



ELSEVIER

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

SCIENCE @ DIRECT®

Applied Animal Behaviour Science 92 (2005) 193–209

APPLIED ANIMAL  
BEHAVIOUR  
SCIENCE

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

# Can we measure human–animal interactions in on-farm animal welfare assessment? Some unresolved issues

Anne Marie de Passillé\*, Jeff Rushen

*Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada,  
PO Box 1000, 6947 Highway 7, Agassiz, BC, Canada V0M 1A0*

Available online 20 June 2005

---

## Abstract

Research has documented the effect that stockmanship has on animal welfare and shown that rough handling of animals can make them fearful of people. Interest has arisen in using measures of animals' responses to people in on-farm welfare assessment. In this article, we discuss some of the unresolved issues that lead us to doubt whether current measures of animals' responses to people can be used effectively in on-farm welfare audits. These include: uncertainty as to the best type of measure to be used, the low level of reliability of some tests, difficulties in establishing a clear cut-off point, and questions about the validity of the measures arising from effects due to the identity of the test person, the location of the test, the influence of motivations other than fear, and poor correspondence with type of handling actually used on farms.

© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Animal welfare; On-farm welfare assessment; Animal welfare audits; Human–animal relationship

---

## 1. Introduction

A large body of research has now documented the effect that stockmanship has on animal welfare and animal productivity (reviewed most recently in [Hemsworth and Coleman, 1998](#); [Hemsworth, 2003](#)). One aspect of stockmanship that has generated much interest is the way that animals are handled, especially the use of aversive handling

---

\* Corresponding author. Fax: +1 604 7960359.

*E-mail address:* [depassilleam@agr.gc.ca](mailto:depassilleam@agr.gc.ca) (A.M. de Passillé).

techniques that may result in animals becoming fearful of their caretakers. Farm animals' fear of the people that care for or handle them can be a major source of stress. There are marked differences between farms in the degree to which the animals are fearful of people and much of the difference is due to the way that the animals are handled.

The importance of good stockmanship for good animal welfare has generated interest in whether some aspect of the human–animal relationship, especially the extent that animals are fearful of people, can be reliably measured in on-farm animal welfare assessments or audits (e.g. Rousing and Waiblinger, 2004). This issue arises at a time when on-farm welfare indices or audits are becoming increasingly used (e.g. Mench, 2003). Bartussek (1999) included information on quality of stockmanship in the animal needs index, while Main et al. (2003) took some measures of flight distance in their on-farm welfare assessments of dairy cattle. However, other welfare audits (e.g. Barnett et al., 2001) have deliberately avoided including measures of human–animal relationships due to difficulties in standardising these measures within the practical constraints of on-farm welfare audits (J.L. Barnett, personal communication, 2004).

In this paper, we discuss some unresolved issues that limit our ability to use these well-documented relationships between handling and animals' responses to people to develop on-farm measures of human–animal relationships. We do not intend a complete review of the literature regarding tests of the human–animal relationship. Instead, we raise issues that are important in the context of on-farm welfare assessments or audits, and that we believe have not been adequately addressed. We illustrate these principally with research done on cattle, although we believe that the constraints we identify apply to other species as well.

To be cost effective in an animal welfare audit, measures of human–animal relationships must be reliable and be able to be taken in one or a few visits of relatively short-duration by people who are not necessarily trained in research. The measures used must be valid indicators, able to predict how the animals are handled or the state of their welfare. Ideally, such measures would have a clear “cut-off” point indicating where overly-aversive handling had been used, or where the welfare of the animals was unacceptably threatened.

## 2. Types of measures

An immediate unresolved issue is what type of measure to use. A number of different ways of measuring animals' fear of people have been used. These can be roughly classified into those that measure the distance that animals keep between themselves and people (“distance measures”), those that assess the animals' response to being handled (“handling measures”), or those that rely on some subjective rating of the animal.

### 2.1. Distance measures

The most common method involves taking some measure of how much the animals approach or avoid people (e.g. Hemsworth et al., 2000; Breuer et al., 2000; Waiblinger et al., 2002); the underlying assumption being that animals that are most fearful of people will keep the greatest distance. These measures are usually taken by placing the animal together with a person in a limited space and taking some measure of the extent that the

animal avoids or approaches the person. A distinction is usually drawn between the extent that an animal will voluntarily approach a stationary person (which we will call “approach” distance), and the distance that an animal will allow a person to approach (which we will call the “flight” distance). Across different studies, there has been little effort to standardise how such measures are taken. For example, approach distance has been measured either by the minimum or average distance that the animal approaches, the latency to make contact, or the amount of time the animal spends in contact or at various distances from the person.

## 2.2. *Handling tests*

A second type of test is to observe the animal’s responses when it is actually being handled. Responses to handling have been observed during normal or routine handling operations, such as milking for dairy cows (e.g. Hemsworth et al., 2000; Breuer et al., 2000), feeding of veal calves (Lensink et al., 2001), or during auction sales of beef cattle (e.g. Lanier et al., 2001). Since these tests include active handling of the animal, some standardisation of the handling is very important, otherwise variation in the handler’s skill or in the type of handling the animals receive could influence the results. However, an impressive diversity of tests have been used. With only a few exceptions, each research group appears to prefer to invent a new way of assessing reactions to handling rather than adopting someone else’s techniques; thus, there is little standardisation of measures and it is difficult to compare and collate results across studies. For example, some researchers use a measure based on the time taken to actually complete the handling task (e.g. Lensink et al., 2000, 2001; Boissy and Bouissou, 1988; Boivin et al., 1992). Other researchers take some measure that might indicate fear in the animals while they are being handled. These include behavioural measures, such as the incidence of kicking or stepping by dairy cows (Hemsworth et al., 2000; Breuer et al., 2000), running or kicking or escape attempts (Le Neindre et al., 1995; Lensink et al., 2001) or even acts of aggression against the handler (Price and Wallach, 1990).

## 2.3. *Rating scales*

A third alternative is to obtain some subjective rating of the animal. Some rating scales are grounded on descriptions of observable behaviour (e.g. Voisinet et al., 1997; Lanier et al., 2001), while other rating scales describe the overall “personality” or “temperament” of the animal and the observers are asked to make a subjective rating of the animal, rather than recording what the animal is actually doing.

In principle, any one of these types of measures could be used in on-farm assessment, although “distance” measures have received most attention in on-farm welfare assessment (e.g. Main et al., 2003; Rousing and Waiblinger, 2004). However, the relationship between the different types of measures and the relative merit of each measure as an indicator of human–animal relationships has not been adequately examined. Nor has there been enough attempts to standardise how these measures are taken by different researchers. This limits our ability to use the accumulated research to support the use of such measures in animal welfare assessment.

### 3. Reliability

On-farm assessment of animals' fear of people requires reliable and valid means of measuring the nature of the relationship between farm animals and people, especially how fearful the animals are of people. Reliability usually refers to the repeatability of the measure; that is if we take the measure more than once, how similar are the results from one time to another? This generally has two components, which can be referred to as inter-observer reliability (which refers to the chance that two different people will produce the same results) and test–retest reliability (which refers to the chance that the same results will be obtained if the test is repeated). In the context of on-farm welfare assessment, each farm will ideally be visited only once, and a number of observers should be able to be used more or less interchangeably.

Reliability can be reduced by differences between observers in how they score the observed behaviour, changes in single observers in how they score the behaviour, and changes in the animals in how they respond from one test to another. The reliability of behavioural tests can be reduced if it is difficult to actually measure the distance between the animals and the people. For example, in “laboratory” settings with video equipment and well-defined landmarks, it is possible to obtain accurate measurements of the distance between an animal and a person, with little risk of reduced reliability due to difficulty in taking the measure. However, in “field” conditions, e.g. on farms, such equipment may not be available and instead subjective judgements may be relied upon more. The fact that a measure of behaviour can be taken under laboratory conditions does not necessarily mean that it can be measured reliably on-farm.

In general, the reliability of measures of animals' fear of people did not seem to be a major concern for most researchers until recently. For example, in a random sample of 30 articles (published before 2003) that attempted to measure human–animal relationships, we found only five that explicitly assessed and reported on the repeatability of the tests used.

To demonstrate that “distance” measures can vary quite markedly over time, or with minor changes in the test situation, [Table 1](#) shows some of our own unpublished data where we repeatedly measured the approach distance of dairy calves at different ages in their home enclosure (1.8 m × 2.0 m), retesting them with two different people on the same day, and with the same person on two consecutive days. We have found low correlations in most cases and at most ages indicating that the repeatability of this test was quite low even

Table 1

Product-moment correlations between measures of the duration of contact between young dairy calves and people, assessed with two different people on the same day (two people), or with one person on two consecutive days (2 days), at ages of one day, two weeks, six weeks and six months

Age	Two people	2 days
Day 1	0.49	0.63
At 2 weeks	0.34	0.56
At 6 weeks	0.66	0.52
At 6 months	0.31 ns	0.32 ns

Except where indicated as ns (non-significant), all correlations are significant ( $P < 0.05$ ).

though we went to great trouble to retest in very similar situations. This admittedly limited data suggest that many of the measures of distance used actually may not be very repeatable.

Recently, two studies have rigorously evaluated the repeatability of measures of human–animal relationships and claim that the measures were repeatable (e.g. [Lensink et al., 2003](#); [Rousing and Waiblinger, 2004](#)). However, a closer inspection of the data leads us to a less sanguine view.

[Lensink et al. \(2003\)](#) measured milk-fed veal calves' fear of people by approaching the calves during the time that milk was being supplied to them. The calves' reaction to the arrival of an unknown person was scored on a two-point scale (withdrawal or not), and their response to an attempt to touch them on the head was assessed on a four-point scale: 1 = no withdrawal, 4 = strong withdrawal. Test–retest repeatability was assessed by repeating the test after a two-day interval. At the arrival of the unknown person, 84% of calves were scored the same way on both instances. This seems encouraging, but it does show that even with a simple two-point scale, 16% of the calves risk being misclassified by a single test. During the attempted touching, the scores on the two days were correlated significantly, but the coefficient was only moderate ( $r = 0.62$ ) indicating that less than 40% of the variance between calves on one day was related to the variance on the other days. The results presented show a number of discrepancies. For example, the calves that were judged as most fearful on day 1 had only a 57% chance of being the most fearful on day 2, and had a 7% chance of being judged the least fearful. The authors claim that this indicates that the measure is “repeatable”, but the question remains as to just how repeatable a measure must be to be used in on-farm assessment.

[Rousing and Waiblinger \(2004\)](#) tested both the inter-observer reliability and the test–retest reliability of two methods of scoring dairy cows' fear of people, an “approach test” where the latency of cows to approach within a defined distance of a stationary person was measured, and an “avoidance” test (which for purposes of consistency we will call a flight distance test), where a person approached the cow and the distance at which the cow moved off was measured.

In general, there seemed few differences between observers. Inter-observer repeatability of the approach test was high (0.97) indicating that there were few differences between the observers. Inter-observer repeatability of the flight distance test was also significant, with weighted kappa values ranging from 0.85 to 0.90, which the authors considered as high. Test–retest values were less encouraging, however. Measures of concordance were significant, indicating that the two days were probably not unrelated. But considerable changes did occur. On the flight distance test, only 52% of the animals were classified the same way on the two days, while 13% of the animals differed by two categories or more (on a 5-point scale). Some of these changes were quite large, for example 15% of the cows that were scored as the most fearful on day 1 (in that they withdrew at a distance of more than 2 m), were classified as among the least fearful (in that they let the person approach to within touching distance) on day 2.

One of the largest difficulties in assessing tests of reliability arises from the lack of clear criteria for deciding when reliability is adequate. Finding statistically significant correlations between two measures simply shows us that the two measures are not likely to be completely unrelated. However, we must not forget that even with a correlation

coefficient of 0.7, less than 50% of the variance in the scores is common between the two tests. Significant and moderate kappa values can be obtained even where a substantial percent of the animals are misclassified. No tests have yet been published examining the repeatability of measures based on farm averages. However, from the data that has been presented on test–retest analyses of the scores of individual animals, we conclude that a number of farms risk being misclassified if scored in a single visit. In situations where the results of a test may impact the livelihood of a farmer (e.g. welfare audits by food retailers) it is essential to have very high reliability, not merely statistical significance.

#### 4. Validity

Validity is much harder to assess than reliability, especially in animal welfare or behavioural research, because we usually lack a well-defined standard with which we can compare our test. Validity of measures of animals' responses to people can be judged by experimentally altering the animals' degree of fear of people, and assessing to what extent the measures are able to detect this. However, often validity is judged circumstantially by the extent that the test is not affected strongly by other variables, and by the extent that different types of tests produce the same conclusion. Validity can be reduced if the measure is sensitive to changes in supposedly minor parameters of the test that should not influence the results. For example, on-farm measures of fearfulness of people should not be strongly affected by the clothing the people wear, or the location of the test. Finding strong effects of such variables would throw doubt on the validity of the test.

In order to judge the validity of a measure, we need to be clear about what exactly the measure is supposed to be measuring. For example, are we measuring the animal's responses to people in general or to one person (the farmer) in particular? Are we assessing the animal's general responses to people, or a particular type of response, e.g. degree of fearfulness, ease of handling, etc.? Are we interested in using the animals' responses to people to assess how they are handled, or as one more indicator of their welfare? Unfortunately, in many studies, the precise goal of the measure is sometimes not made clear.

##### 4.1. *Experimental induction of fear*

The most convincing evidence that the use of aversive handling causes animals to become fearful of people, and that this is apparent in the way the animals react to people, comes from experimental studies in which animals are deliberately handled gently or roughly. Numerous studies on swine, cattle and poultry (reviewed in [Hemsworth and Coleman, 1998](#); [Hemsworth, 2003](#)) show that aversive handling of animals will lead them to avoid people while gentle handling will lead them to approach people. Furthermore, the avoidance of people by aversively handled animals is often associated with physiological indicators of stress, such as elevated corticosteroids. Most recently, for example, [Breuer et al. \(2003\)](#) handled some heifers aversively and others gently. Clear effects of the treatment were noted: aversively handled heifers showed a greater flight distance, increased approach latencies, fewer interactions with the handler, were more agitated in a crush and had higher cortisol concentrations in the presence of the handler.

The relationship is even more convincing where animals have been trained to avoid one person through aversive handling and approach another person through gentle handling. For example, Rushen et al. (1999a) trained cows to recognise two different people, one of whom always treated the cows well and the other always treated the cows roughly. After several treatments, the cows approached the gentle handler closer than the aversive handler. That this indicates some fearfulness of the people was shown by the fact that residual milk and heart rate were higher when the cows were milked in the presence of the aversive handler. This study provides clear evidence that the reaction of an animal to people can predict how the animal was handled and indicate whether or not the animal is likely to show any physiological signs of stress.

However, the relationship between the animals' responses to people and other stress responses can be variable. A subsequent study (Munksgaard et al., 2001), which effectively used the same similar experimental protocol as Rushen et al. (1999a), but less aversive handling, found the same magnitude of effect upon the animals' withdrawal responses from the people but found no effects upon residual milk. Thus, even in highly controlled experiments, the relationship between the animals' behavioural responses to people and other stress measures can be variable and observations of animals' responses to people may not predict their physiological stress responses.

Furthermore, the results of the studies show the difficulties in establishing a standardised "cut-off" point that could be used in welfare audits to determine that the animals' degree of fear of people is sufficiently high to warrant some intervention. Many studies have reported a statistically significant difference between the effect of aversive handling and gentle handling on animals' responses to people. However, a highly significant difference between handling treatments does not mean that the effect of the treatment is large, only that it is highly unlikely to have occurred by chance. When we look at the size of some of the effects of type of handling on animals' responses to people, we discover that these are often quite small. For example, in the study by Breuer et al. (2003) the magnitude of the differences cast some doubts on whether aversively handled heifers were substantially more fearful of people than gently handled ones. Even though approach latencies were higher for aversively handled heifers than for gently handled ones, the aversively handled heifers still took an average of only 165 s to approach within 1 m of the experimenter, compared to only 120 s for gently handled ones. Thus, the magnitude of the results does not give the impression of highly fearful animals, nor of a large difference between gently and aversively handled animals.

We must ask the question whether these differences are large enough to be useful in differentiating between farms where the animals are handled well and those where there is a handling problem that is sufficiently large to reduce the animals' welfare.

#### *4.2. Correlational approach to establish validity*

One way of assessing the validity of measures of animals' fear of people is to examine whether these measures can predict the use of aversive handling techniques by the farmer. This approach has been adopted in a number of studies that have measured dairy cows' fear of people on dairy farms. Three studies in particular provide data on the correlations across farms between various measures of the fear of dairy cows for people, and measures of the

Table 2

Correlations reported in the literature between use of gentle handling (POS) and aversive handling (NEG), (which can be either mild (NEG1) or strongly aversive handling (NEG2)) or neutral handling (NEU) and various measures of approach distance of dairy cows to people

Hemsworth et al. (2000)	Time, 1 m	Time, 3 m	FDA	FDP	FS + FSKon
POS	-0.1	0.02	-0.46*	0.10	-0.27*
NEG1	-0.28*	-0.19	0.13	-0.25*	0.22
NEG2	-0.27*	-0.21	0.13	-0.02	0.22
Breuer et al. (2000)	Time 3 m			FS + FSKon	
POS	0.10			0	
NEG1	0.02			-0.21	
NEG2	-0.28			0.04	
Waiblinger et al. (2002)	FDBarn		SK F		
Collection					
POS	-0.11		-0.12	-0.40*	
NEU	-0.28		0.34	0.03	
NEG	0.02		0.02	-0.10	
Parlour					
POS	-0.46*		0.09	-0.04	
NEU	0.32		0.40*	0.02	
NEG	0.47*		0.32	0.11	

Note: Time 1 m, time 3 m = time the animal spends within 1 or 3 m of the person; FDA, FDP and FDBarn = flight distance of cow measured in an arena, paddock or barn; F, S, K = components of flight/step/kick response, Parlour = measure taken in milking parlour, Collection = measure taken when animals collected for milking.

\* Statistically significant  $P < 0.05$ .

type of handling techniques used. Two studies were done in Australia (Hemsworth et al., 2000—66 farms; Breuer et al., 2000—31 farms) and one in Austria (Waiblinger et al., 2002—30 farms). All of these studies report significant correlations between the measures of the cows' fear of people and the type of handling technique used, which would seem to support the validity of the measures used.

However, the details of the results warn us against too optimistic a conclusion. In Table 2 we present some of the main correlations between handling techniques and cows' fearfulness, trying as far as possible to present results based on comparable measures. First, the correlations are usually small or moderate. None of the correlations are above 0.5, suggesting that at most only 25% of the variance between farms in handling techniques is related to variance in the measures of cows' fearfulness. Furthermore, only a proportion of the correlations are significant. The proportion of the correlations presented between the cows' behaviour and the behaviour of the stockperson which are significant are 4/18 (Breuer et al., 2000), 16/49 (Hemsworth et al., 2000) and 22/70 (Waiblinger et al., 2002). Thus, while we cannot deny that the cows' behaviour is related to that of the stockperson, it would be difficult to make any precise predictions of what types of handling techniques were used solely on the basis of observing the cows.

Second, there are a number of differences between the studies in which exact measures are correlated. For example, Hemsworth et al. (2000) measured flight distance defined as

the distance that the cow first moved away from an approaching person. This was done in the paddock (FDP) and in a specially constructed test arena (FDA). [Table 2](#) shows that flight distance in the arena was negatively correlated with the use of positive handling techniques. Thus, a small flight distance would indicate the use of gentle handling. However, this relationship was not found for flight distance in the paddock, which, rather paradoxically, was negatively correlated with the use of aversive handling techniques. [Waiblinger et al. \(2002\)](#) presented a similar measure of flight distance (although they called it approach distance) which was measured in the barn (FDBarn). In contrast to [Hemsworth et al. \(2000\)](#) they reported ([Table 2](#)) that this measure of flight distance was both negatively correlated with positive handling techniques and positively correlated with aversive handling techniques. However, this was true only for the handling that was done in the milking parlour: flight distance did not appear to relate to the handling used when collecting cows. All three studies also took measures of the cows' behaviour during milking by recording each instance of flinching, kicking or stepping during milking. Again, [Table 2](#) shows substantial discrepancies between the studies in how this measure related to the way the animals were handled.

It is not our purpose to question the general finding of these studies viz. that use of aversive handling techniques is associated with more fearful cows. This general conclusion seems undeniable. The differences between the studies probably reflect relatively small differences in how measures were taken, how handling techniques were classified etc., and are probably not important in affecting the overall conclusions of the research. However, the fact that there are so many differences between the studies in which specific measures of cows' fearfulness are predictive of the handling techniques used leads us to question whether we yet have measures of cows' fearfulness that are sufficiently precise to be used in on-farm welfare audits. For example, it would be difficult to conclude from the three studies whether flight distance or the time in proximity to a person is the best way of measuring cows' fear of people. The differences between the studies suggest that the exact manner in which handling techniques influence the behavioural response of dairy cows to humans can vary greatly according to the context. This would make it difficult to develop a standardised test that could be used with confidence on all dairy farms. Most importantly for our discussion, the low correlations indicate that any measure of cows' apparent fearfulness is only a poor predictor of how the cows are handled. Consequently, the validity of the tests in detecting differences between farms in how the animals are handled is low.

#### *4.3. Tracking changes on farms*

Can measures of human–animal relationships track changes in commercial farms in the way animals are handled? [Hemsworth et al. \(2002\)](#) conducted a large scale intervention study based on a special training program to improve the attitudes of stockpeople to dairy cows. Following the training program, the stockpeople were found to use a lower number of aversive handling treatments and a higher number of gentle or positive handling treatments. The effects were substantial: the number of positive behaviours used on the intervention farms was more than double that on control farms, while the number of forceful negative behaviours was a fifth. The flight distance of the cows was found to be significantly lower on the farms on which the stockpeople had followed the training programs, however, the

effect was small: flight distance on control farms was 4.49 m while on intervention farms it was 4.16 m. Furthermore, there were no significant effects on the occurrence of flinch, step and kick responses during milking. Thus, the cows' behaviour was a poor predictor of changes in the way that the stockperson handled the cows.

A similar study with stockmen on a large commercial piggery (Coleman et al., 2000) found that an appropriate course aimed at targeting attitudes to animals did result in a substantial reduction in use of aversive handling by stockpeople who followed the course. However, there was no significant reduction in the pigs' tendency to withdraw when the stockman approached, although there was a non-significant trend in this direction ( $P = 0.11$ ).

These two intervention studies show that it would be very difficult to track even quite substantial improvements in the stockmen's handling of the animals by using changes in the animals' responses to people. These results support our concern about the validity of using these tests in on-farm welfare assessment or in animal welfare audits.

#### 4.4. *Do different measures measure the same thing?*

If tests are valid, then two tests that claim to measure the same thing should produce the same results. Few comparisons between alternative measures of human–animal relations have been done. Most data available is on the comparison between approach distance and flight distance.

Interestingly, approach distance and flight distance appear not to be alternative ways of measuring the same thing, but rather measures of different aspects of the relationship between people and animals. In beef cattle, Murphey et al. (1981) found that flight distance and approach distance were not correlated across different breeds and herds of cattle. The two different “distances” may also reflect different aspects of how the animals are treated by the stockperson: Hemsworth et al. (2000) found that flight distances of dairy cows tended to be negatively correlated with the extent that the stockperson used positive interactions, while approach distances tended to be correlated with the use of negative handling by the stockperson. Experimental treatments also appear to affect flight and approach distances differently. Jago et al. (1999) found that hand feeding of calves reduced approach distances but had no effect on flight distance. Although, Krohn et al. (2001) reported that extra handling affected both, the approach distance appeared to be more sensitive to the age at which the handling was received than was flight distance.

Clearly, approach distance and flight distance reflect different aspects of the relationship between people and animals, but we have insufficient information, as well as a lack of conceptual models, to determine just what these different aspects may be. Thus, we have insufficient data to judge how other measures of human–animal relationships are interrelated.

#### 4.5. *Effect of motivations other than fear*

It has been clearly demonstrated that rough handling of animals by people leads them to become fearful of people and avoid them. But can the animals' tendencies to approach or avoid people reflect only their degree of fear of them or can it reflect the influence of other

motivations? In many cases, the extent that animals avoid people will likely reflect a mix of motivations including fear and curiosity (Murphey et al., 1980, 1981). Similarly, an animal that is used to being fed by people may approach people closer than animals that have not been fed by people. However, this may not so much indicate the degree of fear as the extent that the animals expect to receive food. Jago et al. (1999) found that calves that had been hand fed by people approached people closer than calves that had been gently patted and stroked. Hand feeding of calves also increases the chance that they will suck or butt at the people (Jago et al., 1999; Krohn et al., 2001, 2003) again suggesting that feeding motivation may be playing a major role in the extent that the animals approach people.

That an animal's response to people may reflect feeding motivation, rather than fear, is particularly problematic for tests that examine animals' response to people in the presence of food. For example, Lensink et al. (2000) tested calves for the tendency to withdraw from feed when a person approached the feed trough. This was used to assess the degree of fear the calf felt for the person, but it may also reflect their level of feeding motivation. This is best demonstrated in the study of Lankin (1997) on sheep. In this study, sheep's responses to people were measured by having the person stand near the feed trough and recording how much the sheep were willing to approach the person in order to eat. The assumption was that fearful sheep would stay away from the person and so eat less. However, Lankin (1997) found that this measure was also (not surprisingly) affected by the sheep's level of hunger. Food deprived sheep approached the person closer than satiated sheep, and sheep approached the person closer when the test was done a long time since the last meal compared to when the test was done soon after the meal.

As well as feeding motivation, it seems probable that an animals responses to people may also be affected by their degree of curiosity. Thus it seems likely that an animal's response to people in such tests will reflect a mix of factors depending on the particular context of the test, and we need to be very careful in assuming that the response simply measures the degree of fear that the animals show. One reason why distance measures are influenced by many factors is that there is often no cost to the animal to avoid people. We have recently examined the possibility of using measures of vigilance to evaluate the quality of the man–animal relationship (Welp et al., 2004). Vigilance behaviour is costly to the animals because they often must stop eating in order to be vigilant. By increasing the cost to the animal of expressing its fear of people in this way, we may more easily interpret changes in vigilance as reflecting changes in animals' fearfulness.

#### 4.6. *Effects of context*

To be valid, the tests should not be strongly affected by minor “contextual” variables that do not strongly impact the animals' degree of fear.

There are many different ways of doing these tests, and some of the differences may affect how the animals react to the people. For example, different studies may use different size enclosures, allow different time latencies for the animals to make contact, either the person or the animal can enter the enclosure first (the latter making it difficult to control for initial distances when the person entered) and the person can be sitting, standing, or even on a horse. These may seem relatively minor points, and if standardised within an experiment, will not necessarily result in any confounding of the results. However, we know little about

how these extraneous variables do affect our measures, and they will limit our ability to evaluate these factors and identify a valid on-farm test. Furthermore it may not be possible to control for these factors on farms, making a standardised test more difficult to develop.

#### 4.6.1. *Effect of location*

The context or location in which the test is applied may play a role. Some studies test animals' reactions in their home area, while others move the animals to another, often unfamiliar area. Some studies do find that the location of the test has no or only slight effects on the tests. For example, the effects of increased early handling on calves' approaches to people were apparent in both the calves' home pens and in an unfamiliar pen (Krohn et al., 2001, 2003). Other studies, however, show a strong effect of location of the test. Calves will approach a person who is gentle to them, and avoid another person who has handled them roughly when tested in the pen where the handling occurred, but may react quite differently if tested in another place (de Passillé et al., 1996). When cows were tested in their own stall and in an unfamiliar stall, the correlation between the distance kept from a person in the two stalls was low and non-significant, even though the stalls were very similar (Munksgaard et al., 1997; Rushen et al., 1998). Similarly, calves that were hand fed interacted more with an unfamiliar person than calves that were not hand fed when they were tested in their home pen, but no differences were found when tested in an unfamiliar pen (Jago et al., 1999).

When animals' fear of people is tested in a novel area, their responses will reflect not only the response to the people but also their response to the novel area. Generally, farm animals isolated in novel areas show responses suggesting increased arousal and a great deal of exploratory behaviour. In some cases, the animals are placed in the arena before the person enters (e.g. Boivin et al., 1998), which may help reduce the effects of novelty of the arena. However, we must be aware that an animal's responses to humans when tested in novel/unfamiliar areas may be unrelated to those in its home environment, because the reactions to a person in a novel area will be a measure of the combination of exploration of the area and fear of the person.

#### 4.6.2. *Identity of the "test" person*

In a welfare assessment, the interest is in how animals are handled by their regular caretaker. However, in many cases it may be necessary to measure animals' responses to another person, with the hope that the animals' responses to the stranger will reflect how they will respond to the caretaker.

Many recent studies show that aversive or gentle handling affects cattle's responses to familiar and unfamiliar people in a similar way (Krohn et al., 2001; Breuer et al., 2003). Lensink et al. (2001) found moderate positive correlations between veal calves' readiness to approach the usual stockperson and an unfamiliar person. Rousing and Waiblinger (2004) found that flight distances of dairy cattle to familiar and unfamiliar people were correlated.

However, a number of studies have shown that cattle and pigs do respond differently in these sorts of tests according to whether the person is familiar or not (reviewed in Rushen et al., 1999b). Certainly, cattle will react very differently to people who treat them differently (Munksgaard et al., 1997, 1999; Rushen et al., 1999a) and evidence is

accumulating that some cattle can use very fine features, such as facial features, to recognise individual people (Rybarczyk et al., 2001). Even a change of clothing can have quite large effects on how cattle react to people (Munksgaard et al., 1999). Thus, it is hazardous to assume that the identity of the person will have no effect.

At this stage we do not fully understand why farm animals sometimes treat all people similarly, while at other times they respond quite differently to different people. Perhaps the number of different people that the animals normally encounter is a factor. But until we understand the factors that do cause animals to react differently to different people, we cannot simply assume that their responses to an unfamiliar person will reflect their responses to their usual caretakers.

#### 4.6.3. *Early rearing conditions*

The degree of contact with conspecifics during early rearing can influence how animals respond to people, independently of how they are handled. For example, Krohn et al. (2003) showed that calves tend to approach people less if they have been reared with other calves, and that rearing them with their dams may limit the effects of extra gentle handling on increasing their approaches to people. Again, these results show that animals' reactions may not reflect the way that the animals have been handled by their caretakers.

Thus, many studies have shown that animals' responses can be (but are not always) affected by the contextual factors that should not strongly influence the animals' fear of people. This throws some doubt on the validity of the measures used. The fact that some contextual factors (such as the identity of the test person) do not always have an effect does not help us, unless we can predict with some certainty whether or not they will be operating in any given test.

## 5. Is animals' fearfulness the most important aspect of stockmanship?

Measures of animals' responses to people measure only one component of stockmanship: the extent that animals are fearful of people. However, there may be other components of stockmanship that are more important for animal welfare. The stockperson can most obviously affect the welfare of animals through the way that routine animal care tasks, such as feeding, cleaning etc., are done. Despite growing recognition of its importance for good animal welfare, as well as good animal productivity, this component of stockmanship has been investigated scientifically and in detail in only a small number of studies.

In one of the most comprehensive studies of stockmanship, Lensink et al. (2001) examined the role of stockmanship in affecting the health and productivity of veal calves on farms in France. Farmers were interviewed, and the researchers scored their attitudes to the animals, e.g. whether or not they believed that calves were sensitive to human contact, and their attitudes to the work routines, e.g. how important cleaning procedures were. The farms were also scored for cleanliness, and the performance of various management routines was noted. In addition, interactions between the farmers and the calves were observed, especially the extent that calves withdrew when the farmers approached, and the extent that the farmer engaged in positive interactions with the calves.

Sizeable correlations between stockmanship variables and the production characteristics of the farm were noted. High producing units (i.e. those with high daily weight gains, good food conversion efficiencies, and low mortality) had healthier calves, tended to be cleaner, had crates disinfected by an external company, had Sunday evening feedings of the calves, and were run by farmers whose own parents had managed a veal unit. The cleanliness of the barns accounted for 19% of the variance between units in daily weight gain and 22% of the variance between units in feed efficiency. The health of the calves was correlated with the attitudes of the farmers, e.g. the more the farmer believed that calves were sensitive to human contact, and the more the farmer felt that cleaning was important, the better the health of the calves. The results show the importance of general stockmanship for the welfare and productivity of the calves, particularly as concerns the care taken in cleaning the facilities. However, the responses of the calves to the stockperson were relatively unimportant in explaining variation between farms in health and productivity of the calves once the effect of these other stockmanship variables were considered.

This study clearly indicates that the quality of stockmanship cannot be adequately assessed by a few measures of animals' behavioural responses to people.

## **6. Alternatives for welfare audits: have stockpeople followed a course in animal handling?**

The way stockpeople handle animals is likely to be a reflection of long-held beliefs about how animals need to be handled and attitudes towards animals in general. The stockpeople's general and specific attitudes to and belief about animals are discussed in [Hemsworth and Coleman \(1998\)](#). A considerable amount of innovative research has now shown that the way that stockpeople handle animals is a reflection of specific beliefs of the stockperson, and that altering these beliefs may be an effective means of improving the way animals are handled ([Hemsworth and Coleman, 1998](#); [Hemsworth, 2003](#)). For example, [Hemsworth et al. \(2002\)](#) examined the effect of a "cognitive-behavioural intervention" on dairy farmers' attitudes towards cows. The intervention clearly improved the attitude of the farmers towards dairy cattle, specifically reducing the belief that considerable force was needed to move dairy cows. Visits to the farms showed that these changes in beliefs resulted in a reduced use of aversive handling techniques, a reduced fearfulness of the cattle, and some evidence of improvement in milk yield.

This study clearly shows the potential for such interventions to improve at least one component of stockmanship, and to improve both the welfare and the productivity of the cattle. A more cost-effective way of auditing stockmanship in on farm visits may, therefore, be to ask whether or not the stockperson has followed a suitable course!

## **7. Conclusions**

From this brief discussion, we suggest that insufficient attention has been paid to the problem of assessing the reliability and the validity of the tests used to assess human-animal relationships within the specific context of on-farm welfare assessments or audits.

The great variety and diversity of tests and measures used by different researchers makes it difficult to develop a standardised test that has solid backing in accumulated research. We have insufficient information to judge whether different measures are measuring the same thing, or whether different measures assess different aspects of the human–animal relationship. The reliability of the measures used is often low, even if statistically significant, and the level of reliability that will be considered adequate in on-farm assessment has not been determined. Appropriate cut-off points have not been established. The results of the measures can be affected by extraneous variables such as the clothing worn, the location of the test, etc., which reduces our confidence in both their validity and reliability. Animals' responses to people are likely to reflect a mix of motivations, and we should not assume that the level of fear is the most important. Clearly more attention needs to be paid to standardising the tests used. The effects of stockmanship on animal welfare are too complex for the adequacy of stockmanship to be assessed by measures of animals' responses to people, and there may be more cost-effective ways of assessing the quality of overall stockmanship.

## Acknowledgements

This paper was presented at the Summer Congress of the International Society for Applied Ethology in Abano Terme, June 2003. We thank all of you who have discussed this issue with us at various times. We particularly thank John Barnett and Michelle Edge for many helpful discussions on the difficulties of conducting animal welfare audits. Jeff Rushen thanks the OECD for travel funds.

## References

- Barnett, J.L., Glatz, P.C., Almond, A., Hemsworth, P.H., Cransberg, P.H., Parkinson, G.B., Jongman, E.C., 2001. A Welfare Audit for the Chicken Meat Industry. Rural Industries Research and Development Corporation, Canberra, Australia.
- 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. *Livest. Prod. Sci.* 61, 179–192.
- Boissy, A., Bouissou, M.-F., 1988. Effects of early handling on heifers' subsequent reactivity to humans and to unfamiliar situations. *Appl. Anim. Behav. Sci.* 20, 259–273.
- Boivin, X., Le Neindre, P., Chupin, J.M., 1992. Establishment of cattle–human relationships. *Appl. Anim. Behav. Sci.* 32, 325–335.
- Boivin, X., Garel, J.P., Mante, A., Le Neindre, P., 1998. Beef cattle react differently to different handlers according to the test situation and their previous interactions with their caretaker. *Appl. Anim. Behav. Sci.* 55, 245–257.
- Breuer, K., Hemsworth, P.H., Barnett, J.L., Matthews, L.R., Coleman, G.J., 2000. Behavioural responses to humans and productivity of commercial dairy cows. *Appl. Anim. Behav. Sci.* 66, 273–288.
- Breuer, K., Hemsworth, P.H., Coleman, G.J., 2003. The effect of positive or negative handling on the behavioural and physiological responses of nonlactating heifers. *Appl. Anim. Behav. Sci.* 84, 3–22.
- Coleman, G.J., Hemsworth, P.H., Hay, M., Cox, M., 2000. Modifying stockperson attitudes and behaviour towards pigs at a large commercial farm. *Appl. Anim. Behav. Sci.* 66, 11–20.
- de Passillé, A.M., Rushen, J., Ladewig, J., Petherick, J.C., 1996. Dairy calves' discrimination of people based on previous handling. *J. Anim. Sci.* 74, 969–974.

- Hemsworth, P.H., Coleman, G.J., 1998. Human-Livestock Interactions: The stockperson and the Productivity and Welfare of Intensively-farmed Animals. CAB International, Oxon, UK.
- Hemsworth, P.H., Coleman, G.J., Barnett, J.L., Borg, S., 2000. Relationships between human-animal interactions and productivity of commercial dairy cows. *J. Anim. Sci.* 78, 2821–2831.
- Hemsworth, P.H., Coleman, G.J., Barnett, J.L., Borg, S., Dowling, S., 2002. The effects of cognitive behavioral intervention on the attitudes and behavior of stockpersons and the behavior and productivity of commercial dairy cows. *J. Anim. Sci.* 80, 68–78.
- Hemsworth, P.H., 2003. Human-animal interactions in livestock production. *Appl. Anim. Behav. Sci.* 81, 185–198.
- Jago, J., Krohn, C.C., Matthews, L.R., 1999. The influence of feeding and handling on the development of the human-animal interactions in young cattle. *Appl. Anim. Behav. Sci.* 62, 137–152.
- Krohn, C.C., Jago, J., Boivin, X., 2001. The effects of early handling on the socialization of young calves to humans. *Appl. Anim. Behav. Sci.* 74, 121–133.
- Krohn, C.C., Boivin, X., Jago, J., 2003. The presence of the dam during handling prevents the socialization of young calves to humans. *Appl. Anim. Behav. Sci.* 80, 263–275.
- Lanier, J.L., Grandin, T., Green, R., Avery, D., McGee, K., 2001. A note on hair whorl position and cattle temperament in the auction ring. *Appl. Anim. Behav. Sci.* 73, 93–101.
- Lankin, V., 1997. Factors of diversity of domestic behaviour in sheep. *Genet. Sel. Evol.* 29, 73–92.
- Le Neindre, P., Trillat, G., Sapa, J., Ménissier, F., Bonnet, J.N., Chupin, J.M., 1995. Individual differences in docility in limousin cattle. *J. Anim. Sci.* 73, 2249–2253.
- Lensink, B.J., Boivin, X., Pradel, P., Le Neindre, P., Veissier, I., 2000. Reducing veal calves' reactivity to people by providing additional human contact. *J. Anim. Sci.* 78, 1213–1218.
- Lensink, B.J., Raussi, S., Boivin, X., Pyykkonen, M., Veissier, I., 2001. Reactions of calves to handling depend on housing condition and previous experience with humans. *Appl. Anim. Behav. Sci.* 70, 187–199.
- Lensink, B.J., van Reenen, C.G., Engel, B., Rodenburg, T.B., Veissier, I., 2003. Repeatability and reliability of an approach test to determine calves' responsiveness to humans: "a brief report". *Appl. Anim. Behav. Sci.* 83, 325–330.
- Main, D.C.J., Whay, H.R., Green, L.E., Webster, A.J.F., 2003. Preliminary investigation into the use of expert opinion to compare the overall welfare of dairy cattle farms in different farm assurance schemes. *Anim. Welfare* 12, 565–569.
- Mench, J.A., 2003. Assessing animal welfare at the farm and group level: a United States perspective. *Anim. Welfare* 12, 493–503.
- Munksgaard, L., de Passillé, A.M., Rushen, J., Thodberg, K., Jensen, M.B., 1997. Discrimination of people by dairy cows based on handling. *J. Dairy Sci.* 80, 1106–1112.
- Munksgaard, L., Rushen, J., de Passillé, A.M., 1999. Dairy cows' use of colour cues to discriminate between people. *Appl. Anim. Behav. Sci.* 65, 1–11.
- Munksgaard, L., de Passillé, A.M., Rushen, J., Herskin, M.S., Kristensen, A.M., 2001. Dairy cows' fear of people: social learning, milk yield and behaviour at milking. *Appl. Anim. Behav. Sci.* 73, 15–26.
- Murphey, R.M., Moura Duarte, F.A., Torres Penedo, M.C., 1980. Approachability of bovine cattle in pastures: breed comparisons and a breed X treatment analysis. *Behav. Genet.* 10, 171–181.
- Murphey, R.M., Moura Duarte, F.A., Torres Penedo, M.C., 1981. Responses of cattle to humans in open spaces: breed comparisons and approach-avoidance relationships. *Behav. Genet.* 11, 37–48.
- Price, E.O., Wallach, S.J.R., 1990. Physical isolation of hand-reared Hereford bulls increases their aggressiveness toward humans. *Appl. Anim. Behav. Sci.* 27, 263–267.
- Rousing, T., Waiblinger, S., 2004. Evaluation of on-farm methods for testing the human-animal relationship in dairy herds with cubicle loose housing systems—test-retest and inter-observer reliability and consistency to familiarity of test person. *Appl. Anim. Behav. Sci.* 85, 215–231.
- Rushen, J., Munksgaard, L., de Passillé, A.M., Thodberg, K., Jensen, M.B., 1998. Location of handling and dairy cows' responses to people. *Appl. Anim. Behav. Sci.* 55, 259–267.
- Rushen, J., Munksgaard, L., de Passillé, A.M., 1999a. Fear of people by cows and its effects on milk yield, behavior and heart rate at milking. *J. Dairy Sci.* 82, 720–727.
- Rushen, J., Taylor, A.A., de Passillé, A.M.B., 1999b. Domestic animals' fear of humans and its effect on their welfare. *Appl. Anim. Behav. Sci.* 6, 285–303.

- Rybarczyk, P., Kobe, Y., Rushen, J., Tanida, H., de Passillé, A.M., 2001. Can cows recognize people by their faces? *Appl. Anim. Behav. Sci.* 74, 175–189.
- Voisinet, B.D., Grandin, T., Tatum, J.D., O'Connor, S.F., Struthers, J.J., 1997. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. *J. Anim. Sci.* 75, 892–896.
- Waiblinger, S., Menke, C., Coleman, G., 2002. The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour of dairy cows. *Appl. Anim. Behav. Sci.* 79, 195–219.
- Welp, T., Rushen, J., Kramer, D.L., Festa-Bianchet, M., de Passillé, A.M.B., 2004. Vigilance as a measure of fear in dairy cattle. *Appl. Anim. Behav. Sci.* 87, 1–13.