INDIVIDUAL DIFFERENCES IN TEMPERAMENT OF DOMESTIC ANIMALS: A REVIEW OF METHODOLOGY

X Manteca[1†] and J M Deag[2]

[1] Department of Physiology, School of Veterinary Science, Universitat Autonoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[2] Institute of Cell, Animal and Population Biology, University of Edinburgh, UK
[†] Contact for correspondence and requests for reprints

Abstract *Animal Welfare 1993, 2:* 247-268

Individual differences in behaviour may be examined at two levels. First, individuals may differ in terms of frequencies, durations and/or patterning of particular measures of their behaviour. Second, individuals may differ in their temperament, ie in the way they react to environmental change and challenge. Individual differences in temperament are particularly relevant to animal welfare studies, for the welfare of an individual largely depends on whether it can cope with environmental challenge. Whereas the study of individual differences in behaviour at the first level may be achieved by using standard behavioural methods, the study of individual differences in temperament requires the use of more unusual methods, namely observers' ratings and behavioural tests. Observers' ratings provide information on subtle aspects of an individual's behaviour that could otherwise be overlooked. Behavioural tests facilitate comparisons between individuals in a more standardized way. It is suggested that both systems should be used together.

Taking individual differences into account when designing experiments may help reduce variability in studies on welfare issues and understanding the causes of individual differences in temperament may allow us to reduce the incidence of some welfare problems.

Keywords: *animal welfare, behaviour, individual differences, temperament*

Introduction

Individual differences in behaviour (ie differences in behaviour that cannot just be attributed to age and sex, and which show consistency over time and/or between contexts), although widely recognized, have not been until recently a topic of general scientific interest (Goodall 1986). Indeed, early ethologists tended to emphasize the similarities across individuals of the same species rather than their differences (Huntingford 1984). The general perception has now changed and individual differences in behaviour are regarded as the rule rather than the exception (Bekoff 1977).

One of the reasons for this current interest in individual differences is that they are relevant to the science of animal welfare (eg Wiepkema *et al* 1987). As a result of this, one goal is to understand the causal links between the numerous stressors acting on animals in captivity and their behavioural responses. Working at the level of individual differences in behaviour will help to tease apart the numerous factors involved in the causation and development of behaviour, and how animals adjust their behaviour to captivity.

If progress is to be made in this latter area, a proper understanding of the methods used in the study of individual differences is required. However, so far most of the work on individual differences in behaviour has been on primates or laboratory animals, and the information on domestic species is scattered and scanty.

This paper reviews the methods used to assess individual differences in behaviour in domestic species. Since some of these methods were originally developed by primatologists, some references to work on primates have also been included. When doing so, however, our aim has not been to provide an updated review of the field of individual differences in primates but only to illustrate how methods developed for work on primates have been applied to domestic species.

Levels of measurement

Individual differences in behaviour may be examined at two levels. First, individuals can differ in terms of frequencies, durations and/or patterning of particular measures of their behaviour (Mendl & Harcourt 1988). For example, a particular individual in a flock of hens may initiate a higher number of aggressive interactions per unit time than any other individual (Cunningham & von Tienhoven 1983).

At the second level, individuals may differ in their temperament. Temperament can be seen as the external manifestation, characteristic of an individual, in which the different behaviours interact temporally and are modulated in intensity. In other words, temperament refers to a more comprehensive view of an individual's behaviour (Lawrence 1981, Feaver *et al* 1986, Mendl & Harcourt 1988). In particular it places a particular emphasis on the way in which individuals react to environmental change and challenge (Lawrence *et al* 1991). Since temperament refers particularly to responses to environmental challenges, the study of temperament is most relevant to animal welfare, for the welfare of an individual depends on whether it can cope with environmental challenge (Broom 1988).

Methods of measurement

There are three main sources of behavioural data for assessing individual differences in behaviour. First, behaviour may be recorded using standard behavioural methods. Second, individual differences may be assessed by using observers' ratings or alternatively scales with predefined scores. Third, behavioural tests may be used (Stevenson-Hinde 1983).

Whereas the study of individual differences in behaviour at the first level may be done using standard behavioural methods, the study of differences in temperament requires the use of more unusual methods, namely observers' ratings or scales with predefined scores, and behavioural tests.

Observers' ratings

Description of the method

This method uses recording forms listing all the individual animals to be rated (typically by several observers acting independently), with a linear scale drawn for each individual animal (Martin & Bateson 1986).

One form is used for each component of behaviour (item). A cross is marked on each scale at a position corresponding to the observer's overall assessment of how much that individual expresses that item. The distance from the left-hand end of the line to the position of each cross is used as the respective numerical score for that individual. The left-hand end of the scale represents the minimum and the right-hand end the maximum degree of expression of the item among all the individuals being rated (Martin & Bateson 1986).

For each item, the numerical scores given by an observer to each individual are ranked and the ranks used to assess inter-observer reliability and validity for the item. Individuals are subsequently compared using the mean of each observer's numerical score obtained for each item. Observers' ratings were used by Stevenson-Hinde *et al* (1980) in a pioneering study on rhesus monkeys. Since then, they have been used on cats (Feaver *et al* 1986), dairy goats (Lyons 1989) and donkeys (French 1991). Before describing and discussing the results of these studies, some methodological problems will be discussed.

The role of the observer

According to Block (1977), the role of an observer when collecting behavioural data is merely that of a kind of recording instrument, for he or she records pre-selected and readily identified behaviour patterns. In contrast, when using a rating scale, the observer is an active instrument that filters, cumulates and integrates the data. It is from this active role of the observer that both the advantages and disadvantages of the method arise.

Advantages

The main advantage of observers' ratings is that they can provide useful information about subtle aspects of an individual's behaviour. For example, in Lyons' study (1989) goats were scored on an item designated 'Tense', that was defined as 'Shows restraint in movement and posture...' Qualities of behaviour like this do not fit well within the structures imposed by conventional recording methods (Lyons 1989). However, since temperament is reflected in an individual's overall behavioural 'style' or emotional 'tone' (Lyons 1989), items like 'Tense' provide useful information to assess temperament. Consequently, observers' ratings add new information to that provided by more conventional methods. In fact, observers' ratings are particularly useful because it is not always possible to quantify the temporal interaction of behaviours and their intensities.

Disadvantages
To use this method effectively, two or more observers should have a detailed knowledge of their subjects acquired usually over several weeks or months and this may have practical limitations (Mendl & Harcourt 1988). Familiarity with the subjects is required not only to rate the individuals but also to select and define the items on which the individuals will be rated (Stevenson-Hinde 1983). For example, although Feaver *et al* (1986) based their study on that of Stevenson-Hinde *et al* (1980), the items were partially modified to better assess individuality of cats. The same was done by Lyons (1989).

Since the observer plays an active role, there is a certain risk of subjectivity. Indeed, Stevenson-Hinde *et al* (1980) referred to the rating method as 'subjective assessment'. Even accepting that any method of observation implies some degree of subjectivity (Martin & Bateson 1986), it seems logical that the more active the role of the observer, the greater the subjectivity of the method. Indeed, the choice of the items on which the animals are rated might reflect the observers' preconceptions and the context of the observations, so that different observers could be using different lists in their attempts to assess the same characteristics in a particular species.

Since the scales are relative (ie the extremes of the scale refer to the extremes within the group of animals under study rather than within the species as a whole) it can be difficult to compare results between studies.

How to assess the effectiveness of the method
No matter what the advantages and disadvantages of a particular method are, its effectiveness should be assessed before accepting any conclusion derived from it. When doing so, two basic issues must be considered, namely reliability and validity (Martin & Bateson 1986).

Reliability concerns the extent to which measurements are repeatable and consistent. For the method under discussion, it usually refers to inter-observer reliability (Stevenson-Hinde *et al* 1980, Feaver *et al* 1986, Lyons 1989). Inter-observer reliability is assessed for each item by calculating the correlation (usually Spearman rank correlation) between each individual's ranks on each item. For example, if there are two observers, each animal will have two ranks on each item (one from each observer) and the correlation between them will be a measure of the inter-observer reliability for that particular item (eg Feaver *et al* 1986).

Validity concerns the extent to which a measurement actually measures those features the investigator wishes to measure. Observers' ratings are usually validated by comparing them with quantitative scores based on the direct observation of behaviour (Stevenson-Hinde *et al* 1980, Feaver *et al* 1986) or with the results of behavioural tests (Lyons 1989). For each item, validity is assessed by calculating the correlation (usually Spearman rank correlation) between the mean rank that each individual has obtained according to the observers' ratings and its rank according to quantitative scores based either on behavioural observations or on the results of behavioural tests (eg Feaver *et al* 1986).

Process of data reduction
As already mentioned, temperament refers to a comprehensive view of an individual's behaviour. However, the immediate outcome of the rating is a list of scores by each observer for each individual, each score corresponding to a particular item. It follows that if an overall picture of the individual's behaviour is to be given, data must be reduced in such a way that the interrelationships between scores are made clear. Two main methods are used (Stevenson-Hinde 1983): principal component analysis (PCA) and the use of an intercorrelation matrices. The reader is referred to Blackith and Reyment (1971) and to Feaver *et al* (1986) for a description of the use of PCA and intercorrelation matrices respectively.

Description of results
To illustrate the application of this method the results of three key studies will be described.

Stevenson-Hinde et al *(1980) on rhesus monkeys.* The temperament of 45 monkeys was assessed on 25 behaviourally defined adjectives (items) over four years by two observers (who were not always the same). A seven-point scale was used. For the last year of rating, inter-observer reliability was significant for 21 items. Unreliable items were discarded. Six of the reliable items were validated against quantitative scores based on behavioural observations and the correlations obtained ranged from moderate to high. Data reduction was done by using PCA as described before.

Thirteen of the reliable items could be explained by three principal components: confident to fearful (C1), excitable to slow (C2) and sociable to solitary (C3). It was found that the individuals in the colony differed greatly in all items. Furthermore, the monkeys tended to be consistent in their scores over the four years of observation, even though they were affected by some traumatic events. For example, sociable scores of some individuals dropped when their preferred companions were removed from the colony. Two factors were found to be important to explain an individual's temperament: early experiences and the temperament of its mother. For example, excitable scores were increased by adverse early experiences. Mothers with high confident scores had confident offspring, excitable mothers had excitable offspring and sociable mothers had sociable offspring.

Feaver et al *(1986) on cats.* Fourteen cats were watched by two observers for three months and scored on 18 items. A continuous scale was used. Inter-observer reliability was high for 7 items. Unreliable items were discarded. Six of the seven items were validated against quantitative scores based on behavioural observation and for five of them correlation ranged from moderate to high. Data reduction was done using an intercorrelation matrix. The seven reliable items were found to fall into three groups: (a) alert = (active + curious)/2, (b) sociable = (sociable with people-fearful of people-hostile to people-tense)/4 and (c) equable. These three groupings seemed to be independent personality dimensions. Individuals that scored similarly on the three groupings had similar temperament. For example, individuals with high scores on all three dimensions were defined as 'sociable, confident, easygoing cats'.

Lyons (1989) on goats. Sixteen goats were scored by two observers on seven items using a continuous scale. Inter-observer reliability was high or very high for all components. The five most reliable items were validated against behavioural tests and the correlations ranged between moderate and high. PCA was used for data reduction. The five most reliable items were explained by a single principal component called timidity: more timid goats were more tense, watchful, excitable and fearful of people, and less friendly towards people. Timidity varied a great deal across individuals, but when dam-reared animals were compared with human-reared animals, the former tended to be more timid. So early experiences had broad effects on an individual's temperament. Consistency over time was found to be high, for scores on timidity replicated those obtained 16 to 22 months earlier.

Discussion of methodology

The effectiveness of observers' ratings is assessed by calculating inter-observer reliability and validity against behavioural observations or tests. All studies discarded unreliable items, which illustrates how important it is to obtain a high reliability on as many items as possible. The percentage of reliable items ranged from about 40 per cent (Feaver *et al* 1986) up to 100 per cent (Lyons 1989). It is likely that the more familiar the observers are with the animals, the higher the reliability. Thus, time is likely to be a serious constraint on the effectiveness of the method. Reliability also depends on the kind of item. For example, the items used in the Lyons study (1989) were mostly related to fearfulness and were thus relatively straightforward.

Neither Stevenson-Hinde *et al* (1980) nor Feaver *et al* (1986) tried to validate all the items, since some could not be compared with direct observations. Some of their items also failed the validation tests. This may have been due to the type of item and the nature of the behavioural observations against which it was validated. For example, the item with the lowest validity in the Stevenson-Hinde *et al* study (1980) was 'excitable'. This item was validated against the observation 'displays; threats directed outside the pen'. However, 'excitable' was defined as 'reacts strongly to changes in the environment'. Since an animal can react in many different ways to a change in the environment, the observation was not equivalent to the item. Therefore, when an item fails to be validated, it is not necessarily the rating that is wrong. What should be done with items that can not be validated? It is likely that the answer to this question depends in part on the scientist's perception of science. To some scientists, objectivity in observation remains the cornerstone of their belief. Others, accepting that complete objectivity is not possible, aim to build and share a perception of what they observe (Fedigan 1982). Indeed, many ethologists have stated that complete objectivity in the observations is never achieved (eg Martin & Bateson 1986, Schaller 1988, Strum 1988).

In fact, the whole idea of using observers' ratings is to measure behavioural attributes that can not easily be measured by more conventional methods.

If we aim to build and share a perception of an animal's individuality, lack of validation is not reason enough to discard a particular item. Indeed, provided inter-observer reliability is high, we are still giving an account of a particular individual's temperament that can be understood by other observers and on which hypotheses can be tested. The problem with this is that two observers can agree with each other and still be wrong. This may happen, for example, if both share the same bias towards a particular individual. One way to avoid this might be to follow the study of Stevenson-Hinde *et al* (1980) in which the two observers were not the same over the whole period.

In any case, the effectiveness of the rating method will only be properly assessed when its predictions are tested by using a different method (Feaver *et al* 1986). Lyons (1989) provides an example: after rating several goats on timidity (see earlier), milk ejection impairment was measured. Since milk ejection impairment is partially due to fear (Goodman & Grosvenor 1983), those animals rated as more timid should have a higher degree of milk ejection impairment. Indeed, a good correlation was found between the measures.

Finally, if we are to share our perception, the way we define each item is important. Despite some contrary opinions (eg Dawkins 1976), if our perception is to be shared, the meaning assigned to a particular adjective (item) should be as close as possible to its common meaning (for a discussion on this topic see Fedigan 1982).

Section summary

Observers' ratings provide a useful method of assessing individuality in behaviour. The active role played by the observer is an interesting feature for it provides information on aspects which might otherwise be omitted. However, the method is not without problems. First, the choice of the items on which the animals will be scored can be biased by the observers' preconceptions. Second, the fact that the scales are relative makes it difficult to compare results between different studies. Consequently, observers' ratings are not a substitute for other methods of assessing individuality but they do provide an additional source of information.

Scales with predefined scores

Some studies on temperament assessment in different domestic species have used scales with predefined scores (eg Tulloh 1961, Dickson *et al* 1970, Kilgour 1975, Agyemang *et al* 1982, McCann *et al* 1987, Purcell *et al* 1988). For example, in the study of McCann *et al* (1987) four observers rated 32 horses on emotionality while the animals were in a chute and upon being released. Escape tendencies, reactivity to people, behaviour after release from the chute and overall emotionality were the categories assessed. Within each category the animals were classified as highly nervous, nervous, normal or quiet. Inter-observer reliability was very high and the scores on emotionality (which varied a great deal across animals) showed a moderate correlation with heart rate measurements in the chute. Thus the method seems useful and relatively simple.

However, some problems exist and they are best illustrated by studies on dairy cattle. For example, Kilgour (1975) assessed temperament of dairy cows using a five-point scale on which point one was defined as 'Very placid, quiet cow...' and point five as 'Flighty, very scary cow...'. The cows were rated by two milkers and the main problem was that no correlation was found between the two of them. Kilgour (1975) concluded that even though only the milkers knew the animals well enough to rate them, their assessments tend to be influenced by personal preferences and experiences with individual animals, or even by the milk production records of the cows. Furthermore, the milkers' impressions affected the way they handled the cows, which in turn affected the cows' behaviour. Indeed, cows which were labelled as restless by the milkers could be very quiet if handled by a different person. In summary, there appeared to be a large man-cow interaction factor operating which limited the reliability of this method (Dickson *et al* 1970). This is likely to be more of a problem in studies on dairy cows, for they are usually rated in the milking parlour (when human-animal interaction is relatively high and difficult to standardize) and by the milkers (whose rating and handling of the animals can be affected by previous experiences and production records). On the contrary, Tulloh (1961) and McCann *et al* (1987) concluded that this method is useful to assess temperament in beef cattle and horses, respectively. In both studies the quality of the human-animal interaction was more likely to be constant throughout the whole rating period and across raters. Furthermore, the raters were less likely to be biased by previous experiences or production records.

In summary, this method can yield useful results. Indeed, the mentioned problems are not due to the method *per se* but rather to the circumstances in which most studies on dairy cows have been done. Furthermore, its simplicity makes this method particularly interesting.

Behavioural tests

General considerations

Studies of temperament place a particular emphasis on the individual's response to environmental challenge. Observers' ratings may provide information on this aspect, but behavioural tests facilitate the comparison of individuals in a standardized way. Indeed, since in behavioural tests the situation is highly controlled, it could be argued they provide the most objective method of assessing individuality. The main problem with this assumption, however, is that one individual's response will depend on its perception of the situation and this perception may be very different between individuals. If this is the case, the differences in behaviour will not be due to differences in temperament *per se* (Stevenson-Hinde 1983). This problem is discussed later.

It is not intended to review all the tests that have been used in studies on individual differences, but to focus on those most commonly used with farm animals. For those interested in a broader approach, Fox (1972) and Goddard and Beilharz (1986) provide a number of examples concerning canids, while Spencer-Booth and Hinde (1969) review tests used on rhesus monkeys.

Types of behavioural tests used with farm animals

Behavioural tests used on farm animals can be divided into three main groups: tests that measure response to environmental challenge, tests of general fearfulness and tests of fearfulness towards humans. These categories may well overlap, but they provide a useful framework for discussion.

Tests that measure response to environmental challenge

This type of test is particularly relevant to the study of individual differences in temperament, for temperament is the individual's basic stance towards environmental change and challenge (Lawrence *et al* 1991). Four types of environmental challenge have been used to measure reactivity, namely: social isolation, handling, introduction of a novel object and food competition.

Response to social isolation. Reactions to separation from conspecifics can be considered to reflect levels of arousal, ranging from indifference to extreme excitement (Syme 1981). In sheep, three types of individuals exist according to their response to short-term visual isolation from conspecifics (Syme 1981): individuals that show no reaction (unresponsive), individuals that respond with the type of vocalization used to call for care and attention (vocal) and individuals that make vigorous attempts to return where the flock is held (physically responsive).

Furthermore, vocal and physically responsive individuals tend to move in the middle and back of the herd (Syme 1981), and when isolated show a higher heart rate than unresponsive individuals (Syme & Elphick 1982). This study is interesting not only because it shows a clear example of individual variability in response to a challenge, but also because of its practical applications. Large variation between individuals is a well-known problem in behavioural research, particularly in studies of welfare-related interest. If subgroups of animals with different reactions are created prior to the study (eg unresponsive, vocal and physically responsive sheep in a study on, say, isolation-induced stress), individual variation could be reduced (Ladewig 1991).

Response to handling. Lawrence *et al* (1991) found a large individual variation in pigs in response to handling. Pigs were rated on latency to leave the pen, order to move from the pen into a corridor, ease of progress down the corridor with an observer walking behind, response to sudden human approach, physical resistance to restraining noose and vocalization response to restraining noose. There was a strong tendency for individual scores across the tests to correlate, suggesting that individuals could be placed on a continuum according to their general responsiveness to handling.

Response to a novel object. This test has been used mainly with pigs (eg Lawrence *et al* 1991, Rasmussen 1991). Both contact with the novel object and attention directed towards the novel object without physical contact are measured. In the study of Lawrence *et al* (1991), individuals that showed more interest in the novel object were those that scored higher on responsiveness to handling.

Food competition test. This type of test has been used with pigs (eg Hunter *et al* 1988, Lawrence *et al* 1991), cats (eg Cole & Shafer 1966), cattle (eg Arnold & Grassia 1983) and horses (eg Arnold & Grassia 1982). Some methods and results will be briefly described.

Hunter *et al* (1988) recorded feeding order and aggressive interactions at the feeder in a group of twenty sows and found that social rank (ie number of sows that were regularly defeated by each particular animal) was positively correlated with feeding order, even though a great variability existed. For example, one sow in the upper half of the social hierarchy ate relatively late. No relationship was found between position in the feeding order or rank and tendency to attack or be attacked. Furthermore, two sows out of twenty were responsible for more than 50 per cent of all attacks, indicating a great individual variability in aggressiveness.

Arnold and Grassia (1982, 1983) used a rather more complicated method. They recorded the individual frequencies of several behaviours (eg leaves the feeder, avoids other individuals, joins others, etc) and subsequently reduced the data by doing principal component analysis (PCA). All the variables could be explained by four principal components, namely: dominance, activity, specific preferences for other individuals in the herd and social attractiveness (in horses), and dominance, activity, isolation seeking and aggressiveness (in cows). This method has the advantage of providing some insight into the relationships between different behavioural traits. For example, they found that high-ranking horses tended to be more active and less attractive (ie were avoided more frequently than low-ranking horses). Furthermore, similarities and differences between individuals in each of the principal components can be found. It was found, for example, that cows similar in rank had very different levels of aggressiveness.

In summary, food competition tests are interesting because they provide a clear example of how much individuals can differ in their behaviour (eg differences in aggressiveness in sows). Furthermore, by using PCA a social profile can be obtained for each individual. However, it must be pointed out that the results are likely to depend on the particular characteristics of each test. For example, Cole and Shafer (1966) found that some cats that were very placid in group tests could become very aggressive when tested in pairs. In other words, the 'individual-situation' interaction is likely to be high.

Tests of general fearfulness

The test most commonly used to measure general fearfulness is the open-field test (OFT) (Gray 1991). This test was originally developed for laboratory animals and consists of a walled arena (in some cases with a battery of lamps and loudspeakers mounted above it). Two types of measures are obtained: the frequency of defecation (defecation score) and scores related to the animal's movements (ambulation scores) (Archer 1973). Position scores (ie near the edge versus away from the edge) are sometimes added (Gray 1991).

It is important to ask whether these two measures actually reflect the animal's fearfulness. In order to answer this question, two aspects should be considered. First, both scores should be reliable in the sense that different tests of the same individual should give similar values. Second, they should be correlated with an independent measure of fear. According to Gray (1991) both types of measures (ie defecation and ambulation) are highly reliable in laboratory rodents. Furthermore, defecation seems to be a fairly general response to a variety of situations that are supposed to be frightening (Gray 1991). The general conclusion is that fear leads to high levels of defecation and low levels of ambulation (Royce 1977). In summary, it seems that despite some reservations (eg Archer 1973), the OFT does measure fearfulness in laboratory rodents. The same conclusion probably applies to rabbits (Meijsser *et al* 1989). Whether the OFT is equally effective at measuring general fearfulness in farm species is not so clear.

Open-field testing of poultry. The OFT has been used to measure fearfulness in poultry in a number of studies (eg Faure 1980a, Jones 1984, Webster & Hurnick 1990). The measures most commonly recorded are latency to move, activity (the arena is divided into a number of sections; activity refers to both the number of sections over which the bird ranges during the test and the number of times it goes from one section to another) and number of distress calls. Fearful birds are characterized in the OFT by a long latency to move, low activity and few distress calls (Jones 1977). At least two lines of evidence support the validity of the OFT as a measure of fearfulness in poultry. First, individuals selected for high ambulation scores in the OFT show lower reaction to novel visual and auditory stimuli in non-open-field situations, while those selected for low ambulation scores show higher reaction to novelty (Faure 1980b). In other words, open-field scores are correlated with an independent measure of fear. Second, the response seems to be consistent over time (see Jones 1987 and references therein). On the other hand, it has also been suggested that open-field behaviour in poultry merely represents a compromise between a tendency to reinstate contact with conspecifics and a tendency to minimize predation (Gallup & Suarez 1980). However, results obtained in subsequent studies tend to deny this hypothesis. For example, birds show higher activity scores when tested in pairs than when tested alone, while Gallup and Suarez's theory would predict the contrary. Also, birds that have been reared alone do not show lower activity scores, as Gallup and Suarez's theory would predict (Jones 1984). Consequently, it seems sensible to conclude that the OFT does measure fearfulness in poultry.

Two factors have been found to account for part of the individual variation in fearfulness in poultry. First, males are more fearful than females (Jones 1977). The reasons for this are not understood. Second, general fearfulness is to a certain extent under genetic control (Jones 1977).

Open-field testing of pigs. The OFT has been used with pigs in several studies (eg Beilharz & Cox 1967, Fraser 1974, 1975). The measures most commonly recorded are ambulation, defecation and different types of vocalizations (Fraser 1974). Indeed, some types of vocalization, ambulation and frequency of defecation are significantly and positively correlated (even though correlations are often only moderate). Furthermore, these measures show some reliability (Fraser 1974). However, a number of problems must be pointed out. First, and perhaps most important, no study provides an independent measure of fear and this makes results difficult to interpret. Indeed, the interpretation of the results cannot be the same as in rodents for frequency of defecation and ambulation are positively correlated in pigs (Fraser 1974) and negatively so in rodents (Royce 1977). Besides, when animals are tested in pairs all measures decrease, suggesting that ambulation could be due to the animal trying to reunite with its companion.

Second, season of the year has a significant influence on ambulation score and probably interacts with sex and breed (Beilharz & Cox 1967). Consequently, care should be taken when extrapolating results from one study to another. Finally, some measures may be due to factors that have nothing to do with temperament *per se*. For example, differences in vocalization in young piglets in an OFT can be due to facial wounding caused by fighting at the udder (Fraser 1975).

In summary, if the OFT is to be used meaningfully, we should understand what aspects of the test situation are responsible for the animals' reaction. This does not seem to be known in pigs (Fraser 1974). Consequently, since we do not know exactly what we are measuring, the OFT is not useful for assessing individual differences in pigs, apart from individual differences in open-field behaviour *per se*.

Open-field testing of ruminants. A number of studies has been carried out on the behaviour of cattle in OFT, not only to assess temperament (eg Kilgour 1975) but also to study the effects of different rearing conditions (eg Warnick *et al* 1977, Arave *et al* 1985, Creel & Albright 1988, Dellmeier *et al* 1990) and to predict milk production (eg Kovalcikova & Kovalcic 1982).

As with pigs, the aspects of behaviour most commonly recorded are vocalizations, ambulation and frequency of defecation (eg Kilgour 1975, Arave *et al* 1985). No study provides any comparison with an independent measure of fear. The correlation between different measures is at best only moderate (Kilgour 1975) and in some studies no correlation was found (Kovalcikova & Kovalcic 1982). Furthermore, since ambulation decreases over successive testings, the interpretation cannot be the same as in rodents (Kilgour 1975).

Apart from this, some of these studies illustrate the problems of using behavioural tests to assess individual differences. For example, in four different studies the behaviour of calves reared in isolation was compared with that of group-reared animals. To start with, three studies reported higher ambulation scores for the isolates (Warnick *et al* 1977, Arave *et al* 1985, Creel & Albright 1988) whilst one study reported the contrary (Le

Neindre 1989). Perhaps more important than this, the interpretations are equally diverse. Arave *et al* (1985) concluded that controls were more disturbed and froze. On the contrary, Warnick *et al* (1977) interpreted the higher ambulation score of the isolates as a sign of nervousness and discomfort. This variety of interpretations emphasizes two points. First, as long as the test results are not validated against independent measures, their interpretation will be very difficult (Le Neindre 1989). Second, comparing individuals with different experimental histories can be even more difficult. For example, novelty can cause fear (Gray 1991) and for an animal reared with conspecifics (but not for one reared in isolation) being alone can be new. Furthermore, if ambulation and vocalizations are caused by the animal trying to reunite with its conspecifics, isolates are likely to show lower scores. This suggestion is in part supported by work on sheep (Price & Thos 1980). On the other hand, if ambulation is due to exploration, isolates may well be more active, for animals kept in barer environments can suffer from a deprivation of exploratory behaviour (Wood-Gush & Vestegaard 1989). Indeed, motivation for highly active locomotor behaviour in general increases with increasing degree and duration of movement restraint, and decreases in response to less restrictive housing (Dellmeier *et al* 1990).

Tests of fearfulness towards humans

Test that measure fear of humans have been used in goats (eg Lyons *et al* 1988a,b, Lyons 1989), pigs (*eg* Hemsworth *et al* 1986a), cattle (Hemsworth *et al* 1987, 1989, Kerr & Wood-Gush 1987, Purcell *et al* 1988) and poultry (eg Murphy & Duncan 1978, Jones 1988). These tests can be divided into three groups: tonic immobility in poultry; flinch, step and kick response in dairy cattle, and approach tests (Gonyou 1991).

Tonic immobility in poultry. Tonic immobility (TI) is an unlearned state of motor inhibition and reduced responsiveness induced by physical restraint (Jones 1986a). The common procedure to induce TI is to place a bird on its back in a U-shaped wooden cradle and restrain it for 15 seconds with one hand on the sternum and other over the head (Jones 1986c). The parameters measured are the number of inductions (ie 15 second periods of restraint) necessary to attain TI lasting at least 10 seconds, the latency to the first alert head movement from the start of TI, the number of such head movements and the duration of TI (ie time elapsed until the bird rights itself) (Jones 1986c). TI in chickens usually lasts for about 10 minutes but may continue for as long as three hours or more (McFarland 1987).

The experimental evidence supporting the validity of TI as a measure of fearfulness is reviewed by Gallup (1979) and Jones (1986b). Tonic immobility appears to be a defensive reaction against predators. When a person grabs a chicken to induce tonic immobility, this simulates a predatory episode (McFarland 1987) and the TI induced is in part a measure of the individual's fear of humans. The duration of tonic immobility is also correlated with general fearfulness as for example measured with the OFT, (Faure 1980a). This suggests that birds that show more general fearfulness are also likely to show more fear of humans. Duration of TI can be selected for, and its heritability is very high (Jones 1986c and references therein).

Individuals are consistent in their tonic immobility response over time (Jones 1987) even though habituation to humans decreases the duration of the response (Jones 1986c). Part of the individual variability in tonic immobility is explained by sex and rank. TI lasts longer in high-ranking animals than in low-ranking ones, while males show longer latencies to the first head movement and fewer head movements than females (Jones 1986c). The effect of rank is not properly understood, but it is noteworthy that in the wild, high-ranking animals are particularly wary and watchful when approaching an area with a high risk of predation (McBride *et al* 1969). The effect of sex may be due to the fact that males are in general more fearful than females (see before). In summary, four main factors account for the individual variability in the duration of tonic immobility: sex, rank, habituation to humans and a genetic component. Furthermore, TI provides a good measure of fearfulness towards humans as long as the tests are carried out in a standardized way (Jones 1986c).

Flinch, step and kick response in dairy cattle. Some dairy cows may display the 'flinch, step and kick' (FSK) response during milking in which they flinch their udder or stomach muscles, shift their weight from one hind foot to the other and kick (Willis 1983). It is thought that fear of humans is a major cause of the FSK response (Hemsworth et al 1987).

Purcell *et al* (1985), using a scale in which kicks scored higher than lifts, counted foot lifts and kicks to obtain an overall score of fearfulness towards humans. Similarly, Hemsworth *et al* (1987, 1989) counted the number of FSK responses during milking to compare animals that had been handled around parturition with those that had not. The former group showed fewer FSK responses during the first 20 weeks of lactation (Hemsworth *et al* 1989).

This method has several advantages, namely: it is easy to carry out, avoids subjectivity and yields results that can be compared across studies. However, several problems exist. First, the behaviour of the milker is likely to influence the number of FSK responses. Second, simply counting the number of responses does not necessarily reflects a cow's reaction to being milked, eg a frightened cow that stands rigidly can receive the same score as a relaxed cow that does not move (Purcell *et al* 1985). This may explain why no correlation was found between approachability in the pen and number of FSK responses during milking by Purcell *et al* (1985). Nevertheless, these problems do not invalidate the use of the FSK response for measuring fearfulness towards humans, but they suggest that it should always be used in conjunction with other methods.

Approach tests. Approach tests measure the behaviour of an individual when in close proximity to a human being. They have been used with pigs (eg Hemsworth *et al* 1986a), goats (eg Lyons *et al* 1988a, b, Lyons 1989), cattle (eg Kerr & Wood-Gush 1987, Purcell *et al* 1988) and poultry (eg Murphy & Duncan 1978, Jones & Faure 1981).

These tests can be divided into two groups: those based on observers' ratings of the animal's response to humans and those that record 'objective' measures. For example, Kerr and Wood-Gush (1987) assessed heifers' docility by scoring their reaction to being touched in a standardized manner. The scores ranged from -2 ('moves away and cannot be coaxed') to +2 ('allows touching and responds by moving forward for more'). For most of the heifers, docility became stable within 20 weeks of age and remained fairly constant up to 90 weeks of age, when the study was discontinued. Furthermore, the heifers tended to show one of two reactions to being touched. One subgroup (55 per cent of the animals) responded by moving forward when touched, while the other subgroup (45 per cent of the animals) responded by moving away. So two bimodally separate populations of individuals were found.

Tests based on 'objective' measures record, for example, flight distance (eg Purcell *et al* 1985), latency to approach humans, latency to interact with humans, time in proximity of humans, number of interactions (eg Hemsworth *et al* 1986a) and amount of time spent moving away from the experimenters (eg Lyons *et al* 1988a).

Approach tests are relatively easy to interpret and have been validated against, for example, physiological measures (eg Hemsworth *et al* 1986a). However it must be pointed out that an animal's response to humans is likely to be influenced by aspects such as the observer's posture, movement, etc (eg Hemsworth *et al* 1986b, Lyons *et al* 1988a) as well as by whether it is tested alone or with conspecifics (eg Lyons *et al* 1988a). Consequently, tests should be carefully standardized (eg Dennison 1985) and care should be taken when extrapolating results from one test to another.

Section summary

The use of behavioural tests for assessing individual differences in temperament has several advantages. First, the results are usually easy to analyse. Second, they can be compared between studies. Furthermore, they provide direct information on aspects of great practical interest (eg fearfulness towards humans). However, several problems exist. Since the particular type of test is likely to influence the results, tests should be carefully standardized and care should be taken when extrapolating results from one type of test to another. Furthermore, in some circumstances the individual-situation interaction is likely to be high. If this is the case, the results may not reflect differences in temperament *per se*. Finally, some aspects of an individual's temperament may be omitted if only tests are used to assess individuality. Therefore, they should be used alongside other methods.

General conclusions

Most studies on individual variation in behaviour have been carried out in non-human primates. However, individual differences are by no means restricted to primates. Indeed, all studies that have addressed this topic in domestic species have shown a great deal of individual variation in many aspects of behaviour (eg Bateson & Turner 1988).

The study of individual variation in behaviour should be tackled by using a two-step process. First, individual differences must be described. Second, their causes must be investigated. Individual differences in behaviour may be described at two different levels of complexity. Of these, differences in temperament are most relevant to animal welfare. Furthermore, describing differences in temperament requires the use of particular methods, namely: observers' ratings or scales with predefined scores, and behavioural tests. While observers' ratings may provide information on subtle aspects of behaviour that could otherwise be omitted, behavioural tests allow individual differences to be studied in a more standardized way. Therefore, it is suggested that both methods should be used in conjunction.

In order to understand the causes of individual variation, both behavioural genetics and the study of behavioural development are of particular interest. Studies of behavioural genetics provide an estimate of the extent to which genetic differences between individuals contribute to individual differences in behaviour (Plomin *et al* 1990). The study of behavioural development can reveal the environmental influences that account for the proportion of individual differences unexplained by genetic factors as well as the way in which the environment and the individual's genetic make-up interact (Bateson 1976). Both sensitive periods and early experiences are particularly relevant to the development of individual differences.

Animal welfare implications

Individual differences are relevant to the study of animal welfare for several reasons. If individual differences are taken into account when designing experiments, variability in the results obtained in studies on welfare issues may be reduced. This is important because variability in the results is one of the main problems encountered in these types of studies. In addition, the welfare of an *individual* (our emphasis) is its state as regards its attempts to cope with its environment (Broom 1988). Since different individuals try to cope with their environment in different ways (eg Bohus *et al* 1987), welfare assessment must take account of individual differences in coping mechanisms. Otherwise, some indicators relevant to animal welfare would be omitted.

Finally, an understanding of the causes of individual differences in temperament may allow us to reduce the incidence of some welfare problems. For example, if fearfulness is to a certain degree under genetic control, a selection program to reduce fearfulness may be undertaken. If experiences during the sensitive period for socialization are also important, changes in human-animal interaction during that period may be recommended.

One of the reasons why interest in individual variation in behaviour has recently increased is that our knowledge of behaviour is now greater and this allows us to realize differences not only between species but also across individuals of the same species (Plomin *et al* 1990). This increase in knowledge comes in part from long-term field studies. Much of what scientists learn in these studies about the individual animals' personalities they do through 'intelligent empathy' (Schaller 1964). At least in part, interest in welfare-related problems and in animal cognition (which is closely related to welfare issues) arises also from a feeling of empathy. Consequently, it is suggested that the emphasis put in these three areas (individual variation, animal welfare and cognition) comes in part from a change in our attitude towards animals. In turn, an awareness of each individual animal's uniqueness is likely to further increase our concern for animal welfare and our tendency to empathise with animals.

Acknowledgements

The authors would like to thank Ute Knierim and Hans Herrmann for invaluable comments and discussion of the first manuscript. Thanks are also due to anonymous referees for their comments. The first author was supported by a studentship granted by the Government of Catalonia, Spain.

References

Agyemang K, Clapp E and van Vleck LD 1982 Components of variance of dairymen's workability traits among Holstein cows. *Journal of Dairy Science 65:* 1334-1338

Arave C W, Mickelson C H and Walters J L 1985 Effects of early rearing experience on subsequent behavior and production of Holstein heifers. *Journal of Dairy Science 68:* 923-929

Archer J 1973 Tests for emotionality in rats and mice: a review. *Animal Behaviour 21:* 205-235

Arnold G W and Grassia A 1982 Ethogram of agonistic behaviour for thoroughbred horses. *Applied Animal Ethology 8:* 5-26

Arnold G W and Grassia A 1983 Social interactions amongst beef cows when competing for food. *Applied Animal Ethology 9:* 239-252

Bateson P 1976 Specificity and the origins of behavior. In Rosenblatt J S, Hinde R A, Shaw E and Beer C (eds) *Advances in the Study of Behavior Volume 6*. Academic Press: New York

Bateson P and Turner D 1988 Questions about cats. In Turner D and Bateson P (eds) *The Domestic Cat. The Biology of its Behaviour* pp 193-201. Cambridge University Press: Cambridge

Beilharz R G and Cox D F 1967 Genetic analysis of open-field behavior in swine. *Journal of Animal Science 26:* 988-990

Bekoff M 1977 Mammalian dispersal and the ontogeny of individual behavioral phenotypes. *American Naturalist 111:* 715-732

Blackith R E and Reyment R A 1971 *Multivariate Morphometrics.* Academic Press: London

Block J 1977 Advancing the psychology of personality: paradigmatic shift or improving the quality of research. In Endler N S and Magnusson D (eds) *Personality at the Crossroads*. Wiley: New York

Bohus B, Benus R F, Fokkema D S, Koolhaas J M, Nyakas C, van Oortmerssen G A, Prins A J A, de Ruiter A J H, Scheurink A J W and Steffens A B 1987 Neuroendocrine states and behavioural and physiological stress responses. In de Kloet E R, Wiegant V M and de Wied D (eds) *Progress in Brain Research Volume 72.* Elsevier: Amsterdam

Broom D M 1988 The scientific assessment of animal welfare. *Applied Animal Behaviour Science 20:* 5-19

Cole D D and Shafer J B 1966 A study of social dominance in cats. *Behaviour 27:* 39-53

Creel S R and Albright J L 1988 The effects of neonatal social isolation on the behaviour and endocrine function of Holstein calves. *Applied Animal Behaviour Science 21:* 293-306

Cunningham D L and von Tienhoven A 1983 Relationship between production factors and dominance in White Leghorn hens in a study on social rank and cage design. *Applied Animal Ethology 11:* 33-44

Dawkins R 1976 *The Selfish Gene.* Oxford University Press: Oxford

Dellmeier G, Friend T and Gbur E 1990 Effects of changing housing on open-field behaviour of calves. *Applied Animal Behaviour Science 26:* 215-230

Dennison S G C 1985 *The Development of Behaviour Patterns and an Assessment of Temperament of Dairy Heifers.* PhD Thesis: University of Edinburgh

Dickson D P, Barr G R, Johnson L P and Weickert D A 1970 Social dominance and temperament of Holstein cows. *Journal of Dairy Science 53:* 904-907

Faure J M 1980a Selection for open-field behaviour in the chicken: the relationship between open-field and social behaviour. *Applied Animal Behaviour Science 6:* 385

Faure J M 1980b To adapt the environment to the bird or the bird to the environment? In Moss R (ed) *The Laying Hen and its Environment* pp 19-30. Martinus Nijhoff: The Hague

Feaver J, Mendl M and Bateson P 1986 A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour 34:* 1016-1025

Fedigan L M 1982 *Primate Paradigms*. Eden Press: Montreal

Fox M W 1972 Socio-ecological implications of individual differences in a wolf litter: a developmental and evolutionary perspective. *Behaviour 41:* 298-313

Fraser D 1974 The vocalizations and other behaviours of growing pigs in an open-field test. *Applied Animal Ethology 1:* 3-16

Fraser D 1975 Vocalizations of isolated piglets I. Sources of variation and relationships among measures. *Applied Animal Ethology 1:* 387-394

French J M 1991 Assessment of donkey temperament and the influence of home environment. In Appleby M C, Horrell R I, Petherick J C and Rutter S M (eds) *Applied Animal Behaviour: Past, Present and Future* pp 47-48. Universities Federation for Animal Welfare: Potters Bar

Gallup G G 1979 Tonic immobility as a measure of fear in the domestic fowl. *Animal Behaviour 27:* 316-317

Gallup G G and Suarez S D 1980 An ethological analysis of open-field behaviour in chickens. *Animal Behaviour 28:* 368-378

Goddard M E and Beilharz R G 1986 Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science 15:* 247-260

Gonyou H W 1991 The interaction of humans with food animals: making husbandry a science. In Appleby M C, Horrell R I, Petherick J C and Rutter S M (eds) *Applied Animal Behaviour: Past, Present and Future* pp 31-33. Universities Federation for Animal Welfare: Potters Bar

Goodall J 1986 *The Chimpanzees of Gombe.* Harvard University Press: Cambridge

Goodman G T and Grosvenor C E 1983 Neuroendocrine control of the milk ejection reflex. *Journal of Dairy Science 66:* 2226-2235

Gray J A 1991 *The Psychology of Fear and Stress 2nd edition.* Cambridge University Press: Cambridge

Hemsworth P H, Barnett J C, Hansen C and Gonyou H W 1986a The influence of early contact with humans on subsequent behavioural response of pigs to humans. *Applied Animal Behaviour Science 15:* 55-63

Hemsworth P H, Gonyou H W and Dziuk P J 1986b Human communication with pigs: the behavioural response of pigs to specific human signals. *Applied Animal Behaviour Science 15:* 45-54

Hemsworth P H, Hansen C and Barnett J L 1987 The effects of human presence at the time of calving of primiparous cows on their subsequent behavioural response to milking. *Applied Animal Behaviour Science 18:* 247-255

Hemsworth P H, Barnett J L, Tilbrook A J and Hansen C 1989 The effects of handling by humans at calving and during milking on the behaviour and milk cortisol concentrations of primiparous dairy cows. *Applied Animal Behaviour Science 22:* 313-326

Hunter E J, Broom D M, Edwards S A and Sibly R M 1988 Social hierarchy and feeder access in a group of 20 sows using a computer-controlled feeder. *Animal Production 47:* 139-149

Huntingford F 1984 *The Study of Animal Behaviour.* Chapman and Hall: London

Jones R B 1977 Sex and strain differences in the open-field responses of the domestic chick. *Applied Animal Ethology 3:* 255-261

Jones R B 1984 The open-field as a stressor: fear in domestic chicks (Abstract). *Applied Animal Behaviour Science 11:* 80-81

Jones R B 1986a Tonic immobility in the domestic fowl: a descriptive profile (Abstract). *Applied Animal Behaviour Science 15:* 182.

Jones R B 1986b Social factors affecting tonic immobility in chickens (Abstract). *Applied Animal Behaviour Science 16:* 96-97

Jones R B 1986c The tonic immobility reaction of the domestic fowl: a review. *World's Poultry Science Journal 42:* 82-96

Jones R B 1987 Repeatability of fear ranks among adult laying hens. *Applied Animal Behaviour Science 19:* 297-304

Jones R B 1988 Fearfulness in chickens: its importance and modification (Abstract). *Applied Animal Behaviour Science 21:* 377

Jones R B and Faure J M 1981 The effects of regular handling on fear responses in the domestic chick. *Behavioural Processes 6:* 135-143

Kerr S G C and Wood-Gush D G M 1987 The development of behaviour patterns and temperament in dairy heifers. *Behavioural Processes 15:* 1-16

Kilgour R 1975 The open-field test as an assessment of the temperament of dairy cows. *Animal Behaviour 23:* 615-624

Kovalcikova M and Kovalcic K 1982 Relationships between parameters of the open-field test of cows and their milk production in loose housing. *Applied Animal Ethology 9:* 121-129

Ladewig J 1991 Development of the concept of stress and its application to farm animal behaviour research. In Appleby M C, Horrell R I, Petherick J C and Hunter S M (eds) *Applied Animal Behaviour: Past, Present and Future* pp 114-117. Universities Federation for Animal Welfare: Potters Bar

Lawrence A B, Terlouw E M C and Illius A W 1991 Individual differences in behavioural responses of pigs exposed to non-social and social challenges. *Applied Animal Behaviour Science 30:* 73-86

Lawrence C W 1981 *Individual Differences in the Mother-Kitten Relationship in the Domestic Cat*. PhD Thesis: University of Edinburgh

Le Neindre P 1989 Influence of rearing conditions and breed on social behaviour and activity of cattle in novel environments. *Applied Animal Behaviour Science 23:* 129-140

Lyons D 1989 Individual differences in temperament of dairy goats and the inhibition of milk ejection. *Applied Animal Behaviour Science 22:* 269-282

Lyons D M, Price E O and Moberg G P 1988a Social modulation of pituitary-adrenal responsiveness and individual differences in behavior of young domestic goats. *Physiology and Behaviour 43:* 451-458

Lyons D M, Price E O and Moberg G P 1988b Individual differences in temperament of domestic dairy goats: constancy and change. *Animal Behaviour 36:* 1323-1333

McBride G, Parer L P and Foenander F 1969 The social organization and behaviour of the feral domestic fowl. *Animal Behaviour Monographs 2:* 127-181

McCann J S, Heird J C, Bell R W and Lutherer L O 1987 Normal and more highly reactive horses I. Heart rate, respiration rate and behavioural observations. *Applied Animal Behaviour Science 19:* 201-214

McFarland D 1987 *The Oxford Companion to Animal Behaviour.* Oxford University Press: Oxford

Martin P and Bateson P 1986 *Measuring Behaviour.* Cambridge University Press: Cambridge

Meijsser F M, Kersten A M P, Wiepkema P R and Metz J H M 1989 An analysis of the open-field performance of sub-adult rabbits. *Applied Animal Behaviour Science 24:* 147-155

Mendl M and Harcourt R 1988 Individuality in the domestic cat. In Turner D and Bateson P (eds) *The Domestic Cat. The Biology of its Behaviour* pp 41-54. Cambridge University Press: Cambridge

Murphy L B and Duncan I J H 1978 Attempts to modify the responses of domestic fowl towards human beings II. The effect of early experience. *Applied Animal Ethology 4:* 5-12

Plomin R, DeFries J C and McClearn G E 1990 *Behavioral Genetics. A primer 2nd edition.* W H Freeman & Company: New York

Price E O and Thos J 1980 Behavioural responses to short term social isolation in sheep and goats. *Applied Animal Ethology 6:* 331-339

Purcell D, Arave C W, Gilbert B J and Walters J L 1985 Relationship of three measures of behavior to milk production. *Journal of Dairy Science 68 (Supplement 1):* 42

Purcell D, Arave C W and Walters J L 1988 Relationship of three measures of behaviour to milk production. *Applied Animal Behaviour Science 21:* 307-313

Rasmussen A A 1991 Behavioural reaction of individual pigs raised in different environments to a novel object and to handling at slaughter in relation to measurements of the adrenal cortex (Abstract). *Applied Animal Behaviour Science 30:* 176

Royce J R 1977 On the construct validity of open-field measures. *Psychological Bulletin 84:* 1098-1106

Schaller G B 1964 *The Year of the Gorilla.* The University of Chicago Press: Chicago

Schaller G B 1988 Foreword. In Strum S C *Almost Human. A Journey into the World of Baboons.* Hamish Hamilton: London

Spencer-Booth Y and Hinde R A 1969 Tests of behavioural characteristics for rhesus monkeys. *Behaviour 33:* 179-211

Stevenson-Hinde J 1983 Individual characteristics: a statement of the problem. Constancy over time. Predictability across situations. In Hinde R A (ed) *Primate Social Relationships: an Integrated Approach.* Blackwell Scientific Publications: Oxford

Stevenson-Hinde J, Stillwell-Barnes R and Zunz M 1980 Subjective assessment of rhesus monkeys over four succesive years. *Primates 21:* 66-82

Strum S C 1988 *Almost Human. A Journey into the World of Baboons.* Hamish Hamilton: London

Syme L A 1981 Social disruption and forced movement orders in sheep. *Animal Behaviour 29:* 283-288

Syme L A and Elphick G R 1982 Heart rate and the behaviour of sheep in yards. *Applied Animal Ethology 9:* 31-35

Tulloh N M 1961 Behaviour of cattle in yards II. A study of temperament. *Animal Behaviour 9:* 25-30

Warnick V D, Arave C W and Mickelsen C H 1977 Effects of group, individual and isolated rearing of calves on weight gain and behavior. *Journal of Dairy Science 60:* 947-953

Webster A B and Hurnick J F 1990 Open-field assessment of behavioural phenotype within genetic stocks of the White Leghorn chicken. *Applied Animal Behaviour Science 27:* 115-126

Wiepkema P R, Van Hellemaond K K, Roessingh P and Romberg H 1987 Behaviour and abomasal damage in individual veal calves. *Applied Animal Behaviour Science 18:* 257-268

Willis G L 1983 A possible relationship between the flinch, step and kick response and milk yield in lactating cows. *Applied Animal Ethology 10:* 287-290

Wood-Gush D G M and Vestegaard K 1989 Exploratory behaviour and the welfare of intensively kept animals. *Journal of Agricultural Ethics 2:* 161-169