Qualitative Assessment of Animal Behaviour as an On-Farm Welfare-Monitoring Tool

Françoise Wemelsfelder
Scottish Agricultural College

Alistair B. Lawrence
Scottish Agricultural College

Recommended Citation

Wemelsfelder, F., & Lawrence, A. B. (2001). Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Acta Agriculturae Scandinavica, Section A-Animal Science, 51(S30), 21-25.

Qualitative Assessment of Animal Behaviour as an On-Farm Welfare-Monitoring Tool

Francoise Wemelsfelder and Alistair B. Lawrence
Scottish Agricultural College

KEYWORDS

animal welfare, on-farm monitoring, free-choice profling, generalised procrustes analysis

ABSTRACT

There is a growing need to monitor the health and welfare of farm animals, and to develop methods which do so efficiently and reliably. A crucial problem with current methods is the integration of separate measurements taken during farm visits into accurate judgements of an animal's overall welfare state. This paper proposes that the qualitative assessment of animal behaviour may serve as an integrative methodology which could guide the interpretation of other, more detailed welfare measurements. Recent research has demonstrated qualitative behaviour assessment to be reliable and repeatable under controlled experimental conditions. The paper concludes with a discussion of ways in which this method may be adapted for practical on-farm use.

Introduction

Consumers are increasingly concerned about the health and welfare of farm animals and request accurate information about the conditions in which animals are kept. There is therefore a need to monitor farm animal health and welfare efficiently and reliably. Different methods of on-farm monitoring have been developed and are being tested in various European countries. Some of these methods focus on physical features of the farm environment and their adequacy for animal management and welfare (e.g. Bartussek, 1998). Others aim primarily to monitor the state of the animal: its physical health (e.g. Alban & Agger, 1996; Busato et al., 1997), its performance of species-specific behaviour and the occurrence of abnormal behaviours (e.g. Veissier et al., 1999). Common to these methods, however, are two needs: first, to collect considerable data during brief farm visits, and second, to integrate these data into balanced health/welfare judgements.

Many different kinds of health and welfare measurement exist, based on physical, physiological and behavioural factors. Current on-farm monitoring methods strive to incorporate many of these, to gain as comprehensive a perspective on welfare as possible. The crucial problem, however, is the integration of the many different measurements into accurate judgements of health and welfare. The requirement is to weigh the importance of different measurements, and in case of contradictory results, to select those that appear most valuable. Essentially we must design a procedure that reliably pieces together the detailed information that we collect, into a larger picture of animal welfare (i.e. an integrated assessment of animal welfare). The various on-farm monitoring methodologies propose different protocols for weighing health

and welfare measurements (e.g. Bracke, 1999; Rousing et al., 1999). These protocols tend to be based on a mixture of research findings, expert knowledge, ethical considerations and common sense, and provide criteria for adding up acquired measurements into a final health/welfare score. However, this procedure leaves considerable leeway for individual variation and bias, and the extent to which amalgamated scores accurately reflect the animal's state should remain a matter of concern.

In this paper we suggest starting-points for a methodology which addresses the problem of integration not just post-hoc, *after* on-farm monitoring has taken place, but as part of the welfare monitoring process itself. The animal, after all, is an integrated being. Animals are not conglomerates of separate physical systems. Through their behaviour, they merge these systems into a dynamic, coherent, adaptive whole. If a methodology were available to assess this integrated level of behaviour, the acquired information could guide the integration and interpretation of other welfare measurements. Conventional reductionist methods may be less suited to this task, because such methods isolate aspects of behaviour for quantification, and break up the dynamic behavioural flow. This may lead to a loss of 'whole-animal' information which cannot be regained at a later stage. We propose therefore that the need exists for an integrative measurement methodology, which, in conjunction with existing reductionist methods, would allow comprehensive welfare-monitoring to take place.

The qualitative assessment of behaviour integrates and summarises the different aspects of an animal's dynamic style of interaction with the environment, using descriptors such as 'calm', 'anxious', 'timid' or 'confident'. Such terms have an expressive connotation which reflects on an animal's experience of a situation, and therefore are directly relevant to evaluations of its welfare (Wemelsfelder, 1997). Animal professionals (producers, veterinarians, handlers) routinely use expressive terms to discuss and manage their animals' health/welfare state, and potentially therefore such assessments provide a suitable integrative on-farm welfare-assessment tool. That qualitative behaviour assessments facilitate the communication and prediction of problems of animal welfare in daily life, suggests that their face- and construct-validity with regard to welfare is high (Streiner, 1993). If such assessments were found to correlate well with conventional measurements of (on-farm) animal welfare, their (construct) validity as biological indicators of welfare would receive further support. However, an obstacle to such research is that behavioural science traditionally distrusts expressive terminologies, fearing these may be based on anthropomorphic guesswork, rather than on reliable assessments of an animal's behavioural state (e.g. Kennedy, 1992). As a first step in using such terminologies as a research tool therefore, it is necessary to test their reliability in a controlled experimental setting, and across a variety of observer groups.

To this purpose we have in the previous few years developed a methodology for the qualitative assessment of animal behaviour (Wemelsfelder et al., 2000, 2001). The aim of this paper is to briefly outline this methodology and our most recent experimental results. In conclusion we will discuss the potential of qualitative behaviour assessment as an on-farm welfare-monitoring tool.

Outline of a methodology for qualitative behaviour assessment

To test the reliability of describing behaviour as an expressive process, we wished to bias observers as little as possible in their judgement of an animal's expression. Rather than give observers a list of pre-fixed terms, we instructed them to create their own descriptive terminologies through an experimental procedure called 'Free-Choice Profiling' (FCP). This method has been developed and tested extensively in the food sciences (Oreskovich et al., 1991), but had not been used in animal science previous to our work. Free-choice conditions, more so than fixed rating scales, are highly conducive to integrative behavior assessment, because the observers themselves have to summarise all perceived details of behaviour in terms which in their view best describe the animals' behavioural expression.

The experimental animals in our most recent experimental work were 10 Large White × Landrace female growing pigs of approximately 15 weeks old, housed together in a straw-bedded pen of 4 × 4 metres. These pigs were trained to be separated from their mates in a test pen, which was identical and directly adjacent to the home pen. At the moment of testing pigs were brought singly into the test pen and given the opportunity to interact with a human person for seven minutes.

Around the test pen, nine observers (academic students who had no previous experience in the observation of pigs) were seated behind wooden screens. These observers were asked to observe each pig for four minutes, and then in the remaining three minutes to write down those terms which in their view best described the pig's behavioural expression. In a second phase, they were instructed to use these self-created terms to quantitatively score the behaviour of the pigs, giving each pig a score on each of their terms. One week after the live sessions had come to an end, observers repeated their assessment from two video-recordings of the live session. One tape showed the pigs in the same order of appearance as the live session, while the other tape showed the pigs (from the same live session) in different order. These tapes were shown to observers on two consecutive days. Thus in total three sets of nine individual observer data-matrices, all describing the same 10 pigs, were obtained.

These data were analysed with Generalised Procrustes Analysis, a multivariate statistical technique associated with FCP (Arnold & Williams, 1985; Oreskovich et al., 1991). This technique transforms individual observer matrices of pig scores into multidimensional configurations, and determines the similarity between these configurations through a process of complex geometric transformation. Thus it finds a 'best-fit' of all observer scoring patterns, generally referred to as the 'consensus profile'. The significance of this consensus profile is calculated against a randomized profile, obtained through re-running GPA with randomised observer data sets (Wakeling et al., 1992).

It should be emphasised that calculation of the consensus profile takes place entirely independently of semantic information provided by observer terminologies. GPA should be thought of as a pattern-detection mechanism, which assumes that even if measurement variables (terms) are not fixed, the distances between samples (pigs) are comparable because the samples are the same. Semantic interpretation of the consensus profile takes place *after* its calculation, by correlating it to the original individual observer data matrices. Thus GPA analysis produces nine individual semantic word-charts (one for each observer), describing the consensus profile in each individual observer's terms. Close comparison of these charts is an important part of investigating observer agreement; in principle a consensus profile, though significant, could show poor convergence in its semantic charts and thus make little sense.

Finally, GPA attributes individual pigs with a score on the main dimensions of the consensus profile. These dimensions can be interpreted with the help of the semantic charts, and the expressive characteristics of individual pigs can thus be identified and compared. In this experiment each individual pig was attributed with three scores, one live- and two video-scores. The correlation between these scores can be used to determine the repeatability of observer assessments. The higher this correlation, the higher the intra-observer reliability of qualitative behavior assessment.

In sum, GPA analysis typically produces three sets of results: (1) 'observer plots', indicating the level of observer agreement for the consensus profile; (2) semantic 'word-charts', reflecting observer interpretations of the consensus profile, and (3) 'animal plots', which give the position of individual pigs on the main dimensions of the consensus profile. For a detailed presentation and discussion of FCP/GPA procedures and their results, the reader is referred to Wemelsfelder et al. (2000, 2001). The present paper offers space only for a brief summary of results.

The reliability and repeatability of qualitative behaviour assessments

The results of our study indicate that qualitative assessments of pig behaviour show high levels of reliability and repeatability. First, the consensus profiles of both the live- and video-sessions were highly significant in comparison to randomised profiles ($p < 0.001$). Thus even though each observer used a different terminology, they showed significant agreement in the way they used their chosen terminologies to record the behavioural expressions of the pigs.

Secondly, the word-charts of the observers showed strong semantic convergence in their characterization of pig behaviour, again in both live and video sessions. Observers distinguished two main dimensions of behavioural expression, which together explained approximately 80% of the variation between pigs. The individual terminologies of observers were mostly highly correlated to these dimensions, describing meaningful, semantically coherent transitions of expression (e.g. con. dent-bold-boisterous-excitable-restless-anxious-timid-calm-relaxed-friendly-confident). Observer word-charts could have been a chaotic mixture of terms; that they were not indicates that observers could use their spontaneously chosen terms systematically, as coherent descriptors of pig behaviour.

Finally, pig scores on the main dimensions of live and video consensus profiles showed highly significant positive correlations (Pearson Correlation, *r*-values ranging from 0.88-0.99, all $p < 0.001$), indicating excellent repeatability. This suggests that the description of behavioural expression is not, as is often assumed, a matter of vague, imprecise guesswork. Observers used their terminologies to locate individual pigs in their expressive charts in a way which could be repeated with great accuracy. Potentially these pig scores can be used to investigate the effect of different housing conditions and management procedures on pig behaviour. It may be hypothesised, for example, that pigs housed in restrictive and barren conditions will obtain significantly lower scores on the timidity-confidence axis than pigs housed under more enriched conditions. We found some evidence of such an effect (Wemelsfelder et al., 2000), but further work is required to confirm this result.

Comparison of assessments across different interest groups

The observers used in the experiment were academic students, a relatively neutral group. The question is whether observers with different professional backgrounds and views would also agree in their assessment of the pigs. We therefore repeated Free-Choice procedures as outlined above with three different observer groups, consisting of 12 pig farmers/stockpeople, 10 large animal veterinarians and 10 animal protectionists. The results of this experiment fully confirm the results reported above. The consensus profiles of the three observer groups were highly significant ($p < 0.001$), both when analysed separately, and in a merged analysis as one large group. Observer word-charts did not diverge significantly from those discussed above, while repeated pig scores again showed highly significant positive correlations ($p < 0.001$). Thus these results suggest that assessments of behaviour in animals are not unduly influenced by personal attitudes, but show a common basis across the various interest groups.

Together the results reported here demonstrate that spontaneous qualitative assessments of pig behavior show high inter- and intra-observer reliability. This indicates that these assessments are based on commonly perceived and systematically applied criteria; the question is on which aspects of behavioural organization these criteria may be based. We suggest that given its integrative nature, qualitative behavior assessment provides access to an integrative, 'whole-animal' level of organisation. On this level we perceive not merely a string of separate behaviours, but we perceive the unity of those behaviours, their focal point of origin, which is the 'behaver'. As an integrated behaver, the animal executes its

behaviours *in a certain manner*, and it is this instrumental relationship that gives behaviour its expressive character (for a more detailed discussion see Wemelsfelder et al., 2001).

From such a point of view of integration, descriptors of behavioural expression (e.g. 'confident', 'anxious') should not be seen as causal states of any kind, physical or mental. They do not describe feelings, genetic predispositions, cognitions, or any factor in isolation, but all of these factors as inclusive aspects of one integrated state. Informally, scientists frequently use 'whole animal' perceptions to guide their interpretation of behavioural and physiological results. Formal recognition of these perceptions could enhance models of animal welfare, and open up new avenues of both fundamental and applied research.

The future development of qualitative behaviour assessment as an on-farm welfare-monitoring tool

The qualitative assessment of behaviour as discussed in this paper took place under controlled experimental experimental conditions. However, given its spontaneous and integrative nature, the proposed Free-Choice methodology seems naturally suitable for on-farm use. Because this method presumes that human observers naturally integrate perceived details of behaviour into qualitative judgements, it is not likely to require extensive training in behavioural observation, nor will it require a detailed scanning of particular aspects of behaviour. Integrative assessment can take place while monitoring of other particular health/welfare aspects is going on; the impressions gained throughout the farm visit can at the end be integrated into qualitative judgements of the animal's overall welfare state. Furthermore integrative behaviour assessment is well suited to farmer participation in welfare assessment schemes. It is the farmer/stockperson who interacts with the animals the most, and who will be in the best position to integrate the expressive details he/she has observed. Integrative assessment could easily fit in with any kind of on-farm work routine, and could increase scrutiny by the farmer of any animal health/welfare problems that may occur.

In conclusion, the next step is to develop FCP procedures for on-farm use, and to test the inter- and intra-observer reliability of the proposed procedures. It will then be possible to cross-validate on-farm assessments of animal behavioural expression with conventional on-farm indicators of welfare, to create a more comprehensive animal welfare monitoring tool.

References

Alban, L. & Agger, J. F. 1996. Welfare in Danish Dairy Herds 1. Disease management routines in 1983 and 1994. Acta Vet. Scand. 37, 49–63.

Arnold, G. M. & Williams, A. A. 1985. The use of generalized Procrustes Techniques in sensory analysis. In: Piggott, J. R. (ed.) Statistical Procedures in Food Research. Elsevier Applied Science, London, pp. 233–253.

Bartussek, H. 1998. Entwicklung und Einsatz des Tiergerechtheitsindex TGI 35L. Tierschutz und Tierhaltung, DVG (Hrsg.), 44-52.

Bracke, M. B. M., Metz. J. H. M. & Spruijt, B. M. 2001. Development of a decision support system to assess animal welfare. Acta Agric. Scand., Sect. A, Animal Sci. Supplementum 30: 17–20.

Busato, A., Steiner, L., Martin, S. W., Shoukri, M. M. & Gaillard, C. 1997. Calf health in cow-calf herds in Switzerland. Prev. Vet. Med. 30 (1), 9–22.

Kennedy, J. S. 1992. The New Anthropomorphism. Cambridge University Press, Cambridge.

Oreskovich, D. C., Klein, B. P. & Sutherland, J. W. 1991. Procrustes Analysis and its applications to Free-Choice and other Sensory Profiling. In: Lawless, H. T. & Klein, B. P. (eds.) Sensory Science: theory and applications in foods. Marcel Dekker Inc, New York, pp. 353–393.

Rousing, T., Bonde, M. & Sorensen, J. T. 2001. How to aggregate welfare indicators into an operational welfare assessment system: a bottom up approach. Acta Agric. Scand., Sect. A, Sci. Supplementum 30: 53–57.

Streiner, D. L. 1993. A checklist for evaluating the usefulness of rating scales. Can. J. Psychiatry 38 (2), 140–148.

Veissier, I., Sarignac, C. & Capdeville, J. 1999. Methods to assess the welfare of domestic animals. Productions Animales 12 (2), 113 –121.

Wakeling, I. N., Raats, M. M. & MacFie, H. J. H. 1992. A comparison of consensus tests for Generalized Procrustes Analysis. J. Sens. Stud. 7, 91–96.

Wemelsfelder, F. 1997. Investigating the animal's point of view; an inquiry into a subject-based method of measurement in the field of animal welfare. In: Dol, M., Kasanmoentalib, S., Lijmbach, S., Rivas, E. & Van den Bos, R. (eds.) Animal Consciousness and Animal Ethics. Van Gorcum, Assen, pp. 73– 89.

Wemelsfelder, F., Hunter, A. T., Mendl, M. T. & Lawrence, A. B. 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative welfare assessment. Appl. Anim. Behav. Sci. 67, 193 –215.

Wemelsfelder, F., Hunter, A. T., Mendl, M. T. & Lawrence, A. B. 2001. Assessing the 'whole animal': a free-choice-profiling approach. Animal Behaviour 61, in press.