Behavioural Processes 82 (2009) 256–264

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

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

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

[a] *Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, UK*
[b] *Royal Veterinary College, Hatfield, UK*

ARTICLE INFO

Article history:
Received 23 March 2009
Received in revised form 10 July 2009
Accepted 10 July 2009

Keywords:
Markov chain analysis
Somersaulting
Starling
Stereotypic behaviour
Sturnus vulgaris
Theme

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,b; 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 may fail to encompass stereotypies that involve movements not normally expressed under semi-free or free-ranging conditions. For example,

---

* Corresponding author at: Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, Henry Wellcome Building for Neuroecology, Framlington Place, Newcastle upon Tyne, Tyne and Wear NE2 4HH, UK.
Tel.: +44 0191 222 6246.
E-mail address: ben.brilot@ncl.ac.uk (B.O. Brilot).

0376-6357/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.beproc.2009.07.003

the somersaulting stereotypy (see below) of caged European starlings (*Sturnus vulgaris*) involves a locomotor pattern that actually adds to the diversity of movement seen in birds kept in free-flight aviaries. However, we agree with their supposition that an increase in the predictability of movement patterns could be a useful method for quantifying stereotypic behaviour (as per Meehan et al., 2004), and further develop this approach in the current paper.

To quantify potential stereotypies in flexible movement patterns, we investigated techniques for identifying patterns in the sequence of locations an animal occupies within its cage. Our rationale was that by using pattern detection algorithms to identify patterns in the sequences of discrete locations an animal visits within its cage, we might be able to objectively quantify route-tracing stereotypies. Moreover, we might be able to use this approach to detect stages in the development of a route-tracing stereotypy before a completely rigid circuit is established, and thus predict animals at risk of developing rigid stereotypies. In order to validate our methods, we asked whether the space use pattern metrics we derived correlated with established measures of abnormal and repetitive behaviour, on the assumption that individuals developing route-tracing stereotypies would also be likely to display other abnormal and repetitive behaviour patterns.

We tested the above ideas using data collected from caged wild-caught European starlings, arguably the most widely used passerine bird species in laboratory research (Asher and Bateson, 2008). In the laboratory, starlings are routinely kept in individual cages, a risk factor for the development of stereotypic behaviour (including route-tracing) in starlings and other bird species (Garner et al., 2003; Meehan et al., 2003; Asher et al., 2009). We recorded the behaviour of the birds over six weeks, thereby aiming to capture the development of abnormal behaviour patterns during the early stages of captivity and record the levels of rigid stereotypies during the later stages.

Some caged starlings develop a “somersaulting” stereotypy in which they repeatedly complete a backwards aerial flip (Greenwood et al., 2004), a behaviour pattern that we hypothesise might develop from a thwarted escape attempt. Somersaulting is readily identifiable and quantifiable as an abnormal stereotypic behaviour pattern; it occurs repeatedly in the same location within the cage, using a set pattern of movements and has no obvious function, indeed it often results in the subject risking damage since on occasion they land on their back. Therefore, we set out to use somersaulting as our standard for categorising and ranking 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.

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 Right wall Rear wall Front wall Ceiling	Bird clinging on to a side of the mesh rectangular cage with its claws.
Left perch Right perch	Bird is perched on one of two natural branch perches positioned towards the top of the cage.
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.

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 behaviour, the current behavioural event is a good predictor of the next behaviour or location, because the sequence is predictable by definition. We performed a Markov chain analysis of sequential dependency using an adapted version of the chi-squared test for first against second-order dependency in sequences of events originally described by Haccou and Meelis (1992). This method uses the transitional probabilities of one event following another event (first order) or pair of events (second order). The more repetitive a sequence, the higher the probability that a given event will follow a certain kind of event or pair of events (see Section 2 and Asher et al. (2009) for full details of this methodology).

We have already demonstrated that this latter technique detects significant differences in the behavioural sequences of starlings housed in cages of different sizes and shapes (Asher et al., 2009) and cages with and without environmental enrichment (Asher et al., unpublish. data). We found that sequential dependency scores derived from Markov chain analysis were highest, indicative of the most repetitive behavioural sequences, in birds that performed somersaulting (Asher et al., 2009). Therefore, we predicted that the subjects that performed somersaulting in this study would similarly have increased tendencies to route-tracing and correspondingly higher sequential dependency scores.

1.2. T-pattern analysis

The software application Theme uses algorithms that detect temporal relationships (termed T-patterns) between pairs of behavioural events (Magnusson, 2000, 2005). When two events co-occur within a critical time interval more often than expected by chance they are designated as a T-pattern. More specifically, the algorithm searches for the shortest (d_1) and longest (d_2) duration between two events (X and Y) and uses these to define a critical time-frame ($t+d_1$ to $t+d_2$). It then tests whether Y appears after X significantly more than expected within this critical time-frame as compared to the situation where Y has a constant probability of occurring throughout the observation period. The process is repeated with newly identified T-patterns treated as single events until no more pairs of events are found (Magnusson, 2000, 2004, 2005). Since T-patterns are identified based on the temporal relationship between events, as opposed to the order of events, a pattern of events can be identified even if it is interrupted by an event that does not form part of the pattern. Thus Theme has a major advantage in being able to detect patterns in sequences of behaviour that would be invisible to Markov chain analysis.

T-pattern analysis has not previously been applied explicitly to the problem of quantifying the development of stereotypic 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 × 2.15 m × 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 × 450 mm × 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 h. 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 3 h 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 min on its first day in the cage and for 30 min 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:

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

where N is the number of occurrences of a particular transition, e.g. N_{XY} is the number of transitions of X to Y. C is calculated for all acts

(A), summed and then compared to chi-squared tables at the 95% level and the relevant degrees of freedom.

The degrees of freedom calculation was based on the transitions that occurred in the data (i.e. the number of cells in the chi table larger than zero). The degrees of freedom were calculated by:

$$df = \sum_{A} (m - kA - 1)\ (m - lA - 1) \tag{2}$$

where k_A is the number of transitions towards A and l_A is the number of transitions from which A cannot occur. This results in two scores: a chi-squared value and a one/zero score which indicates whether the chi-squared value is significant at the level of $p = .05$ level (i.e. particular events follow each other at higher than chance levels). Significant chi-squared values were used and are labelled hereon in as the sequential dependency score; these scores indicate the degree of sequential dependency since higher scores represent more predictable behaviour sequences.

2.3.2. Details of T-pattern analysis

Whereas the Markov chains analysis described above is parameter-free, in Theme the sensitivity of the pattern detection algorithm can be altered using various parameters entered by the experimenter (Magnusson, 2004; Bonasera et al., 2008). Based on some preliminary investigations, we set the parameters as follows: significance level $\alpha = .001$ (this represents the acceptance threshold for the null hypothesis that a pattern could have occurred if the events within it were randomly distributed throughout the observation period), minimum occurrences N_{min} = median (this defines the minimum number of times a pattern has to occur in order to be "detected", median refers to the frequency of all event types). All other settings were left at the defaults specified in Theme (for rationale see Section 4). The final output metrics derived from Theme include: the number of T-patterns; the total number of T-pattern occurrences; and the average number of times that each individual T-pattern occurs. The number of T-patterns is, as described, the number of different types of T-patterns (of varying lengths) found by the pattern detection algorithm. The total number of T-pattern occurrences is the sum total of all occurrences of all T-pattern types. Finally, the average number of times each individual T-pattern occurs is the total number of occurrences of all T-pattern types divided by the number of different types of T-patterns.

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.

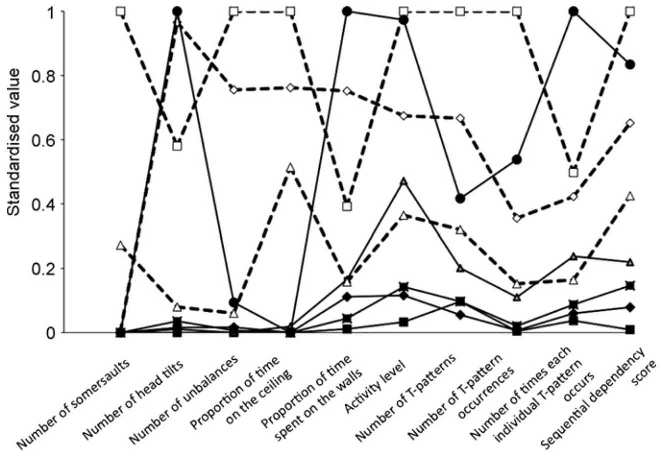

Fig. 1. Relationship between standard behavioural measures, pattern metrics and general activity level for each individual bird. Each variable was standardised by setting the maximum recorded value to one and scaling the remaining values accordingly. This allows for a ranked comparison analogous to the 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.

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

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	.267	.356	.635
	.014*	.389	.251	.043*
Number of head tilts	.322	.714		.691
	.288	.013*		.018*
Number of unbalances	.533	.546		
	.082	.061		

Note: Quoted statistics: upper number represents Kendall’s τ and lower number is the p value.
* Indicates results significant at the $p < .05$ 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	.691	.714	.645	.773
		.085	.018*	.013*	.034*	.024*
Theme analysis	Number of T-patterns	.635	.667	.546	.739	.555
		.043*	.024*	.061	.016*	.153
	Total number of T-pattern occurrences	.535	.691	.714	.645	.588
		.085	.018*	.013*	.034*	.125
	Average number of times each individual T-pattern occurs	.356	.546	.714	.483	.928
		.251	.061	.013*	.111	.001*
Overall activity	Number of transitions	.445	.618	.643	.564	.814
		.152	.034*	.026*	.063	.014*

Note: Quoted statistics: upper number represents Kendall’s τ and lower number is the p value.
* Indicates results that are significant at the $p < .05$ level.

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 $\tau = .966$, $p < .001$; number of T-patterns Kendall’s $\tau = .846$, $p = .008$; total number of T-pattern occurrences, Kendall’s $\tau = .907$, $p = .002$; average number of times each T-pattern occurs, Kendall’s $\tau = .887$, $p = .003$). 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 all) of the variance in abnormal behaviour (Table 4). Overall activity explained no significant variance in abnormal behaviour above and beyond the variance explained by the pattern metrics.

3.4. How does behaviour change over time?

Next, we examined whether there was any change in the various behavioural measures (including “normal” behaviours, abnormal behaviours, pattern metrics and activity levels) over the course of

Table 4
Regression models for each individual abnormal behaviour measure.

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 + average number of times each individual T-pattern occurs	1.327 −.747	35.320 (5,2)	.001	.934
Number of head tilts = average number of times each individual T-pattern occurs	.869	18.461 (6,1)	.005	.755
Number of unbalances = number of T-patterns + sequential dependency score	1.499 −.606	58.052 (5,2)	<.001	.959
Time spent on the ceiling = number of T-patterns + average number of times each individual T-pattern occurs	1.141 −.432	39.184 (5,2)	.001	.940
Time spent on the walls = average number of times each individual T-pattern occurs	.928	37.367 (6,1)	.001	.862

All four of the pattern metrics and activity level were available as independent variables but only those that passed the criterion of $p < .05$ were included in each model in a sequential forward stepwise fashion.

the six-week observation period. Additionally, we asked whether there was a significant difference between subjects that developed a somersaulting stereotypy and those that did not (see Fig. 2a). Since many of the abnormal behaviour measures are correlated, it was necessary to perform a multivariate analysis. However, due to our small sample size and large number of variables we carried out an initial screening procedure, conducting univariate tests for each behavioural variable. As a result, five were included in the multivariate analysis: number of head tilts, proportion of time spent on the cage walls, proportion of time spent on the food bowl, the number of T-patterns, and sequential dependency scores. The multivariate analysis showed an overall significant effect of week number (MANOVA: Wilk's Lambda = .02, $F_{2,5} = 22.64$, $p = .04$), a difference between somersaulting and non-somersaulting birds ($F_{1,6} = 6.73$, $p = .04$) and an interaction between somersaulting and week (Wilk's Lamda = .02, $F_{2,5} = 19.30$, $p = .05$). To understand which group means differ significantly from others, Bonferroni-corrected univariate tests are presented in Fig. 2b–f. The proportion of time spent on the food bowl and sequential dependency scores increased over the six-week period whilst proportion of time spent on the cage walls decreased. Somersaulting birds spent longer on the cage walls and had a higher number of T-patterns. Somersaulting birds also had higher sequential dependency scores in some but not all weeks, reflected by the interaction effect of week and somersaulting. Somersaulting and week effects on head tilting and number of T-patterns were not significant in univariate testing when the Bonferroni correction was applied.

4. Discussion

4.1. Main findings

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

Birds showed no consistent increase in somersaulting across the six-week observation period (Fig. 2a) and did not continue to perform somersaulting behaviour after they were returned to larger free-flight aviaries. This suggests that although the somersaulting stereotypy had developed, it was not fixed and irreversible. With the exception of somersaulting, other abnormal behaviour (such as route-tracing) did not reach a stage where it appeared rigid and stereotypic. To some extent the birds we used might have been buffered against irreversible stereotypy because they were wild-caught and animals raised in barren conditions are more likely to develop irreversible stereotypies (Cooper and Nicol, 1996).

Changes in behaviour over time hint at the aetiology of the somersaulting stereotypy. Time on the cage walls decreased and time on the food bowl increased over time in captivity. This is consistent with the findings of Maddocks et al. (2002) who explained an observed decrease in clinging to cage walls as demonstrative of a reduction in escape motivation as birds became more settled in captivity. As our somersaulting birds were more active than non-somersaulting birds, spending more time on the cage walls particularly during the first four weeks of recordings, we suggest that our data support the hypothesis that somersaulting is linked with escape motivation.

We acknowledge that a more extensive data set is required to establish rigorously how stereotypy and route-tracing behaviour correlate with pattern metrics. Ideally, the data would cover the entire temporal range of the development of stereotypy from its absence to rigid stereotypic behaviour expression. However, whilst this study involved a small sample, it complements previous studies in captive starlings with larger sample sizes that revealed a relationship between somersaulting stereotypies and an increased repetitiveness in movement patterns quantified using Markov chains (Asher et al., 2009; Asher et al., unpublish. data).

4.2. Theme

T-pattern analysis was successful in explaining variation in the levels of abnormal repetitive behaviour expressed. In particular, the average number of times each individual T-pattern occurs was positively (though not always strongly) correlated with the abnormal behaviours and proved to have strong explanatory power in most of the regression analyses. This appears to confirm our initial prediction that stereotypic animals should have a reduced behavioural repertoire with progressively more time devoted to performing stereotypic behaviour (Meehan et al., 2004). However, the number of T-patterns was positively related (and the most

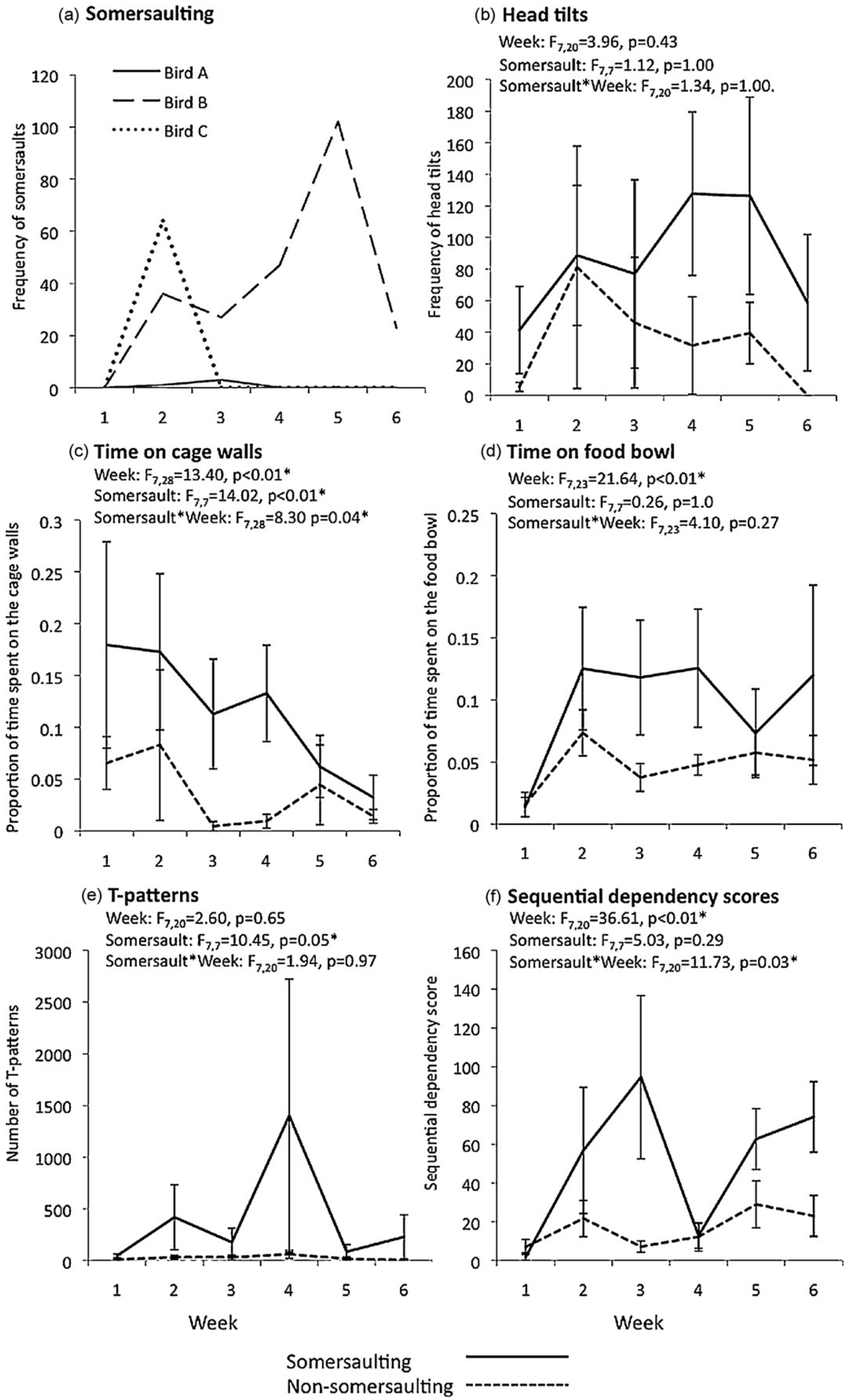

Fig. 2. (a) The number of somersaults performed by individuals across the six weeks. (b–f) Main effects of week and differences between somersaulting and non-somersaulting birds on variables included in multivariate analysis. Each figure lists the Bonferroni-corrected GLM results and shows mean values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

significant explanatory factor) to two of our abnormal behaviour measures: number of unbalances and time spent on the ceiling. This contradicts our prediction that stereotyping individuals should demonstrate fewer different T-patterns as compared to non-stereotyping subjects. We suggest that the particular set of subjects studied might explain this contradiction. Specifically, we had a small sample of individuals that demonstrated early-stage flexible stereotypic patterns or no stereotypies at all.

We hypothesise that stereotyping individuals were more active but in such a way that they performed behaviour in more discrete behavioural bouts. If the behaviour sequence is still flexible (i.e. B does not always follow A) but occurs in discrete behavioural bouts separated by no activity, then a greater number of different T-patterns would be found as compared to an individual who was sequentially flexible but exhibited a constant stream of activity (since this more closely resembles behaviour that occurs with constant probability). If activity is concentrated within larger bouts, there is a higher likelihood that a particular bout would include a greater number of different event types. This would result in Theme finding higher-level T-patterns (i.e. patterns including large numbers of different events) with a corresponding exponential increase in the number of sub-patterns that form the longer, higher-level T-pattern (e.g. the AB, AC, and BC sub-patterns that could form the T-pattern ABC). If stereotyping individuals expressed more discrete behavioural bouts, this would explain the finding that their behaviour contains more T-patterns, more types of T-patterns and that T-patterns occurred more frequently.

Although Theme does have potential as a tool for characterizing stereotypic development, there are drawbacks related to subjective input required from the user. To enable adjustment of the search algorithms as appropriate for the dataset, Theme has a suite of parameters that are set by the experimenter. The two of greatest importance are probably α (the level of significance for accepting a behaviour as occurring within the confidence interval by chance) and N_{min} (the minimum number of times a T-pattern must be detected in order to be counted). Unfortunately, there is no objective approach for setting these parameters (see Bonasera et al., 2008), supplementary information). The Theme manual (Magnusson, 2004) suggests testing multiple settings and deciding upon values best suited for the current task. Theme had not been applied to the behaviour of caged birds prior to this experiment, and we did not feel we should use predictions about the relationship between stereotypic behaviour and the patterns detected to determine our parameter values (indeed our predictions proved incorrect in any case). Since many of our subjects expressed large numbers of behavioural events, we chose a strict value of $\alpha < .001$. This reduced the number of seemingly irrelevant/redundant sub-patterns (e.g. ones involving two events but in reversed order such as A–B and B–A) 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 there is only one pattern metric produced (sequential dependency score) and the computation of this metric is free from any parameter assumptions.

5. Conclusions

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

Acknowledgements

We thank BBSRC for financial support (Project grants BB/E012000/1 and BB/05623/1 to MB); Michelle Waddle for technical assistance in caring for our birds; Magnus Magnusson and Bill Budenberg for useful discussions on the application of Theme; Jeffrey Katz and several anonymous referees for helpful comments.

Contributions: Ben Brilot and Lucy Asher contributed equally to this paper: carrying out the behavioural recording, scoring behaviour, conducting the primary analysis and writing the first draft of the manuscript. Gesa Feenders conducted the T-pattern analysis and wrote portions of the text relevant to this analysis. Melissa Bateson conceived the study; supervised the behavioural recording, scoring and statistical analysis; and edited the draft manuscripts.

References

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

Asher, L., Davies, G.T.O., Bertenshaw, C.E., Cox, M.A.A., Bateson, M., 2009. The effects of cage volume and cage shape on the condition and behaviour of captive European starlings (*Sturnus vulgaris*). Appl. Anim. Behav. Sci. 116, 286–294.

Bateson, M., Matheson, S.M., 2007. Performance on a categorisation task suggests that removal of environmental enrichment induces 'pessimism' in captive European starlings (*Sturnus vulgaris*). Anim. Welfare 16, 33–36.

Blumstein, D.T., Evans, C.S., Daniel, J.C. 2000. JWatcher 0.9. An Introductory User's Guide, 33 pp. http://www.jwatcher.ucla.edu.

Bonasera, S.J., Schenk, A.K., Luxenberg, E.J., Tecott, L.H., 2008. A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: application to *Mus musculus*. Psychopharmacology 196, 591–602.

Cooper, J.J., Nicol, C.J., 1996. Stereotypic behaviour in wild caught and laboratory bred bank voles (*Clethrionomys glareolus*). Anim. Welfare 5, 245–257.

Garner, J.P., Mason, G.J., Smith, R., 2003. Stereotypic route-tracing in experimentally caged songbirds correlates with general behavioural disinhibition. Anim. Behav. 66, 711–727.

Golani, I., Kafkafi, N., Drai, D., 1999. Phenotyping stereotypic behaviour: collective variables, range of variation and predictability. Appl. Anim. Behav. Sci. 65, 191–220.

Greenwood, V.J., Smith, E.L., Goldsmith, A.R., Cuthill, I.C., Crisp, L.H., Walter-Swan, M.B., Bennett, A.T.D., 2004. Does the flicker frequency of fluorescent lighting affect the welfare of captive European starlings? Appl. Anim. Behav. Sci. 86, 145–159.

Haccou, P., Meelis, E., 1992. Statistical Analysis of Behavioural Data: An Approach Based on Time-Structured Models. Oxford University Press, Oxford, p. 404.

Hansen, C.P.B., Jeppesen, L.L., 2001. Swimming activity of farm mink (*Mustela vison*) and its relation to stereotypies. Acta Agric. Scand. Sect. A—Anim. Sci. 51, 71–76.

Lyon, M., Kemp, A.S., 2004. Increased temporal patterns in choice responding and altered cognitive processes in schizophrenia and mania. Psychopharmacology 172, 211–219.

Maddocks, S.A., Goldsmith, A.R., 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 Instrum. Comput. 32, 93–110.

Magnusson, M.S., 2004. Theme: powerful tool for the detection and analysis of hidden patterns in behaviour. In: 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: Anolli, L., Duncan, S., Magnusson, M.S., Riva, G. (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., Rushen, J., 2006. Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare. CABI, Wallingford, pp. 336.

Matheson, S.M., Asher, L., 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., 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., 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., Mason, G., 2004. Stereotypic behavior in Asiatic black and Malayan sun bears. Zoo Biol. 23, 409–430.

Warreyn, P., Roeyers, H., Van Wetswinkel, U., De Groote, I., 2007. Temporal coordination of joint attention behavior in preschoolers with autism spectrum disorder. J. Autism Dev. Disord. 37, 501–512.

Würbel, H., 2002. Behavioral phenotyping enhanced—beyond (environmental) standardization. Genes Brain Behav. 1, 3–8.