The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK

Animal Welfare 2011, 20: 133-143
ISSN 0962-7286

Relationships between multiple welfare indicators measured in individual chickens across different time periods and environments

CJ Nicol*, G Caplen, J Edgar, G Richards and WJ Browne

School of Veterinary Science, University of Bristol, Langford House, Langford BS40 5DU, UK
* Contact for correspondence and requests for reprints: c.j.nicol@bris.ac.uk

Abstract

The assessment of animal welfare requires the collection of multiple indicators of welfare but quantification of their associations in different contexts is lacking. Previous studies have examined correlations between a few indicators, but not relationships between many different indicators, or between indicators taken from the same individuals in more than one environment. We housed 60 hens for six sequential 35-day phases in different pen environments. During each phase, a series of behavioural and physiological measures was taken for every bird: body and plumage condition, surface body temperature, behaviours observed in the home pens and during test periods, tonic immobility, physiological blood profiles, and faecal sample composition. Most variation in nearly all measures was not explained by either individual bird or grouping effects but varied across phases within the birds. Acknowledging this, we examined correlations between all parameters at the phase within-bird level, selecting a conservative P-value. A consistent set of correlations showed that a slow approach response and alert behaviour in the novel object test was associated with higher bodyweight, lower body temperature and lower acute phase protein, heterophil:lymphocyte ratio and blood glucose level. A cluster analysis confirmed these correlations. Other important parameters known to be linked to the hens' environmental preference (eg comfort behaviour) were independent of the set described above. We conclude that statistical techniques can reveal patterns of independence and redundancy in the collection of behavioural and physiological measures of welfare.

Keywords: *animal welfare, cluster analysis, laying hen, multi-level modelling, stress, welfare indicator*

Introduction

Animal welfare science is increasingly seen as a multidisciplinary exercise. No one marker can indicate good or poor welfare (Mason & Mendl 1993) and welfare assessments are more likely to incorporate multiple markers than in the past. For example, information on physiology was obtained for only 2% of 300 laying hen flocks included in the Laywel database in 2005 (http://www.laywel.eu), whereas recent experimental studies have generally taken multiple measures of welfare, including organ weights, white blood cell ratios, plasma, faecal or egg corticosterone levels, tests of immune function and observations of behaviour (Nicol *et al* 2006; Barnett *et al* 2009; Singh *et al* 2009; Tactacan *et al* 2009; Thogerson *et al* 2009). The collection of multiple measures for monitoring hen welfare on farms has also become accepted practice (eg Welfare Quality, www.welfarequality.net).

If different aspects of welfare could be separated into independent, non-overlapping components (eg injury, stress response, resting comfort), and each component unambiguously measured, then just one measure would be needed to represent each component. The relative importance of the components could be weighted, and the measures combined to draw overall conclusions. In reality, though, welfare is not a combination of separate non-overlapping components. Different aspects are inter-related in complex ways (eg injury will affect resting comfort), and the underlying biological systems are hugely complex. As a precaution, scientists take many measures of each aspect of welfare (eg Rodenburg *et al* 2008).

This is a pragmatic approach, but one that risks potentially expensive and time-consuming redundancy in data collection (Richard *et al* 2007). In addition, combining non-independent measures to draw overall conclusions is difficult, especially when they do not co-vary in a consistent way. For example, Nicol *et al* (2006) found that hens housed in single-tier aviaries at low stocking densities had higher mortality and worse plumage condition than hens housed at higher densities, but lower percent liver weights, indicating relatively lower stress levels (eg Thaxton & Puvadolpirod 2000). It is difficult to interpret such findings without knowing whether the lack of co-variance occurs at a flock or an individual level. At a flock level, severe feather pecking could result in mortality and feather damage to a proportion of the flock, but in relatively low stress conditions for the surviving perpetrators. However, if *individual*

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

birds had both poor plumage condition and low liver weights this would be more difficult to explain.

Decisions about which measures to take would be helped by prior statistical information on the relationships between different welfare indicators in different contexts. Currently, very little such information is available for hens. Many studies report informally that a particular treatment, situation or genetic breed difference influences two or more indicators in the same way (eg Cheng & Jefferson 2008; Campo & Prieto 2009; Kjaer 2009; Sherwin *et al* 2010) and, occasionally, the relative time-courses of different welfare indicators in response to a given situation are examined (eg Dawkins *et al* 2004). However, in these studies, as in most others, different measures were taken on different birds, or information from groups of birds was pooled before analysis.

Individual-level explanations require data for both variables to be obtained from the same birds. Studies that have done so have focused mostly on traits correlated with feather pecking (Vestergaard *et al* 1993; Cloutier & Newberry 2002; Albentosa *et al* 2003) and fear (Ghareeb *et al* 2008) and rarely examined a wide range of indicators (though see Webster & Hurnik 1991). Recent work showing that hens fall into different 'types' based on their different patterns of response to behavioural tests (Ghareeb *et al* 2008) or environmental conditions (Nicol *et al* 2009), provides a further reason to ensure that relationships between welfare indicators are examined at a within-bird level. Decisions to streamline data collection should be taken only if the measures are strongly correlated for individual birds.

In a previous paper, we investigated the relationship between welfare indicators and environmental preference in laying hens (Nicol *et al* 2009). The aim of this further analysis was to use our extensive dataset to investigate relationships between the many measures recorded on each bird during the six housing phases of the experiment, ie at a phase within-bird level. We first examined the amount of variation in each indicator that occurred at the various levels in the dataset (phases nested within individual birds nested within groups). Indicators could be primarily socially shared traits, fixed individual traits, or flexible responses that moderate with different environments. The relative variances estimate the likelihood of these three scenarios for each indicator. This is important information to be used in deciding whether future studies need to collect data at an individual or flock level. We then examined correlations between pairs of selected indicators. Finally, we examined how the different indicators were related more globally, using cluster analyses.

Materials and methods

Subjects and housing

Sixty, medium Hyline laying hens were obtained at 18 weeks of age and housed in 24 pens (0.96 × 1.2 × 2.0 m; width × depth × height) either in groups (12 groups of 4 birds) or individually (12 birds) with *ad libitum* feed and water. Individual birds could see other birds only when they were feeding at the externally mounted feed trough. Room temperature was held at between 18 and 22°C, and the birds were on a schedule of 12 h light:12 h dark. The pens could be converted into three different environments by altering the floor type and resources available: Wire floor (W), Shavings floor (Sh), and Peat, Perch and Nest-box (PPN), with additional peat and wood shavings on top of a solid wooden base floor, two perches of diameter 0.035 m fitted across the width of the pen at heights of 0.25 and 0.45 m from floor level and a nest-box. The hens lived in each environment for a 35-day period, before moving to a new environment established in a pen on the opposite side of a corridor. In total, each of the three environments was experienced twice in six sequential phases counterbalanced for order between groups. All work was conducted under UK Home Office licence.

Welfare indicator sampling

During each 35-day housing phase, physical examination of the birds was conducted on days 24 and 32. Faecal samples were collected in the afternoons; one pooled sample was collected over days 29 to 31 from each bird. Blood samples were taken in the mornings from half the birds on day 29 and the other half on day 30. Direct behavioural observations were taken in the mornings on day 16 and 31, tonic immobility (TI) tests were given on the afternoons of days 16 and 30, novel object tests and resource competition tests were conducted throughout day 23. Video recordings were made on days 5, 12, 19, 26 and 33, each for 15 min starting at 0900h and 15 min starting at 1615h. Further details of the techniques used are given in Nicol *et al* (2009).

Physical examination of birds (days 24 and 32 of each phase)

Each hen was weighed, then gently held by an assistant whilst integument and plumage condition was assessed. The condition of the skin and plumage of the head, neck, breast, back, tail, wings and vent were scored on ordinal scales 0–5 (no damage to extensive or severe damage). Ordinal scales were also used to record comb elevation 0–2 (upright, partially floppy, completely floppy), comb colour 0–2 (red, medium, pale), comb size 0–3 (0 < 3.5 cm^2; 1 = 3.5 to 5.5 cm^2, 2 = 5.5–7.5 cm^2; 3 > 7.5 cm^2), foot cleanliness 0–5 (clean to very dirty). Claw length was measured in mm. The temperature of the head, comb, eye and foot was recorded at a distance of 1 m from the bird, within 1 min of removal from the home pen to a holding pen at the same ambient temperature, using a thermal camera (ThermaCAM, FLIR systems, http://www.flir.com/Thermography/eurasia/en/). Foot lesions, crouched posture, lack of alertness, dull eyes and plumage damage (except damage to the breast region) occurred rarely or never and were not analysed further. Inter-observer reliability was checked at intervals throughout the experiment. Differences between the two time points were minimal for the physical indicators variables, and for the statistical analyses reported here, average values from the two measurements collected in each phase were used.

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Direct behavioural observations (days 16 and 31 of each phase)

An observation hatch (27.5 × 50 cm; width × height) at the front of each pen was opened to allow the two observers to see the birds. After 1 min, each bird was observed continuously for 5 min using the ethogram described in Nicol *et al* (2009), and the frequency and duration of all behaviour patterns recorded during morning observation sessions. Pens and hens within pens were observed in a pre-determined random order.

Video behavioural observations (days 5, 19 and 33 of each phase)

Hens were filmed with a camera mounted centrally at the top of each pen. Each hen was observed using focal animal sampling for a continuous 15-min period, taken between 1615 and 1630h. Behavioural analysis of the footage was conducted for each individual bird using Observer Video-Pro 5.0 software (Noldus Information Technology, Wageningen, The Netherlands) by the observers who had taken direct observations, and by two additional observers who were trained in the same way. We collected a total of 270 min behavioural data for 30 hens and, due to a technical failure on day 33 of phase 1, 255 min behavioural data for the other 30 hens. For statistical analysis, we therefore averaged data from the available observations obtained in each phase, and report these as average duration per 15-min period.

Tonic immobility (days 16 and 30 of each phase)

Tonic immobility is usually measured after birds have been exposed to a potentially fear-inducing stimulus. Since we wanted to assess basal fear levels in birds that were well habituated to handling, we had to employ a longer restraint time of 30 s compared with the usual 15 s. Each bird was restrained gently on its side (a procedure we have used before [Cashman *et al* 1989]) with light pressure applied to the body and neck region. The operator's hands were then removed. Three attempts were made to induce TI for every bird. The number of attempts that resulted in a state of TI of more than 15 s was recorded. Birds were allowed to stay in TI for a maximum of 300 s on any occasion, after which they were gently righted. TI durations were markedly skewed with mostly very short durations, so only the number of inductions was used as a measure for further analysis.

Blood sampling and analysis (am on days 29–30 of each phase)

A trained operator removed no more than 4 ml of blood from the wing vein, within 1 min of handling. Heterophil and lymphocyte counts were taken immediately to assess H:L ratio. Samples (0.5 ml) of whole blood were assayed immediately for glucose and lactate (mmol L^{-1}) using the in-house biochemistry laboratory. The remaining 3.5 ml sample was stored on ice overnight, then centrifuged at 4,500 rpm for 10 min to separate serum. A portion of the serum was assayed immediately for osmolality (mmol kg^{-1}) and creatine kinase (μg L^{-1}), whilst the remainder was frozen at –20°C until the end of the experiment. The concentration of the acute phase protein α1-acid glycoprotein, α1AG, in serum was then determined using immunodiffusion kits (Cardiotech Services Inc, KY, USA). Corticosterone concentrations (ng ml^{-1}) were assessed by Cambridge Specialist Laboratories using RIA analysis.

Faecal sampling and analysis (pm on days 29–31 of each phase)

Each hen was placed in a wire cage with a plastic sheet floor, with at least one other hen visible in an adjacent cage, and left for 90 min to produce a faecal deposit. Faecal samples were placed in 50 ml plastic tubes and frozen at –20°C. The dry matter, ash, crude protein, fat and fibre content of each sample were determined using in-house laboratory analysis, using standard proximate analysis techniques. Dry matter digestibility and apparent fat and protein digestibility of the feed was estimated during week 37, using titanium dioxide as an inert marker.

Novel object tests (day 23 of each phase)

Novel object (NO) tests were conducted by opening the observation hatch and allowing birds to settle for 1 min. A novel object was then introduced to the pen and 30 scans of the activity of each bird taken at 10 s intervals. These scans were aggregated to give a total count for each behaviour pattern observed for each bird over the 300 s after the object was introduced. A different novel object was used for each phase of the experiment: striped rubber ball, green apple, inflated balloon, white plastic 200 ml bottle, aluminium can, transparent plastic 2 L bottle.

Mealworm resource competition tests (day 23 of each phase)

Mealworm resource competition (MW) tests were conducted by opening the observation hatch and allowing birds to settle for 1 min. A dish of mealworms was then introduced to the pen and 10 scans of the activity of each bird taken every 10 s. These scans were aggregated to give a total count for each behaviour pattern observed for each bird over the 100-s test.

Selection of behavioural indicators for further analysis

Our aim was to seek correlations and relationships amongst a large set of welfare indicators. However, our initial categorisation of behaviour into 31 separate activities recorded directly (DHP) and by video (VHP) in the home pen, 34 activities recorded in the novel object test and 35 activities recorded in the resource competition test resulted in an over-dominance of behavioural parameters (131 behavioural parameters in all). In our previous work examining associations between welfare indicators and bird preferences, we excluded rare activities but here, to examine relationships between the welfare indicators themselves, we wished to include all activities. We therefore first grouped them into 12 behavioural categories, based on their known functional relationships (Table 1). The duration of time spent in any of the activities within a category was summed to give an overall time spent in each behavioural category.

There were too few occurrences of behaviour in the 15 starred categories in Table 1 for further analysis, leaving us with 21 behavioural duration categories. Two other parameters, latency to approach mealworms and latency to approach the novel object were also included in the subsequent statistical analyses. A secondary aim was to compare the video observations with the direct observations, both for home pen behaviour and for the NO and MW tests. Neither collection method is *a priori* better than the other, but if information obtained using these two methods is comparable then the more convenient method could be used in future.

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Table 1 Grouping of separate activities into 12 behavioural categories, and the number of observations of each category in the four sets of observations taken.

Behavioural category	Activities included	Number of non-zero occurrences of behaviour out of a possible 360 (60 hens × 6 phases)			
		Direct observation (2 × 300 s per phase)	Video observation (3 × 900 s per phase)	Mealworm test (100 s per phase)	Novel object test (300 s per phase)
LOCO	Fly, walk, run	340	359	200	273
COMFORT	Flap wings, stretch wings, dustbathe, feather raise, preen, scratch self, tail wag, shake head	279	354	4*	66
FORAGE	Feed, beak wipe, ground scratch, litter peck, furniture peck, egg peck, approach mealworm dish, eat worms, peck at dish	357	360	351	267
DRINK	Drink	104	349	2*	53
M-PECK-REC	Receive mild feather peck	59	142	1*	1*
M-PECK-GIVE	Give mild feather peck, give beak peck	76	146	2*	2*
S-PECK-REC	Receive severe peck	4*	38	0*	0*
S-PECK-GIVE	Give severe peck	7*	33	0*	0*
INACTIVE	Stand inactive, sit inactive	6*	21	0*	0*
ALERT	Stand alert, sit alert	359	360	280	355
INT-NO	Interaction with novel object, peck at novel object	n/a	n/a	n/a	189
NEST	Nest	0	0	0	0

* Too few observations for further analysis.

Statistical analysis

All indicator variables were either continuous or ordered categorical. For estimating the relative variabilities of group, hen and phase within-hen we fitted 3-level normal response models in MLwiN version 2.16 (Rasbash *et al* 2009). For simplicity, we also treated the variables as continuous when calculating Pearson's correlations and cluster analyses. However, we also calculated non-parametric Spearman correlations to check whether the ordinal scaling of many of the variables affected the relationships found. For the cluster analyses, all variables were standardised prior to clustering, and the clustering was performed using the 'dist' and the 'hclust' commands in the R software package (R Development team 2009). The 'dist' function forms a matrix of distances between the standardised variables and this distance matrix is then the input used in the 'hclust' function which performs a (hierarchical) cluster analysis based on the given distance matrix. We used the default (Euclidean) distance metric in 'dist' and the average method of clustering in the 'hclust' function. For more details, see the R documentation (R Development team 2009).

Results

Indicator variability

It was striking that for most indicators the largest source of variability occurred at the phase within-bird level, demonstrating sensitivity of these indicators to the changing environment and ages of the birds (Table 2). Despite this overarching effect, at least 20% of the variation in indicators, such as bodyweight and condition score, comb characteristics, claw length, posture, dry matter digestibility, creatine kinase levels, responses in the MW test and video observations of alert, foraging, drinking and mild pecking behaviours was ascribed at individual-bird level, demonstrating some stability in these individual bird traits. For birds housed in groups, group-level effects explained some 5 to 16% of the variation in indicators, such as comb size, foot characteristics, condition score, and of many behaviours.

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Table 2 The percentage variation in each indicator that can be ascribed at the group level, the individual bird level, and the phase within-bird level.

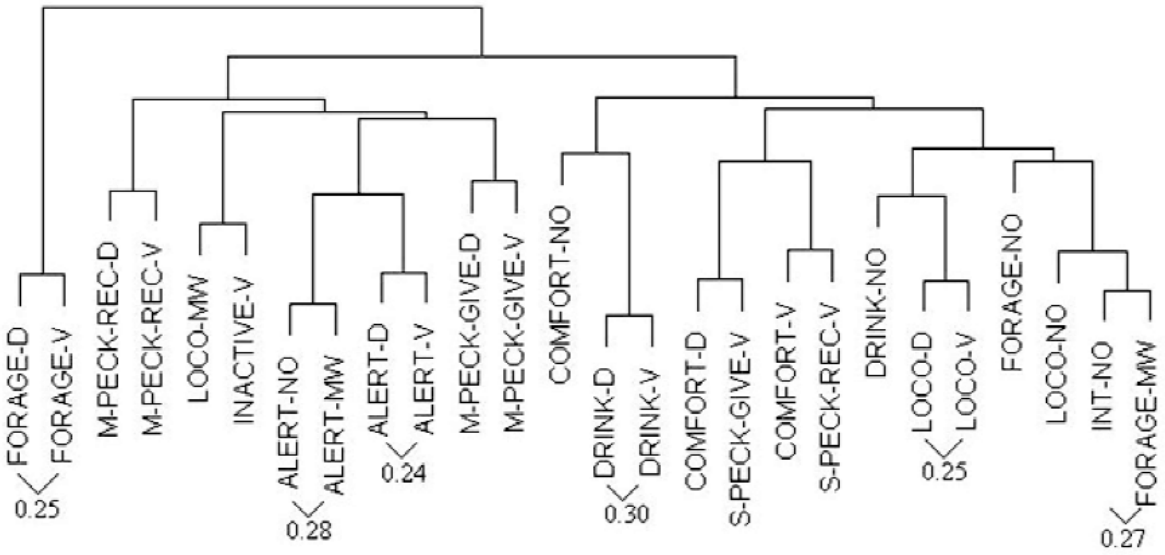
Variable	Description	n	Mean (± SEM)	% group	% bird	% phase within-bird
TI	Tonic immobility	360	0.06 (± 0.01)	0.00	3.56	96.44
BW	Bodyweight	360	1.74 (± 0.02)	0.00	64.20	35.80
FTHD-BR	Feather damage to breast	360	1.17 (± 0.08)	0.00	11.25	88.75
COMB-EL	Comb elevation	360	2.41 (± 0.08)	0.00	69.14	30.86
COMB-SZ	Comb size	360	2.28 (± 0.10)	8.99	68.13	22.88
COMB-COL	Comb colour	360	2.00 (± 0.07)	4.89	55.79	39.31
FEET-C	Feet condition	360	0.77 (± 0.07)	9.06	0.00	90.94
CLAW-L	Claw length	360	15.13 (± 0.28)	13.15	35.50	51.34
POSTURE	Posture	360	0.94 (± 0.02)	0.00	33.61	66.39
COND-S	Condition score	360	1.95 (± 0.07)	12.06	31.65	56.29
TEMP-E	Eye temperature	360	31.71 (± 0.06)	1.62	7.01	91.37
TEMP-F	Foot temperature	360	28.20 (± 0.25)	4.25	12.73	83.03
LAT-MW	Latency to approach mealworms	360	14.24 (± 2.61)	0.02	31.44	68.54
LAT-NO	Latency to approach novel object	359	216.27 (± 11.28)	3.78	0.00	96.22
GLUC	Blood glucose	349	13.04 (± 0.06)	0.00	14.55	85.45
LACT	Blood lactate	349	3.63 (± 0.08)	5.90	7.28	86.82
CK	Blood creatine kinase	301	805.38 (± 62.40)	0.00	34.59	65.41
OSMO	Blood osmolality	324	323.68 (± 0.61)	1.71	6.53	91.77
APP	Acute phase protein α1 AG	347	191.62 (± 5.85)	0.00	0.00	100.00
HL	Heterophil:lymphocyte ratio	352	1.00 (± 0.04)	3.18	2.71	94.11
% WATER	% faecal water	356	74.90 (± 0.37)	4.75	25.68	69.57
% PROT	% faecal protein	354	23.55 (± 0.29)	0.00	11.13	88.87
% FAT	% faecal fat	351	1.49 (± 0.07)	1.80	6.25	91.95
% CARB	% faecal carbohydrate	351	51.05 (± 0.40)	0.35	0.00	99.65
DIGEST-DM	Digestible dry matter	350	59.42 (± 1.22)	4.23	94.96	0.81
DIGEST-F	Digestibility fat	346	76.72 (± 1.41)	3.09	16.76	80.15
CORT	Corticosterone	342	1.09 (± 0.08)	2.07	10.02	87.90
LOCO-D	Locomotion-direct	360	8.95 (± 0.61)	7.35	2.74	89.91
COMFORT-D	Comfort behaviour-direct	360	6.95 (± 1.04)	8.29	7.89	83.82
FORAGE-D	Forage-direct	360	50.66 (± 1.99)	5.44	8.56	85.99
DRINK-D	Drink-direct	360	1.55 (± 0.31)	6.74	0.27	92.99
M-PECK-REC-D	Mild-peck received direct	360	0.30 (± 0.07)	0.00	9.95	90.05
M-PECK-GIVE-D	Mild-peck given direct	360	0.50 (± 0.18)	3.49	8.28	88.24
ALERT-D	Alert direct	360	30.69 (± 1.37)	3.09	10.11	86.80
LOCO-NO	Locomotion in novel object test	359	10.26 (± 1.21)	37.45	0.00	62.55
COMFORT-NO	Comfort behaviour in novel object test	359	3.48 (± 1.16)	28.11	0.00	71.89
FORAGE-NO	Forage in novel object test	359	20.89 (± 1.90)	9.22	0.00	90.78
INT-NO	Interact with novel object	359	9.25 (± 0.91)	2.49	0.00	97.51
DRINK-NO	Drink in novel object test	359	1.38 (± 0.33)	6.37	0.00	93.63
ALERT-NO	Alert behaviour in novel object test	359	55.12 (± 3.08)	13.97	0.00	86.03
LOCO-MW	Locomotion in mealworm test	360	9.30 (± 0.67)	1.33	2.98	95.69
FORAGE-MW	Forage in mealworm test	360	69.34 (± 2.31)	4.77	28.22	67.01

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Table 2 (cont)

Variable	Description	n	Mean (± SEM)	% group	% bird	% phase within bird
ALERT-MW	Alert behaviour in mealworm test	360	20.73 (± 2.27)	7.60	28.51	63.89
LOCO-V	Locomotion-video	360	8.13 (± 0.33)	5.02	14.01	80.97
COMFORT-V	Comfort behaviour-video	360	8.16 (± 0.78)	5.03	6.69	88.28
FORAGE-V	Forage-video	360	63.14 (± 1.55)	2.71	30.55	66.74
DRINK-V	Drink-video	360	4.44 (± 0.27)	0.00	24.30	75.70
M-PECK-REC-V	Mild-peck received-video	360	0.34 (± 0.14)	1.30	23.61	75.09
M-PECK-GIVE-V	Mild-peck given-video	360	0.44 (± 0.20)	5.91	26.37	67.72
S-PECK-REC-V	Severe-peck received-video	360	0.04 (± 0.03)	0.00	0.00	100.0
S-PECK-GIVE-V	Severe-peck given-video	360	0.03 (± 0.02)	3.98	0.00	96.02
INACTIVE-V	Inactive-video	360	0.08 (± 0.04)	1.32	0.00	98.68
ALERT-V	Alert-video	360	15.13 (± 0.82)	0.12	36.80	63.08

Figure 1



Cluster analysis of the behavioural categories recorded in direct observations in home pen (suffix D), video observations in home pen (suffix V), the novel object test (suffix NO) and the mealworm resource competition test (suffix MW). Pearson correlation coefficient values of > 0.24 are also shown.

Comparison of video and direct behaviour observations

An initial cluster analysis examining only the behavioural categories was completed to examine similarities between video and direct observations in all four contexts where behavioural data were collected (ie the DHP, VHP, MW and NO observations). This analysis revealed that the DHP and VHP observations were closely related for the behavioural categories of foraging, alert behaviour, mild pecking, drinking and locomotion (Figure 1).

Relationships between behaviours

Before considering how the behavioural, physiological and physical indicators were related, we examined relationships amongst the behavioural categories in all four contexts where behavioural data were collected (ie the DHP, VHP, MW and NO observations). As animals have a limited time budget, time spent on one activity necessarily reduces the time spent on other categories. We therefore expected, and found, a number of strong negative correlations between

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

different behavioural categories. Spearman and Pearson's correlation analyses gave similar results, except for comfort behaviour, and mild and severe pecking, where the number of observations was low and where Spearman tests gave higher coefficient values. Given the large number of tests performed, we present results only where Pearson's values exceed 0.24. At this value, the correlations are significant at $P < 0.01$ when n = 115. Due to some missing data the number of observations in our tests varied from 301 to 360 resulting in P-values of < 0.001.

The time spent foraging was negatively correlated with time spent alert in all four contexts where behavioural data were collected (DHP, VHP, MW and NO observations) with Pearson's correlation coefficients ranging from –0.74 to –0.89. Foraging in the home pen was also negatively correlated with comfort behaviour for the DHP and VHP observations (–0.47, –0.71, respectively), and with locomotion for the VHP, MW and DHP observations (–0.34, –0.40, –0.44). Alert behaviour in the NO test was negatively correlated with drinking, locomotion and interacting with the novel object (–0.31, –0.45, –0.59). The latency to approach the novel object was positively correlated with time spent alert (0.65) and negatively correlated with time spent drinking (–0.29), foraging (–0.39) and interacting with the novel object (–0.66) during the same test. The latency to approach the mealworms was positively correlated with time spent alert during the same test (0.57).

Positive relationships were examined by both correlations and cluster analysis. In addition to the close relationships between direct and video recordings of the same behavioural categories (reported above and in Figure 1), alert behaviour and mild feather pecking appeared to be related, as did comfort behaviour and severe pecking (Figure 1). Alert behaviour in the DHP and VHP observations clustered relatively well with alert behaviour in the NO and MW test, but in contrast, foraging in the home pen appeared almost entirely unrelated to foraging behaviour in the NO and MW tests.

Relationships among welfare indicators

Our main aim was to examine positive and negative relationships between the behavioural, physiological and physical indicators taken. For the undisturbed home pen observations, we used only the behavioural categories recorded by video, because they represented substantially longer periods of observation than the direct observations, and because we knew that the information obtained from direct and video methods was closely related. To reduce the number of variables, we used eye temperature as a proxy for head and comb temperature measures, as eye temperature was positively correlated with both head (0.36) and comb (0.27) temperature. There were some missing observations for the blood indicators (shown in column 3 of Table 2). We dealt with these by performing mean imputation and replacing missing data with the mean values for these indicators. The low number of missing observations meant there was little effect of this procedure on the resulting cluster analysis and allowed us to use all the observations.

Table 3 Positive correlations between behavioural, physiological and physical indicators (acronyms described in Table 2). Physical measures (green), physiological measures (red), behavioural measures (blue).

Variable 1	Variable 2	Spearman correlation coefficient	Pearson correlation coefficient
BW	COND-S	0.55	0.56
FTHD-BR	ALERT-NO	0.46	0.44
BW	CLAW-L	0.39	0.41
BW	ALERT-NO	0.34	0.36
FTHD-BR F	LAT-NO	0.38	0.35
DIGEST-DM	DIGEST-F	0.39	0.33
TEMP-F	GLUCOSE	0.35	0.32
%WATER	%PROT	0.35	0.32
TEMP-F	%CARB	0.33	0.30
BW	COMB-SZ	0.28	0.29
LACTATE	OSMO	0.44	0.27
TEMP-E	%CARB	0.30	0.27
INT-NO	APP	0.29	0.25
CLAW-L	LAT-NO	0.25	0.27
APP	HL	0.25	0.26
BW	LAT-NO	0.24	0.25
GLUCOSE	LACTATE	0.21	0.25
TEMP-F	APP	0.28	0.24
TEMP-F	HL	0.27	0.24

Positive and negative correlations between the indicators measured, again using cut-offs of 0.24 in magnitude, are shown in Tables 3 and 4. There was generally very good agreement between Spearman and Pearson's correlation coefficients. Relationships between the welfare indicators were examined further by the cluster analysis presented in Figure 2. A joint consideration of the results shown in Table 3 and Figure 2 suggests a cluster of relatively strong links between greater bodyweight, condition score, claw length, breast feather damage and relatively nervous behaviour in the MW and NO tests (indicated by alert behaviour and long latencies to approach). A second important cluster links higher acute phase protein and heterophil:lymphocyte measures with each other, and with relatively confident behaviour in the MW and NO tests, higher body temperature and greater blood glucose concentrations. Lactate and osmolality were closely related (as expected, since osmolality is a measure of blood ion concentration), as were the two digestibility measures. Apparent links between feather pecking and comfort behaviour (Figure 2) were based on low numbers of observations and were not supported by strong correlation coefficient values.

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Table 4 Negative between-test correlations (acronyms described in Table 2). Physical measures (green), physiological measures (red), behavioural measures (blue).

Variable 1	Variable 2	Spearman correlation coefficient	Pearson correlation coefficient
BW	GLUCOSE	–0.38	–0.40
FTHD-BR	APP	–0.47	–0.39
FTHD-BR	INT-NO	–0.39	–0.34
ALERT-NO	APP	–0.38	–0.34
ALERT-NO	FORAGE-MW	–0.35	–0.32
BW	CORT	–0.31	–0.31
LAT-NO	TEMP-F	–0.30	–0.29
CLAW-L	GLUCOSE	–0.31	–0.28
LAT-NO	FORAGE-MW	–0.26	–0.27
LAT-NO	APP	–0.30	–0.26
ALERT-NO	TEMP-F	–0.28	–0.26
ALERT-NO	TEMP-F	–0.28	–0.26
LAT-NO	TEMP-H	–0.26	–0.26
FTHD-BR	DRINK-NO	–0.26	–0.25
LAT NO	TEMP-E	–0.26	–0.25
FTHD-BR	HL	–0.28	–0.24
BW	HL	–0.27	–0.24
ALERT-NO	GLUCOSE	–0.26	–0.24

Discussion

Others have reported intra-situational consistency for individual birds in measures of fearfulness, such as tonic immobility and reaction to a novel object (Jones 1988; Ghareeb *et al* 2008) and we also know that the hens used in this experiment could be grouped into 'personality types' based on differences in their indicator profiles averaged over the entire experimental period (Nicol *et al* 2009). Different types can be maintained in populations because each has advantages under different environmental conditions. More passive (reactive or 'dove') individuals tend to show lower sympathetic reactivity, higher hypothalamic pituitary adrenal (HPA) reactivity and greater behavioural flexibility than more active (proactive or 'hawk') animals (Koolhaas *et al* 1999; Korte *et al* 2009). Given this, the amount of variability in indicator values that occurred at the phase within-bird level was somewhat unexpected. Clearly, the existence of underlying personality traits did not preclude greatly varying responses to housing at different phases of the experiment. However, different behavioural strategies can also be maintained within populations by the same individuals making conditional decisions to play either 'hawk' or 'dove' depending on the context, something not fully explored by Korte *et al* (2009). We simply draw attention to the fact that the response clusters found here do not perfectly map onto the stable bird 'types' identified previously (Nicol *et al* 2009). We considered ranking responses within phases and looking for consistency in bird rankings. However, because of the design of this experiment, different birds experienced different environments in each phase and we were unable to account for changes in bird age and environment simultaneously to examine individual consistency.

Information about the amount of variation in each indicator that occurred at the various levels in the dataset is important in distinguishing how different indicators can be used in a welfare context. Responses that are good indicators of a stable individual personality or body condition could be used to select birds with more desirable traits (Ghareeb *et al* 2008). In this study, the most stable traits were physical indicators, such as comb elevation and colour. More labile responses that vary with age, or as birds react to different environments, are potentially sensitive indicators of how well a bird is faring in its current living environment.

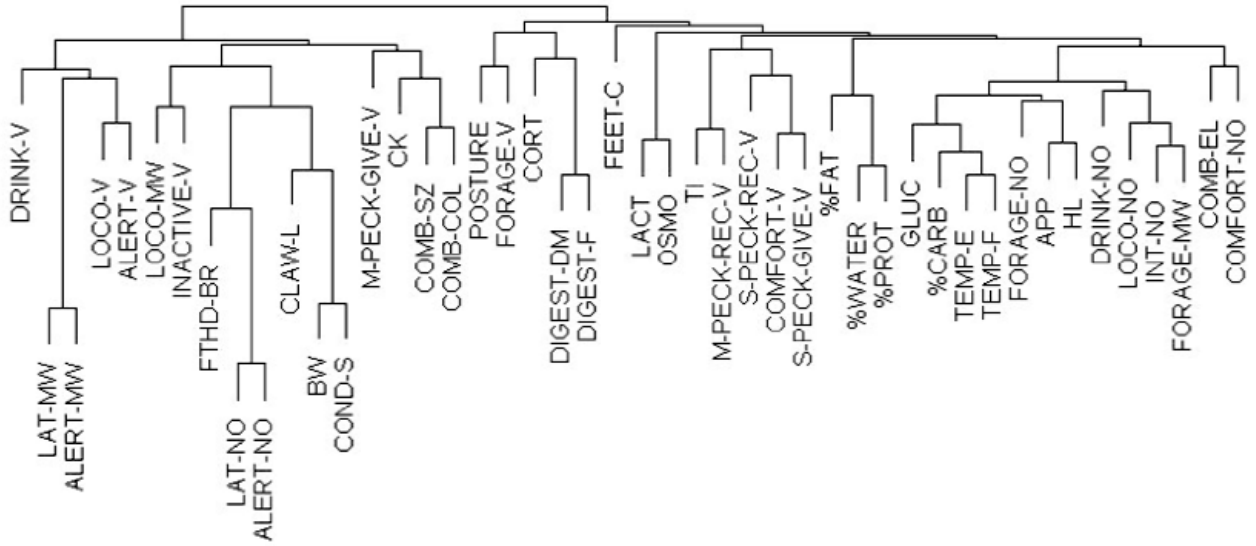
Our analysis reveals considerable lability in most of the behavioural and physiological measures. The behavioural measures recorded during the MW and NO tests had slightly more stable 'individual bird' or 'group' components than the behavioural measures recorded during undisturbed home pen observations, but all measures still varied greatly between experimental phases. Given the correlations between the MW and NO test responses and the physiological measures, we suggest that these behavioural test responses are more useful welfare indicators than previously acknowledged.

There has been much debate about the degree of independence that is likely when observations are taken on animals living in groups, and how non-independence should be treated statistically (Iason & Elston 2002; Knowles & Green 2002; Phillips 2002). By using hierarchical statistical modelling techniques, we found that variability at the group level was rather low, except for locomotion during the NO test. This suggests that the analysis of welfare indicators based on group means to avoid perceived (but not measured) problems of non-independence, can sometimes unnecessarily reduce the power of a study.

Our main aim was to examine correlations between different classes of welfare indicators. If many different indicators are closely linked this suggests redundancy in the data collection process and it could lead to problems in combining non-independent indicators to draw overall conclusions. We found at least two interesting patterns linking the different indicators across a range of different environments and bird ages. First, slow approach responses and alert behaviour during the MW and, especially, the NO test, were linked with higher bodyweight and condition score, longer claws and greater plumage damage. This 'slow response' cluster was negatively associated with indicators of stress — including lower body temperature, APP, H:L ratio and glucose concentration. Unexpectedly, therefore, a greater reluctance to approach

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Figure 2



Cluster analysis of the behaviour recorded in the video observations in the home pen, the novel object test, and the mealworm resource competition test with the physical and physiological indicators recorded in the home pens at each phase.

novel stimuli in a behavioural test situation was strongly and consistently linked with a suite of physiological indicators of a lower stress profile. This conclusion is supported by our previous finding that a longer latency to approach the NO in a particular environment was associated with a positive choice for that same environment (Nicol *et al* 2009) and by more general evidence that there are no clear associations between fear and stress in birds (see Cockrem 2007 for a review). This suggests that novel object responses may be rather good summary indicators of welfare, but may require a very different interpretation from that usually provided.

Some measures, thought to be important for welfare, were seemingly unrelated to others recorded during this study. The most notable example was comfort behaviour, which we know is highly motivated and associated with positive environmental choice in hens (Nicol 1987a,b; Nicol *et al* 2009) but which appeared uncorrelated with other measures. This suggests that comfort behaviour should be included as an independent measure in welfare assessment protocols. Finally, there were some measures that, in this experiment, were not related to the birds' environmental preferences, or to other welfare indicators. These included the number of attempts to induce a state of tonic immobility and basal blood corticosterone concentrations. Basal corticosterone levels can fluctuate rapidly in response to internal and external events and its usefulness as a welfare measure in hens has been questioned previously (D'Eath *et al* 2009), but it remains the most widely used physiological parameter in welfare studies. It would be useful in future work to determine whether corticosterone measured after challenge (eg challenge with different dose levels of ACTH) might show stronger associations with other welfare indicators and/or bird preferences.

Another approach to summarising behavioural variation is to use principal components analysis (PCA). Van Reenen *et al* (2004), for example, used this technique to examine relationships between eight behavioural variables in a dataset examining reactivity of heifer calves. They found principal components (PCs) that loaded heavily on two or three of their eight variables and were fairly interpretable. We also looked at PCA on our larger datasets but the PCs produced were not very easy to interpret and we therefore preferred the cluster diagrams described here. It is our experience that, with larger datasets, PCA is a useful technique for dimension reduction but at a price of interpretability. Also, when using PCA on a set of predictor variables, the PCs produced may explain much of the variability in the predictors themselves, without being any more related to an external response variable than the original predictors. The ideal is to use dimension reduction techniques that capture the variability in the dataset while remaining interpretable. For example, we have recently (Browne *et al* 2010) worked on dimension reduction techniques that use knowledge of the data and modelling to summarise mass spectrograms as a set of parameters representing important peaks in the spectrograms. These are much easier to interpret and use than PCs from a PCA.

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Animal welfare implications

Statistical analysis of large datasets can inform decisions about whether to record welfare-relevant information at an individual animal or group level, and about which indicators are closely related or independent. This can be used to improve, refine and reduce protocols for experimental and on-farm assessments of animal welfare. Finally, there are welfare implications if the interpretation of commonly used novel object tests proves to be incorrect, especially if such tests form the basis of genetic selection programmes.

Conclusion

Consistent results from correlational and cluster analysis revealed that slow approach responses and alert behaviour in novel object tests were associated with lower levels of physiological stress. Comfort behaviour was an independent measure of welfare not linked to this set. The statistical techniques used revealed the level at which most variation occurred in each welfare indicator and highlighted patterns of independence and redundancy in the measures taken.

Acknowledgements

We thank Dr Frances Kim-Madslien for help with some of the video analysis. The work was funded by BBSRC Programme Grant BB/C518949/1.

References

Albentosa MJ, Kjaer JB and Nicol CJ 2003 Strain and age differences in behaviour, fear response and pecking tendency in laying hens. *British Poultry Science 44*: 333-344

Barnett JL, Tauson R, Downing JA, Janardhana V, Lowenthal JW, Butler KL and Cronin GM 2009 The effects of a perch, dust bath, and nest box, either alone or in combination as used in furnished cages, on the welfare of laying hens. *Poultry Science 88*: 456-470

Browne WJ, Dryden IL, Handley K, Mian S and Schadendorf D 2010 Mixed effect modelling of proteomic mass spectrometry data using Gaussian mixtures. *Journal of the Royal Statistical Society, Series C59*: 617-633

Campo JL and Prieto MT 2009 Effects of moist litter, perches, and droppings pit on fluctuating asymmetry, tonic immobility duration, and heterophil-to-lymphocyte ratio of laying hens. *Poultry Science 88*: 708-713

Cashman PJ, Nicol CJ and Jones RB 1989 Effects of transportation on the tonic immobility fear reaction of broilers. *British Poultry Science 30*: 211-221

Cheng HW and Jefferson L 2008 Different behavioral and physiological responses in two genetic lines of laying hens after transportation. *Poultry Science 87*: 885-892

Cloutier S and Newberry RC 2002 Differences in skeletal and ornamental traits between laying hen cannibals, victims and bystanders. *Applied Animal Behaviour Science 77*: 115-126

Cockrem JF 2007 Stress, corticosterone responses and avian personalities. *Journal of Ornithology 148*: S169-S178

Dawkins MS, Edmond A, Lord A, Solomon S and Bain M 2004 Time course of changes in egg-shell quality, faecal corticosteroids and behaviour as welfare measures in laying hens. *Animal Welfare 13*: 321-327

D'Eath RB, Tolkamp BJ, Kyriazakis I and Lawrence AB 2009 'Freedom from hunger' and preventing obesity: the animal welfare implications of reducing food quantity or quality. *Animal Behaviour 77*: 275-288

Ghareeb K, Niebuhr K, Awad WA, Waiblinger S and Troxler J 2008 Stability of fear and sociality in two strains of laying hens. *British Poultry Science 49*: 502-508

Iason GR and Elston DA 2002 Groups, individuals, efficiency and validity of statistical analyses. *Applied Animal Behaviour Science 75*: 261-265

Jones RB 1988 Repeatability of fear ranks among adult laying hens. *Applied Animal Behaviour Science 19*: 297-304

Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW and Blokhuis HJ 1999 Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience and Biobehavioral Reviews* 23: 925-935

Kjaer JB 2009 Feather pecking in domestic fowl is genetically related to locomotor activity levels: implications for a hyperactivity disorder model of feather pecking. *Behavior Genetics 39*: 564-570

Knowles TG and Green LE 2002 Multilevel statistical models allow simultaneous consideration of both individual and group effects. *Applied Animal Behaviour Science 77*: 335-336

Korte SM, Pins J, Vinckers CH and Olivier B 2009 On the origin of allostasis and stress-induced pathology in farm animals: celebrating Darwin's legacy. *Veterinary Journal 182*: 378-383

Mason G and Mendl M 1993 Why is there no simple way of measuring animal welfare? *Animal Welfare 2*: 301-319

Nicol CJ 1987a Effect of cage height and area on the behaviour of hens housed in battery cages. *British Poultry Science 28*: 327-335

Nicol CJ 1987b Behavioural responses of laying hens following a period of spatial restriction. *Animal Behaviour 35*: 1709-1719

Nicol CJ, Brown SN, Glen E, Pope SJ, Short FJ, Warriss PD, Zimmerman PH and Wilkins LJ 2006 Effects of stocking density, flock size and management on the welfare of laying hens in single-tier aviaries. *British Poultry Science 47*: 135-146

Nicol CJ, Caplen G, Edgar J and Browne WJ 2009 Associations between welfare indicators and environmental choice in laying hens. *Animal Behaviour 78*: 413-424

Phillips CJC 2002 Further aspects of the use of individual animals as replicates in statistical analysis. *Applied Animal Behaviour Science 75*: 265-268

R Development Core Team 2009 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Rasbash J, Browne WJ, Healy M, Cameron B and Charlton C 2009 *MLwiN version 2.16*. Centre for Multilevel Modelling: University of Bristol, UK

Richard S, Auperin B, Bolhuis JE, Geverink NA, Jones BC, Lepage O, Mignon-Grasteau S, Mormede P, Prunet P and Beaumont C 2007 Animal welfare: what are the relationships between physiological and behavioural measures of adaptation. *Productions Animals 20*: 29-33

Rodenburg TB, Tuyttens FAM, de Reu K, Herman L, Zoons J and Sonck B 2008 Welfare assessment of laying hens in furnished cages and non-cage systems: assimilating expert opinion. *Animal Welfare 17*: 355-361

Sherwin CS, Richards G and Nicol CJ 2010 A comparison of the welfare of layer hens in four housing systems used in the UK. *British Poultry Science 51*: 488-499

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press

Singh R, Cook N, Cheng KM and Silversides FG 2009 Invasive and non-invasive measurement of stress in laying hens kept in conventional cages and in floor pens. *Poultry Science 88*: 1346-1351

Tactacan GB, Guenter W, Lewis NJ, Rodriguez-Lecompte JC and House JD 2009 Performance and welfare of laying hens in conventional and enriched cages. *Poultry Science 88*: 698-707

Thaxton JP and Puvadolpirod S 2000 Model of physiological stress in chickens 5. Quantitative evaluation. *Poultry Science 79*: 391-395

Thogerson CM, Hester PY, Mench JA, Newberry RC, Pajor EA and Garner JP 2009 The effect of feeder space allocation on behavior of Hy-Line W-36 hens housed in conventional cages. *Poultry Science 88*: 1544-1552

Van Reenen CG, Engel B, Ruis-Heutinck LFM, Van der Werf JTN, Buist WG, Jones RB and Blokhuis HJ 2004 Behavioural reactivity of heifer calves in potentially alarming test situations: a multivariate and correlational analysis. *Applied Animal Behaviour Science 85*: 11-30

Vestergaard KS, Kruijt JP and Hogan JA 1993 Feather pecking and chronic fear in groups of red junglefowl: their relation to dustbathing, rearing environment and social status. *Animal Behaviour 45*: 1127-1140

Webster AB and Hurnik JF 1991 Behavior, production and well-being of the laying hen: individual variation and relationships of behaviour to production and physical condition. *Poultry Science 70*: 421-428

https://doi.org/10.1017/S0962728600002621 Published online by Cambridge University Press