1

Stereotyping starlings are more 'pessimistic'

Ben O. Brilot[1], Lucy Asher[2] and Melissa Bateson[1]

[1]*Centre for Behavior and Evolution, Institute of Neuroscience, Newcastle University, UK.*

[2]*Department of Veterinary Clinical Sciences, Royal Veterinary College, Potters Bar, UK.*

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

E-mail: ben.brilot@ncl.ac.uk

Running head: Stereotypic behaviour in starlings predicts a cognitive bias

2

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 stereotypies 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.

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 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 below, 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 stereotypies, that might also reflect trait anxiety. We present the background to each of these aims in more detail below.

87 **Judgement bias tasks for animals**

88 The tasks developed so far to measure judgment biases in animals are based on

89 the original design of (Harding et al. 2004). Subjects are initially trained to

90 associate one stimulus, S+, with a reward (generally food) and another S- with

91 either a reward of lower value (e.g less food), or a punishment (e.g. white noise

92 or a noxious food item). S+ and S- are chosen to lie on a continuous stimulus

93 spectrum (e.g. a frequency range for auditory stimuli, a spectral range for visual

94 stimuli or a range of directions for spatial stimuli). To measure a cognitive bias,

95 subjects' responses to novel stimuli ('probes') that are intermediate between the

96 trained stimuli are recorded in extinction (i.e. probe trials are not reinforced,

97 avoiding any confound of reinforcement). A subject is regarded as showing a

98 pessimistic judgement bias if it demonstrates a higher probability of exhibiting

99 the response appropriate to the S- stimulus, than either that same subject in a

100 more positive affective state or when compared to control subjects in a more

101 positive state.

102 To date, the majority of published cognitive bias tasks have used a go/no-

103 go procedure (Harding et al. 2004; Bateson and Matheson 2007; Burman et al.

104 2008a). In a go/no-go task the subject is required to respond by performing a

105 behaviour (e.g. lever press) in response to S+, but to refrain from responding to

106 S-. However, interpretation of data from a go/no-go task is complicated by the

107 possibility that negative affective states are often associated with changes in

108 general activity and feeding motivation. Therefore, on a go/no-go task, it is

109 possible that subjects in a more negative affective state show a reduced

110 probability of responding because they are less motivated to exploit a signalled

111 food source, rather than because they interpret the ambiguous stimulus

112 pessimistically. Thus, go/no-go tasks could be measuring a more general

113 response bias as opposed to the assumed biased judgement of the ambiguous

114 stimulus.

115 To address the above problem with go/no-go tasks we, and others, have

116 advocated the use of choice tasks whereby subjects are required to make a

117 different active response to both S+ and S- stimuli (e.g. Matheson et al. 2008;

118 Enkel et al. 2009); see also unpublished studies cited in Mendl et al. 2009). This

119 experimental design allows the effects of a response bias and a judgement bias to

120 be distinguished: the former should result in reduced responding to all stimuli

121 whilst the latter should result in reduced responding only to ambiguous stimuli. In

122 a previous study with starlings we used an operant task in which birds were

123 required to choose (by pecking) a red or green key to classify a light stimulus as

124 S+ or S- (Matheson et al. 2008). However, this task has a number of practical

125 limitations including the length of time taken to train the birds, the requirement

126 for moderate levels of food deprivation and the requirement to catch and handle

127 the birds daily to transfer them to the operant chambers (a potentially anxiety-

128 inducing procedure; Rich and Romero 2005). In the current paper we present a

129 novel choice task that is a modification of the simple lid-flipping task described

130 earlier (Bateson and Matheson 2007). The task was designed to be quick to train,

131 and minimally disruptive to the birds, with all training and experimental

132 procedures conducted in the home cages (c.f. Matheson et al. 2008).

133 In line with previous cognitive bias experiments, we manipulated the

134 affective state of the birds by altering the level of environmental enrichment

135 provided in their cages (Bateson and Matheson 2007; Matheson et al. 2008).

136 There are extensive data from a wide range of species showing that provision of

137 more enrichment in captive animals' cages is associated with better welfare

138 (Young 2003), and we have data from our own laboratory showing that starlings

139 in enriched cages display less abnormal behaviour (Asher 2007). We used a

140 repeated measures design involving a sequential change from environmentally

141 enriched conditions, to non-enriched conditions and then returning to enriched

142 conditions. This design delivers greater statistical power in a study involving low

143 numbers of subjects (a constraint of the intensive training required), and

144 additionally allows us to examine how starlings respond to both reduction and

145 improvement in their environmental conditions. We have previously found that

146 starlings show a greater change in cognitive bias in response to a reduction in

147 environmental conditions than to an improvement (Bateson and Matheson 2007),

148 adding to many results showing that animals' responses to a given situation

149 depend on what they have previously experienced (Flaherty 1996; Bergvall et al.

150 2007; Burman et al. 2008b).

151 We hypothesised that starlings would show more pessimistic judgement

152 biases in the non-enriched conditions compared with the enriched conditions. We

153 also hypothesised that the birds would show a greater response to the removal of

154 environmental enrichment than to its reinstatement in the final stage of the

155 experiment.

156 **Individual differences**

157 The published cognitive bias studies in animals thus far have all examined

158 whether judgement biases are sensitive to relatively short-term manipulations of

159 state (Harding et al. 2004; Bateson and Matheson 2007; Burman et al. 2008a;

160 Matheson et al. 2008). However, the literature in humans suggests that there are

161 also stable individual ('trait') differences in both affect and pessimism (Bar-Haim

162 et al. 2007). How individual animals cope with captivity is not only a matter of

163 animal welfare, but is also of concern for the scientific validity of studies related

164 to cognition. Differences in coping ability might be reflected in an animal's

165 affective state and hence in the choices they make, regardless of the

166 experimental treatment. Repetitive, abnormal behaviour (of which stereotypy is a

167 type) is often regarded as an indicator of poor welfare since this behaviour can be

168 associated with physiological and behavioural measures of stress (Mason and

169 Rushen 2006). The evidence linking stereotypy and poor welfare is, however,

170 mixed: a review of 90 studies in a range of species found that comparing

171 between environments or regimes, those where the subjects stereotyped more

172 invariably also scored lower on additional welfare measures, but within a group of

173 animals under the same husbandry regime, 60% of studies showed that

174 performance of stereotypies was associated with better welfare whilst the

175 remainder showed the opposite (Mason and Latham 2004).

176 Given the above, we hypothesised that the presence of stereotyping

177 behaviour in individual starlings should reflect stable individual differences in

178 affective state (trait anxiety), and hence performance on a judgement bias task.

179 If stereotyping birds are more pessimistic then we argue that stereotypic

180 behaviour is an indicator of poor welfare within starlings sharing the same

181 environment (and vice versa). However, any difference in cognition related to

182 stereotypic behaviour is of importance for future studies using captive birds with

183 stereotypies.

184

Methods

185

Subjects and Husbandry

186

187 The subjects were eight European starlings (four males and four females) caught

188 from the wild under license from Natural England. An equal number of juveniles

189 (birds in their first year) and adults were used. Both sex and age were

190 counterbalanced for position in the experimental laboratory and time of

191 behavioural testing. Prior to the experiment the subjects were group-housed in an

192 indoor aviary (2.4 x 2.15 x 2.3 m) with wood chippings covering the floor, dead

193 trees for perching and cover, and shallow trays of water for bathing. At the start

194 of the experiment, the birds were moved into individual cages (75 cm wide x 45

195 cm deep x 44 cm high) where visual and auditory contact with at least four

196 conspecifics was possible. Previous studies in solitary-housed starlings have

197 shown that differences in cage dimensions and enrichments cause changes in

198 behaviour, condition and affective state (Bateson & Matheson 2007; Matheson et
199 al. 2008; Asher et al. 2009), hence we were confident that the stress of individual
200 housing would not cause a ceiling effect constraint in the present study. During all
201 training phases the cages were furnished with enrichments suggested to improve
202 the welfare of captive starlings, namely: natural branches, a water bath and a
203 tray of bark chippings. The light:dark cycle was maintained at 14:10 hours. At all
204 times, other than those described below, the subjects had *ad libitum* access to
205 Purina kitten food, supplemented with fruit and mealworms (*Tenebrio* larvae).
206 Drinking water was available at all times.

207   The birds were subject to the same daily routine throughout the study:
208 cage husbandry at 8:00 a.m., followed by two hours of food deprivation to
209 increase the subjects' motivation for the learning task, followed by approximately
210 one hour of experimental trials (either learning or performing the cognitive bias
211 task – see below). On completion of the trials, the birds were once more allowed
212 to feed *ad libitum*. Due to the staggering of trials (four birds were tested at a
213 time), all experimental procedures were completed by approximately midday.

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

218 **Cognitive Bias Task**

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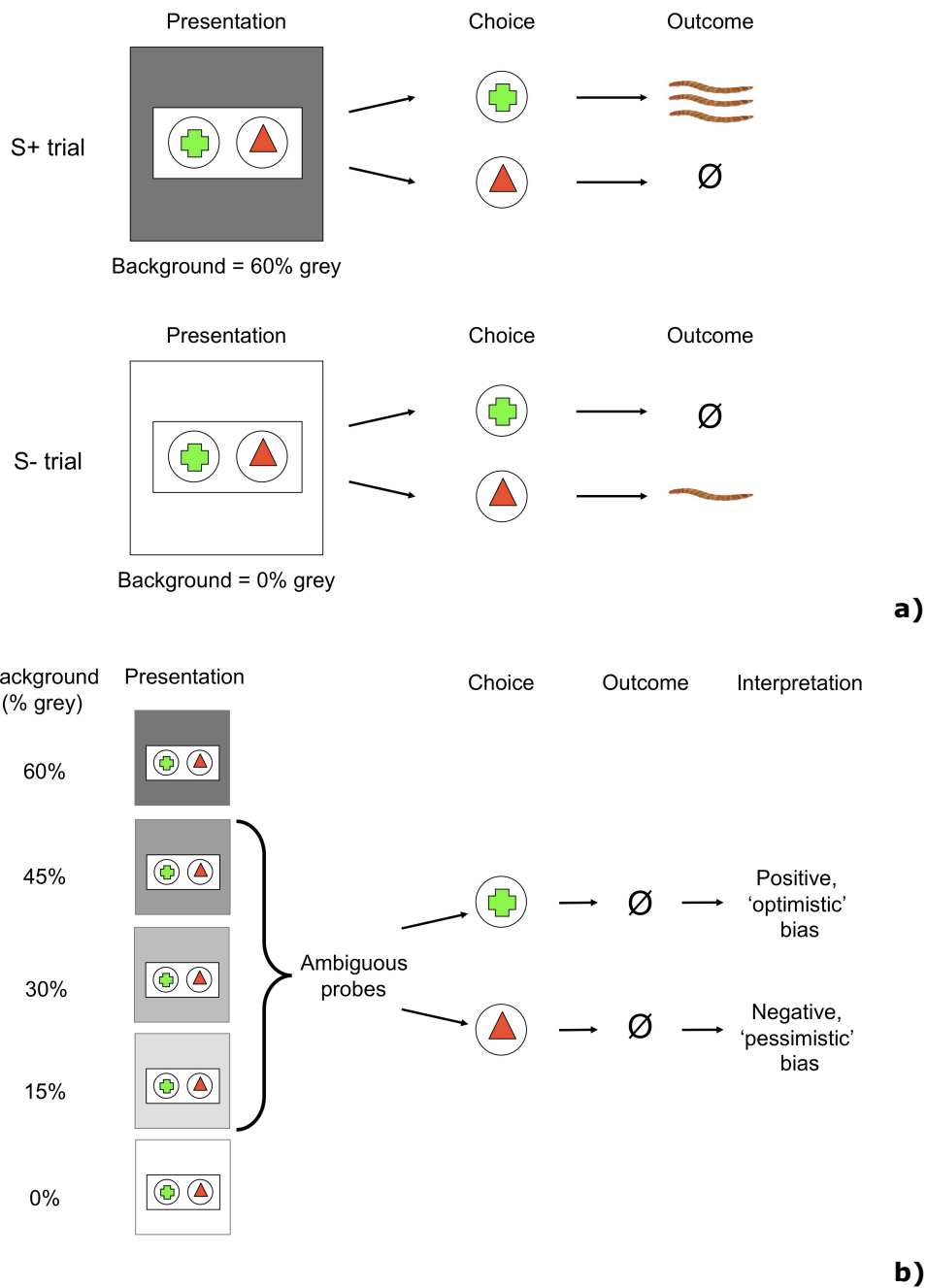232 choice task in starlings.

233

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.

234   **Apparatus.** Two opaque Petri dishes (5 cm diameter x 0.5 cm high) were
235   mounted 3 cm apart on a ceramic tile (15 cm x 15 cm). The background colour of
236   the tile was used as the discriminative stimulus (S+ or S-) and was altered by
237   affixing printed paper to the tile (Fig. 1a). S+ was printed using the settings of
238   Hue 0°; Saturation 0%; and Brightness 40% in Microsoft Powerpoint, and is
239   henceforth referred to as 60% grey. S- was printed using: Hue 0°; Saturation
240   0%; and Brightness 100%, and is henceforth referred to as 0% grey. The
241   intermediate stimuli used as 'probes' for cognitive bias assessment were printed
242   using the same Hue and Saturation values but varying Brightness to 55%, 70%
243   or 85% giving shades henceforth referred to as 45%, 30% and 15% grey
244   respectively (Fig. 1b). All subjects were trained to associate the 60% grey
245   background (i.e. S+) with a three-mealworms reward and the 0% grey
246   background (i.e. S-) with a one-mealworm reward (Matheson's (2007) data
247   showed that there was no effect of whether the higher reward occurred at the
248   dark or light end of the stimulus spectrum). Pilot experiments had established
249   that starlings in the same experimental set-up expressed a significant preference
250   for three mealworms over one mealworm, confirming our assumption that the
251   larger reward is of higher value. The Petri dishes were covered by circular
252   cardboard lids (6.5 cm diameter) printed with one of two distinct stimuli (a red
253   triangle or a green cross) that signalled which of the dishes contained the reward.
254   The stimuli on the lids were also replicated on paper circles that were glued to the
255   inside bottom of the Petri dishes (such that they were visible below the reward
256   once the lid had been removed). Pilot experiments had also established that
257   these stimuli were easily discriminable to the birds. Half of the birds were trained
258   to associate the red triangle with the 3-mealworm reward and the green cross
259   with the 1-mealworm reward whilst the other half were trained to the reverse
260   assignment.
261   Training the birds for the cognitive bias task involved three phases:
262   learning to flip lids, no-choice learning of the S+/S- conditional discrimination and
263   free-choice learning of the full task with no ambiguous intermediates.
264

265   **Training lid-flipping**. A tile with a single, centrally placed Petri dish and a plain
266   yellow cardboard lid were used in the initial lid-flipping training. The birds were
267   rewarded with a mealworm placed on the yellow lid that was in turn placed on the
268   tile. To facilitate learning of the lid-flipping task, the lid was then moved so as to
269   cover more of the Petri dish whilst the mealworm was moved to within the dish
270   (though still visible). In order, the lid was placed: on the ceramic tile, next to the
271   dish which contained the mealworm; leaning against the dish but not obscuring

272 the contents at all; progressively obscuring more of the contents (i.e. ¼ on, ½

273 on, ¾ on) until it covered the entire dish and the contents could only be seen and

274 recovered by moving the lid aside. Sixteen trials were given each day, each

275 lasting for 60 s with a 180-s inter-trial interval (ITI). The bird was required to eat

276 the mealworm within 60 s or the previous training stage was repeated. The bird

277 was considered to have learnt to lid-flip once it had completed 16 consecutive

278 trials with the lid fully covering the dish. This training phase continued until all

279 birds had learnt the task.

280

281 **Training the conditional discrimination**. Next, birds were given a no-choice

282 task using the two backgrounds (S+ and S-) and two lid-types that would be

283 used in the final experiment (red triangle and green cross). They were presented

284 with one open dish and one covered dish. The covered petri dish was always

285 represented by the correct choice given the background context (i.e. if a 60%

286 grey background was shown then only the correct lid would be present on a Petri

287 dish and this would contain the 3-mealworms reward). Trials were separated into

288 blocks such that each bird had six presentations of the 60% grey background in a

289 row followed by six presentations of the 0% grey background with the order of

290 the blocks alternating between days. Each trial lasted 60 s with a 180-s ITI. Upon

291 completion of the 12 trials, the birds were then given six probe trials. In these the

292 birds were presented with a choice of both lids; beneath the correct lid (given the

293 background context) was the appropriate reward. Half of the probe trials used the

294 60% grey background and half used the 0% grey with the order pseudo-

295 randomised. The birds were considered to have learnt the discrimination when

296 they were above significance on the binomial test over the course of three days of

297 testing (14 correct choices out of 18). This phase continued for all birds until the

298 last subject had learnt the discrimination.

299 **Training partial reinforcement**. Next, free-choice trials were given, and the

300 probability of reward was gradually reduced using randomly interspersed

301 unrewarded trials. Fifteen trials were conducted per day in this phase (again with

302 a trial duration of 60 s and ITI of 180 s). Of these trials, all were rewarded on the

303 first day, only 12 were rewarded on the second and third days, and only 10 were

304 rewarded on the fourth day. The last day corresponded exactly to the trials that

305 would be conducted as part of the cognitive bias task: 15 trials, five of which

306 would not be reinforced. Reduction in the reinforcement probability was intended

307 to prolong the conditioned response (CR) during the cognitive bias trials when

308 ambiguous probe trials would be unrewarded and hence would cause the CR to

309 extinguish.

310

311 **Cognitive bias trials**. The experimental cognitive bias test involved one session

312 of 15 trials per day. Of these, five trials were reinforced presentations of the 60%

313 grey background, five trials were reinforced presentations of the 0% grey

314 background, two trials were unreinforced presentations of the 60% and 0% grey

315 backgrounds and the remaining three were unreinforced presentations of each of

316 the intermediate, ambiguous backgrounds (15, 30 and 45% grey). The order of

317 presentation was pseudo-random, although we avoided contiguous unrewarded

318 trials. As in the training trials, if a choice was not expressed within 60 s then the

319 trial was terminated and the usual ITI was observed.

320 In each phase of the experiment the choice made on every trial was

321 recorded for each subject. A choice was either recorded as appropriate for S+

322 (indicative of an optimistic bias) or appropriate for S- (indicative of a pessimistic

323 bias). In addition, the time taken from presentation of the tile until a choice was

324 expressed was also recorded (defined as when the lid was moved such that the

325 bird could observe the reward). Latency has been successfully used as a response

326 variable in previous cognitive bias studies with rats (Burman et al. 2008a), and

327 typically correlates well with choice in previous studies with starlings (e.g.

328 Bateson and Kacelnik 1995).

329 **Housing manipulations**

330 The cognitive bias trials were run daily over the course of three weeks whilst

331 environmental enrichment in the cages was varied each week in a repeated

332 measures design. For the first and third weeks the birds were in environmentally

333 enriched conditions similar to those from prior cognitive bias experiments in

334 starlings (natural wood branches; water for bathing; and a tray filled with bark

335 for natural probing opportunities: Bateson and Matheson 2007). For the second

336 week these enrichments were removed (non-enriched conditions) and the birds

337 were left with the empty water and bark containers and uniform dowel perches.

338 In order to standardise and minimise the disruption caused by the experimenter

339 physically changing the housing conditions, the birds were caught and transferred

340 to new, appropriately furnished (enriched/unenriched) cages. This was done on

341 the day before each week of cognitive bias trials began (i.e. they were transferred

342 three times at weekly intervals).

343 **Scoring stereotypic behaviour**

344 The most easily quantifiable stereotypy in caged starlings is a complete

345 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 six 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, 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 three 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 out of 12 correct) for the subset of trials with the trained backgrounds (0% and 60% grey). Excluded data comprised 8 out of 56 bird days (4 of which were for one subject) for week 1; 9 out 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 one 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±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±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 three 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 three 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 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).
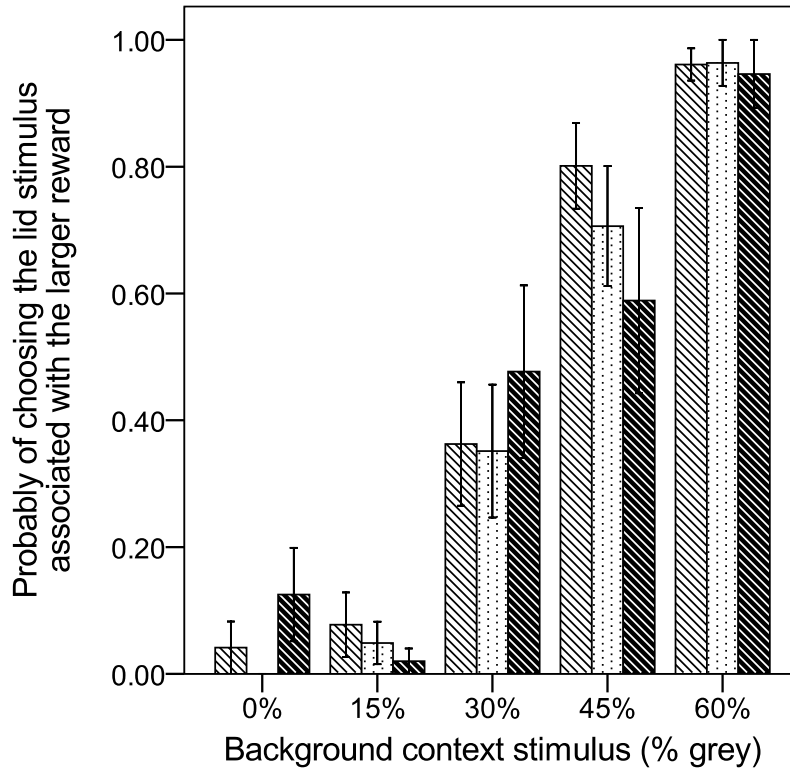
Fig. 2 Probability of choosing the stimulus associated with the higher reward during cognitive bias trials averaged across all subjects. Percentage grey values signify which background context was presented. Light hatched bars represent choices during week 1 (enriched conditions); dotted bars represent choices during week 2 (unenriched conditions); dark hatched bars represent choices during week 3 (enriched conditions). Bars show the mean for the 8 birds±one standard error.

416

417 **Latency to choose**. Although the choice data showed no effect of our housing

418 manipulation, there remained the possibility that the birds' expectancy of reward

419 size in the ambiguous probe trials was reflected in their latencies to respond.

420 Where a bird failed to exhibit a choice within the time allowed, it was allocated

421 the maximum trial duration of 60 s. We calculated the latency to flip a lid for each

422 probe stimulus relative to the 3-mealworm reinforced stimulus (60% grey; Fig.

423 3). We predicted an increase in latency in non-enriched conditions (reflecting

424 pessimism regarding the expected reward), followed by a decrease in latency on

425 return to enriched conditions (reflecting recovered optimism). However,

426 inspection of Fig. 3 suggests that if anything, latencies in the ambiguous probe

427 trials increased across the three weeks of testing. A repeated-measures ANOVA

428 (with week number and probe background as within-subjects fixed factors)

429 showed that latencies differed significantly across the probe background value

430 and across the three weeks of trials (Probe value: $F_{2,14} = 10.22$ $p = 0.002$; Week

431 number: $F_{2,14} = 7.92$ $p = 0.005$). Post-hoc analysis using t-tests (with a

432 Bonferroni correction applied) revealed a significant difference in the latency to

433 respond to the 30% and 45% grey background probes ($p = 0.001$), but all other

434 pairwise comparisons were non-significant ($p > 0.18$ for all). Similarly,

435 Bonferroni-corrected post-hoc analysis revealed a significant difference in the

436 latency to respond for weeks 1 and 3 ($p = 0.039$) but all other pairwise

437 comparisons between weeks were non-significant ($p > 0.15$ for all). There was no

438 significant interaction effect of the probe background value and week number on

439 the latency to choose (Mauchly's test revealed that the assumption of sphericity

440 was not tenable ($\chi_9 = 32.48$, $p < 0.001$), therefore the Greenhouse-Geisser

441 correction was applied: $F_{1.94, 13.55} = 0.64$, $p = 0.537$).

442

443 **Cognitive Bias Task and Individual Behavioural Differences**

444 To ascertain whether presence of stereotypic behaviour (as an indicator of

445 affective state) predicts the probability of choosing the stimulus associated with

446 the larger reward we conducted a repeated-measures ANOVA with probe

447 background value as a within-subjects factor and the presence or not of

448 somersaulting behaviour as a between-subjects factor. We used only the data

449 from the first week of trials for this analysis to minimise the effects of learning

450 observed in the second and third weeks and to avoid any potential confound from

451 the housing manipulation. The data on somersaulting showed that only three out

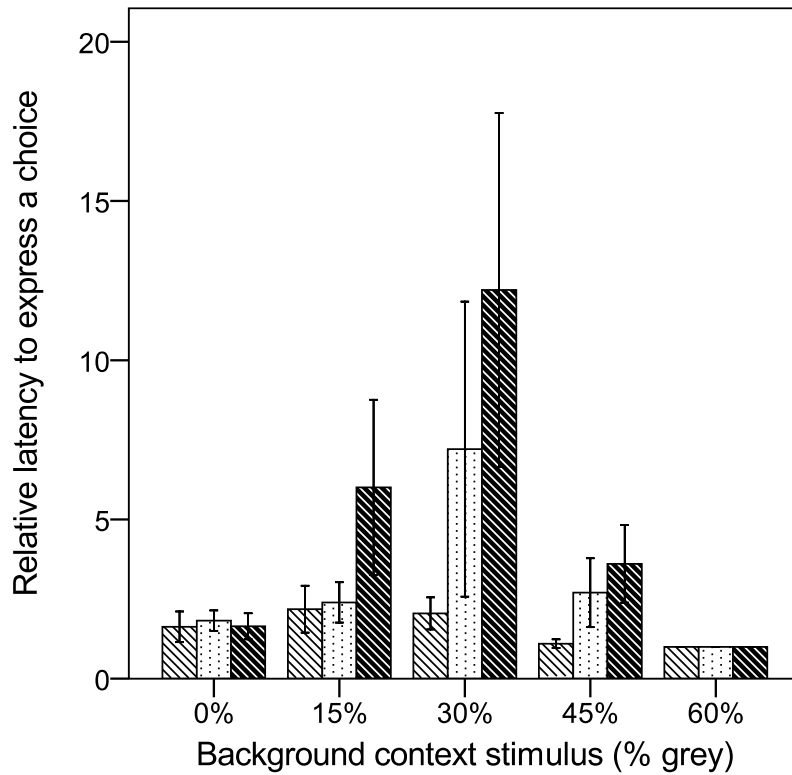452 of eight subjects demonstrated somersaulting behaviour during the period prior

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±one standard error.

14

453   to the cognitive bias trials (see Brilot et al. 2009a). None of the subjects showed

454   somersaulting behaviour during the three-week cognitive bias trial period. The

455   analysis showed that there was a significant effect of the probe background value

456   on the stimulus chosen ($F_{2,12} = 32.33$, $p < 0.001$). Post-hoc analysis (with

457   Bonferroni corrections applied) revealed that there was a significant difference in

458   the response to the 15% vs. 45% grey backgrounds ($p = 0.002$) and the 30% vs.

459   45% grey backgrounds ($p = 0.001$) but there was no significant difference in the

460   response to the 15% vs. 30% grey backgrounds ($p = 0.149$). Somersaulting

461   behaviour had an effect on the choices made, manifested as a significant

462   interaction between probe background value and somersaulting ($F_{2,12} = 4.40$, $p =$

463   0.037; Fig. 4), though there was no significant main effect of somersaulting ($F_{1,7}$

464    $= 1.56$, $p = 0.259$). To establish the meaning of this interaction, we conducted

465   repeated contrasts which revealed a significant interaction when comparing the

466   choices made by stereotyping and non-stereotyping individuals in response to the

467   15% background probe vs. the 30% background probe ($F_{1,6} = 6.36$, $p = 0.045$)

468   and the 30% vs. 45% background probe ($F_{1,6} = 11.54$, $p = 0.015$). Examination

469   of figure 4 confirms this interaction: somersaulting birds were more likely to

470   choose the stimulus associated with the lower reward value, but this difference

471   was only expressed in response to the 30% grey background probe.

472

473

474

475   **Discussion**

476   In this paper we set out first, to develop an improved cognitive bias task for

477   starlings, and, second, to extend previous work in animals by asking whether

478   cognitive biases are correlated with individual differences in the incidence of

479   abnormal behaviour. Although we succeeded in training birds on our novel

480   cognitive bias task, we failed to find the predicted changes in cognitive bias,

481   expected with changes in housing conditions. However, we did find that

482   performance on the task was predicted by individual differences in whether or not

483   birds showed stereotypic somersaulting behaviour. Below we discuss the

484   explanations for these findings and their implications in the context of our original
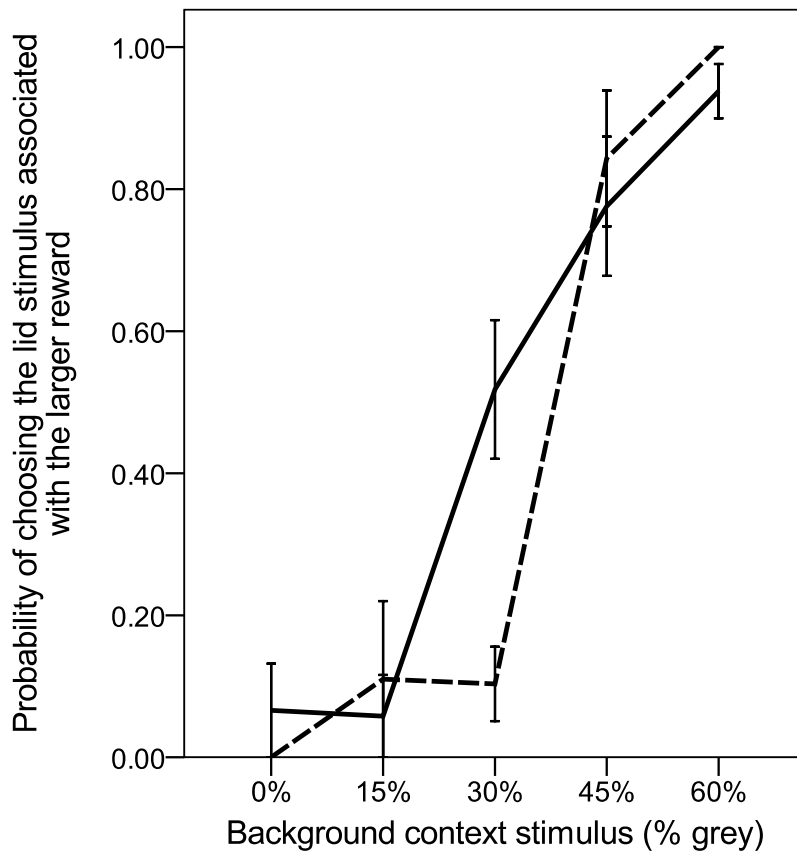
485   aims.

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 six weeks of the training period (dashed line) and those that did not (solid black line). Data points show the mean for the 8 birds±one standard error.

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 three 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.

524   In an attempt to resolve this apparent contradiction we re-examined the
525   data presented in Bateson and Matheson (2007) to investigate whether it could
526   be re-interpreted as the result of learning as opposed to a change in cognitive
527   bias. If the birds learnt that the ambiguous stimuli were never reinforced, this
528   would have resulted in a reduced probability of lid flipping in the second
529   treatment received by the birds, and hence behaviour interpreted as indicating a
530   more pessimistic cognitive bias in the second treatment. In fact, this is exactly
531   what was observed. Fig. 2 of Bateson and Matheson (2007) shows a reduced
532   probability of lid-flipping when the birds moved from enriched to standard
533   conditions. This was interpreted as a cognitive bias shift, since birds in a more
534   negative affective state would be more likely to negatively interpret the stimulus
535   and therefore avoid the lids. However, in the same figure, the birds that received
536   the treatments in the reverse order (i.e. standard to enriched) also showed a
537   (non-significant) reduction in lid-flipping in their second treatment. Taken
538   together with the evidence from the current study showing the same trend, these
539   data strongly suggest that the birds in Bateson and Matheson (2007) were
540   learning that the ambiguous probes were never reinforced as opposed to
541   exhibiting a change in cognitive bias.
542   The possibility of subjects learning rapidly that ambiguous probe trials are
543   unreinforced is therefore a difficulty for experiments designed to detect changes
544   in cognitive bias. The most successful cognitive bias experiments have most likely
545   circumvented this problem by using a between-subjects design with a short
546   duration of testing with ambiguous probes (Harding et al. 2004; Burman et al.
547   2008a). However, even with these designs, the possibility remains that
548   reductions in the probability of responding or latency to respond, interpreted as
549   more pessimistic judgment biases, could actually be attributable to effects of
550   stress on speed of learning. Though the general validity of the Yerkes-Dodson law
551   (that there is an inverted U-response function linking stress and learning speed)
552   is questioned, there is confirmatory evidence linking mild levels of stress and
553   improved memory formation (Mendl 1999). For instance, there is evidence in rats
554   that pharmacologically-induced mild stress (administration of low doses of
555   corticosterone) can enhance learning (Okuda et al. 2004), but only under
556   conditions of emotional arousal (in this case response to a novel object). It is
557   therefore a possibility that experiments aimed at assessing a cognitive bias may
558   be confounded by an additional interaction between stress and learning (as well
559   as stress and cognitive interpretations). In short, individuals under mildly
560   stressful conditions may learn more quickly that ambiguous probes are
561   unreinforced and therefore show a reduced response in both go/go and go/no-go

562 experimental designs. A potential solution to this difficulty lies in the use of

563 paradigms that require only a single exposure to ambiguous, unreinforced probe

564 stimuli where learning cannot be a confound (see Brilot et al. 2009b for a first

565 attempt at such a task).

566 **Cognitive bias and individual behavioural differences**

567 The results from the first week of cognitive bias testing suggest that performance

568 is predicted by whether starlings display stereotypic behaviour in the form of

569 somersaulting. Individuals that performed somersaults demonstrated a

570 significantly more pessimistic interpretation of the most ambiguous (30% grey)

571 probe stimulus than non-stereotyping individuals. Though there proved to be no

572 relationship between responses to the 15% and 45% grey probes and

573 somersaulting behaviour, this is unsurprising given the reduced ambiguity of

574 these probes as compared to the 30% grey background. Fig. 3 shows that these

575 two probes were treated as approximately equivalent to the trained S+ and S-

576 backgrounds as judged by the birds' choice responses. Any sensitivity to

577 individual differences in response was therefore likely overshadowed by a

578 generalised, strong conditioned response to the previously-encountered stimuli.

579 This study examined individual differences in somersaulting behaviour and

580 the relationship between this stereotypy and cognitive bias. Elsewhere we have

581 analysed data on behaviour patterns in the learning phase of the current

582 experiment (Brilot et al. 2009a). This showed that repetitiveness of movement

583 patterns, abnormal stereotypic behaviour (including somersaulting), and the use

584 of abnormal perching locations are all positively correlated in a complex that is

585 suggestive of a behavioural response to caging. Additionally, it is known that an

586 increase in the repetitiveness of behaviour is correlated with the housing

587 conditions of starlings (both with cage type and enrichments: Asher 2007; Asher

588 et al. 2009). There is some evidence to suggest that this may be related to a

589 thwarted escape response, as originally suggested by Maddocks et al. (2002).

590 Our findings here are therefore suggestive that performance on the cognitive bias

591 task, and by implication affective state, relates to this suite of abnormal and

592 repetitive behaviour measures. As outlined in the introduction, it is generally

593 considered that the presence of stereotypic behaviour indicates poor welfare

594 when comparing differing housing regimes. However, the evidence for animals

595 that share the same captive conditions is equivocal, with the majority of studies

596 suggesting that stereotyping individuals actually display indicators of better

597 welfare than non-stereotyping individuals (Mason and Latham 2004; Mason

598 2006). The present study suggests that the presence of stereotypic behaviour in

599   starlings is an indicator of poor welfare, even when comparing individuals who
600   share the same housing conditions.

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

632   In conclusion, our study has revealed that rapid learning of non-reinforced
633   ambiguous probe stimuli can be a problem in cognitive bias tasks. Subjects
634   learning that ambiguous probe trials are never reinforced not only precluded us
635   from detecting changes in affective state with changes in housing conditions in
636   the current experiment, but may also have implications for other studies

637 attempting to establish a cognitive bias where the test phase is not sufficiently

638 short. Performance on the cognitive bias task did, however, reflect behaviour in

639 captivity with regards to the incidence of abnormal repetitive behaviour (namely

640 the somersaulting stereotypy). We suggest that the wider suite of behavioural

641 traits related to repetitive behaviour is indicative of a stress response in captive

642 starlings that also reflects a more negative affective state. The cognitive bias

643 methodology therefore has merit in revealing individual differences in affective

644 state.

645

Acknowledgements

652

653

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. Psychological Bulletin 133:1-24

Bateson M, Kacelnik A (1995) Preferences for fixed and variables 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) JWatcher 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

Eysenck MW, Mogg K, May J, Richards A, Mathews A (1991) Bias in interpretation of ambiguous sentences related to threat in anxiety. Journal of Abnormal Psychology 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 427:312-312

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

702   animal behaviour: fundamentals and applications to welfare, Second edition
703   edn. CABI, Wallingford, pp 325-356
704 Mason G, Rushen J (2006) Stereotypic animal behaviour: fundamentals and
705   applications to welfare. In, Second edition edn. CABI, Wallingford
706 Mason GJ, Latham NR (2004) Can't stop, won't stop: is stereotypy a reliable animal
707   welfare indicator? Anim Welfare 13:S57-69
708 Matheson SM (2007) Using cognitive bias to assess welfare in captive European
709   starlings (*Sturnus vulgaris*). M.Sc. Thesis. Institute of Neuroscience,
710   Newcastle University, Newcastle upon Tyne
711 Matheson SM, Asher L, Bateson M (2008) Larger, enriched cages are associated with
712   'optimistic' response biases in captive European starlings (*Sturnus vulgaris*).
713   Appl Anim Behav Sci 109:374-383
714 Mendl M (1999) Performing under pressure: stress and cognitive function. Appl Anim
715   Behav Sci 65:221-244
716 Mendl M, Burman OHP, Parker RMA, Paul ES (2009) Cognitive bias as an indicator
717   of animal emotion and welfare: Emerging evidence and underlying
718   mechanisms. Appl Anim Behav Sci 118:161-181
719 Okuda S, Roozendaal B, McGaugh JL (2004) Glucocorticoid effects on object
720   recognition memory require training-associated emotional arousal.
721   Proceedings of the National Academy of Sciences of the United States of
722   America 101:853-858
723 Paul ES, Harding EJ, Mendl M (2005) Measuring emotional processes in animals: the
724   utility of a cognitive approach. Neuroscience and Biobehavioral Reviews
725   29:469-491
726 Rich EL, Romero LM (2005) Exposure to chronic stress downregulates corticosterone
727   responses to acute stressors. American Journal of Physiology-Regulatory
728   Integrative and Comparative Physiology 288:R1628-R1636
729 Williams JMG, Watts FN, MacLeod C, Matthews A (1997) Cognitive Psychology
730   and Emotional Disorders. John Wiley & Sons, Chichester
731 Young RJ (2003) Environmental enrichment for captive animals. Blackwell
732   Publishing, Oxford
733
734
735

22

Figure captions

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.

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±one standard error.

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±one standard error.

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 six weeks of the training period (dashed line) and those that did not (solid black line). Data points show the mean for the 8 birds±one standard error.