

Highlighting ethical decisions underlying the scoring of animal welfare in the Welfare Quality[®] scheme

I Veissier^{*†}, KK Jensen[‡], R Botreau[†] and P Sandøe[‡]

[†] INRA, UR1213 Herbivores, F-63122 Saint-Genès-Champagnelle, France

[‡] Department of Food and Resource Economics, Faculty of Life Sciences, University of Copenhagen, Denmark

* Contact for correspondence and requests for reprints: isabelle.veissier@clermont.inra.fr

Abstract

All systems of scoring animal units (groups, farms, slaughter plants, etc) according to the level of the animals' welfare are based inevitably on normative decisions. Similarly, all methods of labelling, in terms of acceptability, are based on choices reflecting ethical values. The evaluative dimension of scoring and labelling does not mean that we should reject them, but it does mean that we need to make the normative and ethical background explicit. The Welfare Quality[®] scoring system is used as a case study in order to highlight the role of underlying value-based decisions. In this scoring system, which was designed in accordance with assessments and judgments from experts in animal and social sciences and stakeholders, we identify value-based decisions at the following five levels. First, there are several definitions of animal welfare (eg hedonist, perfectionist, and preferentialist), and any welfare scoring system will reflect a focus upon one or other definition. In Welfare Quality[®], 12 welfare criteria were defined, and the entire list of criteria was intended to cover relevant definitions of animal welfare. Second, two dimensions can structure an overall evaluation of animal welfare: the individual animals and the welfare criteria (here 12). Hence, a choice needs to be made between the aggregation of information at the individual level (which results in a proportion of animals from the unit in a good vs bad state) and the aggregation at criterion level (which results in a proportion of criteria to which the unit complies vs does not comply). Welfare Quality[®] opted for the second alternative to facilitate the provision of advice to farmers on solving the welfare problems associated with their farms. Third, one has to decide whether the overall welfare assessment should reflect the average state of the animals or give priority to worse-off animals. In the Welfare Quality[®] scoring system the worse-off animals are treated as much more important than the others, but all welfare problems, major or minor, count. Fourth, one has to decide whether good scores on certain criteria can compensate for bad scores on others. In the opinion of most people, welfare scores do not compensate each other. This was taken into account in the Welfare Quality[®] scoring system by using a specific operator instead of mere weighted sums. Finally, a scoring system may either reflect societal demands for high levels of welfare or be based on what can be achieved in practice — in other words, an absolute assessment or a relative one may be proposed. Welfare Quality[®] adopted an intermediate strategy: absolute limits between welfare categories (Not classified, Acceptable, Enhanced, or Excellent level of welfare) were set, but the rules governing the assignment of an animal unit to a category take into account what had been observed on European farms. The scientists behind Welfare Quality[®] are keen to make the value-based choices underlying assessments of animal welfare transparent. This is essential to allow stakeholder groups to understand the extent to which their views are acknowledged and acted upon.

Keywords: animal welfare, ethics, multi-criteria evaluation, science, societal concerns, values

Introduction

Surveys within the European Union suggest that there is considerable interest in animal welfare in member states (European Commission 2005, 2007a,b; Evans & Miele 2007a,b; Kjaernes *et al* 2007). However, European consumers do not believe that they are provided with adequate information about the welfare of livestock, and they also feel unable to take animal welfare into account appropriately when purchasing food (European Commission 2007a,b; Evans & Miele 2007a,b).

In its white paper to the parliament (European Commission 2002), the European Commission launched the idea of a unified labelling system that could be used for bilateral negotiations between countries. It was additionally thought that an information system, possibly translated into a label, could help to raise welfare standards (Polten 2007). Thus, there was a need for a system of animal welfare assessment that would inform consumers and other stakeholders, as well as guide welfare improvements, on the basis of clear identification of welfare problems.

However, to design such a system, value-laden decisions have to be made and this may give rise to dilemmas concerning whose values to consider. One such choice, which serves as an illustrative dilemma, relates to the definition of animal welfare.

Animal welfare is a contested concept. Drawing on philosophical discussion of the nature of human welfare, Appleby and Sandøe (2002) identify three theories of what animal welfare consists in: *hedonism*, according to which welfare is a quality of mental states, with the more pleasurable, or positive, states raising welfare; *perfectionism*, on which welfare involves the living of a life in which species-specific potentials are realised, usually through occupation of an adaptively suitable environment; and *preference satisfaction*, on which welfare accrues to an animal when its preferences are satisfied (again, something most likely to happen in an environment to which the species is adapted). Similar distinctions between differing theories of welfare are made by Fraser (Fraser et al 1997; Fraser 1999).

In the human sphere, preference satisfaction is mostly understood in distinctly non-hedonistic terms: not in a psychological way, but simply as a matter of the preferred state coming about (eg Griffin 1986). However, in animal welfare research, preferences are often seen as reflecting the quality of mental states, on the assumption that pleasure and pain have evolved as guides to animal choices (eg Fraser & Duncan 1998). If the animal's trade-offs between various options are interpreted as its own assessment of the importance of the accompanying mental states, this view is in fact an instance of what is known as preference-hedonism, a position first suggested by Sidgwick (1981, originally published in 1874) as a correction of classical hedonism (Bentham 1948, originally published in 1789). On the latter view, the importance of a feeling was measured in terms of its intensity and duration (independently of preferences); by contrast, the strength of a preference is taken by the preference hedonist to reveal the weight of the underlying mental state (see Sandøe 1996). In practice, therefore, the main conceptions, or theories, of animal welfare boil down to just two: preference-hedonism and perfectionism.

In many cases, these two theories will deliver the same verdict on animal welfare; but in others they will differ. For example, according to perfectionism, but not necessarily according to preference-hedonism, extensive, free-range systems will by definition have an advantage over more intensive indoor systems. And, since many ordinary citizens, unlike many experts, seem to prefer free-range to indoor systems, this may make a real difference (Lassen et al 2006; Evans & Miele 2007a,b). Note that the decision on a theory of animal welfare is essentially evaluative — it is a decision about what counts as a good animal life. The latter cannot be guided solely by factual investigation, since any factual investigation of animal welfare will presuppose a definition of animal welfare.

A number of other value-based decisions will have to be made. One concerns the relative weights of criteria used in measuring animal welfare. In practice, assessments of

animal welfare will have to rely on indicators in multiple dimensions, including health status, physical comfort, expression of behaviour, and so on (Farm Animal Welfare Council 1992). An overall welfare assessment should ideally cover all the relevant dimensions. Still, the importance attributed to each dimension is inherently value-based (Fraser 1995) because it is bound to reflect underlying priorities about what counts in animal welfare. In addition, when welfare is considered at group level, the overall assessment may have to deal with the conflicting interests of the various animals (Fraser 2003). Finally, in judgements of acceptability, animal welfare will have to be balanced against human interests.

It follows that any system of welfare assessment will be based on choices which reflect values, including values relating to animal ethics. This is not a problem, as long as these choices are made in a transparent way so that the users of the assessment can readily see the extent to which the choices accord with their own values. The need for transparency becomes particularly important when information is condensed into a single score (or label), as is the case with the overall assessment of acceptability proposed in Welfare Quality®.

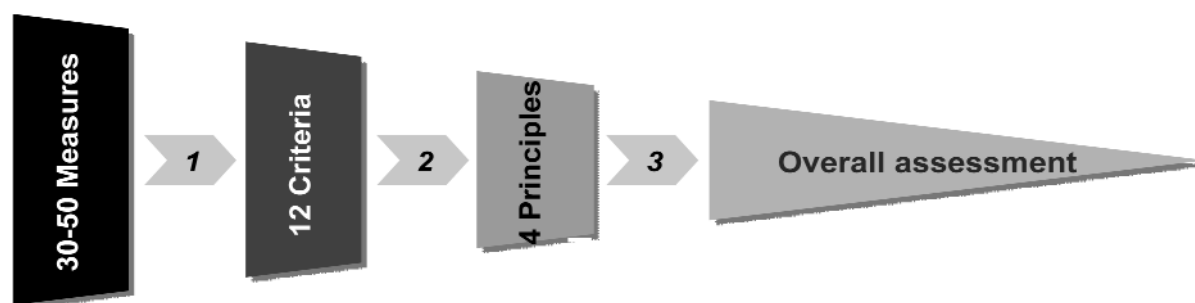
The present paper is based on a constructive debate between animal scientists and philosophers. We provide an *ex post* analysis of the Welfare Quality® scoring model, describing how, exactly, it deals with value choices. The animal scientists were involved in the development of the Welfare Quality® scoring system. The philosophers raised questions about the ethical assumptions made during that development. The combination of the questions raised by the philosophers and the answers provided by the animal scientists will hopefully give the reader an improved understanding of the value assumptions underlying the Welfare Quality® scoring system, and thereby contribute to the transparency of that system.

We begin by briefly describing the Welfare Quality® scoring system (note that we do not aim to describe the methods used to design this system in detail). We then present five ethical issues and seek to explain how the scoring system deals with each. The first issue concerns the definition of animal welfare underlying the system. The next three concern the relative weight attached either to concerns about individual animals or to welfare aspects relative to the total average outcome at different levels of aggregation. The fifth, and last, issue centres on how the line between acceptable and unacceptable levels of animal welfare should be drawn.

Brief description of the Welfare Quality® scoring system

The Welfare Quality® project (<http://www.welfarequality.net/everyone>) aims to develop reliable on-farm monitoring systems and translate them into product information systems. With this aim, a general tool for assessing animal welfare at farm or slaughter plant level at a given time has been produced. A multi-criteria evaluation procedure was designed which relies on methodologies developed in oper-

Figure 1



Hierarchical evaluation model designed in Welfare Quality®.

ational research. Drawing on the scientific literature on animal welfare, Welfare Quality® identified four welfare principles: good feeding, good housing, good health, and appropriate behaviour. These were broken down into 12 welfare criteria (Botreau *et al* 2007a; Veissier & Evans 2007b). All criteria, according to the system, must be taken into account in a full assessment of animal welfare. Researchers then developed 30–50 welfare measures per species (pig, cattle, poultry) in order to check the compliance of animal units, ie farms or slaughter plants, with the 12 criteria (Welfare Quality® 2009a,b,c).

These measures are essentially animal-based, ie they record states of the animals rather than of the environment in which the animals are kept. They include health and behavioural parameters. When no valid animal-based measure with which to check a criterion was available (eg where the animal-based measures were considered to be insufficiently sensitive or reliable) an environmental measure was used instead (eg absence of thirst is generally measured indirectly by the number of drinking points available). These measures generate a substantial quantity of data, and these data need to be interpreted and integrated into a net, or overall, evaluation of the animal unit.

Welfare Quality® designed a hierarchical evaluation model (Figure 1). In the first step of the evaluation, the results from the 30–50 measures are transformed into scores on a value scale (0 = worst, 100 = best) to reflect the compliance of a given farm or slaughter plant with the 12 welfare criteria (Welfare Quality® 2009a,b,c). In a second step, criterion scores are combined to form principle scores. For example, the scores for absence of hunger and absence of thirst are combined to reflect compliance with the principle ‘good feeding’. At each step, the model was designed to best map the level of animal welfare on the farm or slaughterhouse according to the opinion of ‘experts’, these experts being animal scientists and social scientists participating in the Welfare Quality® project. Finally, in a third step, an overall assessment is made which allows farms or slaughter plants to be labelled according to their degree of acceptability. Four categories of attainment are distin-

guished to meet stakeholders’ requirements — animal units are classed as having: a) Excellent welfare; b) Enhanced welfare; c) Acceptable welfare; or d) units that are Not classified.

Welfare Quality® scientists were fully aware that they could not rely solely on animal science in interpreting measures in terms of welfare. Nor could they determine on scientific grounds alone the importance to be attributed to the various criteria and principles, or the compensatory relations between bad scores on some criteria and good scores on others. Finally, it was recognised that it would not be possible, purely on the basis of scientific information, to determine whether, for example, 10% lame cows is a bad result or an acceptable one.

Given a definition of welfare, science can suggest indicators of welfare and answer factual questions about the degree of fulfilment of these indicators. To some extent, it can also answer questions about how the animals choose between alternatives. But questions about how to define welfare, how to interpret measures in terms of welfare, how to aggregate welfare or welfare criteria across individuals, and how to decide on the acceptability of animal welfare levels on the farm are questions which, in different ways, go beyond science.

In Welfare Quality® animal welfare is understood along preference-hedonist lines. It is mostly considered a matter of how the animal perceives its environment, and as far as possible the intention has been to take that point of view into account (for some exceptions to this rule, see the *Discussion*). However, as has been pointed out by Botreau *et al* (2007b), when it comes to synthesising welfare aspects which may, for instance, occur within rather different timescales (eg being sick and being afraid of something) it would seem to be impossible to determine reliably what an animal prefers through preference tests. Inevitably, the assessment of the overall welfare of an animal is to some extent based on human input: we not only decide that animal preferences are what count but also weigh different preferences against each other.

In Welfare Quality® the assessment of the overall welfare of the animal was therefore based on expert opinion. Here, ‘expert’ is understood very broadly; it means ‘having a professional interest in animal welfare’. The experts ranged from animal scientists (recruited for their knowledge of animals and of the measures used for scoring), to social scientists (recruited for their knowledge of expectations of societal groups), to stakeholders who would possibly be using the scoring system in the future (here: the Advisory committee of Welfare Quality® on which sat representatives of farmers, breeders, retailers, veterinarians, animal protectors, and of institutions — EU Commission and OIE). Value issues were not addressed directly by the experts but the researchers and stakeholders involved were asked to give their opinion on datasets or options for synthesising the information. When there was no consensus among the people consulted, the strategy was to reach a compromise. Then, the scoring model was designed to best match the experts’ opinions. The consultation processes are briefly outlined below; they are described in more detail in Miele *et al* (2011). The type of questions asked to experts and how their opinions were used are described in Veissier *et al* (2009).

Ethical issue 1 — Definition of animal welfare: popular vs expert perspective?

As was pointed out in the *Introduction*, there is disagreement about the definition of welfare. There is some evidence that animal scientists tend to understand welfare in hedonistic terms. Certainly, Duncan’s (1993) claim that welfare relates to feelings (mental states) is often quoted by animal scientists. However, there is also evidence that many ordinary consumers or citizens (in reality these are the same people, but under the labels *consumer* and *citizen* they are thought of as occupying rather different roles) tend to understand welfare more in perfectionist terms, placing considerable weight on the notion of a natural life (Lassen *et al* 2006; Veissier & Evans 2007a).

In line with their hedonistic perspective, animal scientists often prefer animal-based indicators which can be interpreted as signs of animal suffering (or animal joy). When lay people are confronted with animal-based information they acknowledge it to some degree. However, for the most part, they are not willing to draw the conclusion that the animals should be kept in indoor housing systems — even when these systems give better opportunities for disease and hygiene control than outdoor alternatives, and even when, in the better of these systems, there is no serious evidence that the animals fare badly in terms of their mental states. Thus, many lay people would continue to insist that a good animal life requires living conditions as close to ‘the way nature intended’ as possible (Lassen *et al* 2006); they would do so even when given assurances that the animals’ avoidance of suffering and accumulation of positive experiences would be as good as they would be in an indoor system. Ultimately, and acknowledging all the facts of the matter, this represents, in part, a disagreement over the definition of welfare, ie over the way that concept should be understood.

There seem to be two strategies for dealing with this problem. One is to deploy just one theory of welfare to guide the scoring system. A single theory would probably give a set of indicators that is theoretically more consistent, which is clearly an advantage. However, the trouble with this strategy is that any single theory of welfare is bound to be controversial; thus some stakeholder groups will view the favoured theory as one that fails to reflect their beliefs regarding welfare.

The other strategy is to use indicators intended to cater for aspects of both main theories. This might lead to more legitimacy and wider acceptance across stakeholder groups. The trouble is that the scoring system might become less sharp in its profile, and there is still a risk that some stakeholder groups will not find their views fully acknowledged. Whatever the chosen strategy, it is, of course, an important task to make the choice and its ethical implications fully transparent.

Welfare Quality® originally pursued the first strategy but ended up opting for the second. The 12 welfare criteria identified by Welfare Quality® (Table 1) and a range of animal-based indicators were proposed to the Advisory committee of the project. They were also discussed in 55 consumer-citizen focus groups across France, Hungary, Italy, The Netherlands, Norway, Sweden and the UK (Miele & Evans 2005). There was a great deal of societal support for the 12 criteria, with consumer-citizens often agreeing that the 12 Welfare Quality® criteria expressed what they mean by animal welfare, but in a more considered way (Miele *et al* 2011).

Nevertheless, some discrepancies arose between the viewpoints of citizens and scientists: citizens view animal welfare both in terms of the state of the animals (eg good health) and in terms of what is provided for animals, and they consistently highlight the importance of providing natural environments for farm animals (Miele *et al* 2011).

Citizens, then, tend to deploy a wider notion of animal welfare than many scientists. The animal scientists in charge of the development of the Welfare Quality® scoring system attempted to tackle the discrepancies between the two views of animal welfare by including important resource-based indicators of welfare (such as outdoor access and the amount of space available), the effects of which cannot be easily measured directly on animals, and by including more indicators relating to the expression of natural behaviours (eg exploration and access to pasture) (Miele *et al* 2011). These decisions were taken during integration meetings where animal scientists and social scientists involved in Welfare Quality® exchanged views — the former focusing on the effects of certain situations on animals or the interpretation of a certain behaviour or symptom, and the latter expressing the views of the stakeholder groups they studied (consumer-citizens, retailers, producers).

The proposed list of welfare measures was also discussed during farmers’ juries (Bock 2009). Moreover, its soundness was recently confirmed in a survey of Belgian citizens and farmers (Tuytens *et al* 2010). The final choice of the criteria and the measures used to check them therefore represented a compromise between the views of the animal scientists and the views of other stakeholders.

Table 1 Welfare principles and criteria identified in Welfare Quality®.

Principles	Criteria	Meaning
Good feeding	Absence of prolonged hunger	Animals should not suffer from prolonged hunger, ie they should have a sufficient and appropriate diet
	Absence of prolonged thirst	Animals should not suffer from prolonged thirst, ie they should have a sufficient and accessible water supply
Good housing	Comfort around resting	Animals should have comfort around resting
	Thermal comfort	Animals should have thermal comfort, ie they should be neither too hot nor too cold
	Ease of movement	Animals should have enough space to move around easily
Good health	Absence of injuries	Animals should be free of physical injuries
	Absence of disease	Animals should be free of disease, ie animal unit managers should maintain high standards of hygiene and care
	Absence of pain induced by management procedures	Animals should not suffer from pain induced by inappropriate management, handling, slaughter or surgical procedures (eg castration, dehorning)
Appropriate behaviour	Expression of social behaviours	Animals should be able to express normal non-harmful, social behaviours (eg grooming)
	Expression of other behaviours	Animals should be able to express other normal behaviours, ie it should be possible to express species-specific natural behaviours, such as foraging
	Good-human animal relationship	Animals should be handled well in all situations, ie handlers should promote good human-animal relationships
	Positive emotional state	Negative emotions, such as fear, distress, frustration or apathy should be avoided, whereas positive emotions, such as security or contentment should be promoted

Ethical issue 2 — Aggregating for individuals or for traits?

When welfare is considered at a given time — whether at farm level, or the level of the slaughter plant, or indeed at the level of any other animal unit — two dimensions structure the welfare issue: on the one hand, there are the individual animals; on the other hand, welfare is multi-dimensional (see the 12 welfare criteria above). This raises the question: how should the information be aggregated in an overall assessment (of the farm, slaughter plant, or other animal unit)? Let us focus on the farm setting, and let us further assume the information can be set out in a grid, as in Table 2.

The information could first be aggregated for each individual. It could then be asked how much each individual is affected by various welfare problems, and then the results could be weighed up across individuals. Alternatively, the information could be aggregated across individuals within each criterion, resulting in an assessment of how an individual farm complies with each criterion. Then, this information would be weighed up across criteria to give an overall assessment of the farm (Figure 2).

In most theories, welfare is a concept that applies to individuals, because only individual animals possess the characteristics (ie mental states or preferences) that make lives go better or worse. When we speak about welfare at farm level, we typically mean the welfare of all of the animals on the farm considered together. It is a natural implication of this that the

Table 2 The structure of welfare assessment of farms.

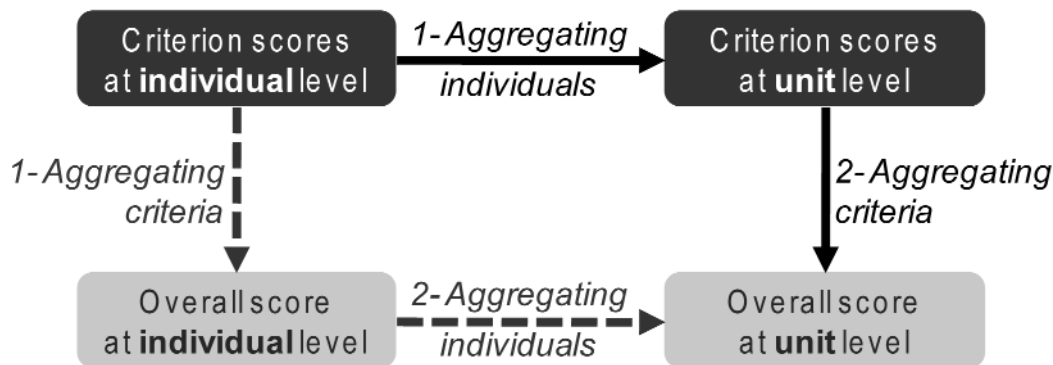
	Criterion 1	Criterion 2	Criterion 3	Criterion 12
Individual 1					
Individual 2					
Individual 3					
.....					
Individual n					

different dimensions of welfare should be given weight according to their importance within the life of an individual.

From this perspective, it would be straightforward to determine the level of welfare for each individual first, and then to base the evaluation of the farm on the distribution of welfare across individuals. This is probably how most people would understand welfare assessment in the first instance: how well off are the individual animals on the farm? It is also the way aggregated human welfare within a group or society normally is conceived (Arrow 1963; Sen 1979; Broome 1991, 2004).

The alternative approach is to base the aggregation on criteria of welfare applying at farm level. On this approach, a given dimension of welfare will be aggregated across indi-

Figure 2



Two ways of aggregating information.

Table 3 Hypothetical case involving two welfare criteria on two farms.

Farm A	Criterion 1	Criterion 2	Farm B	Criterion 1	Criterion 2
Individual 1	Good	Bad	Individual 1	Good	Good
Individual 2	Bad	Good	Individual 2	Bad	Bad

viduals to give an assessment of the way the farm performs against the criterion associated with that dimension. The overall assessment of the farm will then be determined by weighing up across criteria at farm level.

The difference between the two ways of looking at aggregating welfare may well be ethically significant, as is shown by a simple example. Consider Table 3.

Let us suppose that, in Table 3, Criterion 1 is ‘absence of injuries’ and Criterion 2 is ‘positive emotional state’. Judged criterion-by-criterion, Farms A and B emerge as *equally good*, because they score the same for each criterion. On Farm A, however, both of the individual animals have just one welfare problem. On Farm B, by contrast, Individual 2 has two welfare problems (and is therefore likely to be very badly off) and Individual 1 has none (and is therefore likely to be rather better off than either of the individuals on Farm A). An assessment that looks at matters in this second way, rather than criterion-by-criterion, may well conclude that Farm A is *better* overall than Farm B.

Like most other systems for overall assessment of animal welfare at group level (Bartussek 1999; Capdeville & Veissier 2001; Bracke *et al* 2002), Welfare Quality® structured the assessment of welfare at farm level not by individuals but by criteria. A number of considerations led to this approach.

First, on many farms, the animals are frequently exchanged, with some livestock being removed and other livestock arriving. In all meat production units, the turnover of animals is high (eg lifetime of a chicken: a few weeks; pigs: a few months), and this makes it difficult to follow up the welfare problems of particular individuals. Where animals are kept for

a longer period — as happens with horses, for example — a focus on individuals would seem more appropriate.

Second, the Welfare Quality® scoring system is set up to promote welfare improvements in a way that makes problems visible. The results of an assessment form the basis of advice to farmers. Presented in the form ‘n% animals are in a poor condition’, these results will not help farmers to see what problems he or she has, nor will they be readily translated into advice. In order to explain the underlying measures to the farmer, it would be necessary to follow every individual, and this does not seem appropriate when these animals are numerous and stay on the farm only for a limited period (as do broilers). By contrast, when scores are calculated at criterion level they can be used more effectively to highlight problems needing to be addressed. For instance, if a large number of animals are ill, the score for the criterion ‘absence of disease’ will be low. It is relatively clear, given this, that the farmer should be advised to put in place a health plan.

Finally, in principle, the individual approach requires every individual animal to be assessed, and this is not always feasible. Thus, in aviaries, where it is common to have more than 100,000 birds, and even in animal units with hundreds or at most a few thousand animals (as may be the case on pig and cattle farms), separate assessment of each individual will often not be feasible.

For these reasons, then, the Welfare Quality® assessment scoring system is structured by criteria. The scores for each welfare criterion are aggregated across individuals to give an assessment of the farm’s performance regarding this

criterion. An overall assessment of the farm is then determined by weighing up across criteria at farm level. But, as the example in Table 3 shows, this way of structuring the problem implies that each criterion is assessed independently of how the animals are faring on other criteria, and this way of aggregating welfare may conceal ethically significant information about the distribution of problems among the animals. However, it should be emphasised that the main reasons for the choices made here are practical not ethical.

Ethical issue 3 — Aggregating within a criterion: balancing between better-off and worse-off animals?

Suppose now that there are two flocks of animals. They have the same ‘average’ level of welfare. However, on one farm all the animals suffer equally from rather mild welfare problems, whereas in the other 10% of the animals suffer from very severe welfare problems while the rest are well off. Are these farms equally acceptable from an ethical point of view? To many people they are not — many people would find the conditions on the first farm more acceptable than those on the second. Therefore, it is important to consider how the welfare scoring system deals with distribution across individuals.

Briefly comparing the human sphere, the main ethical problem in aggregating welfare across people is indeed posed by distribution. On the one hand, there is the *utilitarian* view that the total, or sum, of welfare is what counts (Sen 1997). Thus, equal and unequal distributions with identical totals of welfare count the same. Opposed to utilitarianism are various distribution-sensitive positions, where individuals receive more weight in the aggregation the worse off they are. On the *priority* view individuals are given more weight, the worse off they are in absolute terms; on the *egalitarian* view individuals get more weight, the worse off they are relative to others (Jensen 2003).

The task in Welfare Quality® is to aggregate the information relating to how well a farm performs on the different welfare dimensions into a score for each criterion. In the human sphere, this can perhaps be compared with evaluation of the distribution across individuals of certain types of goods or resources in society — for instance, the distribution of income, the distribution of education, or the distribution of (privately owned or rented) housing. The difference, however, is that in Welfare Quality® the focus is on the impairment of welfare as measured directly on the animals rather than on the resources that promote it.

Let us consider how to assess the distribution of a specific welfare problem covered by a given criterion — say, lameness in cattle. Lameness can be more or less severe, and in Welfare Quality® three different categories of lameness are identified: not lame, moderately lame and severely lame. The first issue is how these states can be weighted according to their seriousness for the animal.

In order to interpret the data obtained from the various measures to produce criterion scores, Welfare Quality® partners consulted the animal scientists who developed the

measures, because they could interpret them in terms of pain or discomfort. The scientists were asked, first, to rank a number of virtual distributions of the three categories of lameness on a farm from worst to best, and later, to score them on a scale from 0 to 100.

It was assumed that, since a linear aggregation of the percentages of different categories of lameness, weighted with their seriousness, could represent the scientists ranking, the weights which could be derived from this ranking would represent the seriousness of the lameness to the animal (the weights being 0 for not lame, 0.29 for moderately lame and 1 for severely lame).

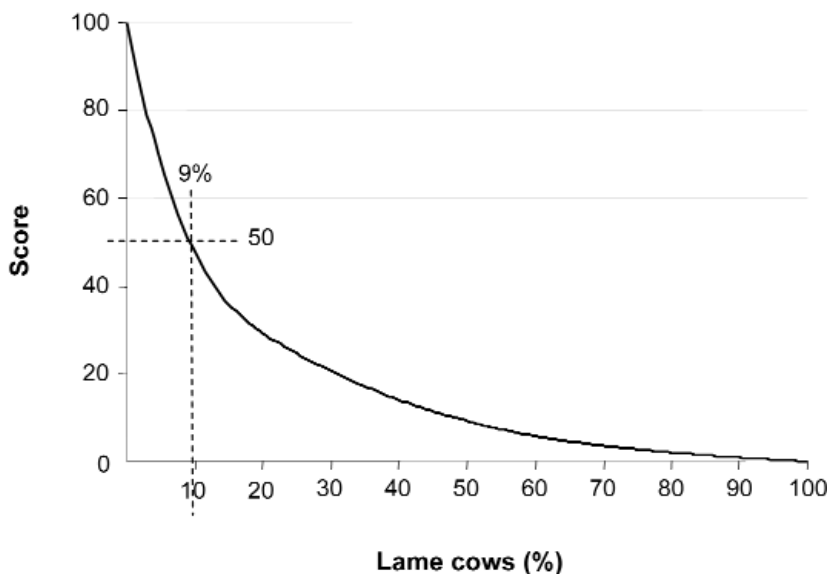
The key ethical issue is how to assess the distribution within the group. One approach would be simply to use the weighted sum, giving equal ethical weight to all individuals. Another would be to attach more ethical weight, the more severe the lameness. The scientists responded by letting all problems count but attaching more importance to the animals in a worse-off condition than to those in a better-off condition. This is reflected in the resulting criterion scores. For instance, as shown in Figure 3, nine per cent lame cows result in a score of 50, although if the same weight had been assigned to animals in bad and good condition, a farm would obtain a score of 50 when 50% animals are lame. The same line of reasoning applies to the other criteria and for other animal types: a score of 50 is given both to ‘absence of hunger’ when 13.5% broiler chickens are emaciated and to ‘good human-animal relationship’ when a panic response at entrance of the observer occurs in 29% pens of fattening pigs. Nevertheless, all problems count. Thus, although a farm with 5% severely lame and 50% moderately lame cows has a smaller number of severely lame cows than a farm with 10% severely lame and no moderately lame cows, the former scores lower than the latter, because there are far more lame cows.

Ethical issue 4 — Aggregating across criteria: compensation allowed or not?

The fourth ethical issue we wish to consider is whether an impairment covered by one criterion can be compensated for by an improvement covered by another. Suppose the principle in question involves two criteria: in fact, this is the case with the principle of ‘good feeding’, which involves the criteria ‘absence of prolonged thirst’ and ‘absence of prolonged hunger’. Assume a farm scores (50, 75), but that, following this, there is a serious drop in ‘absence of prolonged hunger’ to 25 and at the same time a rise in ‘absence of prolonged thirst’ to 100; and that the new scores are therefore (100, 25). How do (50, 75) and (100, 25) compare in terms of the overarching principle ‘good feeding’? Is the impairment in respect of hunger compensated for by the improvement in thirst in such a way that the farm scores equally well on ‘good feeding’ at the earlier and later times? Or is absence of thirst more important than absence of hunger? Or is compensation between these criteria impossible?

If, within a principle, the criteria scores do not interact with each other and all criteria have the same importance, then

Figure 3



Calculation of the partial score for lameness according to the % animals moderately lame and the % animals severely lame (weights: 0.29 for moderate and 1 for severe lameness).

those criterion scores can be added together to calculate the principle score. Here, compensation between criteria is fully allowed. In the example above, the change from (50, 75) to (100, 25) would not affect the score the farm obtains against the principle 'good feeding', because the decrease by 50 in absence of hunger would be compensated for by an increase of the same score in absence of thirst.

When the criteria are ascribed different importance, or weight, a weighted sum will be calculated rather than a simple sum in order to obtain a score at principle level. Full compensation between criteria nevertheless remains possible: if thirst is twice as important as hunger, a decrease by 50 in absence of hunger can be compensated for by an increase by only 25 in absence of thirst.

On the other hand, where the contribution of one criterion to the overarching principle (and hence to the overall assessment of the farm) interacts with the contribution of other criteria, compensation will be restricted. For instance, it could be said that the drop in 'absence of prolonged hunger' is very serious for the overall feeding situation: the fact that there is a significant improvement in thirst helps, but only a little; and the score of 25 for 'absence of prolonged hunger' is so low that the feeding situation as a whole has deteriorated considerably. On this approach, criterion scores cannot simply be summed up to produce an overall assessment.

In Welfare Quality®, questions about the relative importance, and compensatory effects of welfare aspects arose when criterion scores were synthesised into principle scores (Step 2, Figure 1). At that stage, animal and social scientists were consulted. It was clear that they considered some criteria to be more important than others. In most animals, for example, 'absence of disease' is more important than

'absence of injuries', which in turn is more important than 'absence of pain induced by management procedures'. However, the scientists did not allow one score to compensate for another. Thus, absence of disease was not regarded as compensating for injuries, and *vice versa*. A specific operator (Choquet integral) was used to take these two lines of reasoning into account. Briefly, the Choquet integral generalises the notion of a weighted sum when weights are not only attached to each criterion but also possibly to any subset of criteria (Grabisch & Roubens 2000). Examples of principle scores calculated within the Welfare Quality® system are given in Table 4. Essentially, when criterion scores are unequal, the lowest scores prevail and compensation remains very limited.

Ethical issue 5 — The boundary separating overall assessments of acceptable and unacceptable conditions

In its final stage, Welfare Quality® aggregates the scores obtained for each of the four principles into an overall assessment. The farms are sorted into four ordered categories: Excellent, Enhanced, Acceptable and Not classified. These categories are separated by three aspiration levels.

At this stage, the focus changes from an assessment of animal welfare *per se* to classification according to aspiration levels set by stakeholder expectations. Here, the ethical question of how to balance the interests of the farm animals against the interests of citizens and the various other actors of the supply chain, including farmers, needs to be addressed. The most important thing will be the boundary between acceptable and unacceptable conditions on a farm. Which interests should be considered here, and how much weight should they have? Do human interests

always prevail, or should the animals be protected by certain minimal standards? If the latter, how bad could conditions on a farm become and yet still be considered acceptable? In Europe, animals receive minimal protection thanks to European directives (eg Council Directive 98/58/EC). Therefore, Welfare Quality® definitively adopted the second above option (ie took European directives to ensure minimal standards).

There is also a concern about whether all stakeholder expectations can be satisfied at one time. Pressure to differentiate farms into several categories for the sake of a labelling system might bring with it a risk of setting the line between acceptable and unacceptable so low that this line serves, in effect, to classify poor conditions as good because they are widespread, rather than giving reliable information on how poor they are.

Finally, as was indicated at the end of the previous section, the issue of trade-offs between principles arises again at this stage. Could a loss in one principle score (say 'appropriate behaviour') be compensated for by a sufficient increase in another (say 'good housing')? To define the boundaries between categories strictly would be to prevent any trade-off between principles across boundaries. If they are defined less strictly, they will allow trade-offs across boundaries to some extent.

Within Welfare Quality® there was extensive discussion of the question whether the scoring system should be based on absolute or relative boundaries between what can be considered good vs bad. Absolute limits mean that the boundary between 'good' and 'bad' is set in advance, whatever the results obtained by farms. This might be problematic from a practical point of view: minimum values set for good welfare could be too difficult to achieve; again, they might be too easy to achieve and hence not discriminate between farms.

Relative limits mean that the boundary between 'good' and 'bad' corresponds to a certain percentage of farms from a given population in which the system is used. For example, one could stipulate that farms in the top 20% of their population are Excellent, whatever exact level of welfare they have achieved. Relative values seem problematic, however, for several reasons: (a) the best farms might not correspond to an Excellent level of welfare; (b) the classification of farms depends on the population observed — in other words, a farm might be classified as Excellent in a given population because it is in the top 20%, although the same farm would be Enhanced or even only just Acceptable in another population; or (c) it would become impossible to monitor progress on farms: for example, a farm considered just Acceptable that put in place a welfare plan to reduce welfare problems may not be considered a good farm at the next assessment because all of the farms with which it is being compared have improved in the meantime.

Here, Welfare Quality® took an intermediate strategy: it was decided that absolute limits would be used and that checks would be made that these limits were realistic on the farms visited in Europe during the project. This was done in both Step 1 and Step 3. In Step 1, corresponding to the

Table 4 Example of scores for the overall principle 'good health' in fattening pigs given by scores obtained for the criteria 'absence of injuries', 'absence of diseases' and 'absence of pain due to management procedures'.

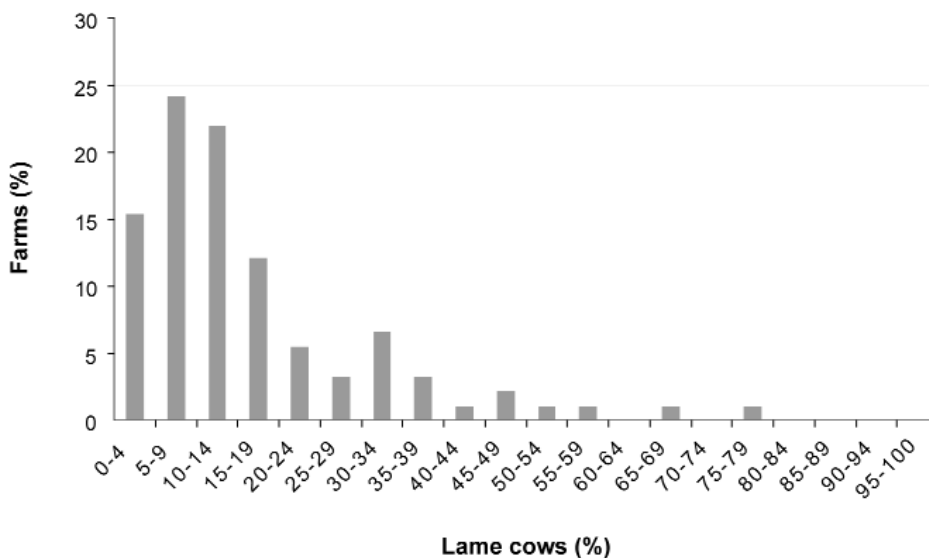
Criteria: absence of...			Principle
Injuries	Disease	Pain	Good health
25	50	75	32
25	75	50	35
50	50	50	50
75	25	50	28
75	50	25	34

calculation of criterion scores, two consultations were run (at least for defining the welfare of dairy cows). The first was run before data were collected on farms. The animal scientists consulted ascribed a value to the prevalence of welfare problems on farms according to their estimates of the severity of these problems for the animals. It would appear that the scientists consulted before data were collected had been rather severe in their requirements. For instance, where lameness was concerned, 50% of farms had more than 12% lame cows, which would result in scores of 30 or lower. The scientists were then presented with the distribution of lameness across European farms (Figure 4) and asked if they were willing to change their initial evaluations. Most of them were less severe in their requirements in the course of this second consultation (Table 5). However, they considered lameness a serious problem for cows, and so 50% farms still scored lower than 44. Expert evaluations during the second consultation were used to design the scoring system. For the other animal types, only one consultation was run because the data collected on farms were available at the time of the consultation.

In Step 3, which corresponds to the final overall scoring, a core group of animal and social scientists proposed aspiration levels for the welfare categories (Excellent, Enhanced, Acceptable, Not classified) according to the purposes for which stakeholders (here represented by the members of the Advisory committee of Welfare Quality®) may use the scoring system (certification, minimum standards, advice, etc). The following aspiration values for each welfare category were then agreed with the Advisory committee.

The aspiration value for the category Acceptable was set at 20, because 20 had a specific meaning on the 0–100 scale used to express criterion scores and principle scores: when scientists were consulted, they were told that if a farm scores below 20 on one criterion or principle, it has a high risk of being excluded from any welfare scheme. Again, 50 had a specific meaning: a farm scoring 50 was said to be neither good nor bad. Hence, the aspiration value for the category Enhanced was set just above this, at 55. Finally, the aspiration value for Excellent was set, symmetrically with that for Acceptable, at 80.

Figure 4



Distribution of lameness across dairy farms visited in Welfare Quality®. The percentage of lame cows is weighted for severity: 0.29 for moderate and 1 for severe lameness.

Table 5 Comparison of scores assigned for lameness by animal scientists before and after being informed of the distribution of lameness across the dairy farms visited.

% lame cows*	Mean expert scores	
	1st consultation	2nd consultation
10	34	46
30	13	20
60	6	6

* Weighted for severity: weights 0.29 for moderate and 1 for severe lameness.

Welfare Quality® then analysed the distribution of dairy farms across the aspiration values for the welfare categories Acceptable, Enhanced, and Excellent for each principle. It turned out that farms would be highly unlikely to achieve 80 for any welfare principle; in point of fact, that value was obtained only for the principle ‘Good feeding’. Nevertheless, a wide range of results was obtained for the other principles, from 0 to 70. Welfare Quality® thus decided to retain the three aspiration values as initially defined (20, 55, and 80), mindful that if the system is put in place improvements should be encouraged and scores of 80 or more might then be obtained.

Then, in order to be able to determine that a farm performs well enough to fulfil the aspiration value of a category, several membership rules were tested. A first, very intuitive

rule is ‘unanimity’, which means that a farm needs to reach the aspiration value of a given category for all welfare principles to be assigned to that category. For instance, to receive an Excellent classification, a farm would need to score at least 80 on all principles. Such a rule seemed unrealistic, since half the dairy farms visited in Welfare Quality® would then emerge as not classified and the other half would be classified merely as Acceptable. Welfare Quality® thus investigated alternative rules — that is, rules given which a farm needs to score higher than the aspiration value of a category only on some principles (3 or 2 out of the 4) while not falling below the aspiration value of the next lower category for any principle. The likely distribution of farms given by such rules was described.

Welfare Quality® also compared the classification obtained by farms according to the various rules tested with the general impression made by observers who visited the farms (expressed on a Visual Analogue Scale). The most appropriate rule appeared to be the following (Botreau et al 2009):

A farm is considered Excellent if it scores more than 55 on all principles and more than 80 on at least two of them. It is considered Enhanced if it scores more than 20 on all principles and more than 55 on at least two of them. Farms with Acceptable levels of animal welfare score more than 10 on all principles and more than 20 on at least three of them. Farms that do not reach these minimum standards are Not classified.

At the conclusion of this work, a balance between theoretical aspirations and what can be realistically achieved in practice was struck in the Welfare Quality® scoring system. The boundaries between categories are defined strictly, but some flexibility is introduced in the way farms are categorised.

The Welfare Quality® scientists were aware that if a farm obtains low scores on all principles then the absence of compensation does not encourage the farmer to improve at least one or two criteria (since the overall assessment of that farm would remain very low unless it were the case that, in the long term, all principle scores are expected to improve). Equally, however, the introduction of substantial compensation between criteria might result in a farm with a very low score on one criterion still being acceptable because it scores quite well on the other criteria.

Animal welfare implications and conclusion

Welfare Quality® is a model for the overall evaluation of animal welfare. Animal and social scientists were consulted in setting up the model, and as a consequence the final model reflects their assessments and ethical reasoning (and in particular their response to the question whether to focus on the average welfare of a group or the welfare of animals in poor conditions, and the question whether to allow for relations of compensation between welfare criteria). The model also involves a classification that has been adjusted to consider stakeholder expectations: a group of stakeholders was consulted, and the evaluation model was adjusted to reflect, to some extent, their (mostly implicit) ethical reasoning.

However, the proposal of an evaluation model may suggest that answering ethical questions is a purely scientific exercise. Some may therefore argue that Welfare Quality® could instead have outlined different possible proposals and left it to the political sphere, or to other stakeholders, to develop answers. Welfare Quality® decided to go beyond that stage and proposed a final evaluation, mapped in a formal model. This decision was made for two reasons. First, a precise formal representation of assessments and ethical decisions involves techniques that are not always accessible to the broad public. For example, few people are conversant with the Choquet integral used in Welfare Quality®. Second, there was general agreement that it would be best to suggest an answer based on expert consultations and stakeholder consultations and at the same time try to make explicit the assumptions underlying the choices made.

Methodologies relevant to the present topic have been developed in operational research, a discipline that merges management, mathematics, and computer science in order to help and support decisions (Roy 1993; Bouyssou *et al* 2000). Decision theory can help to formalise the problems linked to multi-criteria evaluation tasks, and provide a mathematical justification of integration procedures for a comprehensive evaluation. Welfare Quality® illustrates how such methodologies can be put in practice to assess animal welfare. More details on the techniques used can be found in Botreau *et al* (2008, 2009). The most important goal, in building the Welfare Quality® evaluation model, was to highlight the reasoning of people when evaluating the overall welfare of farm animals. The technique only provided a support for modelling that very reasoning correctly.

There is evidence from social science research in Welfare Quality® that some consumers, or citizens, would not fully agree with the Welfare Quality® assessment system, because it does not treat 'the living of a natural life' as a prerequisite of welfare (Evans & Miele 2007a,b; Miele & Evans 2010). In addition, many people may not be happy with the compromise struck in Welfare Quality® between absolute and relative standards. Levels of welfare categorised as acceptable by Welfare Quality® scientists may well be perceived as 'unacceptable, but could be worse' by consumers (Miele & Evans 2009).

This notwithstanding, the Welfare Quality® scoring system seems to represent a considerable improvement on previous scoring systems in as much as it incorporates (or, at any rate, attempts to incorporate) societal expectations about animal welfare. It achieves this goal by considering a wide range of welfare criteria, which reflect both scientific and societal concerns. Furthermore, it treats all animals as important, but sees that importance as something that grows as the welfare conditions in a given group worsen. Finally, it offers limited compensation between different welfare criteria.

On the whole, the Welfare Quality® assessment system must be taken as just one option, one attempt to balance some of the ethical issues underlying animal welfare. However, it is also true that the ethical decisions made within the Welfare Quality® project seem to reflect widely accepted views on animal welfare in Europe.

By not only providing an overall assessment of welfare, but also keeping available the results obtained at the level of principle, criterion, and measure, the Welfare Quality® system appears to be well suited to identifying both poor and good welfare situations, and so it could serve to certify farms (and give them access to certain markets). Moreover, it provides a strong basis for advice to farmers and a reasoned basis on which to draw their attention to specific problems that might require action.

Both animal scientists and philosophers agree, of course, that transparency in relation to the value-based and ethical decisions made within the Welfare Quality® project is essential for its successful implementation. A considerable effort has been made to explain in detail how the scores are calculated, and to clarify the assumptions that lie behind the scoring system (Veissier *et al* 2009; Welfare Quality® 2009a,b,c). This effort should continue, and it will be essential to check in the future that the ethical decisions underlying the scoring system are still widely accepted.

Acknowledgements

Welfare Quality® is co-financed by the European Commission, within the 6th Framework Programme, Contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission, who will not be liable for any use made of such information.

References

- Appleby M and Sandøe P** 2002 Philosophical debate on the nature of well-being: implications for animal welfare. *Animal Welfare* 11: 283-294
- Arrow K** 1963 *Social Choice and Individual Values*. John Wiley & Sons: New York, USA
- 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
- Bentham J** 1948 (originally published in 1789) *An Introduction to the Principles of Morals and Legislation*. Hafner: New York, USA
- Bock B** 2009 Farmers' perspectives. *Conference Delivering Animal Welfare and Quality: Transparency in the Food Production Chain* pp 73-75. 8-9 October 2009, Uppsala, Sweden
- Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ** 2007a Definition of criteria for overall assessment of animal welfare. *Animal Welfare* 16: 225-228
- Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, van Reenen CG and Veissier I** 2007b Aggregation of measures to produce an overall assessment of animal welfare: Part 2: Analysis of constraints. *Animal* 1: 1188-1197
- Botreau R, Capdeville J, Perny P and Veissier I** 2008 Multicriteria evaluation of animal welfare at farm level; an application to MCDA methodologies. *Foundations of Computing and Decision Sciences* 33: 287-316
- Botreau R, Veissier I and Perny P** 2009 Overall assessment of cow welfare: strategy adopted in Welfare Quality[®]. *Animal Welfare* 18: 363-370
- 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, The Netherlands
- Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP** 2002 Decision support system for overall welfare assessment in pregnant sows: A Model structure and weighting procedure. *Journal of Animal Science* 80: 1819-1834
- Broome J** 1991 *Weighing Goods: Equality, Uncertainty and Time*. Blackwell: Oxford, UK
- Broome J** 2004 *Weighing Lives*. Oxford University Press: Oxford, UK
- Capdeville J and Veissier I** 2001 A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. *Acta Agriculturae Scandinavica Section A-Animal Science Supplement* 30: 62-68
- Duncan IJH** 1993 Welfare is to do with what animals feel. *Journal of Agricultural and Environmental Ethics* 6: 8-14
- European Commission** 2002 *Communication from the European commission to the council and the European parliament on animal welfare legislation on farmed animals in third countries and the implications for the EU* p 626. EU: Brussels, Belgium
- European Commission** 2005 Attitudes of consumers towards the welfare of farmed animals. *Special Eurobarometer 229/Wave 63.2 - TNS opinion and social*. EU: Brussels, Belgium. http://ec.europa.eu/public_opinion/archives/ebs/ebs_229_en.pdf
- European Commission** 2007a Attitudes of consumers towards the welfare of farmed animals: Wave 2. *Special Eurobarometer 229(2)/Wave 64.4 - TNS Opinion & Social*. EU: Brussels, Belgium. http://ec.europa.eu/food/animal/welfare/survey/sp_barometer_fa_en.pdf
- European Commission** 2007b Attitudes of EU citizens towards Animal Welfare. *Special Eurobarometer 270/Wave 66.1 - TNS Opinion & Social*. EU: Brussels, Belgium. http://ec.europa.eu/food/animal/welfare/survey/sp_barometer_fa_en.pdf
- Evans A and Miele M** 2007a Consumers' views about farm animal welfare - Part II: European comparative report based on focus group research. *Welfare Quality Reports*. Cardiff University: Cardiff, UK
- Evans A and Miele M** 2007b Consumers' views about farm animal welfare - Part I: National reports based on focus group research. *Welfare Quality Reports*. Cardiff University: Cardiff, UK
- Farm Animal Welfare Council** 1992 FAWC updates the five freedoms. *The Veterinary Record* 17: 357
- Fraser D** 1995 Science, values and animal welfare: exploring the 'inextricable connection'. *Animal Welfare* 4: 103-117
- Fraser D** 1999 Animal ethics and animal welfare science: bridging the two cultures. *Applied Animal Behaviour Science* 65: 171-189
- Fraser D** 2003 Assessing animal welfare at the farm and group level: the interplay of science and values. *Animal Welfare* 12: 433-443
- Fraser D and Duncan IJH** 1998 'Pleasures', 'pains' and animal welfare: toward a natural history of affect. *Animal Welfare* 7: 383-396
- Fraser D, Weary DM, Pajor EA and Milligan BN** 1997 A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6: 187-205
- Grabisch M and Roubens M** 2000 Application of the Choquet integral in multicriteria decision making. In: Grabish M, Murofushi T and Sugeno M (eds) *Fuzzy Measures and Integrals: Theory and Applications; Studies in Fuzziness and Soft Computing*. Physica-Verlag: Heidelberg, Germany
- Griffin J** 1986 *Well-Being: Its Meaning, Measurement and Moral Importance*. Clarendon Press: Oxford, UK
- Jensen KK** 2003 What is the difference between (moderate) egalitarianism and prioritarianism? *Economics and Philosophy* 19: 89-110
- Kjaernes U, Miele M and Roex J** 2007 Attitudes of Consumers, Retailers and Producers to Animal Welfare. *Welfare Quality Reports*. Cardiff University: Cardiff, UK
- Lassen J, Sandøe P and Forkman B** 2006 Happy pigs are dirty! Conflicting perspectives on animal welfare. *Livestock Science* 103: 221-230
- Miele M and Evans A** 2005 European consumers' views about farm animal welfare. *1st Welfare Quality[®] Stakeholder Conference: Science and Society Improving Animal Welfare*. 17-18 November 2005, Brussels, Belgium
- Miele M and Evans A** 2009 Citizen Juries, a comparative analysis of the public opinions about the Welfare Quality on farm assessment system. *Welfare Quality Reports*. Cardiff University: Cardiff, UK
- Miele M and Evans A** 2010 When foods become animals, ruminations on ethics and responsibility in care-full spaces of consumption. *Ethics, Place and Environment*, submitted
- Miele M, Veissier I and Evans A** 2011 Establishing a dialogue between science and society. *Animal Welfare* 20: 103-117
- Polten R** 2007 *Proceedings of the Workshop on Animal Welfare Improving by Labelling?* 28 March 2007, Brussels, Belgium
- Roy B** 1993 Decision science or decision-aid science? *European Journal of Operation Research* 66: 184-204
- Sandøe P** 1996 Animal and human welfare: are they the same kind of thing? *Acta Agriculturae Scandinavica Section A-Animal Science* 27: 11-15
- Sen AK** 1979 *Collective Choice and Social Welfare*. *Advanced Text Books in Economics*. North Holland: Amsterdam, The Netherlands
- Sen AK** 1997 *On Economic Inequality*. *Expanded Edition*. Oxford University Press: Oxford, UK
- Sidgwick H** 1981 (originally published in 1874) *The Methods of Ethics*. Hackett Publishing: Indianapolis, USA

- Tuytens FAM, Vanhonacker F, Van Poucke E and Verbeke W** 2010 Quantitative verification of the correspondence between the Welfare Quality[®] operational definition of farm animal welfare and the opinion of Flemish farmers, citizens and vegetarians. *Livestock Science* 131(1): 108-114
- Veissier I and Evans A** 2007a Rationale behind the Welfare Quality[®] assessment of animal welfare. *Second Welfare Quality[®] stakeholder conference, Assuring Animal Welfare: from Societal Concerns to Implementation* pp 19-22. 3-4 May 2007, Berlin, Germany
- Veissier I and Evans A** 2007b *Principles and Criteria of Good Farm Animal Welfare*. Welfare Quality[®] Fact Sheet Welfare Quality[®]: Lelystad, The Netherlands
- Veissier I, Botreau R and Perny P** 2009 Scoring animal welfare: difficulties and Welfare Quality[®] solutions. In: Keeling L (ed) *An Overview of the Development of the Welfare Quality[®] Assessment Systems*, Cardiff University: Cardiff, UK
- Welfare Quality[®]** 2009a *Welfare Quality[®] Assessment Protocol for Cattle (Fattening Cattle, Dairy Cows, Veal Calves)*. Welfare Quality[®] Consortium: Lelystad, The Netherlands
- Welfare Quality[®]** 2009b *Welfare Quality[®] Assessment Protocol for Pigs (Sows and Piglets, Growing and Finishing Pigs)*. Welfare Quality[®] Consortium: Lelystad, The Netherlands
- Welfare Quality[®]** 2009c *Welfare Quality[®] Assessment Protocol for Poultry (Broilers, Laying Hens)*. Welfare Quality[®] Consortium: Lelystad, The Netherlands