animal

Aggregation of measures to produce an overall assessment of animal welfare. Part 2: analysis of constraints

R. Botreau[1,2†], M. B. M. Bracke[3], P. Perny[4], A. Butterworth[5], J. Capdeville[2]
C. G. Van Reenen[3] and I. Veissier[1]

[1]INRA, UR1213 Herbivores, Site de Theix, F-63122 Saint-Genès-Champanelle, France; [2]Institut de l'Elevage, BP18, Castanet Tolosan F-31321, France; [3]Animal Sciences Group, Wageningen University and Research Centre, POB 65, Lelystad NL-8200 AB, The Netherlands; [4]Laboratoire Informatique de Paris 6, Université Pierre et Marie Curie, 8 rue du Capitaine Scott, Paris F-75015, France; [5]University of Bristol Clinical Veterinary Science, Langford, Bristol BS40 5DU, UK

The overall assessment of animal welfare is a multicriterion evaluation problem that needs a constructive strategy to compound information produced by many measures. The construction depends on specific features such as the concept of welfare, the measures used and the way data are collected. Welfare is multidimensional and one dimension probably cannot fully compensate for another one (e.g. good health cannot fully compensate for behavioural deprivation). Welfare measures may vary in precision, relevance and their relative contribution to an overall welfare assessment. The data collected are often expressed on ordinal scales, which limits the use of weighted sums to aggregate them. A sequential aggregation is proposed in the Welfare Quality® project, first from measures to welfare criteria (corresponding to dimensions with pre-set objectives) and then to an overall welfare assessment, using rules determined at each level depending on the nature and number of variables to be considered and the level of compensation to be permitted. Scientific evidence and expert opinion are used to refine the model, and stakeholders' approval of general principles is sought. This approach could potentially be extended to other problems in agriculture such as the overall assessment of the sustainability of production systems.

Introduction

During the last decade, numerous trade groups (producers, processors, retailers and restaurant chains) have developed certification systems with their suppliers. Many of these inspection schemes include a section on animal welfare (e.g. *IKB* (*Integrale KetenBeheersing*) by the Dutch meat industry, *Swedish Broiler Control*, *Filières Qualité Carrefour* in France and McDonald's Europe). However, there is no common standard for assessing animal welfare and for providing consumers with the relevant information.

One objective of the European Welfare Quality® project is to design an overall assessment of animal welfare that will help provide consumers with information on the products they buy (Blokhuis *et al.*, 2003). For this purpose, we need a system that (i) can be used routinely and on many different farms (throughout Europe), (ii) is sensitive to fluctuations in the welfare status of animals on these farms, (iii) reflects the welfare status of the herd as a whole,

(iv) remains transparent for stakeholders (producers, retailers, consumers and citizens) and (v) corresponds to the current state of the art in animal welfare science.

Welfare is a multidimensional concept. Several requirements that have to be met to ensure welfare have been identified (e.g. the classical five freedoms (Farm Animal Welfare Council (FAWC), 1992)). This multidimensionality implies that welfare is most adequately assessed through a number of measures, each linked to a specific welfare dimension (or to several welfare dimensions). In turn, measures must be integrated to support an overall judgement (e.g. Bartussek, 1999; Capdeville and Veissier, 2001; Bracke *et al.*, 2002a). Although science can help assign a relative importance to welfare dimensions, the decision is inherently value based (Fraser, 1995). Nevertheless, science can help define welfare indices, and formalise the judgement made on their relative importance by societal groups.

In Part 1 of the present dissertation (Botreau *et al.*, 2007a), a review of the methods currently proposed to construct an overall welfare assessment from several measures is presented. To date, since each single method

presents advantages and disadvantages, none seems fully adequate to form the basis of a European standard for evaluating welfare in animal units (farms and slaughter plants) and to supply information on animal welfare to end-users. This review highlighted that (i) non-formal aggregation by experts has to be performed by the same expert assessors to generate comparable results, (ii) sum of ranks generates comparisons only within definite populations and (iii) weighted sums may be problematic because they allow full compensation between welfare aspects, which might conflict with the multidimensional nature of animal welfare.

Before constructing a tool for an overall assessment of animal welfare, the specific features linked to animal welfare assessment that constrain the aggregation of welfare measures should be considered. These features may be linked to the concept of welfare *per se*, to the interpretation of measures in terms of welfare, or to data collection. Here, we identify these specific features and suggest solutions to overcome current difficulties. These solutions are currently being investigated in the Welfare Quality® project, which will be briefly described.

Specific features linked to the concept of welfare

Welfare is a multidimensional concept
Welfare is a multidimensional concept. For example, the FAWC (1992) defined five basic requirements for animal welfare: freedom from hunger and thirst, freedom from discomfort, freedom from pain, injury or disease, freedom to express normal behaviour and freedom from fear and distress. Other arrangements of the basic requirements of animals have been proposed, e.g. the classification proposed by Fraser (1993), which consists of freedom from suffering, high levels of biological functioning and the existence of positive experiences. These dimensions are sometimes considered to be more or less independent, i.e. their fulfilment depends on different aspects of the environment that are likely to be experienced differently by animals. For instance, feeling sick is likely to be unrelated to feeling frustrated (for example, due to some behavioural activity being thwarted). The dimensions listed above may be, in turn, subdivided into separate independent items. For instance, feeling sick (nausea) and being injured have distinct causes and consequences for the animal despite being both categorised as elements of 'health'. Hence, many authors have advocated the consideration of more numerous aspects of welfare.

In a literature review, Bracke *et al.* (1999b) formulated 13 basic welfare needs: ingestion (including the need for food and water), rest, social contact, reproduction-related needs (including sexual interaction, nest building, maternal needs), kinesis/movement, exploration (including exploring novelty and foraging), play, body care, evacuation (defecation/urination), thermoregulation, respiration, health (including illness and injury-related needs) and safety (including the perception of danger and social conflict). These needs are considered to correspond to different

biological control systems in which welfare-relevant emotions/feelings such as hunger, feeling ill, feeling pain, fear, etc. play a role. Capdeville and Veissier (2001) defined a similar list of 16 basic needs (no hunger, no thirst, no malnutrition, no physical stress, no climatic stress, no disease, no injury, feeding behaviour, locomotion, resting, social behaviour, sexual behaviour, maternal behaviour, no frightening events, opportunity for avoidance, good contacts with humans). Scientific reports on the welfare of certain animals produced by the European Food Safety Authority generally start with a list of the needs of the animals under consideration. For instance, in calves, 14 needs are identified ranging from needs essential for life (need to breath, need to sleep, etc.) to needs to perform certain behaviours like exploration or mastication of food (Algers *et al.*, 2006).

As with any multidimensional evaluation model, it is important to define a clear set of criteria, i.e. dimensions with set objectives (values to be achieved, or optimised), which can be used to check the welfare of an animal. The set of criteria should be exhaustive (i.e. all important aspects should be considered) and minimal (i.e. only aspects necessary to take a decision are included), and it should be possible to interpret any aspect separately from the other aspects. Additionally, the methods should have the approval of the stakeholders (Bouyssou, 1990). As regards farm animal welfare, these stakeholders may not only be scientists but may also be other people with an interest in animal welfare (consumers or citizens in general, producers, etc.). Based on these considerations, a set of 12 principles was proposed to develop systems for welfare monitoring in the Welfare Quality® project: absence of prolonged hunger, absence of prolonged thirst, comfort around resting, thermal comfort, ease of movement, absence of injuries, absence of disease, absence of pain induced by management procedures, expression of social behaviour, expression of other behaviour, good human–animal relationship, absence of general fear. Each item is assumed to be perceived by the animal differently from the others and (or) to be linked to different causal factors (for discussion see Botreau *et al.*, 2007c).

Animal welfare is defined at the individual level, whereas an overall assessment is usually produced at farm level
Animal welfare is generally defined at individual level. For instance, Broom (1986) defined the welfare of an animal in regards to the attempts this individual makes to cope with its environment. Dawkins (1980) stresses that production parameters, e.g. growth, morbidity or mortality, are not relevant to animal welfare, partly because they are taken at farm level: the production of a farm can be satisfactory even if some animals are in poor conditions.

The question of compensation between individuals has been addressed in animal ethics, with the two major theories – *utilitarianism* v. *deontology* (or *rights theory*) – likely to result in opposite interpretations.

Utilitarianism seeks 'the greatest good for the greatest number' (from Joseph Priestley, read by Bentham, 'father' of utilitarianism, in 1768). Hence, solutions that offer the best balance between the sum of satisfactions and the sum of frustrations are to be preferred. Some philosophers like Singer (1990), extended utilitarianism to animals. Utilitarianism does not imply good welfare for everyone, for while some animals may be living in good welfare conditions, a minority of others may suffer from bad welfare conditions. In this sense, utilitarianism allows compensation between animals.

By contrast, *deontology* stresses that any individual has basic rights (e.g. Feinberg, 1980) and means that, in principle, it is not possible to justify good results obtained using means which violate the rights of any single individual (Regan, 1992). In a deontological approach, compensation between individuals is not seen as ethical.

A pragmatic approach would consider animal suffering to be of prime importance while accepting that a certain percentage of animals will suffer. In an attempt to assess how this can work in practice, we consulted scientists working on animal welfare and asked them to score farms where various proportions of animals in more or less severe conditions related to lameness, injuries and body condition score were presented. In all cases, their answers showed that their judgement was more positive when all animals were in medium conditions than when some animals were in very poor conditions and some others in excellent conditions (e.g. an increase in animals which were not lame never outbalanced an increase of the same extent in animals which were severely lame) (Botreau *et al.*, 2007b).

To make an evaluation at farm level, when welfare can be considered to have its effect on individuals it is proposed not to recommend the use of average values taken at farm level alone, but rather to try to describe also the variation within the farm by, for example, considering the standard variation of a continuous variable or by splitting measures into classes differing in severity.

Welfare dimensions may not fully compensate for each other

As already mentioned in Part 1 of the present dissertation (Botreau *et al.*, 2007a), the various dimensions of animal welfare may not compensate each other.

Utilitarianism and deontology are both aimed at whole populations and can be used to debate whether compensation between individuals should be, from an ethical point of view, permitted or not. However, they can be both translated at the individual level to determine whether compensation between welfare dimensions should or should not be allowed. Transposed at individual level, utilitarianism leads to the conclusion that individuals try to maximise their state of wellbeing, expressed by the surplus of pleasure over distress. Hence, animals should be able to compensate bad situations on some welfare aspects by good situations on other aspects. Aerts *et al.* (2006), who

shared this view, explained this by the capacity of animals to adapt to different situations. In the deontology perspective, each individual must have all his basic needs fulfilled, and the fact that this individual is far above basic requirements for some needs cannot compensate for basic requirements corresponding to other needs being not satisfied.

The most common approach to compensation between welfare dimensions lies between utilitarianism and deontology, allowing some compensation but never full compensation. For instance, Heleski *et al.* (2005) conducted a survey among veterinarians from 27 US veterinary colleges. They showed that 71% of respondents described their attitude toward farm animal welfare as 'we can use animals for the greater human good but have an obligation to provide for the majority of the animals' physiologic and behavioral needs'. Such a statement is clearly somewhere between *utilitarianism* (justifying the use of animals by the fact that this benefits to humans) and *deontology* (recognising that animals have rights). This intermediate view can result in a consideration that the level of animal welfare on a farm is good if a great majority of the animals present on the farm are living in good welfare conditions. At individual level, it could be possible to consider that compensations between welfare dimensions should be limited as they correspond to basic needs. This is illustrated in the first report on calf welfare by the Scientific Veterinary Committee of the European Commission (Broom *et al.*, 1995, p. 97). This report concluded that the welfare of calves was very poor in small individual crates despite the lower incidence of disease. The update of this report stresses the difficulties in balancing the provision of social contacts against the increase in health problems (Algers *et al.*, 2006, p. 61).

To check whether compensation could be allowed between the 12 welfare principles defined in the Welfare Quality® project (see above), we consulted 20 researchers involved in the project, 14 animal scientists and six social scientists. They were considered expert in animal welfare according to their previous researches and their contribution to the project. We asked them to produce welfare scores from various combinations of the fulfilment of the 12 principles. All except one considered that compensations should be highly limited between welfare dimensions (unpublished data).

Of course, the final judgement on whether we can allow compensation or not between welfare dimensions should come from animals themselves. Studies designed to produce an overall assessment of animal welfare have only recently started to appear in the literature. When several factors are likely to affect animals in a similar way (i.e. likely to influence the same variables measured on animals), interactions between factors can be analysed and if there is no interaction, then this suggests that items cannot compensate each other. For instance, in the work by Raussi *et al.* (2003), calves were exposed to different amounts of social and/or human contacts, with the hypothesis that human contacts would be more effective for

calves deprived of social contacts. However, no interactions between these two types of contact were found, and the authors concluded that human contacts cannot compensate for the lack of social contacts with animals of the same kind, and vice versa. To complete this approach, it would be very useful to design experiments where animals are offered several alternatives to check to what extent one alternative can compensate for the lack of another. Operant conditioning might help in this regard: two reinforcements each consisting of various proportions of two rewards could be presented in a concurrent schedule.

Taken together, ethical, societal and animal studies suggest that dimensions of welfare may not compensate for each other (or at least not fully). Therefore, a cautious approach is to use methods that can (but not necessarily) limit compensation between welfare dimensions.

To avoid having to fix whether compensations between welfare dimensions have to be limited or not, the process of aggregation can be stopped at criterion level. For instance, Beyer (1998) considered three different dimensions (housing system, animal care and management of the exercise yard) and stopped the aggregation process at this level. Thus, instead of producing one overall evaluation, this system yields three different scores, one per dimension. To limit compensation, while still providing an overall assessment, Capdeville and Veissier (2001) introduced specific rules to aggregate information so that a good score could not fully balance a bad one. Similarly, Mellor and Reid (1994) defined five component scores (one for each freedom) and suggested setting the overall score at the lowest component score. However, in this case, an improvement in any component other than the worst one will not increase the overall score, which may deter farmers from making improvements. For instance, a pig on slatted flooring that is seriously ill has poor welfare, but if this pig is provided with a comfortable straw-bed, its welfare will improve, at least to some extent, even though the main problem is the illness, not the floor. Another way to limit compensation is to add constraints by defining thresholds below which a value cannot be compensated for (minimum requirements) (Bracke et al., 2002a; Spoolder et al., 2003) or to use a more sophisticated algorithm that assigns higher weightings to lower component scores (e.g. Yager, 1988).

The welfare of an animal is interpreted by humans
Animal welfare should be assessed from the point of view of the animal (Dawkins, 1990). When a single aspect of welfare is considered, the animal's point of view may perhaps be obtained using measures of preferences (e.g. demand curves). However, it is much more difficult, although maybe not totally unrealistic, to determine how an animal would rank very different aspects of welfare occurring at different time scales, for instance, being afraid of something and being sick. The assessment of an animal's welfare as a whole will of course always be to some extent an assessment from a human point of view (see discussion in Fraser, 2003). The list of principles chosen in Welfare

Quality® (see above) received preliminary agreement from focus groups of consumers (49 in all from seven European countries) and from the Advisory Committee of Welfare Quality® composed of representatives of main stakeholder groups (consumers, retailers, producers, animal welfare advocates and policy makers). The Advisory Committee will also be asked to give its views on the aggregation of criteria.

For an overall assessment model to be understood and accepted by all stakeholders (producers, consumers, citizens, scientists, etc.), the model should ideally lead to a consensus about what matters from the animal's point of view. Thus, a major challenge is identifying and reconciling points of difference between a scientific assessment of animal welfare and the opinion of the different stakeholders, in order to be able to communicate a welfare standard. This requires the principles of the methods used to aggregate information to be understandable, and understood, by lay people so that they can give their informed opinion and may impact on the fine-tuning of the assessment model.

Specific features linked to the welfare interpretation of the measures

Welfare is a prolonged mental state, resulting from how the animal experiences its environment over time (Dawkins, 1980; Duncan, 1996; Bracke et al., 1999a). Measures used to assess animal welfare are not direct measures of mental state but only indices that need to be interpreted in terms of welfare. This introduces specific constraints.

The validity of measures in relation to animal welfare needs appraisal
Two types of measures are used to assess animal welfare; those measuring aspects of the animals' environment and those measuring aspects of the animals themselves (also known as design criteria and performance criteria) (Anonymous, 2001). The relationships between environment-based measures and their possible effects on animals are not always straightforward. For instance, a farmer may respond to suboptimal air conditions in a barn with better surveillance and health care. Animal-based measures such as behavioural and health parameters are generally considered to be more closely linked to the welfare of animals (Capdeville and Veissier, 2001; Winckler et al., 2003; Whay et al., 2003c). Ideally, the relation between animal-based measures and actual mental states of animals should be checked. However, the identification of mental states in animals is an emerging area of research (Desire et al., 2002) and these relations are still unknown today. In practice, welfare measures to be used on farms are validated by comparing with more sophisticated methods (concurrent validity), by studying the effects of specific treatments (predictive validity), or alternatively by checking experts

agree on considering these measures as valid (validity based on scientific consensus).

In Welfare Quality®, concurrent validity will be given priority. When no reference method exists, predictive validity will be looked. However, in many cases we will be able to obtain only consensus validity for the time being of the project.

Some measures may be linked to several welfare dimensions

Different measures are needed to assess the different dimensions of welfare, e.g. body condition score may be used as an index of hunger, and the level of abnormal behaviour as an indicator of the animal's inability to express natural behaviour. Some measures may provide information about more than one welfare dimension. For example, a low body condition score can be due to prolonged lack of food (hunger) or a chronic disease (health). Similarly, elevated cortisol levels may be indicative of a wide range of welfare problems. The same problem (of one measure being indicative of multiple welfare dimensions) also arises with environment-based measures. For instance, space affects numerous aspects of welfare, including mobility and social contact. Consequently, this measure can have a greater impact on the overall assessment than a measure linked to only one dimension (Horning, 2001). Unwarranted double counting, however, has to be avoided. This can be limited when the interpretation of the measure in terms of welfare is different from one dimension to another, as in Bracke *et al.* (2002a), where space for exercise and space for social interactions were distinguished.

Measures do not all have the same importance for animal welfare

The assignment of relative weightings to determine the contribution of a measure to overall welfare is always a critical point. There is a general risk of assigning less importance to dimensions that are described by fewer measures. For instance, it is common to assess the comfort of the resting area through several indices (e.g. for cows: difficulties in lying down and in getting up, dog sitting positions, position of the cows in a stall or cubicle, animals lying outside the lying area), while lameness may be assessed with only one measure (e.g. % of lame animals) as in Capdeville and Veissier (2001). Physical comfort should not be given a greater importance than lameness in the overall assessment solely because it requires more measures. Lameness probably affects animals more than discomfort around resting since it is a painful condition. To overcome this problem, the welfare dimensions that need to be covered (i.e. welfare criteria and possibly subcriteria) must be clearly identified and assigned a relative importance. Only then, for each dimension covered by several measures, should the relative contribution of the different measures to that dimension be identified.

Again, the point of view of animal should be taken into account when this is available. However, most experiments designed to assess the preferences of animals offer several alternatives of the same nature: several foods or several types of housing, etc. (e.g. Klopper *et al.*, 1981; Barber *et al.*, 2004). It would be of value to consider which aspect of welfare the animal prioritises as more important, e.g. health, comfort around resting or the possibility to express behaviour, but experiments providing such information have not yet been carried out.

Expert opinion can be used (i) when no study has been yet run to address a specific point (but related studies can help form an opinion of what is most probable to be) and/or (ii) when scientific evidence alone cannot solve a problem (Roqueplo, 1997). Point (i) is highly relevant for animal welfare due to lack of information on the importance that animals may attribute to the various welfare aspects, but the consequences of the non-fulfilment of each aspect can be assessed thanks to indirect parameters such as mortality, morbidity, expression of behaviour (Algers *et al.*, 2006). Point (ii) is also very relevant for animal welfare. As described by Fraser (1995), animal welfare cannot be addressed in an entirely objective way and the importance attributed to the various dimensions of welfare is inevitably value based. As a consequence, the weighting of various welfare aspects can be based on expert opinion.

Butterworth *et al.* (2004) and Haslam and Kestin (2003) used conjoint analyses to define weightings: a limited number of measures were selected with pre-set possible levels, and permutations between possible values of measures were then presented to experts who were asked to give their opinion. Weightings were then extracted assuming a linear combination of measures. This method is only possible with a limited number of measures and with pre-set values for these measures (e.g. Yes/No or few ordinal levels). When these conditions are not met, experts can be consulted using the Delphi method as in Anonymous (2001) or Whay *et al.* (2003b), where the arguments and opinions of experts are taken into account in several proposals following an iterative procedure. Alternatively, weighting coefficients can be derived from a classification of scientific evidence (Bracke *et al.*, 2002a) and further validated by comparison with expert opinion (Bracke *et al.*, 2002b). In general, it is not recommended to ask experts directly to assign weightings to measures because weightings depend on the method used to aggregate the measures. Experts can be asked to give their opinion on situations or data sets, and weightings can then be calculated to match their answers. This is the strategy followed in Welfare Quality®.

The choice of experts can have a large impact on the results. For instance, veterinarians might attribute more importance to health factors, where ethologists may attribute more importance to the expression of normal behaviour. Hence, each time experts are consulted, one has to ensure that experts are chosen according to their knowledge of the field and that various points of views are balanced within the group of experts (Roqueplo, 1997).

Relations may exist between measures

Besides the characteristics of the data, any aggregation process should take into account the links between measures (Bouyssou, 1990).

Welfare measures may be affected by external factors which do not directly affect welfare *per se*, and such measures may need to be corrected for a meaningful welfare interpretation to be made. For instance, in dairy cows, body condition scores may need to be corrected for the stage of lactation. Similarly, when assessing pain, lameness scores may need to be corrected for the walking requirements imposed by the housing system. For example, lameness may have more severe consequences for cows at pasture than when they are in tie-stalls, because of the different walking requirements, e.g. to get food and water.

Measures used to assess welfare may also affect one another. For example, cows that suffer from lameness have shorter avoidance distances, a measure that is often taken to assess cows' fear of humans (Špinka *et al.*, 2005). A solution to avoid misinterpretation could be to exclude lame animals when assessing fear responses.

When constructing a multicriterion evaluation, such relationships between measures must be identified in order to avoid double counting (Bouyssou, 1990). When links between measures are unavoidable then those measures should be considered jointly (i.e. in the same criterion).

Specific features linked to the collection of data

Data can be collected on different types of scale depending on the measures

Measures to assess animal welfare are generally expressed with units on several types of scale (Scott *et al.*, 2001). Some measures may be expressed on cardinal (i.e. quantitative) scales, as for example the frequency of aggressive behaviour during a defined period, or the flight distance of an animal when a human approaches. Other measures, however, are expressed on ordinal scales, i.e. observations are assigned to ordered categories. For instance, we may consider that an animal has mild, moderate or strong reactions to handling. 'Strong' is greater than 'moderate' and 'moderate' is greater than 'mild'. For ordinal scales, average scores may be misleading. In the above example, if all the animals respond moderately the same 'average' is obtained as if one half of the group does not respond and the other half responds strongly. But, since the difference between 'mild' and 'moderate' may be smaller (or larger) than the difference between 'moderate' and 'strong', the first group may actually be on average less (or more) responsive than the second.

Furthermore, the relation between a measure and its interpretation in terms of animal welfare is often not proportional. For instance, two animals that flee when the experimenter is 8 and 6 m away will be considered similar in terms of fear of humans. By contrast, an animal that flees when the experimenter is 2 m away will be considered more frightened by humans than an animal that accepts being touched (i.e. flight distance of 0 m). Yet, in both cases the difference in flight distance is actually the same (2 m).

In some cases, optimal values may be neither minimum nor maximum values, but somewhere in between. For instance, for gregarious species, both social isolation and being in a very large group may be detrimental to animal welfare: being in a small group, which corresponds to what is observed in natural conditions may be most appropriate. In poultry, feather pecking is much more frequent in large groups (>100 hens) than in small groups of 15 to 60 birds (Bilcik and Keeling, 2000).

Hence, when raw data are converted onto a value scale, from 'poor welfare' to 'good welfare', the conversion may not necessarily be a linear function. Again this function may be derived from expert opinion.

When a welfare assessment is performed, raw data are often converted into scores on a value scale composed of discrete values. The scores are then often processed as cardinal data (interval or ratio), whereas in some cases it would be more appropriate to consider them as ordinal data and avoid calculating means or sums.

Measures differ in precision

Measures used to assess welfare may range in precision. For example, there can be variations between observers and between days of observation. Environment-based measures often tend to be more reliable than animal-based measures. For instance, it is easy to measure the length of a stall, and it is unlikely that this will vary greatly between observers and from one day to another. By contrast, measures taken on animals tend to be subject to variation. For instance, when estimating the reaction of an animal facing a human on a predefined ordinal scale (e.g. no/mild/moderate/strong/very strong reaction), the observer may hesitate between two successive levels of the scale. In addition, the behaviour of an animal is never exactly the same between repetitions of the test and so it is essential to determine the reliability of each welfare measure. An aggregation model should take into account variation in reliability between measures. This can be done, for example, by setting discrimination thresholds, below which differences between values are not considered significant (Perny, 1998 or Bouyssou *et al.*, 2000, p. 180), or by the use of fuzzy logic (Lacroix *et al.*, 1998).

In Welfare Quality®, the method used to construct criteria varies depending on the precision of measures, and, in the case of measures with low precision, it is the intention to make as many measures as is practically possible to derive a criterion score.

Data are sometimes missing

A problem inherent to measures taken under practical conditions, as opposed those taken in experimental conditions, is the difficulty in recording them, and this can result in missing data. For instance, health records may not be kept accurately, so that disease prevalences cannot be

assessed properly. Missing values may be handled by substituting other related data, however, in some cases, it may prove impossible to carry out the assessment because critical data are missing.

It is sometimes difficult to assess the range of variation of a measure within a population

Another problem for overall welfare assessment arises when the range of variation of a measure is not known. Observations in experimental conditions are usually not sufficient to obtain such information and it may be necessary to run observations on a large scale on farms (or during transport, or at slaughter). These observations are very demanding and are rarely carried out to identify the distribution of all measures. In Welfare Quality®, surveys in a range of European countries are planned and it is hoped this will result in the required information on minimum, maximum, mean and standard error for cardinal measures, and information on percentiles for qualitative and ordinal measures. This information is needed to adjust the aggregation model to the characteristics of the measures, so that the overall assessment remains sensitive (i.e. farms that apparently differ do not obtain the same final assessment).

Discussion

An ideal model for an overall assessment of animal welfare should deal adequately with all the requirements linked to welfare assessment. The most challenging requirement is that welfare is a multidimensional concept and dimensions may not fully compensate for one another. Methods used to synthesise information can be more or less compensatory, and the choice of a method shall depend on the level of compensation one wants to allow. In addition, welfare measures vary in precision, relevance and importance, and the type of data collected have prompted us to search for calculations other than simple (and intuitively appealing) weighted sums.

Assessing welfare involves describing how well the animals experience their world based on the best possible judgement of their situation. This judgement requires detailed knowledge of the available scientific information. Such information is necessary to avoid errors in interpreting a given measure (e.g. wallowing in pigs could be interpreted solely as a sign of contentment, whereas this behaviour is often displayed when the animal is overheated (Baldwin and Ingram, 1967)). But this judgement cannot solely be based on science and on the data collected from experiments. As mentioned by Fraser (1995), the assignment of a relative importance to welfare dimensions is at least partly subjective. Researchers need to be confident that these dimensions and their relative importance match the expectations of societal groups involved in the keeping, selling or protection of farm animals. In addition, science cannot tell us what is socially acceptable or not – and so threshold values are usually set according to expert opinion (e.g. disease prevalence values above which welfare is

considered to be poor and where remedial measures are required at herd level can be set from veterinary advice) (Whay *et al.*, 2003a). Finally, ethical, economical and political issues may also come into play (Commission for the European Communities, 2002).

The great variability of evaluation problems encountered in practice has led scientists with different backgrounds (management sciences, mathematical psychology, economics, operations research and computer sciences) to develop a variety of formal models and methodologies in decision theory to support evaluation tasks and decision making activities, e.g. the outranking approach based on ordinal aggregation methods, multiattribute utility theory (MAUT; e.g. Roy, 1996) based on cardinal aggregation methods (additive or non-additive, e.g. Choquet integrals (Grabisch, 1996) and generalised additive independence utility (Gonzales and Perny, 2005), etc.). These models and methodologies concern different important issues relevant to multidimensional evaluation such as measurement problems (preferences, perceptions and performance by numerical information), aggregation problems (overall evaluations from multidimensional and possibly conflicting viewpoints), uncertainty and imprecision modelling (Vincke, 1992; Roy, 1996; Bouyssou *et al.*, 2000).

The objectives of Welfare Quality® are (i) to construct a standardised overall welfare assessment for cattle, pigs and poultry for use at a European scale and (ii) to establish a standardised Europe-wide communication system for information on products (Blokhuis *et al.*, 2003). The overall assessment produced by Welfare Quality® could be included in an information system (assigning animal units to a few predefined categories, from low welfare to high welfare), on a voluntary basis. To have a complete evaluation of the welfare level of animals throughout their lives, farms, hauliers and slaughterhouses will have to be assessed.

To make an overall assessment of animal welfare, it is planned to follow a hierarchical aggregation process, going first from the measures (performed in the field, e.g. on farms) to the 12 welfare principles identified in Welfare Quality® (Botreau *et al.*, 2007b). These principles are absence of prolonged hunger, absence of prolonged thirst, comfort around resting, thermal comfort, ease of movement, absence of injuries, absence of disease, absence of pain induced by management procedures, expression of social behaviour, expression of other behaviour, good human–animal relationship and absence of general fear. To facilitate the communication with consumers, these 12 elements have been grouped into four main criteria: feeding, housing, health and behaviour (and called the 12 welfare basic elements 'subcriteria'). These four criteria will subsequently be aggregated to form an overall assessment (Table 1).

The subcriteria will be constructed using mathematical methods (e.g. weighted sums, comparison with minimal requirements) chosen according to the number of measures contained in each subcriterion, their nature and the precision with which it is assumed they can be made.

Table 1 *Hierarchical structure designed in Welfare Quality$^{®}$ to integrate numerous measures ($\sim$ 30) into an overall welfare assessment*

Measures (examples)	Principles	Criteria	Overall assessment

	Compensations		
Body condition score	Absence of prolonged hunger	Good feeding	
Quantity of water provision	Absence of prolonged thirst		
Cleanliness, abnormal rising	Comfort around resting		
Panting	Thermal comfort	Good housing	
Tethering, slipping	Ease of movement		
Injuries, lameness	Absence of injuries		
Mastitis, diarrhoea	Absence of disease	Good health	Overall assessment
Dehorning, tail docking	Absence of pain induced by management procedures		
Agonistic behaviours	Expression of social behaviours		
Stereotypical behaviours	Expression of other behaviours	Appropriate behaviour	
Avoidance distance	Good human-animal relationship		
Reaction facing a novel situation	Absence of general fear		

First, 12 principles will be checked thanks to a combination of relevant measures. Second, the information will be compounded into four criteria and finally aggregated to form one overall assessment. Different mathematical methods will be used to process the information allowing decreasing the level of compensation along the hierarchical structure.

The appropriate subcriteria will be combined to evaluate each criterion. The method chosen to aggregate the sub-criteria into criteria will limit compensations when this appears to better match evaluation given by experts, by assigning more importance to the lowest subcriterion-scores, thus hopefully encouraging producers to correct the more severe problems first. The aggregation of criteria, to create an overall assessment, will be performed using comparisons with pre-set profiles in order to be able to limit further compensations. Hence, the higher the aggregation is in the hierarchical structure the more limited may be the compensation between components.

Stakeholders are involved during the construction of the assessment method to enhance its potential for further implementation in practice. The information resulting from this evaluation will need to be expressed in a compounded way to inform consumers. Finally, the method will have to remain flexible enough to follow the evolution of farming, transport and slaughter conditions, and that of societal concerns, and, last but not least, to be updated according to the state of the art in animal welfare science and societal expectations.

An ideal method for the overall assessment of animal welfare should satisfy all the specific features and recommendations presented here. Even though this may not prove totally feasible, we intend to produce a 'best possible' method for the overall assessment of animal welfare based on scientific evidence, expert opinion, and stakeholders' views. This work is in progress and the final model of overall welfare assessment will be described in forthcoming papers.

In conclusion, the routine overall assessment of animal welfare needs a formal model of multicriterion evaluation. The construction of such a model requires bridging animal sciences, social sciences and methodologies developed in decision theory. The design of the strategy outlined in Welfare Quality$^{®}$ may also be applicable to a wide range of similar problems found in animal production such as defining an overall model for the assessment of sustainability of farming systems.

Acknowledgements

References

Aerts S, Lips D, Spencer S, Decuypere E and Tavernier J 2006. A new framework for the assessment of animal welfare: integrating existing knowledge from a practical ethics perspective. Journal of Agricultural and Environmental Ethics 19, 67–76.

Algers B, Broom D, Canali E, Hartung J, Smulders F, van Rennen CG and Veissier I 2006. Scientific opinion on the risk of poor welfare in intensive calf farming systems: an update of the Scientific Veterinary Committee Report on the Welfare of Calves. The EFSA Journal 366, 1–144.

Anonymous 2001. Scientists' assessment of the impact of housing and management on animal welfare. Journal of Applied Animal Welfare Science 4, 3–52.

Baldwin BA and Ingram DL 1967. Behavioural thermoregulation in pigs. Physiology and Behavior 2, 15–16.

Barber CL, Prescott NB, Wathes CM, Le Sueur C and Perry G 2004. Preferences of growing ducklings and turkey poults for illuminance. Animal Welfare 13, 211–224.

Bartussek H 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. Livestock Production Science 61, 179–192.

Beyer S 1998. Konstruktion und Überprüfung eines Bewerungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit. Wissenschftlicher Fachverlag, Giessen.

Bilcik B and Keeling LJ 2000. Relationship between feather pecking and ground pecking in laying hens and the effect of group size. Applied Animal Behaviour Science 68, 55–66.

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. Animal Welfare 12, 445–455.

Botreau R, Bonde M, Butterworth A, Perny P, Bracke MBM, Capdeville J and Veissier I 2007a. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods. Animal 1, 1179–1187.

Botreau R, Perny P, Capdeville J and Veissier I 2007b. Construction of product information from animal welfare assessment. Proceedings of the Second Welfare Quality® Stakeholder Conference, pp. 33–36.

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007c. Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, 225–228.

Bouyssou D 1990. Building criteria: a prerequisite for MCDA. In Readings in multiple criteria decision-aid (ed. CA Bana e Costa), pp. 58–80. Springer Verlag, Heidelberg.

Bouyssou D, Marchant T, Pirlot M, Perny P, Tsoukias A and Vincke P 2000. Evaluation and decision models – a critical perspective. Kluwer Academic Publishers, Dordrecht.

Bracke MBM, Spruijt BM and Metz JHM 1999a. Overall animal welfare assessment reviewed. Part 1: is it possible? Netherlands Journal of Agricultural Science 47, 273–291.

Bracke MBM, Spruijt BM and Metz JHM 1999b. Overall animal welfare reviewed. Part 3: welfare assessment based on needs and supported by expert opinion. Netherlands Journal of Agricultural Science 47, 307–322.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a. Decision support system for overall welfare assessment in pregnant sows A: model structure and weighting procedure. Journal of Animal Science 80, 1819–1834.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002b. Decision support system for overall welfare assessment in pregnant sows B: validation by expert opinion. Journal of Animal Science 80, 1835–1845.

Broom DM 1986. Indicators of poor welfare. The British Veterinary Journal 142, 524–526.

Broom DM, Blokhuis WJ, Canali E, Dijkhuizen AA, Fallom R, Le Neindre P, Saloniemi H and Webster AJF 1995. Report of the scientific veterinary committee, Animal welfare section on the welfare of calves. Commission of the European Communities.

Butterworth A, Sadler L, Knowles TG and Kestin SC 2004. Evaluating possible indicators of insensibility and death in Cetacea. Animal Welfare 13, 13–17.

Capdeville J and Veissier I 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 62–68.

Commission for the European Communities 2002. Communication from the commission to the Council and the European Parliament on animal welfare legislation on farmed animals in third countries and the implications for EU.

Dawkins MS 1980. Animal suffering: the science of animal welfare. Chapman and Hall Ltd, London.

Dawkins MS 1990. From an animal's point of view: motivation, fitness, and animal welfare. Psychological Science 13, 1–61.

Desire L, Boissy A and Veissier I 2002. Emotions in farm animals: a new approach to animal welfare in applied ethology. Behavioural Processes 602, 165–180.

Duncan IJH 1996. Animal welfare defined in terms of feelings. Acta Agriculturae Scandinavica, Section A, Animal Science, Supplementum 27, 29–35.

Farm Animal Welfare Council 1992. FAWC updates the five freedoms. The Veterinary Record 17, 357.

Feinberg J 1980. Human duties and animal rights. In Rights, justice and the bounds of liberty (ed. J Feinberg), pp. 185–206. Princeton University Press, Princeton.

Fraser D 1993. Assessing animal well-being: common sense, uncommon science. Proceedings of the food animal well-being-conference proceedings and deliberations (ed. USDA and Purdue University Office of Agricultural Research Programs), pp. 37–54. Purdue University Office of Agricultural Research Programs, West Lafayette, Indiana.

Fraser D 1995. Science, values and animal welfare: exploring the 'inextricable connection'. Animal Welfare 4, 103–117.

Fraser D 2003. Assessing animal welfare at the farm and group level: the interplay of science and values. Animal Welfare 12, 433–443.

Gonzales C, Perny P 2005. GAI networks for decision making under certainty. Proceedings of the 19th international joint conference on artificial intelligence – multidisciplinary workshop on advances in preference handling (ed. R Brafman and U Junker), pp. 100–105.

Grabisch M 1996. The application of fuzzy integrals in multicriteria decision making. European Journal of Operational Research 89, 445–456.

Haslam SM and Kestin SC 2003. use of conjoint analysis to weight welfare assessment measures for broiler chickens in UK husbandry systems. Animal Welfare 12, 669–675.

Heleski CR, Mertig AG and Zanella AJ 2005. Results of a National Survey of US Veterinary College Faculty regarding attitudes toward farm animal welfare. Journal of the American Veterinary Medical Association 226, 1538–1546.

Horning B 2001. The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 42–47.

Klopper FD, Kilgour R and Matthews LR 1981. Paired comparison analysis of palatabilities of twenty foods to dairy cows. Proceedings of the New Zealand Society of Animal Production 41, 242–247.

Lacroix R, Strasser M, Kok R and Wade KM 1998. Performance analysis of a fuzzy decision-support system for culling of dairy cows. Canadian Agricultural Engineering 40, 139–152.

Mellor DJ and Reid CSW 1994. Concepts of animal well-being and predicting the impact of procedures on experimental animals. In Proceedings of the improving the well-being of animals in the research environment (ed. RM Baker, G Jenkin and DJ Mellor), pp. 3–18. ANZCCART, Sydney.

Perny P 1998. Multicriteria filtering methods based on concordance and non-discordance principles. Annals of Operations Research 80, 137–165.

Raussi S, Lensink BJ, Boissy A, Pyykkönen M and Veissier I 2003. The impact of social and human contacts on calves' behaviour and stress responses. Animal Welfare 12, 191–203.

Regan T 1992. Pour les droits des animaux. Les Cahiers Antispécistes 5.

Roqueplo P 1997. Entre savoir et décision, l'expertise scientifique. INRA editions, Paris, France.

Roy B 1996. Multicriteria methodology for decision aiding. Kluwer Academic, Dordrecht.

Scott EM, Nolan AM and Fitzpatrick JL 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 5–10.

Singer P 1990. The significance of animal suffering. Psychological Science 13, 9–12.

Špinka M, Dembele I, Panamá J and Stìhulová I 2005. Lame dairy cows have shorter avoidance distances. Proceedings of the 39th international congress of the International Society for Applied Ethology, Sagamihara, Japan.

Spoolder H, De Rosa G, Horning B, Waiblinger S and Wemelsfelder F 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529–534.

Vincke P 1992. Multicriteria decision aid. Wiley, New York, USA.

Whay HR, Main DCJ, Green LE and Webster AJF 2003a. assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records. The Veterinary Record 153, 197–202.

Whay HR, Main DCJ, Green LE and Webster AJF 2003b. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. Animal Welfare 12, 205–217.

Whay HR, Main DCJ, Green LE and Webster AJF 2003c. An animal-based welfare assessment of group-housed calves on UK dairy farms. Animal Welfare 12, 611–617.

Winckler C, Capdeville J, Gebresenbet G, Horning B, Roiha U, Tosi M and Waiblinger S 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Animal Welfare 12, 619–624.

Yager RR 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190.