Quantification of abnormal repetitive behaviour in captive European starlings (*Sturnus vulgaris*)

Ben, O. Brilot[a]*, Lucy Asher[b], Gesa Feenders[a] and Melissa Bateson[a].

[a]Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, UK.

[b]Royal Veterinary College, Hatfield, UK.

*Corresponding author: Dr Ben Brilot, Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, Henry Wellcome Building for Neuroecology, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.

E-mail: ben.brilot@ncl.ac.uk.
Telephone: +44 (0)191 222 6246.

Abstract

Stereotypies are repetitive, unvarying and goalless behaviour patterns that are often considered indicative of poor welfare in captive animals. Quantifying stereotypies can be difficult, particularly during the early stages of their development when behaviour is still flexible. We compared two methods for objectively quantifying the development of route-tracing stereotypies in caged starlings. We used Markov chains and T-pattern analysis (implemented by the software package, Theme) to identify patterns in the sequence of locations a bird occupied within its cage. Pattern metrics produced by both methods correlated with the frequency of established measures of stereotypic behaviour and abnormal behaviour patterns counted from video recordings, suggesting that both methods could be useful for identifying stereotypic individuals and quantifying stereotypic behaviour. We discuss the relative benefits and disadvantages of the two approaches.

1. Introduction

Stereotypic behaviour patterns, defined as behaviour patterns that are repetitive, unvarying and with no apparent function, have been described in a wide range of captive mammalian and avian species including farm, zoo, companion and laboratory animals. Stereotypies are most common in animals housed in barren and or spatially restricting cages, and are generally considered indicative of poor welfare (Mason, 1991a; 1991b; Mason and Rushen, 2006). Understanding the proximate and ultimate causes of stereotypies is an important area of pure and applied ethological research (Mason and Rushen, 2006). However, progress is limited by the crude and time-consuming methods currently employed for quantifying stereotypic behaviour. Our aim in this paper is to apply and compare two novel methods for identifying patterns in an animal's use of space. We argue that these methods could be sensitive, easily automated methods for objectively quantifying route-tracing locomotor stereotypies.

The development of stereotypic behaviour patterns in caged animals is characterized by four stages: first, ritualisation, in which behaviour becomes less variable; second, emancipation, in which a behavioural sequence is elicited by progressively more environmental stimuli; third, establishment, in which the sequence becomes more difficult to reverse; and finally, escalation, in which the sequence begins to take up progressively more of the animal's time (Meehan et al., 2004). It would be useful to be able to identify vulnerable animals during the first ritualisation stage, before behavioural sequences become difficult to reverse. However, quantification of flexible behaviour patterns can be difficult, particularly in the case of route-tracing stereotypies, where ritualisation is characterized by a gradual reduction in the variability of the route the animal traces around its cage (Garner et al., 2003).

Current techniques for quantifying stereotypies involve counting individual incidences of complete iterations of a stereotypy, such as a circuit around a cage, or an abnormal behavioural event such as a somersault. Defining these sequences of behaviour objectively can be difficult, and often relies upon a subjective judgement as to whether a behaviour sequence is abnormal in quality or quantity (Mason, 1991a; Würbel, 2002). Most importantly, such methods cannot be used during ritualisation when sequences of behaviour are still flexible (Meehan et al., 2004).

Golani et al. (1999) attempted to quantify stereotypic behaviour in laboratory rats on the basis of the types and ranges of physical movement expressed. They defined a stereotypy as a reduction in the number and range of "collective variables" (i.e. movements expressed) together with an increase in the predictability of movement sequences. However, their methodology

88   may fail to encompass stereotypies that involve movements not
89   normally expressed under semi-free or free-ranging conditions. For
90   example, the somersaulting stereotypy (see below) of caged
91   European starlings (*Sturnus vulgaris*) involves a locomotor pattern
92   that actually adds to the diversity of movement seen in birds kept in
93   free-flight aviaries. However, we agree with their supposition that
94   an increase in the predictability of movement patterns could be a
95   useful method for quantifying stereotypic behaviour (as per Meehan
96   et al., 2004), and further develop this approach in the current
97   paper.
98   To quantify potential stereotypies in flexible movement
99   patterns, we investigated techniques for identifying patterns in the
100 sequence of locations an animal occupies within its cage. Our
101 rationale was that by using pattern detection algorithms to identify
102 patterns in the sequences of discrete locations an animal visits
103 within its cage, we might be able to objectively quantify route-
104 tracing stereotypies. Moreover, we might be able to use this
105 approach to detect stages in the development of a route-tracing
106 stereotypy before a completely rigid circuit is established, and thus
107 predict animals at risk of developing rigid stereotypies. In order to
108 validate our methods, we asked whether the space use pattern
109 metrics we derived correlated with established measures of
110 abnormal and repetitive behaviour, on the assumption that
111 individuals developing route-tracing stereotypies would also be
112 likely to display other abnormal and repetitive behaviour patterns.
113   We tested the above ideas using data collected from caged
114 wild-caught European starlings, arguably the most widely used
115 passerine bird species in laboratory research (Asher and Bateson,
116 2008). In the laboratory, starlings are routinely kept in individual
117 cages, a risk factor for the development of stereotypic behaviour
118 (including route-tracing) in starlings and other bird species (Garner
119 et al., 2003; Meehan et al., 2003; Asher et al., 2009). We recorded
120 the behaviour of the birds over six weeks, thereby aiming to
121 capture the development of abnormal behaviour patterns during the
122 early stages of captivity and record the levels of rigid stereotypies
123 during the later stages.
124   Some caged starlings develop a "somersaulting" stereotypy in
125 which they repeatedly complete a backwards aerial flip (Greenwood
126 et al., 2004), a behaviour pattern that we hypothesise might
127 develop from a thwarted escape attempt. Somersaulting is readily
128 identifiable and quantifiable as an abnormal stereotypic behaviour
129 pattern; it occurs repeatedly in the same location within the cage,
130 using a set pattern of movements and has no obvious function,
131 indeed it often results in the subject risking damage since on
132 occasion they land on their back. Therefore, we set out to use
133 somersaulting as our standard for categorising and ranking
134 stereotypic tendencies in our birds. However, since not all caged

starlings develop somersaulting, we also counted the frequency of other discrete abnormal behaviour patterns including: perching on cage walls and ceiling, head tilting and unbalancing (see Table 1 for definitions). Spending time on the cage walls and ceiling has previously been suggested to be indicative of escape attempts and is associated with other measures of poor welfare (Maddocks et al., 2002). Head tilting and unbalancing are behavioural events that may be related to hyper-vigilance behaviour triggered by the acute stress caused by the initial transfer from free-flight aviaries to individual cages. This supposed link with vigilance means that it is harder to categorise these behavioural events as functionless, and hence stereotypic. They commonly occur wherever the bird is perched within the cage and are more variable in terms of identifiable movement patterns than the somersaulting stereotypy. However, these behaviours are abnormal in the sense that they have only been observed in starlings kept in cages (as opposed to free-living starlings or those held in free-flight aviaries). We hypothesise that they might represent precursors to the full somersaulting stereotypy. Therefore, we describe head tilting and unbalancing as abnormal behavioural events as distinct from stereotypies.

In the first part of our statistical analysis we investigated correlations between somersaulting, perching on walls and ceiling, head tilting and unbalancing to test our hypothesis that these behaviours are functionally related, perhaps via escape motivation. If this hypothesis was confirmed, then all of these simple behaviours could be used to identify individual birds with stereotypic tendencies.

Although rigid route-tracing stereotypies have not been previously described in caged starlings, the early stages of the development of such stereotypies have been reported in starlings housed in cages for as little as one week (Asher et al., 2009). To objectively quantify route-tracing, we recorded the time at which a bird arrived at each new location within the cage, and subjected the sequences of data obtained to two pattern detection algorithms: a method based on Markov chain analysis; and T-pattern analysis implemented in the software package Theme (Magnusson, 2000). Both methods have the potential to identify the early flexible stages in the ritualisation of a route-tracing stereotypy that are characterized by increased predictability in the sequence of locations visited by a bird. We describe these methods and our predictions in more detail below.

1.1 Markov chain analysis

A Markov chain is a stochastic process, comprising a finite set of events, where the next event depends only on the previous event (or previous few events). In an animal performing stereotypic

182   behaviour, the current behavioural event is a good predictor of the
183   next behaviour or location, because the sequence is predictable by
184   definition. We performed a Markov chain analysis of sequential
185   dependency using an adapted version of the chi-squared test for
186   first against second-order dependency in sequences of events
187   originally described by Haccou and Meelis (1992). This method uses
188   the transitional probabilities of one event following another event
189   (first order) or pair of events (second order). The more repetitive a
190   sequence, the higher the probability that a given event will follow a
191   certain kind of event or pair of events (see Methods and Asher et al.
192   (2009) for full details of this methodology).
193   We have already demonstrated that this latter technique
194   detects significant differences in the behavioural sequences of
195   starlings housed in cages of different sizes and shapes (Asher et al.,
196   2009) and cages with and without environmental enrichment (Asher
197   et al., unpublish. data). We found that sequential dependency
198   scores derived from Markov chain analysis were highest, indicative
199   of the most repetitive behavioural sequences, in birds that
200   performed somersaulting (Asher et al., 2009). Therefore, we
201   predicted that the subjects that performed somersaulting in this
202   study would similarly have increased tendencies to route-tracing
203   and correspondingly higher sequential dependency scores.
204
205   *1.2 T-pattern analysis*

206   The software application Theme uses algorithms that detect
207   temporal relationships (termed T-patterns) between pairs of
208   behavioural events (Magnusson, 2000; Magnusson, 2005). When
209   two events co-occur within a critical time interval more often than
210   expected by chance they are designated as a T-pattern. More
211   specifically, the algorithm searches for the shortest (d_1) and longest
212   (d_2) duration between two events (X and Y) and uses these to
213   define a critical time-frame ($t+d_1$ to $t+d_2$). It then tests whether Y
214   appears after X significantly more than expected within this critical
215   time-frame as compared to the situation where Y has a constant
216   probability of occurring throughout the observation period. The
217   process is repeated with newly identified T-patterns treated as
218   single events until no more pairs of events are found (Magnusson,
219   2000; 2004; Magnusson, 2005). Since T-patterns are identified
220   based on the temporal relationship between events, as opposed to
221   the order of events, a pattern of events can be identified even if it is
222   interrupted by an event that does not form part of the pattern. Thus
223   Theme has a major advantage in being able to detect patterns in
224   sequences of behaviour that would be invisible to Markov chain
225   analysis.
226   T-pattern analysis has not previously been applied explicitly to
227   the problem of quantifying the development of stereotypic
228   behaviour in either humans or other animals. However, there is

some evidence to suggest that T-pattern analysis could be useful for differentiating individuals with abnormal behaviour or behavioural stereotypies such as for example preschool children with autistic spectrum disorder (Warreyn et al., 2007) and psychiatric patients with schizophrenia and mania (Lyon and Kemp, 2004). T-pattern analysis has also been used to successfully differentiate mice treated with different doses of the dopamine transporter inhibitor GBR-12909, a drug known to induce locomotor stereotypies (Bonasera et al., 2008).

We predicted that development of route-tracing in starlings would be associated with an initial increase in the number of T-patterns, as starlings start to follow flexible variations of the same route. However, as route-tracing becomes more stereotyped, the number of different T-patterns would decrease as a smaller number of patterns are performed a greater number of times.

2. Methods

2.1 Subjects and husbandry

The subjects were eight starlings (four males and four females) caught from the wild under license from Natural England. Four of the birds designated as "juveniles" were caught in the summer of their first year, whereas the other four "adults" were at least one year of age at the time of catching. Both sex and age were counterbalanced for position in the laboratory and time of behavioural recording.

Prior to the experiment the birds were group-housed in an indoor aviary (2.4 m x 2.15 m x 2.3 m) with wood chippings covering the floor, dead trees for perching and cover, and shallow trays of water for bathing. At the start of the experiment, the birds were moved into individual cages (750 mm x 450 mm x 440 mm) where visual and auditory contact with four or five conspecifics was possible dependent upon location in the laboratory. The cages were furnished with natural bark branches of varying thickness and angles; a water bath; and a tray of bark chippings, enrichments suggested to improve the welfare of captive starlings (Bateson and Matheson, 2007; Matheson et al., 2008).

The light:dark cycle was maintained at 14:10 hours. At all times, other than those described below, the subjects had *ad libitum* access to Purina kitten food, supplemented with fruit and mealworms (*Tenebrio* larvae). Drinking water was available at all times.

The birds were subject to daily learning task trials associated with another study. These trials involved cleaning of the cages at 08:00 followed by 2 h of food deprivation and guano collection,

followed by approximately 1 h of experimental trials. The trials required the subjects to learn a colour/shape discrimination by flipping coloured cardboard lids off a petri dish to obtain a variable mealworm reward. On completion of the trials, the subjects were once more allowed to feed *ad libitum*. All experimental procedures were completed by approximately 12:00, allowing the birds three hours to settle for the behavioural recording required by the current study (see below).

Our study adhered to the Association for the Study of Animal Behaviour's Guidelines for the Use of Animals in Research and also passed internal ethical review. Birds were released back into free-flight aviaries after the experiment, and following completion of our studies they were released back to the wild at the site of original capture.

2.2 Behavioural recording

The birds' behaviour in the absence of the experimenter was recorded using two Sony DCR-SR32 video recorders. Recording always took place between 15:00 and 16:00. Each bird was recorded for 30 minutes on its first day in the cage and for 30 minutes at weekly intervals for a total of six weeks.

We manually analysed the video tapes using the freeware behaviour analysis program J-Watcher version 1.0 (Blumstein et al., 2000). We scored the location of the bird in the cage as a state variable (whereby recording the arrival of a bird in a new location had the effect of cancelling the previous location) and abnormal/repetitive behaviour patterns as discrete events. The details of the different locations and events we scored are given in Table 1. The location data were used to compute the proportion of time spent by the bird in different locations of the cage.

2.3 Pattern detection and statistical analysis

For the purposes of the pattern detection analysis we separated the location data from the discrete behavioural events. The sequences of locations visited by each bird were analysed using both our Markov chains method and Theme version 5.0 (Noldus Information Technology, Wageningen, Netherlands) in order to quantify the development of potential route-tracing behaviour.

2.3.1 Details of Markov chain analysis

For the purposes of the Markov chains analysis we analysed just the sequence of locations occupied by a bird using a custom written programme that automated the following calculations. For three locations XYZ the transition probability of Z following Y, ($P_{YZ} = N_{YZ} | N_Y$) uses the chi-squared test statistic:

Table 1. Definitions of locations and behavioural events scored.

Behaviour	Description
Abnormal repetitive behaviour (events)	
Head-tilt	The bird tilts its head back such that its bill breaks the vertical plane. Each time the head was brought down and then the bill again passed through the vertical plane was counted as a separate tilt.
Unbalance	Any wing movement required to correct the bird's posture back to upright once it had tilted its head/body backwards.
Somersault	A somersault was defined as the subject leaving the floor/perch and its feet passing over its head (unless it held on to the ceiling during the motion).
Location (states)	
Left wall	Bird clinging on to a side of the mesh rectangular cage with its claws.
Right wall	
Rear wall	
Front wall	
Ceiling	
Left perch	Bird is perched on one of two natural branch perches positioned towards the top of the cage
Right perch	
Food bowl	Bird perched on or in a small round food dish positioned on the floor of the cage approximately in the centre
Foraging tray	Bird perched on or in a rectangular tray filled with bark chippings.
Water bath	Bird perched on or in a shallow circular dish filled with water.
Floor	Bird anywhere on the floor of the cage not containing the food bowl, foraging tray or water bath.

320    $$C_Y = \sum_X \sum_Z \frac{(N_{XYZ} - N_{XY}P_{YZ})^2}{N_{XY}P_{YZ}}$$ (1)

321   where N is the number of occurrences of a particular transition, e.g.
322   N_{XY} is the number of transitions of X to Y. C is calculated for all acts
323   (A), summed and then compared to chi-squared tables at the 95%
324   level and the relevant degrees of freedom.
325   The degrees of freedom calculation was based on the
326   transitions that occurred in the data (i.e. the number of cells in the
327   chi table larger than zero). The degrees of freedom were calculated
328   by:

329    $$df = \sum_A (m - k_A - 1)(m - l_A - 1)$$ (2)

330   where k_A is the number of transitions towards A and l_A is the number
331   of transitions from which A cannot occur. This results in two scores:
332   a chi-squared value and a one/zero score which indicates whether
333   the chi-squared value is significant at the level of $p=0.05$ level (i.e.
334   particular events follow each other at higher than chance levels).
335   Significant chi-squared values were used and are labelled hereon in
336   as the sequential dependency score; these scores indicate the
337   degree of sequential dependency since higher scores represent
338   more predictable behaviour sequences.
339
340   *2.3.2 Details of T-pattern analysis*
341   Whereas the Markov chains analysis described above is parameter-
342   free, in Theme the sensitivity of the pattern detection algorithm can
343   be altered using various parameters entered by the experimenter
344   (Magnusson, 2004; Bonasera et al., 2008). Based on some
345   preliminary investigations, we set the parameters as follows:
346   significance level α = 0.001 (this represents the acceptance
347   threshold for the null hypothesis that a pattern could have occurred
348   if the events within it were randomly distributed throughout the
349   observation period), minimum occurrences N_{min} = median (this
350   defines the minimum number of times a pattern has to occur in
351   order to be "detected", median refers to the frequency of all event
352   types). All other settings were left at the defaults specified in
353   Theme (for rationale see Discussion). The final output metrics
354   derived from Theme include: the number of T-patterns; the total
355   number of T-pattern occurrences; and the average number of times
356   that each individual T-pattern occurs. The number of T-patterns is,
357   as described, the number of different types of T-patterns (of varying
358   lengths) found by the pattern detection algorithm. The total number
359   of T-pattern occurrences is the sum total of all occurrences of all T-
360   pattern types. Finally, the average number of times each individual
361   T-pattern occurs is the total number of occurrences of all T-pattern
362   types divided by the number of different types of T-patterns.
363

2.3.3 Overall activity
We defined an activity metric as the overall number of location transitions during the observation period. The longer a string of recorded behaviour, the more likely it is that correlations will occur between randomly occurring behaviours due to chance. Though both Markov chain analysis and T-pattern analysis have partial controls for this effect, we included a measure of general activity level to examine its relationship to route-tracing and stereotypic behaviour.

2.3.4 Statistical analysis
All other statistical analyses were conducted using SPSS 16.0 for Mac (SPSS Inc.) or SAS 9.1. Parametric and non-parametric methods were used as appropriate with all assumptions checked. When multiple post hoc tests were conducted the Bonferroni correction was applied.

  We began by exploring the relationship between abnormal repetitive behaviours and proportion of time spent in abnormal cage locations (on the walls and ceiling). We then explored the correlation between the pattern metrics, activity level and the above behavioural measures. To ascertain which pattern metric was the best predictor of stereotypic behaviour, we conducted a forward stepwise regression analysis establishing which of the pattern metrics predicted significant amounts of variability for each abnormal behaviour measure. For all of the above analyses the values for each bird across the six weeks of the experiment were averaged to avoid pseudoreplication.

  To establish whether the pattern metrics could be used to predict the emergence of stereotypic behaviour in particular subjects, we categorised them into somersaulting and non-somersaulting individuals. We used a General Linear Model with week number as a within-subjects factor and somersaulting behaviour as a between-subjects factor to conduct univariate and multivariate analyses. Since many of the variables were correlated, changes in behaviour patterns over time were analysed using a doubly multivariate analysis. All behavioural measures were screened individually for time effects using a univariate analysis where time (weeks 1-6) was a within-subjects factor. Only those variables that were statistically significant were included in the multivariate analysis.

3. Results

3.1 Do different abnormal behaviours correlate with each other?

Since only three of our eight birds performed somersaults, it was first necessary to establish whether there was any connection to discrete behaviours we recognised as stereotypic or abnormal that were performed by more of our subjects. The number of somersaults (the most widely accepted and easily recognised stereotypy in starlings) correlated significantly with the number of unbalancing events (six out of eight birds performed unbalances), which in turn correlated significantly with the number of head tilt events (seven out of eight birds performed head tilts) (see Fig. 1 and Table 2). The number of somersaults also correlated significantly with the proportion of time spent on the ceiling. However, the number of head tilts correlated significantly with the proportion of time spent on the cage walls.

3.2 Do pattern metrics correlate with abnormal behaviour?

We wanted to establish whether the pattern metrics computed using Markov chain and T-pattern analysis as well as general activity level correlated with number of somersaults. Since only three birds performed somersaults (resulting in a high number of tied ranks), we also examined the correlation of the pattern metrics with the other abnormal repetitive behaviours and abnormal perching locations (which were performed by more subjects and showed moderate to high levels of inter-correlation).

The pattern metrics computed using T-pattern and Markov chain analysis all correlated significantly with at least two measures of abnormal behaviour/location (see Fig. 1 and Table 3). The number of T-patterns was the only pattern metric to correlate significantly with somersaulting.

Our simpler measure of overall activity also correlated significantly with three measures of abnormal behaviour (and was moderately to strongly correlated with the remaining two), as well as all four pattern metrics (sequential dependency score, Kendall's τ = .966, p < 0.001; number of T-patterns Kendall's τ = .846, p = 0.008; total number of T-pattern occurrences, Kendall's τ = .907, p = 0.002; average number of times each T-pattern occurs, Kendall's τ = .887, p = 0.003). These results therefore raise the question of whether the more complex pattern metrics reveal anything more than overall activity levels?

3.3 What accounts for most variance in abnormal behaviour?

Since activity correlates with the occurrence of abnormal behaviour, we conducted a separate regression analysis for each of the abnormal behaviour measures to establish whether our pattern metrics or overall activity explains the most variance in the frequency of abnormal behaviour. Regression analyses conducted for each measure of abnormal behaviour show that the metrics from T-pattern analysis account for the majority (and sometimes nearly
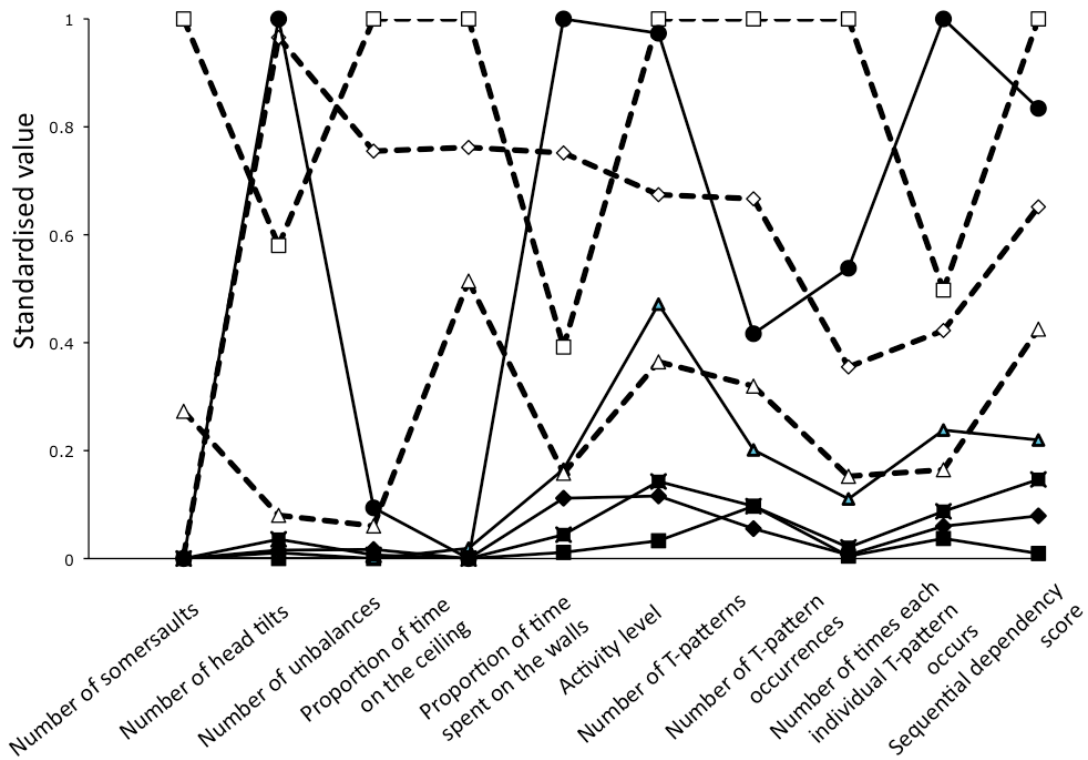
Fig. 1. Relationship between standard behavioural measures, pattern metrics and general activity level for each individual bird. Each variable was standardised by setting the maximum recorded value to one and scaling the remaining values accordingly. This allows for a ranked comparison analogous to the statistical values presented in Table 2. Note that each line represents a subject but they do not imply any extrapolation between data points. Those three birds that exhibited somersaulting are marked by open symbols.

Table 2. Correlations between the frequency of abnormal repetitive behaviours and time spent perching in "abnormal" locations.

Abnormal repetitive behaviours	Abnormal location / behaviour			
	Proportion of time spent on the ceiling	*Proportion of time spent on the walls*	*Number of head tilts*	*Number of unbalances*
Number of somersaults	.804 .014*	.267 .389	.356 .251	.635 .043*
Number of head tilts	.322 .288	.714 .013*		.691 .018*
Number of unbalances	.533 .082	.546 .061		

Note: Quoted statistics: upper number represents Kendall's τ and lower number is the p value. * indicates results significant at the $p < 0.05$ level.

Table 3. Correlations between pattern/activity metrics and the frequency of abnormal repetitive behaviours/time spent in "abnormal" locations.

Measurement method	Pattern metric	Number of somersaults	Number of unbalances	Number of head-tilts	Time spent on the ceiling	Time spent on the walls
Markov chain analysis	*Sequential dependency score*	.535 .085	.691 .018*	.714 .013*	.645 .034*	.773 .024*
Theme analysis	*Number of T-patterns*	.635 .043*	.667 .024*	.546 .061	.739 .016*	.555 .153
	Total number of T-pattern occurrences	.535 .085	.691 .018*	.714 .013*	.645 .034*	.588 .125
	Average number of times each individual T-pattern occurs	.356 .251	.546 .061	.714 .013*	.483 .111	.928 .001*
Overall activity	*Number of transitions*	.445 .152	.618 .034*	.643 .026*	.564 .063	.814 .014*

Note: Quoted statistics: upper number represents Kendall's τ and lower number is the p value. * indicates results that are significant at the $p < 0.05$ level.

456   all) of the variance in abnormal behaviour (Table 4). Overall activity
457   explained no significant variance in abnormal behaviour above and
458   beyond the variance explained by the pattern metrics.
459

460   *3.4 How does behaviour change over time?*

461   Next, we examined whether there was any change in the various
462   behavioural measures (including "normal" behaviours, abnormal
463   behaviours, pattern metrics and activity levels) over the course of
464   the six-week observation period. Additionally, we asked whether
465   there was a significant difference between subjects that developed a
466   somersaulting stereotypy and those that did not (see Fig. 2a). Since
467   many of the abnormal behaviour measures are correlated, it was
468   necessary to perform a multivariate analysis. However, due to our
469   small sample size and large number of variables we carried out an
470   initial screening procedure, conducting univariate tests for each
471   behavioural variable. As a result, five were included in the
472   multivariate analysis: number of head tilts, proportion of time spent
473   on the cage walls, proportion of time spent on the food bowl, the
474   number of T-patterns, and sequential dependency scores. The
475   multivariate analysis showed an overall significant effect of week
476   number (MANOVA: Wilk's Lambda =0.02, $F_{2,5}$= 22.64, p=0.04), a
477   difference between somersaulting and non-somersaulting birds
478   ($F_{1,6}$=6.73, p=0.04) and an interaction between somersaulting and
479   week (Wilk's Lamda =0.02, $F_{2,5}$=19.30, p=0.05). To understand
480   which group means differ significantly from others, Bonferroni-
481   corrected univariate tests are presented in Fig. 2 (b-f). The
482   proportion of time spent on the food bowl and sequential
483   dependency scores increased over the six-week period whilst
484   proportion of time spent on the cage walls decreased.
485   Somersaulting birds spent longer on the cage walls and had a
486   higher number of T-patterns. Somersaulting birds also had higher
487   sequential dependency scores in some but not all weeks, reflected
488   by the interaction effect of week and somersaulting. Somersaulting
489   and week effects on head tilting and number of T-patterns were not
490   significant in univariate testing when the Bonferroni correction was
491   applied.
492
493

4. Discussion

495
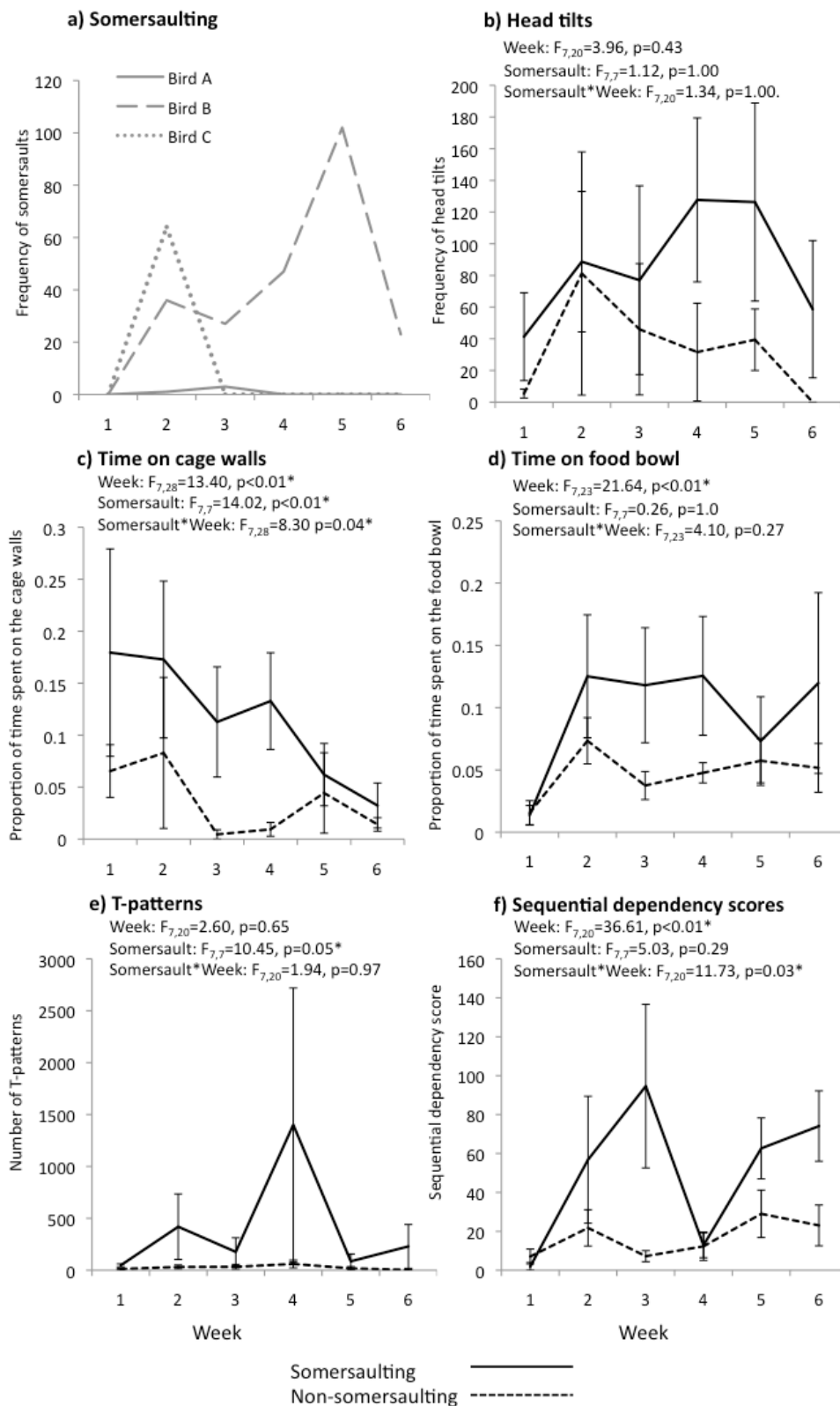
4.1 Main findings

497   The behavioural scores for recognised stereotypies (number of
498   somersaults) and putative related abnormal behaviours (number of
499   head tilts; number of unbalances; proportion of time spent on the
500   walls and ceiling) are all moderately to highly correlated with each

Table 4. Regression models for each individual abnormal behaviour measure. All four of the pattern metrics and activity level were available as independent variables but only those that passed the criterion of p<0.05 were included in each model in a sequential forward stepwise fashion.

Dependent variable and significant predictors	Standardised beta coefficient	F-value (degrees of freedom)	Significance	R-square of overall model
Number of somersaults =				
Total number of T-pattern occurrences +	1.327	35.320 (5,2)	0.001	.934
Average number of times each individual T-pattern occurs	-0.747			
Number of head tilts =				
Average number of times each individual T-pattern occurs	0.869	18.461 (6,1)	0.005	.755
Number of unbalances =				
Number of T-patterns +	1.499	58.052 (5,2)	<0.001	.959
Sequential dependency score	-0.606			
Time spent on the ceiling =				
Number of T-patterns +	1.141	39.184 (5,2)	0.001	.940
Average number of times each individual T-pattern occurs	-0.432			
Time spent on the walls =				
Average number of times each individual T-pattern occurs	0.928	37.367 (6,1)	0.001	.862

a) Somersaulting

Bird A
Bird B
Bird C

Frequency of somersaults

b) Head tilts

Week: $F_{7,20}=3.96$, p=0.43
Somersault: $F_{7,7}=1.12$, p=1.00
Somersault*Week: $F_{7,20}=1.34$, p=1.00.

Frequency of head tilts

c) Time on cage walls

Week: $F_{7,28}=13.40$, p<0.01*
Somersault: $F_{7,7}=14.02$, p<0.01*
Somersault*Week: $F_{7,28}=8.30$ p=0.04*

Proportion of time spent on the cage walls

d) Time on food bowl

Week: $F_{7,23}=21.64$, p<0.01*
Somersault: $F_{7,7}=0.26$, p=1.0
Somersault*Week: $F_{7,23}=4.10$, p=0.27

Proportion of time spent on the food bowl

e) T-patterns

Week: $F_{7,20}=2.60$, p=0.65
Somersault: $F_{7,7}=10.45$, p=0.05*
Somersault*Week: $F_{7,20}=1.94$, p=0.97

Number of T-patterns

Week

f) Sequential dependency scores

Week: $F_{7,20}=36.61$, p<0.01*
Somersault: $F_{7,7}=5.03$, p=0.29
Somersault*Week: $F_{7,20}=11.73$, p=0.03*

Sequential dependency score

Week

Somersaulting ——————
Non-somersaulting - - - - - - - -

Fig. 2. (a) The number of somersaults performed by individuals across the six weeks. (b)-(f) Main effects of week and differences between somersaulting and non-somersaulting birds on variables included in multivariate analysis. Each figure lists the Bonferroni corrected GLM results and shows mean vales ± one standard error.

other in our caged starlings (Table 2). Though we recognise that these findings derive from a small sample, we regard them as suggestive evidence that these behaviour patterns are functionally related, perhaps via escape motivation. We therefore used these behaviours as a standard of abnormal behaviour against which to validate our novel pattern detection methods. The application of the pattern metrics (derived from Markov chains and T-pattern analysis) for objectively quantifying route-tracing stereotypic behaviour is supported by strong correlations between these metrics and the measures of stereotypic and abnormal behaviour established above (Table 3).

The high level of correlation between the pattern metrics, the established measures of abnormal behaviour and the general activity levels of the birds suggests a single underlying cause. It is important to establish whether all of our behavioural measures are simply different proxies for activity. Stereotypic behaviour has been previously linked to activity (Hansen and Jeppesen, 2001; Vickery and Mason, 2004), and has been hypothesised to be an active response to an eliciting stimulus (Mason, 1991b). Higher activity levels translate directly into longer sequences of locations from which to calculate the pattern metrics. Statistically, a longer sequence would result in more patterns than a shorter sequence, even if sequences were random. However, our regression analyses show that the pattern metrics explain high levels of variance in the scores of abnormal behaviour, above and beyond that explained by overall activity. This is particularly the case for the T-pattern metrics, which are better predictors of abnormal behaviour than the sequential dependency scores derived from Markov chain analysis. None of the final regression models used to predict abnormal behaviour included activity level as a significant predictor variable (Table 4). Pattern metrics therefore explain more variance in the standard measures than activity levels, and we can be confident that T-pattern analysis (and to a lesser extent Markov chain analysis) are not merely detecting differences in the overall amount of behaviour performed.

Pattern metrics, abnormal behaviour scores and other behaviours showed changes over the six-week time course of the experiment that differed between somersaulting and non-somersaulting birds. However, since somersaulting appeared as early as the second session of behavioural recording we were unable to test whether the pattern metrics could be used to predict which individuals would develop stereotypies in the future. We believe that the pattern metrics we used have the potential to predict the development of stereotypic behaviour, but in order to study this in captive starlings it would be necessary to record behaviour more often during the first two weeks of caging before somersaulting emerges.

548   Birds showed no consistent increase in somersaulting across
549   the six-week observation period (Fig. 2a) and did not continue to
550   perform somersaulting behaviour after they were returned to larger
551   free-flight aviaries. This suggests that although the somersaulting
552   stereotypy had developed, it was not fixed and irreversible. With
553   the exception of somersaulting, other abnormal behaviour (such as
554   route-tracing) did not reach a stage where it appeared rigid and
555   stereotypic. To some extent the birds we used might have been
556   buffered against irreversible stereotypy because they were wild-
557   caught and animals raised in barren conditions are more likely to
558   develop irreversible stereotypies (Cooper and Nicol, 1996).
559   Changes in behaviour over time hint at the aetiology of the
560   somersaulting stereotypy. Time on the cage walls decreased and
561   time on the food bowl increased over time in captivity. This is
562   consistent with the findings of Maddocks et al. (2002) who
563   explained an observed decrease in clinging to cage walls as
564   demonstrative of a reduction in escape motivation as birds became
565   more settled in captivity. As our somersaulting birds were more
566   active than non-somersaulting birds, spending more time on the
567   cage walls particularly during the first four weeks of recordings, we
568   suggest that our data support the hypothesis that somersaulting is
569   linked with escape motivation.
570   We acknowledge that a more extensive data set is required to
571   establish rigorously how stereotypy and route-tracing behaviour
572   correlate with pattern metrics. Ideally, the data would cover the
573   entire temporal range of the development of stereotypy from its
574   absence to rigid stereotypic behaviour expression. However, whilst
575   this study involved a small sample, it complements previous studies
576   in captive starlings with larger sample sizes that revealed a
577   relationship between somersaulting stereotypies and an increased
578   repetitiveness in movement patterns quantified using Markov chains
579   (Asher et al., 2009; Asher et al., unpublish. data).
580
581   *4.2 Theme*

582   T-pattern analysis was successful in explaining variation in the
583   levels of abnormal repetitive behaviour expressed. In particular, the
584   average number of times each individual T-pattern occurs was
585   positively (though not always strongly) correlated with the
586   abnormal behaviours and proved to have strong explanatory power
587   in most of the regression analyses. This appears to confirm our
588   initial prediction that stereotypic animals should have a reduced
589   behavioural repertoire with progressively more time devoted to
590   performing stereotypic behaviour (Meehan et al., 2004). However,
591   the number of T-patterns was positively related (and the most
592   significant explanatory factor) to two of our abnormal behaviour
593   measures: number of unbalances and time spent on the ceiling.
594   This contradicts our prediction that stereotyping individuals should

595   demonstrate fewer different T-patterns as compared to non-
596   stereotyping subjects. We suggest that the particular set of subjects
597   studied might explain this contradiction. Specifically, we had a small
598   sample of individuals that demonstrated early-stage flexible
599   stereotypic patterns or no stereotypies at all.
600   We hypothesise that stereotyping individuals were more
601   active but in such a way that they performed behaviour in more
602   discrete behavioural bouts. If the behaviour sequence is still flexible
603   (i.e. B does not always follow A) but occurs in discrete behavioural
604   bouts separated by no activity, then a greater number of different
605   T-patterns would be found as compared to an individual who was
606   sequentially flexible but exhibited a constant stream of activity
607   (since this more closely resembles behaviour that occurs with
608   constant probability). If activity is concentrated within larger bouts,
609   there is a higher likelihood that a particular bout would include a
610   greater number of different event types. This would result in Theme
611   finding higher-level T-patterns (i.e. patterns including large
612   numbers of different events) with a corresponding exponential
613   increase in the number of sub-patterns that form the longer,
614   higher-level T-pattern (e.g. the AB, AC, and BC sub-patterns that
615   could form the T-pattern ABC). If stereotyping individuals expressed
616   more discrete behavioural bouts, this would explain the finding that
617   their behaviour contains more T-patterns, more types of T-patterns
618   and that T-patterns occurred more frequently.
619   Although Theme does have potential as a tool for
620   characterising stereotypic development, there are drawbacks
621   related to subjective input required from the user. To enable
622   adjustment of the search algorithms as appropriate for the dataset,
623   Theme has a suite of parameters that are set by the experimenter.
624   The two of greatest importance are probably α (the level of
625   significance for accepting a behaviour as occurring within the
626   confidence interval by chance) and N_{min} (the minimum number of
627   times a T-pattern must be detected in order to be counted).
628   Unfortunately, there is no objective approach for setting these
629   parameters (see Bonasera et al., 2008), supplementary
630   information). The Theme manual (Magnusson, 2004) suggests
631   testing multiple settings and deciding upon values best suited for
632   the current task. Theme had not been applied to the behaviour of
633   caged birds prior to this experiment, and we did not feel we should
634   use predictions about the relationship between stereotypic
635   behaviour and the patterns detected to determine our parameter
636   values (indeed our predictions proved incorrect in any case). Since
637   many of our subjects expressed large numbers of behavioural
638   events, we chose a strict value of $\alpha < 0.001$. This reduced the
639   number of seemingly irrelevant/redundant sub-patterns (e.g. ones
640   involving two events but in reversed order such as A-B and B-A)
641   and reduced the large levels of variance between individuals (since

activity levels were similarly highly variable). The Theme manual (Magnusson, 2004) suggests that N_{min} is generally set to 3 (i.e. a pattern has to occur a minimum of 3 times in order to be kept) or "median" (the median of the overall frequency of events). We employed the latter since this provided an additional way of reducing the large variation in activity (and hence number of behavioural events) that occurred between our subjects. It seemed most appropriate to set the other parameters at the default levels as there was no reason to restrict the pattern detection algorithm any further.

One further difficulty with Theme lies in the validation of T-pattern detection for larger data sets (since more patterns will occur by chance in longer strings of recorded behaviour). Magnusson (2000) has discussed methods of validation, but Theme is limited to a graphical comparison of T-pattern levels of the data set and a randomised version of the same data. It is suggested that the search algorithm parameters are adjusted such that no T-patterns are found in the randomised data. However, this reduces the number of T-patterns also found in the real data, with no discrimination made between T-patterns that are part of normal, functional behaviour and those that are behaviourally functionless (and hence, by definition, stereotypies). A formal statistical comparison of Theme outputs from the actual data against the randomised data set would be highly desirable in order to provide objective confirmation that detected T-patterns have biological significance.

4.3 Markov chain analysis

The sequential dependency scores produced from Markov chain analysis did not explain as much variation in the standard measures of abnormal behaviour as T-pattern analysis. They did, however, correlate with our standard measures of abnormal behaviour. The differences between Theme metrics and sequential dependency might have related to the extra temporal dimension that is incorporated into detection of T-patterns. Whereas sequential dependency describes the degree to which contiguous events occur more than expected by chance, Theme uses the relative position in time, and can therefore detect a pattern of events even if it is interrupted by an unrelated event. As with Theme, higher sequential dependency scores are expected as the number of events (and hence general activity) increases. Developing a Monte-Carlo or bootstrapping validation using random permutations could provide a formal control for this phenomenon.

Despite the drawbacks outlined above, the sequential dependency method provides results that are simpler to interpret than Theme. A score is judged as showing evidence of significant sequential dependency based on the chi-square statistic. In addition

689 there is only one pattern metric produced (sequential dependency
690 score) and the computation of this metric is free from any
691 parameter assumptions.
692
693

694 *4.4 Conclusions*

695 Our results show that T-pattern analysis in Theme, and to a lesser
696 extent Markov chain-based methods, can be used to quantify
697 individual differences in animals' use of space. Pattern metrics
698 derived from Theme were the best predictors we found of a range of
699 abnormal behaviour patterns in starlings including the
700 somersaulting stereotypy. These results suggest that space use
701 pattern metrics could be useful for identifying individuals with a
702 tendency towards stereotypic behaviour. When combined with
703 technologies for automatically recording the spatial location of an
704 animal within a cage (e.g. using Noldus' Ethovision), the pattern
705 metrics we describe could be used to fully automate the
706 quantification of complex route-tracing stereotypies.
707
708
709

710   Acknowledgements

717 *Author note*
718 Ben Brilot and Lucy Asher contributed equally to this paper:
719 carrying out the behavioural recording, scoring behaviour,
720 conducting the primary analysis and writing the first draft of the
721 manuscript. Gesa Feenders conducted the T-pattern analysis and
722 wrote portions of the text relevant to this analysis. Melissa Bateson
723 conceived the study; supervised the behavioural recording, scoring
724 and statistical analysis; and edited the draft manuscripts.
725
726
727

728   References

729 Asher, L. and Bateson, M., 2008. Use and husbandry of captive
730 European starlings (*Sturnus vulgaris*) in scientific research: a review
731 of current practice. Lab. Anim., 42: 111--126.
732

733 Asher, L., Davies, G.T.O., Bertenshaw, C.E., Cox, M.A.A. and
734 Bateson, M., 2009. The effects of cage volume and cage shape on
735 the condition and behaviour of captive European starlings (*Sturnus*
736 *vulgaris*). Appl. Anim. Behav. Sci., 116: 286--294.
737
738 Bateson, M. and Matheson, S.M., 2007. Performance on a
739 categorisation task suggests that removal of environmental
740 enrichment induces 'pessimism' in captive European starlings
741 (*Sturnus vulgaris*). Anim. Welfare, 16: 33--36.
742
743 Blumstein, D.T., Evans, C.S. and Daniel, J.C. 2000. JWatcher 0.9.
744 An Introductory User's Guide, 33 pp.
745
746 Bonasera, S.J., Schenk, A.K., Luxenberg, E.J. and Tecott, L.H.,
747 2008. A novel method for automatic quantification of
748 psychostimulant-evoked route-tracing stereotypy: application to
749 *Mus musculus*. Psychopharmacol., 196: 591--602.
750
751 Cooper, J.J. and Nicol, C.J., 1996. Stereotypic behaviour in wild
752 caught and laboratory bred bank voles (*Clethrionomys glareolus*).
753 Anim. Welfare, 5: 245--257.
754
755 Garner, J.P., Mason, G.J. and Smith, R., 2003. Stereotypic route-
756 tracing in experimentally caged songbirds correlates with general
757 behavioural disinhibition. Anim. Behav., 66: 711--727.
758
759 Golani, I., Kafkafi, N. and Drai, D., 1999. Phenotyping stereotypic
760 behaviour: collective variables, range of variation and predictability.
761 Appl. Anim. Behav. Sci., 65: 191--220.
762
763 Greenwood, V.J., Smith, E.L., Goldsmith, A.R., Cuthill, I.C., Crisp,
764 L.H., Walter-Swan, M.B. and Bennett, A.T.D., 2004. Does the flicker
765 frequency of fluorescent lighting affect the welfare of captive
766 European starlings? Appl. Anim. Behav. Sci., 86: 145--159.
767
768 Haccou, P. and Meelis, E. 1992. Statistical analysis of behavioural
769 data: an approach based on time-structured models. Oxford
770 University Press, Oxford, 404 pp.
771
772 Hansen, C.P.B. and Jeppesen, L.L., 2001. Swimming activity of farm
773 mink (*Mustela vison*) and its relation to stereotypies. Acta
774 Agriculturae Scandinavica Section a-Animal Science, 51: 71--76.
775
776 Lyon, M. and Kemp, A.S., 2004. Increased temporal patterns in
777 choice responding and altered cognitive processes in schizophrenia
778 and mania. Psychopharmacol., 172: 211--219.
779

Maddocks, S.A., Goldsmith, A.R. and Cuthill, I.C., 2002. Behavioural and physiological effects of absence of ultraviolet wavelengths on European starlings *Sturnus vulgaris*. J. Avian Biol., 33: 103--106.

Magnusson, M.S., 2000. Discovering hidden time patterns in behavior: T-patterns and their detection. Behav. Res. Methods Instrumen. Comput., 32: 93--110.

Magnusson, M.S. 2004. Theme: Powerful tool for the detection and analysis of hidden patterns in behaviour. Reference Manual Version 5.0. Noldus Information Technology, Wageningen, 221 pp.

Magnusson, M.S., 2005. Understanding social interaction: discovering hidden structure with model and algorithms. In: L. Anolli, S. Duncan, M.S. Magnusson and G. Riva (Eds.), The hidden structure of interaction: from neurons to culture patterns, IOS Press, Amsterdam, pp. 3--22.

Mason, G.J., 1991a. Stereotypies and suffering. Behav. Process., 25: 103--115.

Mason, G.J., 1991b. Stereotypies - a critical review. Anim. Behav., 41: 1015--1037.

Mason, G.J. and Rushen, J., 2006. Stereotypic animal behaviour: fundamentals and applications to welfare. CABI, Wallingford, 336 pp.

Matheson, S.M., Asher, L. and Bateson, M., 2008. Larger, enriched cages are associated with 'optimistic' response biases in captive European starlings (*Sturnus vulgaris*). Appl. Anim. Behav. Sci., 109: 374--383.

Meehan, C.L., Garner, J.P. and Mench, J.A., 2003. Isosexual pair housing improves the welfare of young Amazon parrots. Appl. Anim. Behav. Sci., 81: 73--88.

Meehan, C.L., Garner, J.P. and Mench, J.A., 2004. Environmental enrichment and development of cage stereotypy in Orange-winged Amazon parrots (*Amazona amazonica*). Dev. Psychobiol., 44: 209--218.

Vickery, S. and Mason, G., 2004. Stereotypic behavior in Asiatic black and Malayan sun bears. Zoo Biol., 23: 409--430.

Warreyn, P., Roeyers, H., Van Wetswinkel, U. and De Groote, I., 2007. Temporal coordination of joint attention behavior in

827   preschoolers with autism spectrum disorder. Journal of Autism and
828   Developmental Disorders, 37: 501--512.
829
830   Würbel, H., 2002. Behavioral phenotyping enhanced – beyond
831   (environmental) standardization. Genes Brain and Behavior, 1: 3--
832   8.
833
834
835
836