pleasure (1), persistent (1), joyful (1), intimate (1), inviting (1), grateful (1), communicative (1), respect-seeking (1), affectionate (1), emotional (1)
Loadings varying from 0.06 to 0.67	Loadings varying from –0.23 to –0.80

Values in brackets give the number of observers using the specific terms.

‘calm’ and ‘content’ versus ‘aggressive’, ‘bullying’ and ‘dominant’. The terms used most frequently to characterise the second dimension of the consensus profile were ‘passive’, ‘cautious’ and ‘indifferent’ versus ‘playful’, ‘sociable’ and ‘interested’. These lists of terms show that 23 of the 144 highest loading terms were used by two or more observers. Where terms differ between observers, the meaning of these terms was either very close (e.g. ‘calm’/‘at ease’/‘relaxed’, or ‘aggressive’/‘bullying’/‘threatening’) or they reflected complementary aspects of the expressive repertoire. For example, ‘friendly’ is not the same as ‘calm’ nor is ‘bored’ the same as ‘passive’, but ‘calm’ cows may well also appear ‘friendly’, just as cows that appear very ‘passive’ may also appear ‘bored’. Thus, the terms describing the two main consensus dimensions converge meaningfully in a semantic tone, and provide a transparent characterisation of the two dimensions of social behavioural expression.

### 3.3. Intra-observer reliability

Fig. 3 reflects the expressive scores of the cows on the two main dimensions of the merged consensus profile. This figure shows that on the whole the scores of the two repeat studies (represented by dots marked with same letter) are located closely together. This closeness is reflected in the high and significant Spearman rank correlations between repeat scores, of 0.96 for the first dimension ( $P < 0.001$ ) and 0.95 for the second dimension ( $P < 0.001$ ). These results indicate that observers repeated their qualitative assessments of the cows’ social activity with considerable accuracy.

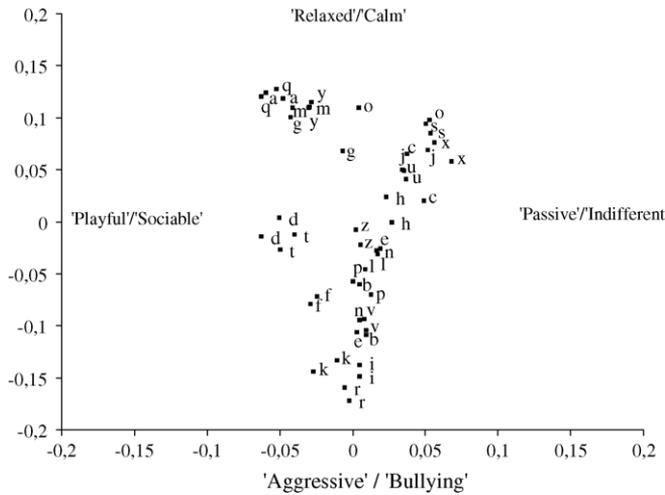


Fig. 3. Score plot of the social events. The first and second axes reflect GPA scaling values for the relative distance between the assessments of the 25 social events in the repeated studies. Dots marked with the same letter indicate that these are the same video clips.

### 3.4. The relationship between qualitative and quantitative behaviour measures of social behaviour

The Spearman rank correlations between qualitative expressive scores of the merged data analysis and quantitative behaviour measures taken from the same video clips are given in Table 3. The first dimension ('relaxed'/'calm' versus 'aggressive'/'bullying') correlated significantly and positively to the quantitative score social licking (frequency and duration;  $r = 0.68$ ,  $P < 0.001$ ), and significantly and negatively to the quantitative scores head butting and response to pushing or head butting (head butting; frequency  $r = -0.55$ ,  $P < 0.001$ ; duration  $r = -0.62$ ,  $P < 0.001$ , as well as response to pushing or head butting; frequency  $r = -0.42$ ,  $P < 0.001$ ). The second dimension ('passive'/'indifferent' versus 'playful'/'sociable') correlated significantly and negatively to the quantitative score social licking (frequency  $r = -0.58$ ,  $P < 0.001$ ; duration  $r = -0.59$ ,  $P < 0.001$ ).

## 4. Discussion

The results reported in this study indicate that the qualitative assessment of cows' social behaviour based on a free choice profiling methodology shows high levels of inter- and intra-observer reliability. Based on the qualitative assessment of social interactions around the drinker, we found two main dimensions of behavioural expression in dairy cattle (dimension 1: 'relaxed'/'calm' versus 'aggressive'/'bullying'; dimension 2: 'passive'/'indifferent' versus 'playful'/'sociable'), which together explain 74% of the variation between assessments of cows' social behaviour. The scores attributed to the cows on the two main dimension were found to correlated meaningfully to the quantitative behaviour

Table 3

Spearman rank correlations between qualitative expressive scores and quantitative measures of social behaviour of cows in the same 25 video clips

Quantitative behaviour measures	Qualitative expressive scores			
	Dimension 1 (relaxed/calm–aggressive/bullying)		Dimension 2 (passive/indifferent–playful/sociable)	
	Frequency	Duration	Frequency	Duration
Social licking	0.68***	0.68***	–0.58**	–0.59**
Sniffing head	0.01	0.23	0.03	0.09
Sniffing body	–0.23	–0.13	0.27	0.36
Gentle push	–0.16	–0.19	0.23	0.23
Moderate push	–0.29	–0.29	–0.10	–0.09
Total push	–0.29	–0.34	0.17	0.17
Head butting	–0.55**	–0.62***	0.15	0.12
Fight	–0.28	–0.28	–0.29	–0.33
Response <sup>a</sup> to pushing or head butting	–0.42*	–	–0.25	–

<sup>a</sup> ‘Receiving’ cow/other cows of the interactions responding with one of the mentioned behaviour elements and/or subsequently did ‘withdraw’ (defined as moving away from the drinker).

\*  $P < 0.05$ .

\*\*  $P < 0.01$ .

\*\*\*  $P < 0.001$ .

measures in as much as cows in interactions perceived as ‘relaxed’/‘calm’ (dimension 1) showed significantly more social licking, while cows in interactions perceived as ‘aggressive’/‘bullying’ (dimension 1) showed significantly more head butting and antagonistic behaviour. With respect to dimension 2 (‘passive’/‘indifferent’ versus ‘playful’/‘sociable’), cows in interactions perceived as ‘playful’/‘sociable’ showed significantly more social licking.

These results suggest that it is possible to reliably provide qualitative summaries of cow social behaviour in terms that are in accordance with other existing investigations of cow behaviour. Plusquelle and Bouissou (2001) for example concluded from quantitative behaviour measurements of fighting and dominance structure that cows were more or less dominant, fearful, aggressive, tolerant or socially motivated. Furthermore, ‘mock fighting’ is often included in ethograms of social behaviour of cattle, not only in calves but also in adult cattle (Plusquelle and Bouissou, 2001; Kabuga et al., 1992). This behaviour is described as a playful intention among adult semi-wild living cattle by e.g. Reinhardt et al. (1986).

The distinguishing characteristic of the present study compared to other studies is that observers were able to directly assess the quality of the cows’ social interaction, rather than infer these characteristics retrospectively from quantitative data as is customary. This is likely to provide a greater level of accuracy, because it takes into account various aspects of behaviour and its context that are difficult to quantify, and integrates the many observed aspects of behaviour while still observing the animal rather than from memory.

Thus, qualitative assessment should be expected to support traditional quantitative assessment, particularly in cases where observed behaviour requires an interpretation in e.g. welfare assessment.

## 5. Conclusion

The results of the present study suggest that qualitative behaviour assessment may be a reliable method for the on-farm assessment of social interactions in dairy cows. A more extended cross-validation of qualitative assessments of cow social behaviour with quantitative welfare indicators would help to further investigate the validity of this approach. The question then arises how it may be possible to operationalise qualitative behaviour assessment and apply it to for example the practice of herd welfare management. One possibility is to combine qualitative and quantitative assessments to identify 'key' indicators that may assist in the ongoing monitoring of social behaviour and its relevance for animal welfare.

## Acknowledgements

The authors would like to thank Marie Haskell for her invaluable help with the selection and visiting of farms, the 12 observers for their time and effort in assessing the videos, Tony Hunter for his expert assistance with GPA analysis, Jan Tind Sørensen for his comments on the experimental design, and Laura Rennie, Verity Bowell, and Marianne Farish for their help with the experimental procedures.

## References

- Arnold, G.M., Williams, A.A., 1985. The use of generalised procrustes techniques in sensory analysis. In: Pigott, J.R. (Ed.), *Statistical Procedures in Food Research*. Elsevier Applied Science, London, pp. 233–253.
- Bouissou, M.-F., Boissy, A., Le Neindre, P., Veissier, I., 2001. The social behaviour of cattle. In: Keeling, L.J., Gonyou, H.W. (Eds.), *Social Behaviour in Farm Animals*. CAB International, Wallingford, Oxon, UK, pp. 113–145.
- Feaver, J., Mendl, M., Bateson, 1986. A method for rating the individual distinctiveness of domestic cats. *Anim. Behav.* 34, 1016–1025.
- Gower, J.C., 1975. Generalized procrustes analysis. *Psychometrika* 40, 35–51.
- Gower, J.C., Dijksterhuis, G.B., 1994. Multivariate analysis of coffee images: a study in the simultaneous display of multivariate quantitative and qualitative variables for several assessors. *Qual. Quan.* 28, 165–184.
- Haskell, M.J., Rennie, L.J., Bowell, V.A., Wemelsfelder, F., Lawrence, A.B., 2003. On-farm assessment of the effect of management and housing type on behaviour and welfare of dairy cattle. *Anim. Welf.* 12, 553–556.
- Kabuga, J.D., Gyawu, P., Asarce, K., Karikari, P.K., Appiah, P., Awunyo, K., Kwarteng, F.A., 1992. Social behaviour at and around oestrus of N'dama cows treated with prostaglandin F<sub>2</sub>α analogue, Cloprostenol. *Appl. Anim. Behav. Sci.* 35, 103–114.
- Kessler, M.R., Turner, D.C., 1999. Effects of density and cage size on stress in domestic cats (*Felis silvestris catus*) housed in animal shelters and boarding catteries. *Anim. Welf.* 8, 259–267.
- Plusquellec, P., Bouissou, M.F., 2001. Behavioural characteristics of two dairy breeds of cows selected (Herens) or not (Brune des Alpes) for fighting and dominance ability. *Appl. Anim. Behav. Sci.* 72, 1–21.
- Reinhardt, C., Reinhardt, A., Reinhardt, V., 1986. Social behaviour and reproduction performance in semi-wild Scottish highland cattle. *Appl. Anim. Behav. Sci.* 15, 125–136.
- Stevenson-Hinde, J., 1983. Individual characteristics: a statement of the problem. In: Hinde, R.A. (Ed.), *Primate Social Relationships: An Integrated Approach*. Blackwell Scientific Publications, Oxford, pp. 28–34.
- Wemelsfelder, F., 1997. Investigation the animal's point of view; an inquiry into a subject-based method of measurement in the field of animal welfare. In: Dol, M., Kasamoentalib, S., Lijmbach, S., Rivas, E., Van den Bos, R. (Eds.), *Animal Consciousness, Animal Ethics*. Van Gorcum, Assen, pp. 73–89.

- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A., 2000. The spontaneous qualitative assessment of behavioural experiences in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Appl. Anim. Behav. Sci.* 67, 139–215.
- Wemelsfelder, F., 2001. The inside and outside aspects of consciousness: complementary approaches to the study of animal emotion. *Anim. Welf.* 10, 129–139.
- Wemelsfelder, F., Lawrence, A.B., 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. *Acta Agric. Scand. Sect. A Anim. Sci. (Suppl. 30)*, 21–25.
- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A., 2001. Assessing the ‘whole-animal’; a free-choice-profiling approach. *Anim. Behav.* 62, 209–220.
- Winckler, C., Bühnemann, A., Seidel, K., 2002. Social behaviour of commercial dairy herds as a parameter for on-farm welfare assessment. In: Koene, P. (Ed.), *Proceedings of the 36th International Society for Applied Ethology*, Wageningen, The Netherlands, p. 86.