

WHY IS THERE NO SIMPLE WAY OF MEASURING ANIMAL WELFARE?

G Mason[†] and M Mendl¹

Sub-department of Animal Behaviour, University of Cambridge, Madingley,
Cambridge CB3 8AA, UK and Department of Clinical Veterinary Medicine,
University of Cambridge

¹ Present address: Scottish Agricultural College, Edinburgh, UK

[†] Contact for correspondence and requests for reprints

Abstract

Animal Welfare 1993, 2: 301-319

Although the physiological and behavioural changes that can indicate poor welfare are generally agreed upon, using these measures in practice sometimes yields results that are hard to interpret. For example, different types of measure may suggest quite different things about an animal's welfare. Such contradictions are often due to the differing properties of the variables being measured. How each variable responds to a stressor can be affected by several factors - the type of unpleasant stimulus to which the animal is exposed; when and for how long exposure occurs; the animal's psychological state, eg does it feel that it is in control?; and the time at which the measurement is made, relative to the stressor. Typical responses also often differ between species and between individuals, and may even change in a single individual over time. Furthermore, some responses used to assess welfare lack specificity: they can be elicited by neutral or even pleasant events as well as by aversive ones. Appreciating these factors is vital when designing experiments, when choosing what to measure along with each welfare variable, and when interpreting results. Even after taking these factors into consideration, interpreting a result can still be difficult. One approach then is to consider the effects on welfare of the changes measured, eg if there is immunosuppression, does the animal succumb to disease? Another is to use the animal's behaviour to indicate its preference for, or aversion to, particular environments. Ultimately, however, interpreting welfare measures involves subjective judgements which will be influenced by the nature of our concern for the animal under consideration. By raising these problems, we hope that this review will highlight and clarify the apparent contradictions that sometimes emerge in scientific studies of animal welfare, and help researchers improve the designs of their experiments for the benefit of the animals concerned.

Keywords: *animal welfare, experimental design, measurement, stress*

Introduction

Decisions about animal welfare issues are ethically and politically important and often generate emotional debate. Scientific studies have attempted to add objectivity to this

decision-making process, but to those who want quick and unambiguous answers to questions about animal welfare, the results produced by scientists must often seem confused and contradictory. There is a variety of reasons for this, ranging from philosophical controversies about how animal welfare should be defined, to more scientific problems concerning how to measure it.

A major problem is deciding what 'animal welfare' actually means. In the scientific literature, animal welfare means different things to different researchers. Some equate welfare with biological fitness, claiming that welfare is only reduced if the animal's ability to survive and reproduce is diminished (Barnett & Hemsworth 1990). Broom (eg 1991) agrees that physical condition is important, and so considers that a badly injured animal's welfare is poor even if sleep or anaesthesia prevent it from suffering. However, he adds that an animal's welfare may also be poor in the absence of physical problems, for example if it is frightened, anxious, frustrated or bored. To other researchers, feelings like these are of paramount importance: to them, an animal's welfare is only impaired if it is experiencing an unpleasant mental state (Dawkins 1980, 1990, Duncan & Petherick 1989, Sandøe & Simonsen 1992). For them, an animal with a tumour it cannot feel does not have a welfare problem, even if it does have a health problem.

The exact way in which scientists define welfare will clearly influence the types of measure they use to attempt to assess welfare objectively. Assessing welfare is relatively simple for those who think breeding and physical health are the definitive measures to use. However, measurement is less easy for those to whom an animal's feelings are the most important determinants of its welfare. There have been two main ways of finding measurable changes in an animal's behaviour or physiology that plausibly correlate with its subjective feelings. One way has been to subject animals to 'stressors', ie stimuli scientists feel sure must be unpleasant, such as hunger, electric shocks, sensory deprivation and so on, and then to record the changes in these animals' behaviour and physiology. If similar physiological or behavioural responses occur in another context, it can then be inferred that the animal finds its situation correspondingly unpleasant. The second approach is to record the changes that occur in humans who feel scared, worried, in pain, etc, and to look for similar changes in the animals whose welfare we are trying to assess. This latter approach relies on the 'argument by analogy' that mental suffering in animals is accompanied by the same sorts of behavioural and physiological changes that are seen in suffering humans (Dawkins 1990, Sandøe & Simonsen 1992).

Both approaches suggest very similar measures, so much so that despite the continuing debates about definitions of welfare, particular measures crop up repeatedly in the animal welfare literature, suggesting a degree of consensus about which are the most useful and practicable. Commonly used indicators of poor welfare (eg as listed by Fraser & Broom 1990) include gastric ulceration; poor functioning of the immune system; breeding problems such as reduced fertility, and behavioural problems such as extreme apathy, stereotypy, and some forms of infanticide. The activity of the pituitary-adrenocortical system, as reflected by levels of circulating corticosteroids such as cortisol (eg Selye 1979), also features prominently in animal welfare research. Levels of these hormones

rise in many situations that probably involve suffering (eg Mason 1971, Lundberg & Forsmann 1979). Another hormone used as a welfare indicator is prolactin. Raised prolactin levels are found in, for example, people who are depressed (Cohen-Cole *et al* 1981) and medical students facing exams (Herbert 1987).

The information gained using indicators such as these has often helped improve animal welfare. For example, the discovery of elevated cortisol levels in tethered sows revealed the hitherto unsuspected problem of unresolved aggression between neighbours, thus indicating a need to re-design their housing (Barnett *et al* 1987). However, as we discuss below, in other cases, the use of similar indicators yields results that are at best inconclusive, at worst, downright contradictory. Different measures made on the same animal sometimes give apparently contradictory information about its welfare, and measures of the relative welfare of animals in two sorts of environment may produce disparate results. Assessing welfare by measuring behavioural and physiological responses to situations is clearly not always easy.

In this short review, our aim is to try to tackle these problems in the measurement of welfare by highlighting the biological reasons why interpreting results is often difficult. Some of the studies we will discuss have involved causing suffering to the animals used. We do not necessarily condone these studies and it is unlikely that they all would be allowed to occur in the UK today. However, the studies have been done and some of them have uncovered important points and principles that are of direct relevance to animal welfare research. We feel that rather than ignoring or attempting to rediscover these principles all over again, animal welfare researchers should use this information in a constructive way to advance knowledge and understanding of what is undoubtedly a difficult area of scientific research. The issues we discuss in this review are not new - they have been discussed by many scientists in a wide range of journals. But by bringing them together, we hope to show how appreciating the problems can refine the methodology and interpretation of future animal welfare research.

Problems with measuring welfare

Three main problems emerge when trying to assess welfare. One is that different measures do not always co-vary. For example, when an animal is moved to a new cage, some measures might suggest that its welfare has become poorer, while others might suggest that it is unchanged or even improved. A second problem is that the significance of some measures is difficult to interpret. In some aversive situations they may rise, while in other apparently equally unpleasant situations they may show no change, or even fall. Some measures may also change in situations that do not seem aversive at all. Therefore it can be difficult to decide whether, for example, a rise in adrenalin reflects anxiety, or pleasurable excitement. A final problem is that even when one study gives rise to an unambiguous conclusion about the welfare of a particular animal, a repeat of that study might yield exactly the opposite result, perhaps because the precise conditions or the characteristics of the individual animals are different.

When different measures suggest contradictory things about an animal's welfare, it can be very difficult to draw a conclusion. For example, individual animals may differ in their corticosteroid levels or behavioural signs of anxiety but have similarly extensive gastric ulceration (Wiepkema *et al* 1987, Ödberg 1989). Decreasing the size of battery cages may have no effect on hens' corticosteroid levels even though egg production falls and mortality rises (reviewed by Rushen 1991). Bredbacka (1988) found that in a strain of hens, birds which took the longest to recover from tonic immobility (interpreted as revealing a greater level of fear) produced more eggs than seemingly bolder hens. Similarly, the egg weights of battery hens were greater than those of deep litter birds, although their body-weights were lower (Wegner 1983).

Interpreting results is also difficult when measures seem to be inherently unpredictable. In conditions that seem to be sub-optimal, the values of some measures may sometimes increase and sometimes decrease. For example, the length of time an animal takes to emerge from shelter, in a strange situation, is often used to estimate how afraid it is: the slower the emergence, the greater its fear. But as Duncan (1985) points out, a scared animal may sometimes leap from an enclosure as quickly as possible, while a more relaxed animal may saunter out slowly. Similarly, the exploratory behaviour shown in such a situation can either increase or decrease with increasing fear (Archer 1973). The effect of prolonged exposure to unpleasant stimuli on the activity of the adrenal cortex is also unpredictable. Amongst human students, individuals whose responses to various psychological tests indicated high chronic stress and anxiety had lower cortisol levels, before and after a stressful oral examination, than a 'low stress' group of students (Dorian *et al* 1981). However, in veal calves the opposite appears to happen. Calves housed in the confinement of crates showed a more pronounced adrenocortical response to the acute stressors of handling and transport than those housed in groups (Trunkfield *et al* 1991). In other studies of cattle, restraint has been found to either increase or decrease their corticosteroid responses to acute challenge (reviewed by Rushen 1991). Heart rate can increase on exposure to stressors, because of increased activity of the sympathetic nervous system (eg Dantzer *et al* 1983a,b, Ladewig & von Borell 1988), but it can also fall (eg Campos 1976), particularly if the animal 'freezes' (reviewed by Manser 1992, p 90). As a final example, Kennes and de Rycke (1988) found certain changes in the white blood cell counts of stressed voles. Levels of neutrophils rose and lymphocytes and eosinophils fell in animals subjected to repeated electric shocks to the feet for five minutes a day, but the opposite occurred after ten days of this treatment.

Finally, doing the same experiment afresh can sometimes lead to different conclusions. For example, repeated comparisons of battery and pen-housed hens have yielded contradictory findings concerning their levels of corticosteroids (reviewed by Rushen 1991). This suggests that replicate experiments are perhaps carried out in subtly different ways and that this somehow affects the variable being measured.

Despite the problems that evidently exist in this field, the reasons for these discrepancies are often quite simple. So why is it that the picture is often so confusing?

Factors underlying the problems of welfare measurement

The type of aversive stimulus

The nature of the aversive stimulus or situation may influence the animal's reaction to it. For example, rats exposed to a localized aversive stimulus may attempt to flee, but if exposed to a diffuse one such as chronic cold (Robbins *et al* 1990), or distant or uncertain threats of predation (Archer 1979b), they may stay very still. Likewise, although fear and anxiety are generally believed to cause corticosteroid levels to rise (eg Mason 1971), pain does not seem reliably to result in a corticosteroid response (Rushen 1986, Bateson 1991). There is also some evidence that long-term social stress results in increased adrenocortical activity, as indicated by increased cortisol response to an adrenocorticotrophic hormone (ACTH) challenge, while situations such as confinement have less clear-cut effects (see Mendl *et al* 1992). Lastly, sustained cold or prolonged forced running lowers prolactin levels, while enforced immobilization raises them (reviewed by Manser 1992).

The different responses to different unpleasant stimuli often make functional sense. For example, it probably enhances a rat's likelihood of survival if it freezes rather than runs when it cannot tell the safest direction to take. Many unpleasant events, including social ones, probably require an animal to be prepared for activity, so it makes functional sense to secrete catabolic hormones such as glucocorticoids (reviewed by Manser 1992) to aid the mobilization of fat and carbohydrate 'fuels'. This secretion may, however, be a far less appropriate response to pain and to situations such as confinement, because increased muscular activity is here either undesirable or impossible. As a final example, dehydration in sheep does not result in a corticosteroid increase, presumably because it would not be adaptive to show a response that results in, amongst other things, increased body temperature (Broom & Johnson *in press*). So, the nature of the stressor can determine which welfare indicators are activated, often because the response is functionally appropriate.

The animal's psychological state

Whatever the physical nature of an aversive stimulus, how it is perceived by an animal can have a major effect on how that animal reacts to it. Hence the response to a particular unpleasant stimulus can vary from one situation to another. For example, an animal which is familiar with a particular situation or procedure may differ in its response from an animal encountering it for the first time (Pfister 1979, Dantzer & Mormède 1985, McCune 1992), and an animal gently handled when very young will be less disturbed by the presence of observing humans (eg Pederson & Jeppesen 1990). The situation is exactly the same, but the animals perceive it in quite different ways. Experienced animals may also affect the responses of their fellows. In social animals such as goats, the presence of confident, unafraid companions can decrease corticosteroid responses to the presence of a human (Lyons *et al* 1988). Experienced animals may habituate to the situation because they learn to control or predict it. It is well known that the degree to

which an event can be predicted or controlled may have a profound effect on the animal's response (eg Wiepkema 1987). For example, Weiss (1968, 1971) showed that rats given electric shocks that they could turn off, and hence partially control, developed three times less gastric ulceration than rats given exactly the same number of shocks, of exactly the same duration, but which were unable to control their exposure. Inducing two unpleasant states, for instance hunger and fear, such that an animal is in a state of conflict, reduces its welfare to a greater extent than if conflict was not present. For instance, rats required to decide when to negotiate an aversive stimulus to reach food showed higher levels of gastric ulceration than those exposed to the same aversive stimulus and food restriction without the conflict of decision making (Sawrey *et al* 1956).

Various factors clearly modulate an animal's psychological state and perception of a particular aversive stimulus. For example, prior experience of the stimulus may make it appear less alarming. The beneficial effects of predictability also show that being prepared in some way is a help. An appreciation of these factors can help to explain why the same situation can give rise to quite different responses in different animals.

The timing and duration of an aversive stimulus

Exactly when an animal is exposed to an aversive situation, and for how long, can affect the value of measurements. In some cases, the result can depend on when the stressor interacts with the system being measured. For example, a stressful event after exposure to a pathogen may depress the immune response, but the same stressor prior to the pathogen can enhance it (reviewed by Martin 1989). In many species, female reproduction seems much more susceptible to disruption before implantation of the blastocysts (Moberg 1985a) and in mink, females are most prone to disturbance-induced infanticide in the first few days after giving birth (Rice 1967). Likewise, the behavioural response to a startling stimulus may vary according to the animal's ongoing behaviour at the time (Fentress 1976).

The length of exposure to the stressor may also alter the response. For example, chronic stress may lead to habituation. This is evident in the waning response of prolactin to repeated stressors (reviewed by Manser 1992). Also, in contrast to acute stress, chronic stress may leave basal corticosteroid levels unchanged (eg Kristensen & Jeppesen 1988, Ladewig 1990). Prolonged stress may also, paradoxically, sometimes enhance the immune response to infection (reviewed by Martin 1989). However, habituation does not always affect all systems equally. For example, the corticosteroid levels of newly-tethered bulls return to normal after a month (Ladewig & Smidt 1989), but they maintain a lowered response to a challenge of adrenocorticotrophic hormone (ACTH). This indicates that the adrenal cortex has habituated, as it is demonstrating reduced responsiveness to the hormone that usually activates it. However, the central nervous system may well be still maintaining a high output of ACTH (Ladewig & von Borell 1988). Ladewig and von Borell therefore suggest that measuring the episodic ACTH secretion may provide a much more reliable way of assessing welfare than measuring resting levels of corticosteroids.

In terms of studies of animal welfare, it is therefore clearly important to take into account exactly when an unpleasant stimulus or situation is imposed on the animal, and for how long.

The time-course of the response being measured

Some indicators of welfare react more quickly or are more sensitive to challenging situations than others, and some take longer to return to resting levels. Often the measures generally agreed to be indicative of serious and 'pathological' states take the longest time to appear. These differences in time-course can mean, once again, that measures do not co-vary. For example, sympathetico-adrenomedullary responses such as increases in heart rate and levels of catecholamines may occur within seconds of an aversive stimulus, while elevations in other welfare indicators such as corticosteroids and prolactin may take minutes to occur (reviewed by Manser 1992). As peak values occur at different times, data collected immediately after the stressor is presented will differ from those collected only minutes later. On a still longer time-scale, some signs of stress may take years to develop (eg Fox 1984, p 180), partly because they only develop in response to chronic exposure to aversive situations, but also because they are structural changes that simply cannot happen rapidly, eg the roughening of arteries that accelerates the development of arteriosclerosis (Henry *et al* 1971, MacSween & Whaley 1992).

Therefore, the timing of sampling relative to the aversive stimulus can affect the results obtained. Knowing more about the functional role played by the parameter being measured can help in understanding, and even predicting, these differences in time-course. For example, acute exposure to an aversive stimulus in general leads to a fall in anabolic hormones such as insulin and testosterone, so that the storage of sugars and fats is reduced and their availability as ready sources of fuel is enhanced. However, such hormones may then rebound to higher levels than normal afterwards, in an 'anabolic recovery phase' (reviewed by Archer 1979a).

Species differences

Species may differ from each other in their responses to some situations, while resembling each other in others. For example, the interpretation of responses in an open-field test differs in mice and rats (eg Archer 1973): the urination and defecation of mice in this situation is more akin to scent-marking than the sign of fear and associated sympathetic nervous system activity seen in rats. Other species differences in physiological responses to aversive situations also occur. Bulls showed a decreased corticosteroid response to injected ACTH after five weeks of tethering, compared with control animals kept loose on straw (Ladewig & Smidt 1989), whereas male pigs showed an increased responsiveness to ACTH after a similar period of tethering (von Borell & Ladewig 1989). Species also differ in their behavioural responses, for example to pain (reviewed by Bateson 1991). Bateson speculates (after Fraser & Broom 1990, p 269 *et seq*) that certain 'prey' species have evolved not to show signs of weakness nor readily to make sounds of distress, even when they are injured. For example the apparently

stoical reactions of sheep to injury are in sharp contrast to those of an injured dog. Some species, when faced with a strongly aversive stimulus, will freeze, while others will flee (Bolles 1970, Archer 1979b). Strains within a species can also differ. Strains of hen with the greatest peak in heart rate in response to alarming visual stimuli also have the quickest return to normal (Duncan 1985), making it difficult to assess whether their welfare is poorer than that of a slow-to-recover strain.

Taking the species' natural history into account can help in the interpretation of cases such as these. An understanding of the species' biology helps us to predict the nature of its emergency responses, and to select and evaluate our measures of welfare accordingly. Consider the great concern one would feel for, let us say, a sheep that screamed, compared with the lesser concern one would feel for a shrieking pig or monkey. To illustrate further, small rodents often show freezing behaviour in the natural environment, which allows them to avoid detection by predators. Therefore, their freezing response in the open-field test is a readily-performed, appropriate indicator of fear in these species (Daly 1977). Freezing is a less common response in carnivores such as cats, however; when this reaction occurs in these species it might be thought of as indicating a more severe welfare problem.

Age and sex differences

The age of the animals being studied can have a pronounced effect on the welfare measures being assessed. For example, *in vitro* tests of immune function, such as mitogenesis, suggest that immune system activity declines with age in certain species (eg Manser 1992, p 139). The stereotypies developed in animals that are very young when exposed to particular conditions often tend to be more severe, in the sense of being less reversible, than those developed in older animals (reviewed by Mason 1991a, 1993). Conversely, the responsiveness of the adrenal glands of pigs to an ACTH injection appears to increase with age (Cooper *et al* 1989, Parrott *et al* 1989, Mendl *et al* 1992).

Sex differences in responses to threatening or challenging situations also occur. For example, a large amount of work on rats suggests that females show lower levels of freezing and defecation in open-field tests than males; appear to be less prone to develop ulcers in conflict-inducing situations, and are better at learning active avoidance behaviour than males (reviewed by Gray 1987, see also Archer 1979b).

Animals of a different age or sex may thus respond to the same situation in quite different ways, primarily because of prior differences in their underlying physiological and neurological systems, rather than differences in the impact which the situation has on their welfare.

Individual differences and individual characteristics

Many of the factors that we have considered so far can help to explain why different types of individuals or types of stressor produce different responses. However, many

differences are seen when individuals of one type are exposed to one particular situation, and these are not easily attributable to known causal factors. Consequently, researchers often invoke concepts such as individuality, temperament or behavioural style to account for these differences (Mendl & Harcourt 1988). Although these terms are often used in a loose and confused way, an implicit assumption underlying their use is that there is some degree of internal consistency in the ways in which a given individual responds to situations. Evidence for this is now accumulating from a variety of studies.

In threatening, unfamiliar situations, some young rhesus monkeys are slow to explore and show high heart rates and a large rise in plasma cortisol, while others appear much less behaviourally inhibited and show much smaller physiological responses (Suomi 1987). Similar individual differences are also seen in human children (Kagan *et al* 1988), and in both species there is evidence of cross-time consistency in these individual characteristics.

Amongst male tree shrews, some individuals react to a dominant by actively avoiding him and showing a pronounced activation of the adrenomedullary-sympathetic system, while others submit and succumb to inertia, lowered immunity, and chronically elevated corticosteroids (von Holst *et al* 1983). Similar findings come from extensive studies of rats and mice. Some rats (Katzev & Mills 1974, Dantzer *et al* 1988) and house mice (Benus *et al* 1990a) freeze when given electric shocks, while others attempt to flee. The same is true of rats (Dantzer *et al* 1988) and mice (Benus *et al* 1992) faced with an aggressive conspecific. Benus and co-workers (1989, 1990b) coined the terms 'active' and 'passive' for these two extreme styles of reacting to adversity. They also found that these styles were correlated with a number of other features of the animal's behavioural organization (see also McCune 1992). Active mice were found to be more aggressive and to be less sensitive to changes in the environment than passive mice (Benus *et al* 1987, 1988). There is some evidence that differences in brain chemistry may partly explain some of the differences in the response styles of these mice (Benus *et al* 1991), and these differences may be genetic in origin (van Oortmerssen & Bakker 1981).

Thus, there is mounting evidence that in some species at least, individuals have their own consistent modes of response to challenging situations. In addition, clusters of characteristics tend to occur together within individuals and represent alternative, and sometimes perhaps equally successful, strategies for coping with adversity (eg Mendl *et al* 1992). It is often useful to try to assess the characteristics of subject animals before initiating welfare studies of their response to specific situations or procedures (von Borell & Ladewig 1992), so as to try to control for any effects that their characteristics may have on the responses recorded. If, prior to an experiment, subjects in one group have consistently different individual characteristics from those in another group, these individual differences will confound the results (eg see Archer's discussion [1979a] of Brady's 'executive monkeys' experiment).

Welfare measures do not always measure welfare: functional significance and specificity of responses

So far we have considered why different welfare variables respond in diverse ways to aversive conditions. In this last section, we turn to the separate issue of whether these welfare measures reliably indicate that the situation is in fact aversive to the animal. Some measures, indeed most, may also change in response to non-aversive events: neutral and even pleasant things may affect them. This is so for at least four common welfare measures: corticosteroids, prolactin, weight loss and heart rate.

Corticosteroid levels have been shown to increase during coitus in some male rats (Szechtman *et al* 1974); just before laying in hens (Beuving 1983); while nursing in mammals (Walker *et al* 1992), and in situations of chronic cold even when psychological stress appears to be absent (Mason 1971). Furthermore they vary with diurnal rhythms (Ladewig 1984, 1987) and are released in a pulsatile or episodic manner (Ladewig 1984, 1987). Thus their levels fluctuate even when the animal is not faced with unpleasant stimuli.

Prolactin varies not only with an animal's welfare, but also with time of year. For example, if using this hormone as a welfare measure in mink, one would have to take into account its seasonal involvement in the moult (Martinet *et al* 1982, 1985). Prolactin levels also depend on the stage of oestrus, for example in the female rat (reviewed by Manser 1992, p 58). Heart rate, too, varies with factors other than fear or anxiety. A heart rate increase can reflect metabolic demands, circadian rhythms or, in humans, that the subject smokes (Burdick 1978, Dantzer *et al* 1983a,b).

Differences in body-weight should also be interpreted with caution. Although stress often suppresses growth (eg Christian & Davis 1964, Fowler 1986) and encourages weight loss (eg Weiss 1968, 1971), perhaps due to the catabolic effects of corticosteroids, weight will also be influenced by diet, by activity levels, and possibly by other factors too. Dawkins (1980) points out that a turkey might gain weight quickly, simply because it is in such close confinement that it can expend little energy in exercise; therefore one cannot infer much about welfare from measures of weight alone.

So, it is necessary to be aware of what other factors could influence measures of welfare, and why. Once again, understanding the function of the system being measured can particularly help in interpretation. It is unsurprising if corticosteroids are released when activity levels or environmental temperatures demand an increase in metabolic rate. Likewise, the hormone prolactin has specific functions that influence when it is released, irrespective of whether the animal is facing a threatening or challenging situation. If a variable is affected by a range of events, it is vital to check for these or to control for these when assessing welfare.

Discussion

Our aim in this review has been to point out a number of reasons why the scientific measurement of animal welfare is often a complicated and confusing process, if welfare

is considered more than just physical health and reproductive ability. This is not meant to be a counsel of despair. Studies of animal welfare have sometimes come up with relatively clear-cut conclusions, that have stood the test of replication and have demonstrated good agreement between a variety of welfare measures. The effect of the quality of stockmanship on welfare and production measures in pigs is a good example (Hemsworth *et al* 1986, 1987). However, there is still a need to try to identify the factors that have caused other studies to produce conflicting and equivocal results. As Dawkins put it 'it is necessary to be aware of the limitations of your materials before you start building a house' (Dawkins 1980, p 108). Our brief survey has shown the importance of considering the animal as a whole, and also its context. The factors which we think are particularly important can be summarized under four main headings.

First, an understanding of the mode of action and functional significance of the welfare measure being used will facilitate good experimental design and will also increase the ease with which results can be interpreted (Rushen 1991). For example, heart rate changes reflect physical activity as well as emotional state. Therefore, experiments which use this measure and also control for the animal's activity levels (Baldock & Sibly 1990) are more likely to produce information that really indicates emotional state than studies which do not. In addition, this sort of knowledge will help us to select appropriate welfare measures according to the nature of the stressor that we are studying. For instance, if we are examining whether the temperature in a housing system is too cold for the animals, corticosteroids would not be a good welfare measure to use because they have a homeostatic thermogenic function which is brought into play irrespective of whether or not the animal is suffering (Mason 1971). Nor are stereotypes likely to be appropriate, as animals often react to sustained cold with huddled inactivity rather than with behaviourally active responses (Robbins *et al* 1990, Mason 1991b). Other measures should be used, and in addition, an assessment of an animal's preference for identical environments at different temperatures, or a measure of how hard it will work to avoid the cold temperature (Dawkins 1983, 1988, 1990) may yield more useful information about welfare.

Second, species differences mean that cross-species generalizations are often not valid. Different species will be adapted to different ways of life and, therefore, they will be differentially affected by the same stressor, and their methods of responding to this may also be quite different. Consequently, certain measures of welfare will be more appropriate for certain species.

Third, knowledge of the individuals used in welfare studies will also aid interpretation of results. As we have seen, individual characteristics such as sex, age and experience, and the individual's current state, can affect the responses it makes to a challenging situation. Therefore, attempts to control for these factors, or at least to make some baseline or pre-trial assessment of them, will increase the chances of reliable results being obtained (von Borell & Ladewig 1992).

Finally, details of experimental design, such as the timing of presentation of the stressor and the duration of exposure, can have a pronounced effect on the animal's responses which may overshadow or confound any changes which are due to the relative aversiveness of the situations being studied. It is possible that factors such as these partly explain the diverse changes in the activity of the adrenocortical system, as measured by ACTH challenge, which have been found by different researchers examining the effects of chronic confinement, restraint and heat stress (see Rushen 1991, Mendl *et al* 1992).

Clearly, even the most well designed and tightly controlled experiment is unlikely to be able to deal with all the potential problems listed above. Nevertheless, if welfare studies take at least some of these factors into account before experimental work is started, the chance of obtaining reliable, consistent results will be increased. Also, it is clear that more knowledge is required about the precise actions and functions of many of the responses we assess as measures of welfare, even the very commonly used measures of stereotypy (Mason 1991a,b) and pituitary-adrenocortical activity (Rushen 1991). There is a strong case for arguing that welfare research should include more fundamental work in these areas.

Even when a superbly designed experiment has been carried out and the results seem consistent across several types of measure, we are still faced with the problem of how to interpret the measures we have collected in terms of animal welfare. One approach is to consider the effects of the measured changes, as well as their magnitude. Some changes in behaviour and physiology themselves cause states which are deleterious to welfare; one might therefore choose to give the most weight to these measures (eg Moberg 1985b). For example, sustained low immunity may result in an animal succumbing to disease; some persistent stereotypies can result in an animal injuring itself (reviewed by Mason 1991a), and gastric ulcers are a product of chronic stress (eg Moody *et al* 1976) that are very likely to cause discomfort or pain. Prolonged high levels of corticosteroids may also cause or indicate metabolic, immunological and reproductive problems (Cross 1989, Barnett & Hemsworth 1990), though not all agree with this; see Rushen and de Passillé 1992. An additional technique is to supplement the assessment of an animal's welfare with measurement of its preferences for particular stimuli or situations (for a full discussion of the pros and cons of such methods see eg Dawkins 1990).

Although many different measures of animal welfare are important, in the end it is always difficult to relate them precisely to what the animal is actually feeling (Dawkins 1990). There is no simple way of knowing how much weight to give each of the various measures in this respect, nor at what level a measure indicates suffering. Indeed, as there are no doubt degrees of good and poor welfare, it is often inappropriate to impose specific cut-off values ('above value *x* welfare is poor, below value *x* welfare is fine') on our measures (Mendl 1991). Science can never 'prove' that an animal is or is not suffering, because we can never really access the private world of another's mind (though cf Wemelsfelder 1993). But what science can be used for is the collection of evidence from which to make inferences (much like those made by the clinician who uses symptoms to make a judgement about a disease). Even when all the problems with

measuring animal welfare have been ironed out, and our judgements are as well-informed as possible, these judgements will remain ultimately subjective. They will be influenced by the type of ethical concerns that motivate us (eg Bateson 1991); our consideration for other factors like economic or environmental costs (Sandøe and Simonsen 1992), and the extent to which we are prepared to give the animal the benefit of the doubt.

Acknowledgements

The authors thank Don Broom, Sandra McCune, Liz Paul, Peter Raffan and Michael Simpson for their useful comments on this manuscript. Georgia Mason was supported by a Clare College (University of Cambridge) Research Fellowship and Mike Mendl by a British Veterinary Association Research Fellowship.

References

- Archer J 1973 Tests for emotionality in rats and mice: a review. *Animal Behaviour* 21: 205-235
- Archer J 1979a *Animals Under Stress*. Edward Arnold: London
- Archer J 1979b Behavioural aspects of fear. In Sluckin W (ed) *Fear in Animals and Man* pp 56-85. Van Nostrand Reinhold: New York
- Baldock N M and Sibly R M 1990 Effects of handling and transportation on the heart rate and behaviour of sheep. *Applied Animal Behaviour Science* 28: 15-39
- Barnett J L and Hemsworth P H 1990 The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science* 25: 177-187
- Barnett J L, Hemsworth P H and Winfield C G 1987 The effects of design of individual stalls on the social behaviour and physiological responses related to the welfare of pregnant pigs. *Applied Animal Behaviour Science* 18: 133-142
- Bateson P P G 1991 Assessment of pain in animals. *Animal Behaviour* 42: 827-839
- Benus R F, Koolhaas J M and Oortmerssen G A van 1987 Individual differences in behavioural reaction to a changing environment in mice and rats. *Behaviour* 100: 105-122
- Benus R F, Koolhaas J M and Oortmerssen G A van 1988 Aggression and adaptation to the light-dark cycle: role of intrinsic and extrinsic control. *Physiology and Behaviour* 43: 131-137
- Benus R F, Koolhaas J M and Oortmerssen G A van 1992 Individual strategies of aggressive and non-aggressive male mice in encounters with trained residents. *Animal Behaviour* 43: 531-540
- Benus R F, Bohus B, Koolhaas J M and Oortmerssen G A van 1989 Behavioural strategies of aggressive and non-aggressive male mice in active shock avoidance. *Behavioural Processes* 20: 1-12

- Benus R F, Bohus B, Koolhaas J M and Oortmerssen G A van 1990b Behavioural strategies of aggressive and non-aggressive male mice in response to inescapable shock. *Behavioural Processes* 21: 127-141
- Benus R F, Bohus B, Koolhaas J M and Oortmerssen G A van 1991 Behavioural differences between artificially selected aggressive and non-aggressive mice: response to apomorphine. *Behaviour and Brain Research* 43: 203-208
- Benus R F, Daas S den, Koolhaas J M and Oortmerssen G A van 1990a Routine information and flexibility in social and non-social behaviour of aggressive and non-aggressive male mice. *Behaviour* 112: 176-193
- Beuving G 1983 Corticosteroids in welfare research of laying hens. In Smidt D (ed) *Indicators Relevant to Farm Animal Welfare* pp 47-53. Martinus Nijhoff: The Hague
- Bolles R C 1970 Species-specific defense reactions and avoidance learning. *Psychological Review* 77: 32-48
- Borell E von and Ladewig J 1989 Altered adrenocortical response to acute stressors or ACTH (1-24) in intensively housed pigs. *Domestic Animal Endocrinology* 6: 299-309
- Borell E von and Ladewig J 1992 Relationship between behaviour and adrenocortical response pattern in domestic pigs. *Applied Animal Behaviour Science* 34: 195-206
- Bredbacka P 1988 Relationships between fear, welfare, and productive traits in caged white leghorn hens. In Unshelm J, Putten G van, Zeeb K and Ekesbo I (eds) *Proceedings of the International Congress on Applied Ethology in Farm Animals, Skara* pp 74-79. KTBL: Darmstadt
- Broom D M 1991 Assessing welfare and suffering. *Behavioural Processes* 25: 117-123
- Broom D M and Johnson K (in press) *Stress and Animal Welfare*. Chapman and Hall: London
- Burdick J A 1978 A review of heart rate variability and evaluation. *Perception and Motor Skills* 47: 95-105
- Campos J J 1976 Heart rate: a sensitive tool for the study of emotional development in the infant. In Lipsitt L P (ed) *Developmental Psychology - the Significance of Infancy* pp 1-31. Lawrence Erlbaum: Hillsdale, New Jersey
- Christian J J and Davis D E 1964 Endocrines, behavior, and population. *Science* 146: 1550-1560
- Cohen-Cole S, Cogen R, Stevens A, Kirk K, Gaitan E, Hain J and Freeman A 1981 Psychosocial, endocrine and immune factors in acute necrotizing gingivitis. *Psychosomatic Medicine* 43: 91 (Abstract)
- Cooper T R, Trunkfield H R, Zanella A J and Booth W D 1989 An enzyme-linked immunosorbent assay for cortisol in the saliva of man and domestic farm animals. *Journal of Endocrinology* 123: R13-R16

- Cross B 1989 The physiological (neurological) basis of fear. In Gibson T E (ed) *The Detection and Control of Fear in Animals, Proceedings of the BVA Animal Welfare Foundation 8th Symposium* pp 5-14. BVA Animal Welfare Foundation: London
- Daly M 1977 Early stimulation of rodents: a critical review of present interpretations. *British Journal of Psychology* 64: 435-460
- Dantzer R and Mormède P 1985 Stress in domestic animals: a psycho-neuroendocrine approach. In Moberg G P (ed) *Animal Stress* pp 81-95. American Physiological Society: Bethesda
- Dantzer R, Mormède P and Henry J P 1983a Physiological assessment of adaptation in farm animals. In Baxter S H, Baxter M R and McCormack J A D (eds) *Farm Animal Housing and Welfare* pp 8-19. Martinus Nijhoff: The Hague
- Dantzer R, Mormède P and Henry J P 1983b The significance of physiological criteria in assessing animal welfare. In Smidt D (ed) *Indicators Relevant to Assessing Farm Animal Welfare* pp 29-37. Martinus Nijhoff: The Hague
- Dantzer R, Terlouw C, Tazi A, Koolhaas J M, Bohus B, Koob G F and Moal M le 1988 The propensity for schedule-induced polydipsia is related to differences in conditioned avoidance behaviour and in defence reactions in a defeat test. *Physiology and Behaviour* 43: 269-272
- Dawkins M 1980 *Animal Suffering: The Science of Animal Welfare*. Chapman and Hall: London
- Dawkins M 1983 Battery hens name their price: consumer demand theory and the measurement of ethological 'needs'. *Animal Behaviour* 31: 1195-1205
- Dawkins M 1988 Behavioural deprivation: a central problem in animal welfare. *Applied Animal Behaviour Science* 20: 209-225
- Dawkins M 1990 From an animal's point of view: motivation, fitness and animal welfare. *Behavioural and Brain Sciences* 13: 1-61
- Dorian B J, Keystone E, Garfinkel P E and Brown G M 1981 Immune mechanisms in acute psychological stress. *Psychosomatic Medicine* 43: 84 (Abstract)
- Duncan I J H 1985 How do fearful birds respond? In Wegner R M (ed) *Proceedings of the Second European Symposium on Poultry Welfare* pp 96-106. World Poultry Science Association: Celle
- Duncan I J H and Petherick J C 1989 Cognition: the implications for animal welfare. *Applied Animal Behaviour Science* 24: 81 (Abstract)
- Fentress J C 1976 Dynamic boundaries of patterned behaviour: interaction and self-organisation. In Bateson P P G and Hinde R A (eds) *Growing Points in Ethology* pp 135-168. Cambridge University Press: Cambridge
- Fowler M E 1986 Stress. In Fowler M E (ed) *Zoo and Wild Animal Medicine, 2nd edition* pp 34-35. Saunders: Philadelphia

- Fox M W 1984 *Farm Animals: Husbandry, Behavior and Veterinary Practice*. University Park Press: Baltimore
- Fraser A F and Broom D M 1990 *Farm Animal Behaviour and Welfare*. Baillière-Tindall: London
- Gray J A 1987 *The Psychology of Fear and Stress, 2nd edition*. Cambridge University Press: Cambridge
- Hemsworth P H, Barnett J L and Hansen C 1986 The influence of handling by humans on the behaviour, reproduction and corticosteroids of male and female pigs. *Applied Animal Behaviour Science* 15: 303-314
- Hemsworth P H, Barnett J L and Hansen C 1987 The influence of inconsistent handling by humans on the behaviour, growth and corticosteroids of young pigs. *Applied Animal Behaviour Science* 17: 245-252
- Henry J P, Ely D L, Stephens O P M, Ratcliffe H L, Santisteban G A and Shapiro A P 1971 The role of psychosocial factors in the development of atherosclerosis in CBA Mice. *Atherosclerosis* 14: 203-218
- Herbert J 1987 Neuroendocrine responses to social stress. *Baillière's Clinical Endocrinology and Metabolism* 1: 467-490
- Holst D von, Fuchs E and Stohr W 1983 Physiological changes in male *Tupaia belangeri* under different types of social stress. In Dembroski T M, Schmidt T H and Blumchen G (eds) *Biobehavioral Bases of Coronary Heart Disease* pp 382-390. Karger: Basel
- Kagan J, Reznick J S and Snidman N 1988 Biological bases of childhood shyness. *Science* 240: 167-171
- Katzev R D and Mills S K 1974 Strain differences in avoidance conditioning as a function of the classical CS-US contingency. *Journal of Comparative Physiology and Psychology* 87: 661-671
- Kennes D and Rycke P H de 1988 The influence of the performance of stereotypies on plasma corticosterone and eosinophil levels in bank voles (*Clethrionomys glareolus*). In Unshelm J, Putten G van, Zeeb K and Ekesbo I (eds) *Proceedings of the International Congress on Applied Ethology in Farm Animals, Skara* pp 238-240. KTBL: Darmstadt
- Kristensen M P and Jeppesen L L 1988 Effects of experimentally-induced stress on cortisol, blood cell parameters and exploratory behaviour in farmed foxes. *Scientificur* 12: 169-176
- Ladewig J 1984 The effect of behavioural stress on the episodic release and circadian variation of cortisol in bulls. In Unshelm J, Putten G van and Zeeb K (eds) *Proceedings of the International Congress on Applied Ethology in Farm Animals, Kiel* pp 339-342. KTBL: Darmstadt

- Ladewig J 1987 Endocrine aspects of stress: evaluation of stress reactions in farm animals. In Wiepkema P R and Adrichem P W M van (eds) *The Biology of Stress in Farm Animals: An Integrated Approach* pp 13-25. Martinus Nijhoff: Dordrecht
- Ladewig J 1990 Physiological stress responses in large farm animals. Paper given at the *Workshop on Stress and Abnormal Behaviour, Elsinor, Denmark, Sept 1990*
- Ladewig J and Borell E von 1988 Ethological methods alone are not sufficient to measure the impact of environment on animal health and animal well-being. In Unshelm J, Putten G van, Zeeb K and Ekesbo I (eds) *Proceedings of the International Congress on Applied Ethology in Farm Animals, Skara* pp 95-102. KTBL: Darmstadt
- Ladewig J and Smidt D 1989 Behaviour, episodic secretion of cortisol and adrenocortical reactivity in bulls subjected to tethering. *Hormones and Behaviour* 23: 344-360
- Lundberg U and Forsmann L 1979 Adrenal-medullary and adrenal cortical responses to understimulation and overstimulation: a comparison between Type A and Type B persons. *Biological Psychology* 9: 79-89
- Lyons D M, Price E O and Moberg G P 1988 Social modulation of pituitary-adrenal responses and individual differences in behaviour of young domesticated goats. *Physiology and Behaviour* 43: 451-458
- McCune S 1992 *Temperament and the Welfare of Caged Cats*. PhD thesis, University of Cambridge
- MacSween R N M and Whaley K (eds) 1992 *Muir's Textbook of Pathology, 13th edition*. Edward Arnold: London
- Manser C E 1992 *The Assessment of Stress in Laboratory Animals*. RSPCA: Horsham
- Martin P 1989 Psychoimmunology - relations between brain, behaviour and immune function. In Bateson P P G and Klopfer P H (eds) *Whither Ethology?, Volume 8 of Perspectives in Ethology* pp 173-214. Plenum Press: New York
- Martinet L, Allain D and Chabi Y 1985 Pineal denervation by cervical sympathetic ganglionectomy suppresses the role of photoperiod on pregnancy or pseudopregnancy, body weight and moulting periods in the mink (*Mustela vison*). *Journal of Endocrinology* 107: 31-39
- Martinet L, Ravault J P and Meunier M 1982 Seasonal variations in mink (*Mustela vison*) plasma prolactin measured by heterologous radioimmunoassay. *General and Comparative Endocrinology* 48: 71-75
- Mason G J 1991a Stereotypies: a critical review. *Animal Behaviour* 41: 1015-1037
- Mason G J 1991b Stereotypies and suffering. *Behavioural Processes* 25: 103-115
- Mason G J 1993 Forms of stereotypic behaviour. In Lawrence A B and Rushen J (eds) *Stereotypic Animal Behaviour: Fundamentals and Applications to Animal Welfare* pp 7-40. CAB: Wallingford

- Mason J W** 1971 A re-evaluation of the concept of 'non-specificity' in stress theory. *Journal of Psychiatric Research* 8: 323-333
- Mendl M** 1991 Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146
- Mendl M and Harcourt R** 1988 Individuality in the domestic cat. In Turner D C and Bateson P (eds) *The Domestic Cat: The Biology of its Behaviour* pp 41-54. Cambridge University Press: Cambridge
- Mendl M, Zanella A J and Broom D M** 1992 Physiological and reproductive correlates of behavioural strategies in female domestic pigs. *Animal Behaviour* 44: 1107-1121
- Moberg G P** 1985a Influence of stress on reproduction: measure of well-being. In Moberg G P (ed) *Animal Stress* pp 245-267. American Physiological Society: Bethesda
- Moberg G P** 1985b Problems in defining stress and distress in animals. In Moberg G P (ed) *Animal Stress* pp 26-49. American Physiological Society: Bethesda
- Moody F G, Cheung L Y, Simons M A and Zalewsky C** 1976 Stress and the acute gastric mucosal lesion. *Digestive Diseases* 21: 148-154
- Ödberg F** 1989 Adaptation: a boon and a nuisance. Paper given at the *International Ethological Congress, Utrecht*
- Oortmerssen G A van and Bakker T C M** 1981 Artificial selection for short and long attack latencies in wild *Mus musculus domesticus*. *Behavioural Genetics* 11: 115-126
- Parrott R F, Misson B N and Baldwin B A** 1989 Salivary cortisol in pigs following adrenocorticotrophic hormone stimulation: comparison with plasma levels. *British Veterinary Journal* 145: 362-366
- Pederson V and Jeppesen L L** 1990 Effects of early handling on later behaviour and stress responses in the silver fox (*Vulpes vulpes*). *Applied Animal Behaviour Science* 26: 383-393
- Pfister H P** 1979 The glucocorticosterone response to novelty as a psychological stressor. *Physiology and Behaviour* 23: 649-652
- Rice R P** 1967 The mink: a short outline of general management. *Journal of the Institute of Animal Technicians* 18: 66-79
- Robbins T W, Mittleman G, O'Brien J and Winn P** 1990 The neurophysiological significance of stereotypy induced by stimulant drugs. In Cooper S J and Dourish C T (eds) *Neurobiology of Stereotyped Behaviour* pp 25-63. Clarendon: Oxford
- Rushen J** 1986 Some problems with the physiological concept of 'stress'. *Australian Veterinary Journal* 63: 359-361
- Rushen J** 1991 Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science* 28: 381-386
- Rushen J and de Passillé A M B** 1992 The scientific assessment of the impact of housing on animal welfare. *Canadian Journal of Animal Science* 72: 721-743

- Sandøe P and Simonsen H P 1992 Assessing animal welfare: where does science end and philosophy begin? *Animal Welfare 1*: 257-267
- Sawrey W L, Conger J J and Turrell E S 1956 An experimental investigation of the role of psychological factors in the production of gastric ulcers in rats. *Journal of Comparative Physiology and Psychology 49*: 457-461
- Selye H 1979 Stress. In Day S B (ed) *A Companion to the Life Sciences, volume 1*. Van Nostrand Reinhold: New York
- Suomi S J 1987 Genetic and maternal contributions to individual differences in rhesus monkey biobehavioral development. In Krasnegor N A, Blass E M, Hofer M A and Smotherman W P (eds) *Perinatal Development: A Psychobiological Perspective*. Academic Press: New York
- Szechtman H, Lambrou P J, Caggiula A R, and Redgate E S 1974 Plasma corticosterone levels during sexual behaviour in male rats. *Hormones and Behaviour 5*: 191-200
- Trunkfield H R, Broom D M, Maatje K, Wierenga H K, Lambooy E and Kooijman J 1991 Effects of housing on responses of veal calves to handling and transport. In Metz J H M and Groenestein C M (eds) *New Trends in Veal Calf Production* pp 40-43. Wageningen Pudoc: Wageningen
- Walker C D, Lightman S L, Steele M K and Dallaman M F 1992 Suckling is a persistent stimulus to the adreno-cortical system of the rat. *Endocrinology 130*: 115-125
- Wegner R M 1983 Production performance in laying hens kept under different housing systems. In Smidt D (ed) *Indicators Relevant to Farm Animal Welfare* pp 186-196. Martinus Nijhoff: The Hague
- Weiss J M 1968 Effects of coping responses on stress. *Journal of Comparative Physiology and Psychology 65*: 251-260
- Weiss J M 1971 Effects of coping behaviour in different warning signal conditions on stress pathology in rats. *Journal of Comparative Physiology and Psychology 77*: 1-13
- Wemelsfelder F 1993 *Animal Boredom: Towards an Empirical Approach of Animal Subjectivity*. PhD thesis, University of Leiden
- Wiepkema P R 1987 Behavioural aspects of stress. In Adrichem P W M van and Wiepkema P R (eds) *The Biology of Stress in Farm Animals: An Integrated Approach* pp 113-134. Martinus Nijhoff: Dordrecht
- Wiepkema P R, Hellemond K K van, Roessingh P and Romberg H 1987 Behaviour and abomasal damage in individual veal calves. *Applied Animal Behaviour Science 18*: 257-268