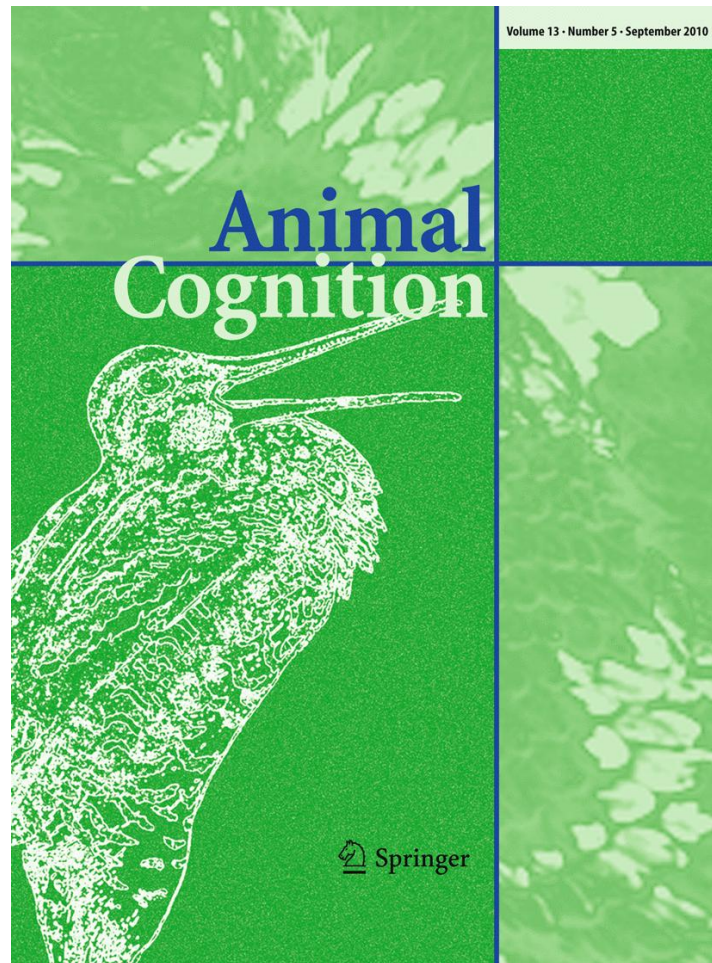


ISSN 1435-9448, Volume 13, Number 5



**This article was published in the above mentioned Springer issue.
The material, including all portions thereof, is protected by copyright;
all rights are held exclusively by Springer Science + Business Media.
The material is for personal use only;
commercial use is not permitted.
Unauthorized reproduction, transfer and/or use
may be a violation of criminal as well as civil law.**

Stereotyping starlings are more ‘pessimistic’

Ben O. Brilot · Lucy Asher · Melissa Bateson

Received: 14 September 2009/Revised: 10 March 2010/Accepted: 21 April 2010/Published online: 13 May 2010
© Springer-Verlag 2010

Abstract Negative affect in humans and animals is known to cause individuals to interpret ambiguous stimuli pessimistically, a phenomenon termed ‘cognitive bias’. Here, we used captive European starlings (*Sturnus vulgaris*) to test the hypothesis that a reduction in environmental conditions, from enriched to non-enriched cages, would engender negative affect, and hence ‘pessimistic’ biases. We also explored whether individual differences in stereotypic behaviour (repetitive somersaulting) predicted ‘pessimism’. Eight birds were trained on a novel conditional discrimination task with differential rewards, in which background shade (light or dark) determined which of two covered dishes contained a food reward. The reward was small when the background was light, but large when the background was dark. We then presented background shades intermediate between those trained to assess the birds’ bias to choose the dish associated with the smaller food reward (a ‘pessimistic’ judgement) when the discriminative stimulus was ambiguous. Contrary to predictions, changes in the level of cage enrichment had no effect on ‘pessimism’. However, changes in the latency to choose and probability of expressing a choice suggested that birds learnt rapidly that trials with ambiguous stimuli were unreinforced. Individual differences in performance of

stereotypes did predict ‘pessimism’. Specifically, birds that somersaulted were more likely to choose the dish associated with the smaller food reward in the presence of the most ambiguous discriminative stimulus. We propose that somersaulting is part of a wider suite of behavioural traits indicative of a stress response to captive conditions that is symptomatic of a negative affective state.

Keywords *Sturnus vulgaris* · European starling · Stereotypic behaviour · Cognitive bias · Environmental enrichment · Anxiety

Introduction

There is an extensive body of literature in human psychology showing that emotions can influence cognitive processes (Williams et al. 1997). For example, negative affective states such as anxiety can cause increased attention to threatening stimuli (Bar-Haim et al. 2007), and can increase the likelihood that ambiguous information will be interpreted pessimistically (Eysenck et al. 1991). These cognitive biases are sensitive both to short-term changes in anxiety (i.e. state anxiety) and stable individual differences in anxiety (i.e. trait anxiety) (Bar-Haim et al. 2007). Similar cognitive biases have also been shown to occur in animals whose states have been manipulated in various ways. Studies in laboratory rats (*Rattus norvegicus*; Harding et al. 2004; Burman et al. 2008a) and captive wild-caught European starlings (*Sturnus vulgaris*; Bateson and Matheson 2007; Matheson et al. 2008) have demonstrated that changes in husbandry that are likely to engender negative affective states cause ‘pessimistic’ biases in the animals’ interpretation of ambiguous stimuli, i.e. they have an expectancy of a more negative outcome. For example, in a

B. O. Brilot (✉) · M. Bateson
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

L. Asher
Department of Veterinary Clinical Sciences,
Royal Veterinary College, Hawkshead Lane, Potters Bar,
Hertfordshire AL9 7TA, UK

previous experiment, we trained starlings on a go/no-go task to discriminate between white and dark grey cardboard lids associated respectively with palatable and unpalatable mealworms hidden underneath. Once the birds had learnt to flip the white lids and avoid the dark grey lids, we measured their judgement biases (a form of cognitive bias) by presenting them with ambiguous lids of intermediate shades of grey. When the birds were housed in un-enriched cages, they were more reluctant to approach and flip the ambiguous lids than when they were housed in enriched cages (Bateson and Matheson 2007). We interpreted this result as a pessimistic judgement bias in birds housed in environmental conditions known to be associated with poorer welfare. Cognitive bias tasks have been tentatively supported as a new tool for diagnosing negative affective states in captive animals (Paul et al. 2005; Mendl et al. 2009). However, as we will explain later, a number of theoretical and empirical issues remain (see also Mendl et al. 2009). Our aims in the current paper are (1) to develop a novel judgement task for measurement of cognitive bias in starlings intended to improve on previous tasks and (2) to extend previous work in animals, by asking whether cognitive biases are correlated with individual differences in the incidence of abnormal behaviour, specifically stereotypes, that might also reflect trait anxiety. We present the background to each of these aims in more detail.

Judgement bias tasks for animals

The tasks developed so far to measure judgment biases in animals are based on the original design of (Harding et al. 2004). Subjects are initially trained to associate one stimulus, S+, with a reward (generally food) and another S– with either a reward of lower value (e.g. less food) or a punishment (e.g. white noise or a noxious food item). S+ and S– are chosen to lie on a continuous stimulus spectrum (e.g. a frequency range for auditory stimuli, a spectral range for visual stimuli or a range of directions for spatial stimuli). To measure a cognitive bias, subjects' responses to novel stimuli ('probes') that are intermediate between the trained stimuli are recorded in extinction (i.e. probe trials are not reinforced, avoiding any confound of reinforcement). A subject is regarded as showing a pessimistic judgement bias if it demonstrates a higher probability of exhibiting the response appropriate to the S– stimulus, than either that same subject in a more positive affective state or when compared to control subjects in a more positive state.

To date, the majority of published cognitive bias tasks have used a go/no-go procedure (Harding et al. 2004; Bateson and Matheson 2007; Burman et al. 2008a). In a go/no-go task, the subject is required to respond by performing a behaviour (e.g. lever press) in response to S+, but to

refrain from responding to S–. However, interpretation of data from a go/no-go task is complicated by the possibility that negative affective states are often associated with changes in general activity and feeding motivation. Therefore, on a go/no-go task, it is possible that subjects in a more negative affective state show a reduced probability of responding because they are less motivated to exploit a signalled food source, rather than because they interpret the ambiguous stimulus pessimistically. Thus, go/no-go tasks could be measuring a more general response bias as opposed to the assumed biased judgement of the ambiguous stimulus.

To address the above problem with go/no-go tasks we, and others, have advocated the use of choice tasks whereby subjects are required to make a different active response to both S+ and S– stimuli (e.g. Matheson et al. 2008; Enkel et al. 2009; see also unpublished studies cited in Mendl et al. 2009). This experimental design allows the effects of a response bias and a judgement bias to be distinguished: the former should result in reduced responding to all stimuli whilst the latter should result in reduced responding only to ambiguous stimuli. In a previous study with starlings, we used an operant task in which birds were required to choose (by pecking) a red or green key to classify a light stimulus as S+ or S– (Matheson et al. 2008). However, this task has a number of practical limitations including the length of time taken to train the birds, the requirement for moderate levels of food deprivation and the requirement to catch and handle the birds daily to transfer them to the operant chambers (a potentially anxiety-inducing procedure; Rich and Romero 2005). In the current paper, we present a novel choice task that is a modification of the simple lid-flipping task described earlier (Bateson and Matheson 2007). The task was designed to be quick to train, and minimally disruptive to the birds, with all training and experimental procedures conducted in the home cages (c.f. Matheson et al. 2008).

In line with previous cognitive bias experiments, we manipulated the affective state of the birds by altering the level of environmental enrichment provided in their cages (Bateson and Matheson 2007; Matheson et al. 2008). There are extensive data from a wide range of species showing that provision of more enrichment in captive animals' cages is associated with better welfare (Young 2003), and we have data from our own laboratory showing that starlings in enriched cages display less abnormal behaviour (Asher 2007). We used a repeated measures design involving a sequential change from environmentally enriched conditions to non-enriched conditions and then returning to enriched conditions. This design delivers greater statistical power in a study involving low numbers of subjects (a constraint of the intensive training required) and additionally allows us to examine how starlings

respond to both reduction and improvement in their environmental conditions. We have previously found that starlings show a greater change in cognitive bias in response to a reduction in environmental conditions than to an improvement (Bateson and Matheson 2007), adding to many results showing that animals' responses to a given situation depend on what they have previously experienced (Flaherty 1996; Bergvall et al. 2007; Burman et al. 2008b).

We hypothesised that starlings would show more pessimistic judgement biases in the non-enriched conditions compared with the enriched conditions. We also hypothesised that the birds would show a greater response to the removal of environmental enrichment than to its reinstatement in the final stage of the experiment.

Individual differences

The published cognitive bias studies in animals thus far have all examined whether judgement biases are sensitive to relatively short-term manipulations of state (Harding et al. 2004; Bateson and Matheson 2007; Burman et al. 2008a; Matheson et al. 2008). However, the literature in humans suggests that there are also stable individual ('trait') differences in both affect and pessimism (Bar-Haim et al. 2007). How individual animals cope with captivity is not only a matter of animal welfare, but is also of concern for the scientific validity of studies related to cognition. Differences in coping ability might be reflected in an animal's affective state and hence in the choices they make, regardless of the experimental treatment. Repetitive, abnormal behaviour (of which stereotypy is a type) is often regarded as an indicator of poor welfare since this behaviour can be associated with physiological and behavioural measures of stress (Mason and Rushen 2006). The evidence linking stereotypy and poor welfare is, however, mixed: a review of 90 studies in a range of species found that comparing between environments or regimes, those where the subjects stereotyped more invariably also scored lower on additional welfare measures, but within a group of animals under the same husbandry regime, 60% of studies showed that performance of stereotypies was associated with better welfare whilst the remainder showed the opposite (Mason and Latham 2004).

Given the above, we hypothesised that the presence of stereotyping behaviour in individual starlings should reflect stable individual differences in affective state (trait anxiety), and hence performance on a judgement bias task. If stereotyping birds are more pessimistic, then we argue that stereotypic behaviour is an indicator of poor welfare within starlings sharing the same environment (and vice versa). However, any difference in cognition related to stereotypic behaviour is of importance for future studies using captive birds with stereotypies.

Methods

Subjects and husbandry

The subjects were eight European starlings (four males and four females) caught from the wild under license from Natural England. An equal number of juveniles (birds in their first year) and adults were used. Both sex and age were counterbalanced for position in the experimental laboratory and time of behavioural testing. Prior to the experiment, the subjects were group-housed in an indoor aviary (2.4 × 2.15 × 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 (75 cm wide × 45 cm deep × 44 cm high) where visual and auditory contact with at least four conspecifics was possible. Previous studies in solitary-housed starlings have shown that differences in cage dimensions and enrichments cause changes in behaviour, condition and affective state (Bateson and Matheson 2007; Matheson et al. 2008; Asher et al. 2009); hence, we were confident that the stress of individual housing would not cause a ceiling effect constraint in the present study. During all training phases, the cages were furnished with enrichments suggested to improve the welfare of captive starlings, namely: natural branches, a water bath and a tray of bark chippings. The light: dark cycle was maintained at 14:10 h. At all times, other than those described in the following text, 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 the same daily routine throughout the study: cage husbandry at 8:00 a.m., followed by 2 h of food deprivation to increase the subjects' motivation for the learning task, followed by approximately 1 h of experimental trials (either learning or performing the cognitive bias task). On completion of the trials, the birds were once more allowed to feed ad libitum. Due to the staggering of trials (four birds were tested at a time), all experimental procedures were completed by approximately midday.

The birds' behaviour in the absence of the experimenter was recorded using two video cameras. Four birds were recorded per half hour between 3:00 and 4:00 p.m. The order of recording was counterbalanced such that each bird was recorded alternately from 3:00–3:30 or 3:30–4:00 on each day.

Cognitive bias task

We used a visual conditional discrimination task with differential rewards whereby the birds had to attend to the

colour of the background (S+ or S−) to predict which of two visually distinct dishes placed on it contained a hidden treat and which was empty (see Fig. 1a). In S+ trials, the treat was of a higher value than in S− trials. This difference in the level of reinforcement was required to ensure that active responses would be given to both the S+ and S− stimuli but that these responses could be differentiated by the subjects' motivation to exploit the reward. Once birds had learnt this discrimination, the test of cognitive bias involved presenting intermediate backgrounds between S+ and S− and recording which of the two dishes the bird chose in extinction (see Fig. 1b). We predicted that a more pessimistic bird would be more likely to interpret the ambiguous background as S−, and would therefore be more likely to choose the dish reinforced in S− trials. Matheson (2007) has previously piloted a version of this choice task in starlings.

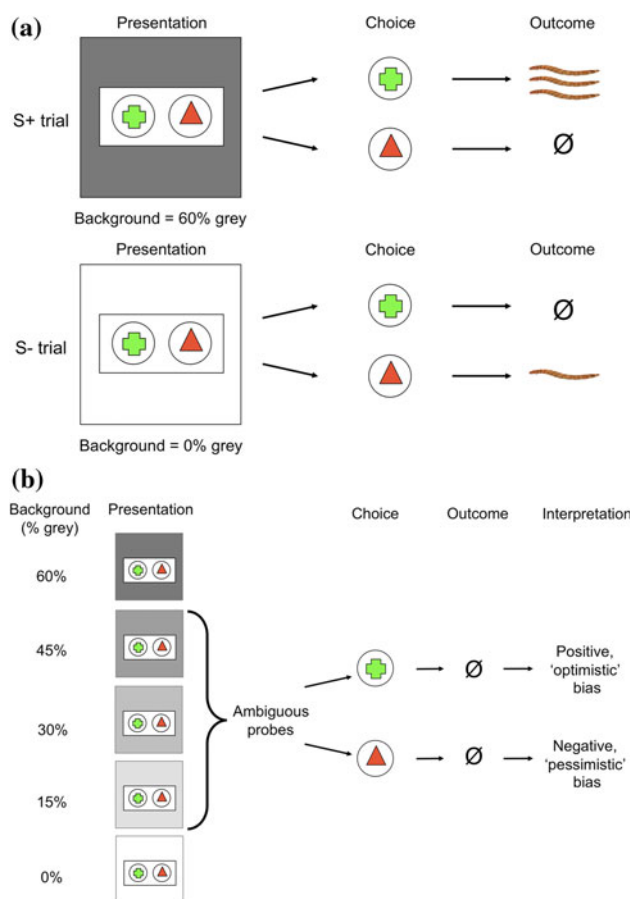


Fig. 1 **a** Details of the conditional discrimination task. The reward for a correct decision in the S+ trials was three mealworms, in the S− trials it was one mealworm. **b** Details of the cognitive bias test showing the three ambiguous probe background shades and our interpretation of the birds' choices

Apparatus

Two opaque Petri dishes (5 cm diameter × 0.5 cm high) were mounted 3 cm apart on a ceramic tile (15 × 15 cm). The background colour of the tile was used as the discriminative stimulus (S+ or S−) and was altered by affixing printed paper to the tile (Fig. 1a). S+ was printed using the settings of Hue 0°; Saturation 0%; and Brightness 40% in Microsoft Powerpoint, and is henceforth referred to as 60% grey. S− was printed using the following: Hue 0°; Saturation 0%; and Brightness 100%, and is henceforth referred to as 0% grey. The intermediate stimuli used as 'probes' for cognitive bias assessment were printed using the same Hue and Saturation values but varying Brightness to 55, 70 or 85% giving shades henceforth referred to as 45, 30 and 15% grey, respectively (Fig. 1b). All subjects were trained to associate the 60% grey background (i.e. S+) with a three mealworms reward and the 0% grey background (i.e. S−) with a one mealworm reward (Matheson's (2007) data showed that there was no effect of whether the higher reward occurred at the dark or light end of the stimulus spectrum). Pilot experiments had established that starlings in the same experimental set-up expressed a significant preference for three mealworms over one mealworm, confirming our assumption that the larger reward is of higher value. The Petri dishes were covered by circular cardboard lids (6.5 cm diameter) printed with one of two distinct stimuli (a red triangle or a green cross) that signalled which of the dishes contained the reward. The stimuli on the lids were also replicated on paper circles that were glued to the inside bottom of the Petri dishes (such that they were visible below the reward once the lid had been removed). Pilot experiments had also established that these stimuli were easily discriminable to the birds. Half of the birds were trained to associate the red triangle with the 3 mealworm reward and the green cross with the 1 mealworm reward whilst the other half were trained to the reverse assignment.

Training the birds for the cognitive bias task involved three phases: learning to flip lids, no-choice learning of the S+/S− conditional discrimination and free-choice learning of the full task with no ambiguous intermediates.

Training lid-flipping

A tile with a single, centrally placed Petri dish and a plain yellow cardboard lid was used in the initial lid-flipping training. The birds were rewarded with a mealworm placed on the yellow lid that was in turn placed on the tile. To facilitate learning of the lid-flipping task, the lid was then moved so as to cover more of the Petri dish whilst the mealworm was moved to within the dish (though still visible). In order, the lid was placed on the

ceramic tile, next to the dish which contained the mealworm; leaning against the dish but not obscuring the contents at all; progressively obscuring more of the contents (i.e. $\frac{1}{4}$ on, $\frac{1}{2}$ on, $\frac{3}{4}$ on) until it covered the entire dish and the contents could only be seen and recovered by moving the lid aside. Sixteen trials were given each day, each lasting for 60 s with a 180-s inter-trial interval (ITI). The bird was required to eat the mealworm within 60 s or the previous training stage was repeated. The bird was considered to have learnt to lid-flip once it had completed 16 consecutive trials with the lid fully covering the dish. This training phase continued until all birds had learnt the task.

Training the conditional discrimination

Next, birds were given a no-choice task using the two backgrounds (S+ and S−) and two lid-types that would be used in the final experiment (red triangle and green cross). They were presented with one open dish and one covered dish. The covered petri dish was always represented by the correct choice given the background context (i.e. if a 60% grey background was shown, then only the correct lid would be present on a Petri dish and this would contain the 3 mealworms reward). Trials were separated into blocks such that each bird had six presentations of the 60% grey background in a row followed by six presentations of the 0% grey background with the order of the blocks alternating between days. Each trial lasted 60 s with a 180-s ITI. Upon completion of the 12 trials, the birds were then given six probe trials. In these, the birds were presented with a choice of both lids; beneath the correct lid (given the background context) was the appropriate reward. Half of the probe trials used the 60% grey background and half used the 0% grey with the order pseudo-randomised. The birds were considered to have learnt the discrimination when they were above significance on the binomial test over the course of 3 days of testing (14 correct choices of 18). This phase continued for all birds until the last subject had learnt the discrimination.

Training partial reinforcement

Next, free-choice trials were given, and the probability of reward was gradually reduced using randomly interspersed unrewarded trials. Fifteen trials were conducted per day in this phase (again with a trial duration of 60 s and ITI of 180 s). Of these trials, all were rewarded on the first day, only 12 were rewarded on the second and third days, and only 10 were rewarded on the fourth day. The last day corresponded exactly to the trials that would be conducted as part of the cognitive bias task: 15

trials, five of which would not be reinforced. Reduction in the reinforcement probability was intended to prolong the conditioned response (CR) during the cognitive bias trials when ambiguous probe trials would be unrewarded and hence would cause the CR to extinguish.

Cognitive bias trials

The experimental cognitive bias test involved one session of 15 trials per day. Of these, five trials were reinforced presentations of the 60% grey background, five trials were reinforced presentations of the 0% grey background, two trials were unreinforced presentations of the 60 and 0% grey backgrounds and the remaining three were unreinforced presentations of each of the intermediate, ambiguous backgrounds (15, 30 and 45% grey). The order of presentation was pseudo-random, although we avoided contiguous unrewarded trials. As in the training trials, if a choice was not expressed within 60 s, then the trial was terminated and the usual ITI was observed.

In each phase of the experiment, the choice made on every trial was recorded for each subject. A choice was either recorded as appropriate for S+ (indicative of an optimistic bias) or appropriate for S− (indicative of a pessimistic bias). In addition, the time taken from presentation of the tile until a choice was expressed was also recorded (defined as when the lid was moved such that the bird could observe the reward). Latency has been successfully used as a response variable in previous cognitive bias studies with rats (Burman et al. 2008a), and typically correlates well with choice in previous studies with starlings (e.g. Bateson and Kacelnik 1995).

Housing manipulations

The cognitive bias trials were run daily over the course of 3 weeks whilst environmental enrichment in the cages was varied each week in a repeated measures design. For the first and third weeks, the birds were in environmentally enriched conditions similar to those from prior cognitive bias experiments in starlings (natural wood branches; water for bathing; and a tray filled with bark for natural probing opportunities: Bateson and Matheson 2007). For the second week, these enrichments were removed (non-enriched conditions) and the birds were left with the empty water and bark containers and uniform dowel perches. In order to standardise and minimise the disruption caused by the experimenter physically changing the housing conditions, the birds were caught and transferred to new, appropriately furnished (enriched/unenriched) cages. This was done on the day before each week of cognitive bias trials began (i.e. they were transferred three times at weekly intervals).

Scoring stereotypic behaviour

The most easily quantifiable stereotypy in caged starlings is a complete backwards aerial flip (or somersault; Greenwood et al. 2004; Brilot et al. 2009a). In previous work on stereotypic behaviour from this data set, we counted the number of somersaults for each bird, classifying a somersault as being any movement where the bird's feet passed above its head (Brilot et al. 2009a). These counts were scored, using J-Watcher v1.0 (Blumstein et al. 2000), from one half hour recording per week for each subject for the 6 weeks of the training period prior to the cognitive bias trials. Since not all birds exhibited somersaulting, we classified each as having exhibited somersaulting behaviour or not. We know from a previous study using data from these subjects that somersaulting behaviour is closely related to other abnormal repetitive behaviours and is associated with more repetitive movement patterns and with higher activity levels (Brilot et al. 2009a). Somersaulting therefore acts as a useful proxy measure for generally abnormal and repetitive behaviour.

Statistical analysis

All statistical analyses were carried out using SPSS 16.0 for Mac (SPSS Inc., Chicago, IL, U.S.A.). All data were modelled using repeated measures general linear models (GLMs), with assumptions being checked and the data being transformed prior to analysis where appropriate. Some of the birds developed a side bias during the 3 weeks of cognitive bias trials; we reduced the effect of the bias by discarding data from a bird for any day when it failed to reach criterion (at least 10 of 12 correct) for the subset of trials with the trained backgrounds (0 and 60% grey). Excluded data comprised 8 of 56 bird days (4 of which were for one subject) for week 1; 9 of 56 bird days for week 2 (spread across four birds); and none in the last week (one bird day comprises data from one subject for 1 day). Since we artificially reduced the variance in the response to the trained unambiguous stimuli, we excluded the data from them in our analyses.

Ethical note

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 inspected on a daily basis by the experimenter, were released back into free-flight aviaries after the experiment and showed no signs of adverse effects. Following completion of our studies, they received a full health inspection by a qualified veterinarian prior to release at the original capture site.

Results

Cognitive bias task

Training

The birds took 4.38 ± 2.13 days (mean \pm SD) to learn the lid-flipping task. All subjects had learnt the task by day seven. The birds took an additional 13.25 ± 4.33 (mean \pm SD) days to reach criterion on the conditional discrimination task. All birds had learnt the task after 20 days of discrimination training. In the last 3 days of discrimination training, before partial reinforcement was introduced, there was a difference in the latency of the birds to make a choice with 0 and 60% grey backgrounds, with birds being slower in the 0% grey background trials where the reward was only one mealworm compared with 60% grey background trials where the reward was three mealworms (paired *t*-test: $t_7 = 2.463$, $P = 0.043$).

Probability of choosing the stimulus associated with the larger reward

To establish whether cognitive bias was altered by our housing manipulation, we compared the probability of the subjects choosing the lid stimulus associated with the larger reward for the three ambiguous probe shades in each of the 3 weeks of the test (see Fig. 2 which plots the data from all unreinforced trials across all five shades, both ambiguous probes and non-ambiguous trained stimuli, to allow a baseline comparison). We used probe background

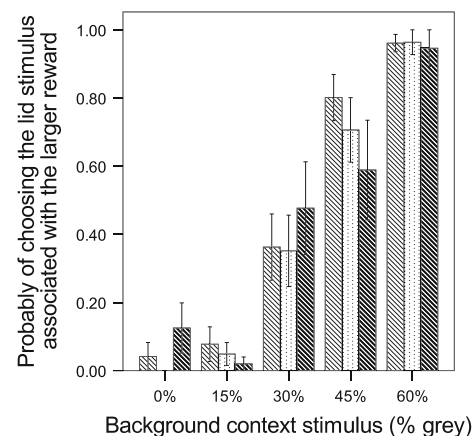


Fig. 2 Probability of choosing the stimulus associated with the higher reward during cognitive bias trials averaged across all subjects. Percentage grey values signify which background context was presented. Light hatched bars represent choices during week 1 (enriched conditions); dotted bars represent choices during week 2 (unenriched conditions); dark hatched bars represent choices during week 3 (enriched conditions). Bars show the mean for the 8 birds \pm one standard error

value and week number as categorical within-subjects factors in a repeated measures ANOVA, with the probability of choosing the lid stimulus associated with the larger reward as the dependent variable. There was a significant effect of the probe background shade on the birds' choices ($F_{2,8} = 58.90$, $P < 0.001$) but there was no significant effect of either the week of testing ($F_{2,8} = 0.14$, $P = 0.871$) or the interaction between the week of testing and probe background shade ($F_{4,16} = 0.01$, $P = 1.00$). Since three of the subjects did not respond to at least one probe background value for at least one of the weeks of testing, only five subjects could be included in a repeated measures ANOVA. This statistical test is therefore likely to be conservative. To reduce the likelihood of a type II error, we re-ran the analysis using the data from each probe background value in turn and included week as the only independent variable. This revealed that week of testing still had no significant effect on the choices expressed (Probe 15% grey background: $F_{2,12} = 0.24$, $P = 0.792$; Probe 30% grey background: $F_{2,8} = 0.02$, $P = 0.982$; Probe 45% grey background: $F_{2,14} = 1.01$, $P = 0.390$; changes in the degrees of freedom represent changes in sample size since some subjects failed to give a response to the probe in a given week).

Latency to choose

Although the choice data showed no effect of our housing manipulation, there remained the possibility that the birds' expectancy of reward size in the ambiguous probe trials was reflected in their latencies to respond. Where a bird failed to exhibit a choice within the time allowed, it was allocated the maximum trial duration of 60 s. We calculated the latency to flip a lid for each probe stimulus relative to the 3 mealworm reinforced stimulus (60% grey; Fig. 3). We predicted an increase in latency in non-enriched conditions (reflecting pessimism regarding the expected reward), followed by a decrease in latency on return to enriched conditions (reflecting recovered optimism). However, inspection of Fig. 3 suggests that if anything, latencies in the ambiguous probe trials increased across the 3 weeks of testing. A repeated measures ANOVA (with week number and probe background as within-subjects fixed factors) showed that latencies differed significantly across the probe background value and across the 3 weeks of trials (Probe value: $F_{2,14} = 10.22$, $P = 0.002$; Week number: $F_{2,14} = 7.92$, $P = 0.005$). Post hoc analysis using t-tests (with a Bonferroni correction applied) revealed a significant difference in the latency to respond to the 30 and 45% grey background probes ($P = 0.001$), but all other pairwise comparisons were non-significant ($P > 0.18$ for all). Similarly, Bonferroni-corrected post hoc analysis revealed a significant difference in

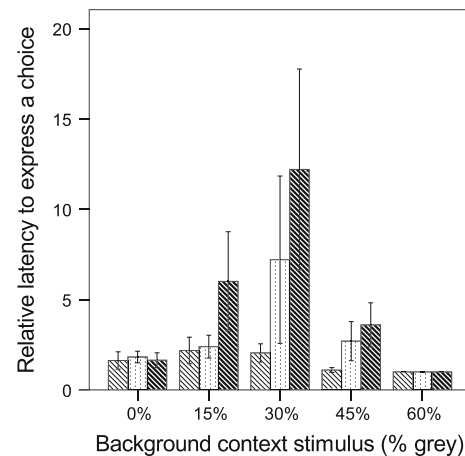


Fig. 3 Latency to approach and flip the lid averaged for each background context stimulus. The latency is corrected for each individual bird by dividing the actual mean latency by the mean latency to flip the lid of the rewarded “three-mealworms” stimulus (i.e. the 60% grey background trials) during the same week of trials. The percentage grey values signify which background context was presented. Light hatched bars represent latencies during week 1 (enriched conditions); dotted bars represent latencies during week 2 (unenriched conditions); dark hatched bars represent latencies during week 3 (enriched conditions). Bars show the mean for the 8 birds \pm one standard error

the latency to respond for weeks 1 and 3 ($P = 0.039$) but all other pairwise comparisons between weeks were non-significant ($P > 0.15$ for all). There was no significant interaction effect of the probe background value and week number on the latency to choose (Mauchly's test revealed that the assumption of sphericity was not tenable ($\chi_9 = 32.48$, $P < 0.001$), therefore the Greenhouse-Geisser correction was applied: $F_{1.94, 13.55} = 0.64$, $P = 0.537$).

Cognitive bias task and individual behavioural differences

To ascertain whether the presence of stereotypic behaviour (as an indicator of affective state) predicts the probability of choosing the stimulus associated with the larger reward, we conducted a repeated measures ANOVA with probe background value as a within-subjects factor and the presence or not of somersaulting behaviour as a between-subjects factor. We used only the data from the first week of trials for this analysis to minimise the effects of learning observed in the second and third weeks and to avoid any potential confound from the housing manipulation. The data on somersaulting showed that only three of eight subjects demonstrated somersaulting behaviour during the period prior to the cognitive bias trials (see Brilot et al. 2009a). None of the subjects showed somersaulting behaviour during the 3-week cognitive bias trial period. The analysis showed that there was a significant effect of

the probe background value on the stimulus chosen ($F_{2,12} = 32.33$, $P < 0.001$). Post hoc analysis (with Bonferroni corrections applied) revealed that there was a significant difference in the response to the 15 vs. 45% grey backgrounds ($P = 0.002$) and the 30 vs. 45% grey backgrounds ($P = 0.001$) but there was no significant difference in the response to the 15 vs. 30% grey backgrounds ($P = 0.149$). Somersaulting behaviour had an effect on the choices made, manifested as a significant interaction between probe background value and somersaulting ($F_{2,12} = 4.40$, $P = 0.037$; Fig. 4), though there was no significant main effect of somersaulting ($F_{1,7} = 1.56$, $P = 0.259$). To establish the meaning of this interaction, we conducted repeated contrasts which revealed a significant interaction when comparing the choices made by stereotyping and non-stereotyping individuals in response to the 15% background probe vs. the 30% background probe ($F_{1,6} = 6.36$, $P = 0.045$) and the 30 vs. 45% background probe ($F_{1,6} = 11.54$, $P = 0.015$). Examination of Fig. 4 confirms this interaction: somersaulting birds were more likely to choose the stimulus associated with the lower reward value, but this difference was only expressed in response to the 30% grey background probe.

Discussion

In this paper, we set out first to develop an improved cognitive bias task for starlings, and, second, to extend

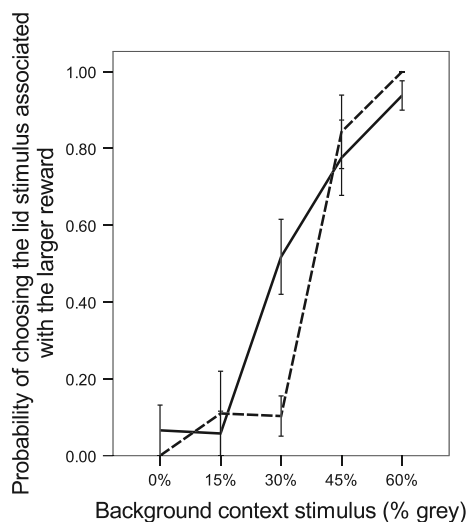


Fig. 4 Probability of choosing the stimulus associated with the larger reward for each background context stimulus in the first week of cognitive bias trials (enriched conditions). The subjects are divided into those that exhibited somersaulting behaviour at some stage during the first 6 weeks of the training period (*dashed line*) and those that did not (*solid black line*). Data points show the mean \pm one standard error

previous work in animals by asking whether cognitive biases are correlated with individual differences in the incidence of abnormal behaviour. Although we succeeded in training birds on our novel cognitive bias task, we failed to find the predicted changes in cognitive bias, expected with changes in housing conditions. However, we did find that performance on the task was predicted by individual differences in whether or not birds showed stereotypic somersaulting behaviour. In the text later, we discuss the explanations for these findings and their implications in the context of our original aims.

Cognitive bias tasks and learning

The birds' judgment biases, as measured by their choice of which lid to flip during ambiguous probe trials, were not affected in any consistent manner by our manipulation of their housing conditions (Fig. 2). The cognitive bias task therefore failed to detect any changes in affective state that might have been induced by the change in environmental conditions. However, this is unsurprising given the additional data on the increase in latencies to choose across the 3 weeks of cognitive bias testing (Fig. 3). This increase is inconsistent with a cognitive bias interpretation, and instead suggests that the birds were learning that their choice in ambiguous probe trials was never rewarded with mealworms. Indeed, by the third week of the testing, two birds completely failed to make a choice in the 30% grey background probe trials. We therefore conclude that the birds learnt quickly that the intermediate probe stimuli were never associated with reinforcement, thus rendering the probe trials unambiguous by the second week of testing, and the task ineffective for detecting changes in affective state. Ours is the first cognitive bias experiment to find evidence for such rapid learning and loss of ambiguity in probe trials raising the question of why this occurred.

The experiment presented is the only cognitive bias task, that we are aware of, that has employed this specific repeated measures methodology (i.e. from condition A to condition B and return to condition A). The rationale that this allowed each bird to be its own control was justified given the large range of inter-individual variability we found in response to the ambiguous probes in the initial stages. A between-groups design would have required greater sample sizes to detect similar effects given this noise from individual differences. However, the repeated measures design also meant that learning became a significant factor in reducing the sensitivity of the cognitive bias measure. The cognitive bias testing lasted 21 days with the birds having 21 exposures to each probe stimulus over this time. In the two previous cognitive bias experiments on starlings (Bateson and Matheson 2007; Matheson et al. 2008), the test phases lasted for 10 and 20 days

respectively and the birds had 36 and up to 80 exposures to each probe stimulus over this time, respectively. It is not possible to compare the latter study with the current study since the stimuli used were entirely different. However, the former study (Bateson and Matheson 2007) used similar stimuli and training techniques to the current experiment. In fact, the stimuli used in the current study were actually drawn from a smaller range than in Bateson and Matheson (2007) and therefore we would have predicted that if anything, the ambiguous stimuli would have been harder to distinguish from the trained S+ and S− stimuli, not easier.

In an attempt to resolve this apparent contradiction, we re-examined the data presented in Bateson and Matheson (2007) to investigate whether it could be re-interpreted as the result of learning as opposed to a change in cognitive bias. If the birds learnt that the ambiguous stimuli were never reinforced, this would have resulted in a reduced probability of lid-flipping in the second treatment received by the birds, and hence behaviour interpreted as indicating a more pessimistic cognitive bias in the second treatment. In fact, this is exactly what was observed. Figure 2 of Bateson and Matheson (2007) shows a reduced probability of lid-flipping when the birds moved from enriched to standard conditions. This was interpreted as a cognitive bias shift, since birds in a more negative affective state would be more likely to negatively interpret the stimulus and therefore avoid the lids. However, in the same figure, the birds that received the treatments in the reverse order (i.e. standard to enriched) also showed a (non-significant) reduction in lid-flipping in their second treatment. Taken together with the evidence from the current study showing the same trend, these data strongly suggest that the birds in Bateson and Matheson (2007) were learning that the ambiguous probes were never reinforced as opposed to exhibiting a change in cognitive bias.

The possibility of subjects learning rapidly that ambiguous probe trials are unreinforced is therefore a difficulty for experiments designed to detect changes in cognitive bias. The most successful cognitive bias experiments have most likely circumvented this problem by using a between-subjects design with a short duration of testing with ambiguous probes (Harding et al. 2004; Burman et al. 2008a). However, even with these designs, the possibility remains that reductions in the probability of responding or latency to respond, interpreted as more pessimistic judgment biases, could actually be attributable to effects of stress on speed of learning. Though the general validity of the Yerkes–Dodson law (that there is an inverted U-response function linking stress and learning speed) is questioned, there is confirmatory evidence linking mild levels of stress and improved memory formation (Mendl 1999). For instance, there is evidence in rats that pharmacologically induced mild stress (administration of low

doses of corticosterone) can enhance learning (Okuda et al. 2004), but only under conditions of emotional arousal (in this case response to a novel object). It is therefore a possibility that experiments aimed at assessing a cognitive bias may be confounded by an additional interaction between stress and learning (as well as stress and cognitive interpretations). In short, individuals under mildly stressful conditions may learn more quickly that ambiguous probes are unreinforced and therefore show a reduced response in both go/go and go/no-go experimental designs. A potential solution to this difficulty lies in the use of paradigms that require only a single exposure to ambiguous, unreinforced probe stimuli where learning cannot be a confound (see Brilot et al. 2009b for a first attempt at such a task).

Cognitive bias and individual behavioural differences

The results from the first week of cognitive bias testing suggest that performance is predicted by whether starlings display stereotypic behaviour in the form of somersaulting. Individuals that performed somersaults demonstrated a significantly more pessimistic interpretation of the most ambiguous (30% grey) probe stimulus than non-stereotyping individuals. Though there proved to be no relationship between responses to the 15 and 45% grey probes and somersaulting behaviour, this is unsurprising given the reduced ambiguity of these probes when compared to the 30% grey background. Figure 3 shows that these two probes were treated as approximately equivalent to the trained S+ and S− backgrounds as judged by the birds' choice responses. Any sensitivity to individual differences in response was therefore likely overshadowed by a generalised, strong conditioned response to the previously encountered stimuli.

This study examined individual differences in somersaulting behaviour and the relationship between this stereotypy and cognitive bias. Elsewhere, we have analysed data on behaviour patterns in the learning phase of the current experiment (Brilot et al. 2009a). This showed that repetitiveness of movement patterns, abnormal stereotypic behaviour (including somersaulting), and the use of abnormal perching locations are all positively correlated in a complex that is suggestive of a behavioural response to caging. Additionally, it is known that an increase in the repetitiveness of behaviour is correlated with the housing conditions of starlings (both with cage type and enrichments: Asher 2007; Asher et al. 2009). There is some evidence to suggest that this may be related to a thwarted escape response, as originally suggested by Maddocks et al. (2002). Our findings here are therefore suggestive that performance on the cognitive bias task, and by implication affective state, relates to this suite of abnormal and repetitive behaviour measures. As outlined in the introduction, it

is generally considered that the presence of stereotypic behaviour indicates poor welfare when comparing differing housing regimes. However, the evidence for animals that share the same captive conditions is equivocal, with the majority of studies suggesting that stereotyping individuals actually display indicators of better welfare than non-stereotyping individuals (Mason and Latham 2004; Mason 2006). The present study suggests that the presence of stereotypic behaviour in starlings is an indicator of poor welfare, even when comparing individuals who share the same housing conditions.

There are a number of reasons why stereotypic behaviour might be an indicator of negative affective state and therefore of poor welfare (Mason and Latham 2004). We suggest that the typical starling stereotypy, somersaulting, observed in our study, fulfils the criteria in Table 2 of Mason and Latham (2004) for a stereotypic behaviour that is an index of poor welfare (specifically an index of frustration: Table 3, Mason and Latham 2004). First, the stereotypy is not a suitable replacement for the natural activity. Since we hypothesise that the behaviour patterns and stereotypic behaviour are indicators of a thwarted escape response, there is no likelihood that they act as a suitable substitute. Second, it seems unlikely that this behaviour has a 'mantra effect', i.e. a positively reinforcing ability to reduce stress, though the present data do not allow us to exclude this possibility. Third, stereotypic behaviour in our study was embedded within a suite of flexible behaviours. The individuals that demonstrated somersaulting behaviour were still able to attend to and complete all training tasks. There was no negative relationship between stereotypic behaviour and the length of training across the subjects as might be expected if stereotyping individuals were unwilling or unable to attend to external stimuli. Fourth and finally, stereotypic behaviour seems to have been elicited 'appropriately' within the context of an escape response. Somersaulting behaviour was expressed most prominently during the first 3 weeks of captivity (Brilot et al. 2009a) and subsequently decreased. However, though somersaulting decreased over time during the experimental video recordings (when no humans were present), it was still stimulated to an extent by the presence of the experimenter during daily cognitive bias training and husbandry (personal observations). This suggests that the thwarted escape response was heightened by the presence of a perceived threat and therefore stereotypic behaviour was manifested. Given that the stereotypic somersaulting behaviour of starlings fits the criteria for a good indicator of poor welfare, we suggest that the present study indicates that starlings that display more repetitive behaviour patterns and stereotypic behaviours are also suffering from a more negative affective state (as measured by the cognitive bias task).

In conclusion, our study has revealed that rapid learning of non-reinforced ambiguous probe stimuli can be a problem in cognitive bias tasks. Subjects learning that ambiguous probe trials are never reinforced not only precluded us from detecting changes in affective state with changes in housing conditions in the current experiment, but may also have implications for other studies attempting to establish a cognitive bias where the test phase is not sufficiently short. Performance on the cognitive bias task did, however, reflect behaviour in captivity with regards to the incidence of abnormal repetitive behaviour (namely the somersaulting stereotypy). We suggest that the wider suite of behavioural traits related to repetitive behaviour is indicative of a stress response in captive starlings that also reflects a more negative affective state. The cognitive bias methodology therefore has merit in revealing individual differences in affective state.

Acknowledgments We thank Michelle Waddle for technical help. We thank Jim Clapp, Domhnall Jennings, Stephanie Matheson, Mike Mendl, Jeroen Minderman, Liz Paul and several anonymous referees for help and advice. This work was supported by two grants awarded to MB from the UK's Biotechnology and Biological Sciences Research Council (BB/E012000/1 and BB/05623/1).

References

- Asher L (2007) Welfare of captive European starlings (*Sturnus vulgaris*). Ph.D. Thesis. Institute of Neuroscience, Newcastle University, Newcastle upon Tyne
- Asher L, Davies GTO, Bertenshaw CE, Cox MAA, 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
- Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, van Ijendoorn MH (2007) Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychol Bull* 133:1–24
- Bateson M, Kacelnik A (1995) Preferences for fixed and variable food sources—variability in amount and delay. *J Exp Anal Behav* 63:313–329
- Bateson M, Matheson SM (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
- Bergvall UA, Rautio P, Luotola T, Leimar O (2007) A test of simultaneous and successive negative contrast in fallow deer foraging behaviour. *Anim Behav* 74:395–402
- Blumstein DT, Evans CS, Daniel JC (2000) *J Watcher* 0.9. An introductory user's guide
- Brilot BO, Asher L, Feenders G, Bateson M (2009a) Quantification of abnormal repetitive behavior in captive European starlings (*Sturnus vulgaris*). *Behav Process* 82:256–264
- Brilot BO, Normandale CL, Parkin A, Bateson M (2009b) Can we use starlings' aversion to eyespots as the basis for a novel 'cognitive bias' task? *Appl Anim Behav Sci* 118:182–190
- Burman OHP, Parker R, Paul ES, Mendl M (2008a) A spatial judgement task to determine background emotional state in laboratory rats, *Rattus norvegicus*. *Anim Behav* 76:801–809

- Burman OHP, Parker RMA, Paul ES, Mendl M (2008b) Sensitivity to reward loss as an indicator of animal emotion and welfare. *Biol Lett* 4:330–333
- Enkel T, Gholizadeh D, von Bohlen Und Halbach O, Sanchis-Segura C, Hurlemann R, Spanagel R, Gass P, Vollmayr B (2009) Ambiguous-cue interpretation is biased under stress- and depression-like states in rats. *Neuropsychopharmacol*. Advance online publication. doi: [10.1038/npp.2009.204](https://doi.org/10.1038/npp.2009.204)
- Eysenck MW, Mogg K, May J, Richards A, Mathews A (1991) Bias in interpretation of ambiguous sentences related to threat in anxiety. *J Abnorm Psychol* 100:144–150
- Flaherty CF (1996) *Incentive relativity*. Cambridge University Press, Cambridge
- Greenwood VJ, Smith EL, Goldsmith AR, Cuthill IC, Crisp LH, Walter-Swan MB, Bennett ATD (2004) Does the flicker frequency of fluorescent lighting affect the welfare of captive European starlings? *Appl Anim Behav Sci* 86:145–159
- Harding EJ, Paul ES, Mendl M (2004) Animal behavior: cognitive bias and affective state. *Nature* 427:312
- Maddocks SA, Goldsmith AR, Cuthill IC (2002) Behavioural and physiological effects of absence of ultraviolet wavelengths on European starlings *Sturnus vulgaris*. *J Avian Biol* 33:103–106
- Mason G (2006) Stereotypic behaviour in captive animals: fundamentals and implications for welfare and beyond. In: Mason G, Rushen J (eds) *Stereotypic animal behaviour: fundamentals and applications to welfare*, 2nd edn. CABI, Wallingford, pp 325–356
- Mason GJ, Latham NR (2004) Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? *Anim Welfare* 13:S57–S69
- Mason G, Rushen J (2006) *Stereotypic animal behaviour: fundamentals and applications to welfare*, 2nd edn. CABI, Wallingford
- Matheson SM (2007) Using cognitive bias to assess welfare in captive European starlings (*Sturnus vulgaris*). M.Sc. Thesis. Institute of Neuroscience, Newcastle University, Newcastle upon Tyne
- Matheson SM, 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
- Mendl M (1999) Performing under pressure: stress and cognitive function. *Appl Anim Behav Sci* 65:221–244
- Mendl M, Burman OHP, Parker RMA, Paul ES (2009) Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. *Appl Anim Behav Sci* 118:161–181
- Okuda S, Roozendaal B, McGaugh JL (2004) Glucocorticoid effects on object recognition memory require training-associated emotional arousal. *Proc Natl Acad Sci USA* 101:853–858
- Paul ES, Harding EJ, Mendl M (2005) Measuring emotional processes in animals: the utility of a cognitive approach. *Neurosci Biobehav Rev* 29:469–491
- Rich EL, Romero LM (2005) Exposure to chronic stress downregulates corticosterone responses to acute stressors. *Am J Physiol Regul Integr Comp Physiol* 288:R1628–R1636
- Williams JMG, Watts FN, MacLeod C, Matthews A (1997) *Cognitive psychology and emotional disorders*. Wiley, Chichester
- Young RJ (2003) *Environmental enrichment for captive animals*. Blackwell Publishing, Oxford