



HAL
open science

Does song repertoire size in Common Blackbirds play a role in an intra-sexual context?

Nana Hesler, Roger Mundry, Torben Dabelsteen

► To cite this version:

Nana Hesler, Roger Mundry, Torben Dabelsteen. Does song repertoire size in Common Blackbirds play a role in an intra-sexual context?. *Journal für Ornithologie = Journal of Ornithology*, 2010, 152 (3), pp.591-601. 10.1007/s10336-010-0618-5. hal-00647967

HAL Id: hal-00647967

<https://hal.science/hal-00647967>

Submitted on 4 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Does song repertoire size in common blackbirds play a role in an intra-sexual context?

2

3 Nana Hesler¹, Roger Mundry², Torben Dabelsteen¹

4

5 ¹ University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark

6 ² Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig,

7 Germany

8

9 **Abstract**

10 Bird song is thought to have a function in both inter- and intra-sexual contexts with song
11 complexity serving as an honest signal of male quality. Theory predicts that males use
12 repertoire sizes to estimate rivals' fighting ability. Here we tested whether element
13 repertoire size plays a role in an intra-sexual context in the common blackbird *Turdus*
14 *merula*, a songbird with a large repertoire. In a territory intrusion playback experiment
15 we broadcast common blackbird song with different element repertoire sizes to territorial
16 males. The test birds did not respond differently to the different repertoire sizes. This was
17 also the case when we considered the test birds' own repertoire sizes in the analysis.
18 Thus, we found no evidence for the hypothesis that males use repertoire size for
19 estimating rivals' fighting force. However, this does not exclude the possibility that
20 repertoire size plays a role in different intra-sexual contexts in common blackbirds. In the
21 discussion, we address alternative scenarios.

22

23

24 **Zusammenfassung**

25 **Spielt die Größe des Gesangsrepertoires bei Amseln eine Rolle in einem intra-**
26 **sexuellen Kontext?**

27 Vogelgesang spielt sowohl in inter- als auch intra-sexuellen Kontexten eine Rolle, wobei
28 angenommen wird, dass die Gesangskomplexität die Qualität des singenden Männchens
29 anzeigt. Männchen könnten die Konkurrenzstärke von Rivalen an der Größe ihres
30 Gesangsrepertoires abschätzen. Wir testeten hier an der Amsel *Turdus merula*, einem
31 Singvogel mit einem großen Repertoire, ob die Größe des Elementrepertoires eine Rolle
32 in einem intra-sexuellen Kontext spielt. In einem *Playback*-Experiment mit territorialen
33 Männchen spielten wir Amselgesang mit verschiedenen Elementrepertoiregrößen ab. Die
34 Testvögel reagierten nicht unterschiedlich auf die verschiedenen Repertoiregrößen, was
35 auch dann der Fall war, wenn wir die Repertoiregröße der Testtiere selber
36 berücksichtigten. Folglich fanden wir keine Bestätigung der Hypothese, dass Männchen
37 gesangliche Repertoiregrößen benutzen, um die Konkurrenzstärke von Rivalen
38 einzuschätzen. Dies schließt jedoch nicht die Möglichkeit aus, dass Repertoiregrößen bei
39 Amseln in anderen intra-sexuellen Kontexten eine Rolle spielen. Wir diskutieren deshalb
40 verschiedene alternative Szenarien.

41

42

43 **Introduction**

44 In most oscine bird species, male song is composed of different song types and/or song
45 elements. These repertoires are generally seen as the outcome of sexual selection driven
46 by female choice and male-male competition, with repertoire size serving as a signal of
47 male quality (review e.g. in Collins 2004, Catchpole & Slater 2008). Even though this
48 hypothesis is widely accepted and supported by several studies, the meaning of repertoire
49 sizes in different contexts remains unclear.

50 In classic speaker replacement experiments, territorial males were removed and
51 speakers playing back conspecific song were installed in their territories. In both great tits
52 (*Parus major*, Krebs et al. 1978) and red-winged blackbirds (*Agelaius phoeniceus*,
53 Yasukawa, 1981), playback containing a repertoire of different song types was more
54 efficient in keeping out intruders than playback comprised of a single song type.
55 However, a common problem of these studies is the fact that the single song type
56 playbacks contained single songs from one individual while the repertoire playbacks
57 consisted of several song types from several different individuals. Assuming that
58 individuality may be encoded in song (Weary & Krebs 1992, but see also Gentner et al.
59 2000) and can be decoded by a receiver, the test birds of these experiments may have
60 perceived the repertoire playback as song originating from multiple birds rather than one
61 bird with a large repertoire. As a quality signal, repertoire size could be used by males to
62 estimate other males' fighting ability in actual encounters. Only a few studies have tested
63 the direct response to different repertoire sizes, leading to ambiguous results. In a two
64 speaker experiment, male territorial field sparrows *Spizella pusilla* approached closer to
65 the speaker playing larger repertoires (Searcy 1983), and male song sparrows *Melospiza*

66 *melodia* reacted more persistently to playback of larger repertoires (Stoddard et al. 1988).
67 However, in other experiments – even one conducted with song sparrows – the reaction
68 of male birds did not vary in accordance with different repertoire sizes (song sparrow,
69 Searcy 1983; sedge warbler *Acrocephalus schoenobaenus*, Catchpole 1989). In species
70 with discontinuous song (i.e. song which is structured into discrete songs and pauses),
71 repertoire size is usually measured as the number of different song types (a song type
72 always contains the same element types, which are the shortest song unit). In the
73 chaffinch *Fringilla coelebs*, a species with a small song type repertoire, within-song
74 complexity is meaningful to both females and males (Leitão et al. 2006). Also, Balsby &
75 Dabelsteen (2001) found that male whitethroats *Sylvia communis* did not respond to
76 differences in repertoire size, but song length, and suggested that in this species the
77 evolution of repertoire size is not driven by male competition.

78 These inconsistent results may have several reasons. First, a territory owner's
79 response towards an intruder may not only depend on the intruder's fighting ability, but
80 also on the territory owner's fighting ability relative to that of the intruder (Searcy &
81 Beecher 2009). Therefore a bird with a small repertoire may react highly aggressively
82 towards a small repertoire intruder, but choose a different strategy when a large repertoire
83 male is intruding, who would probably be a superior rival in a fight. A solution to this
84 problem is to take the test birds' own repertoire sizes into consideration when evaluating
85 their response to playbacks of different repertoire sizes, as was by done Balsby &
86 Dabelsteen (2001). Second, different songbird species may use a diversity of singing
87 styles, ranging from continuous to discontinuous singing, with repertoire sizes at different
88 levels of song organisation (song vs. element). To date, most studies on song complexity

89 concern species with either small song type repertoires or no song types at all [for
90 example: sedge warbler 30-55 syllables, great reed warbler *Acrocephalus arundinaceus*
91 10-20 syllables (Catchpole 1980); great tit 2-5 song types, red-winged blackbird 2-6 song
92 types, song sparrow 5-11 song types (Beecher et al. 2000); but also common nightingale
93 *Luscinia megarhynchos* 150-230 song types (Kipper et al. 2004)]. However, it is possible
94 that a more clear effect of repertoire size can be seen in species with large repertoires.

95 Therefore, we chose to investigate the common blackbird *Turdus merula* (from
96 now on called ‘blackbird’), which has a large repertoire and a complex singing style
97 without fixed song types. Full song is organised into single songs (Strophen) of about 2
98 sec, separated by silent intervals of ca. 3 sec. A typical blackbird’s full song starts with a
99 long-ranging whistle part consisting of 3-5 low-frequency elements (1.5-3 kHz) and ends
100 with a twitter part comprising a more variable number of elements with a broader
101 frequency range and on average higher frequency (2-ca.10 kHz) and a considerably lower
102 sound pressure level (SPL; Todt 1981; Dabelsteen 1984; Fig. 1). Entire songs are only
103 occasionally repeated, but there are ‘motifs’, that is, fixed sequences of elements, which
104 are always sung in the same order (Dabelsteen 1984, 2005; Rasmussen & Dabelsteen
105 2002). These motifs are combined in a flexible way within songs, and the fact that they
106 repeatedly occur in different songs of a male gives them some resemblance with the song
107 types of other species. Thus, in terms of song organisation, blackbirds are between a
108 discontinuous singing style with song type repertoires and a continuous singing style
109 without any song types. Previous studies have shown that there are large inter-individual
110 differences in motif repertoire size and that song is used in both intra- and inter-sexual
111 contexts (Rasmussen & Dabelsteen 2002). Furthermore, in another study we found

112 correlations between body size and repertoire size in the blackbird, indicating that
113 repertoire size is a potential signal of male quality in this species (Hesler 2010).

114 In this study we investigate the importance of repertoire sizes in the blackbird in
115 an intra-sexual context. In a playback experiment, we broadcast song with different
116 element repertoire sizes (small versus large) to territorial males, and compared their
117 responses, taking the test birds' own repertoire sizes into account. If males use repertoire
118 size to estimate a rival's fighting ability, we expect test subjects to react differently to the
119 different repertoire sizes, presumably with the strength of reaction increasing as the
120 repertoire size broadcast increases. Furthermore, we predicted that the response may
121 depend on the bird's own repertoire size, i.e. its own quality (Balsby & Dabelsteen 2001,
122 Searcy & Beecher 2009). For example, a bird with a large repertoire size might react
123 more strongly to playback of large repertoires, because they expect a stronger opponent.
124 On the other hand, birds with small repertoires may react less strongly to playback of
125 large repertoires, as they would expect an opponent who is stronger than themselves.

126

127 Methods

128 Study population

129 The study was conducted on the island of Helgoland, Germany. The blackbird population
130 there consists of ca. 80 breeding pairs, most of whom breed in the residential areas of the
131 island and are thus very well habituated to humans. Most of the population was colour-
132 ringed with permission given by the Vogelwarte Helgoland (Sacher et al. 2006).

133

134 Song recordings and repertoire analyses

135 All recordings used were taken from colour-ringed males during dawn chorus ca. 1hr
136 before sunrise (03:40 – 06:30) in the early breeding seasons (8. April – 10. June) of 2006
137 – 2008. All recorded males were in their third calendar year or older. These birds were
138 distinguishable from second-calendar year males by a moult border in the greater coverts
139 (Svensson 1992).

140 We used a directional microphone (Sennheiser K6/ME67) and a Marantz PMD
141 670 solid state recorder with a sample rate of 22050 Hz (2006) or 44100 Hz (2007, 2008)
142 to record songs. Recordings in 44100 Hz were resampled to 22050 Hz. Avisoft-SASLab
143 Pro 4.4 – 4.52 was used for all sound analyses. Recordings were filtered (butterworth,
144 high pass 1.5 kHz, low pass 10 kHz) and cut into single songs. Printed spectrograms were
145 used for repertoire analyses (settings: FFT 512, window Hamming, bandwidth 56 Hz,
146 resolution 43 Hz, frame size 100%, overlap 75%, resolution 5.8 ms; graphical scale on
147 prints: 1 sec = 36 mm, 1 kHz = 6 mm). All repertoire analyses were carried out by the
148 same person by visual inspection of the spectrograms (N.H.). We classified all elements
149 in the recordings according to their overall appearance on the spectrogram (frequency and
150 temporal characteristics). An element was defined as a single continuous trace in time
151 without visible sound free pauses, i.e. pauses longer than ca. 0.015 sec (Fig. 1). Intra-
152 individual pause durations between two element types can be highly variable. If elements
153 appeared at least once with a distinguishable pause, they were always considered as two
154 elements, even if they sometimes appeared as one continuous element. Repertoire sizes
155 were determined by counting the number of different elements in 100 consecutive songs,
156 including both whistle and twitter elements. We reasoned that it should be possible to

157 adequately assess inter-individual differences on the basis of 100 songs, even if this does
158 not always reflect a blackbird's full repertoire (Fig. 2).

159

160 Construction of test files for playback

161 We constructed sound files for playback with large and small element repertoire sizes. In
162 order to avoid confounding influence by song parameters other than that we were
163 interested in (repertoire size), we did not use song from birds with naturally small and
164 large repertoires. Instead, we created large and small repertoires using recordings from
165 the same bird. It is possible that identity of the singer is encoded in voice characteristics
166 in the song and can be identified by a listening bird (Weary & Krebs 1992), which could
167 influence the response to the playback. Therefore we decided to use playbacks from two
168 different stimulus birds for each test bird. We ran the experiment with 16 test birds using
169 recordings from 8 stimulus birds which meant that the songs from each stimulus bird
170 were used four times (twice as small and twice as large repertoire playback). For
171 instance, test birds 1 and 9 would receive playback of a large repertoire from stimulus
172 bird A and small repertoire from stimulus bird B whereas test birds 2 and 10 would
173 receive large repertoire from bird B and small from bird A and so on. In order to control
174 for pseudoreplication we included the stimulus ID in the statistical analysis (see below).

175 To create a large repertoire for playback, we first analysed a recording of ≥ 100
176 songs from a stimulus bird and then selected a sequence of 50 songs with the largest
177 possible element repertoire. This was achieved by deleting songs from the sequence
178 which contained only or mostly element types also present in other songs. The repertoire
179 size in the derived sequence determined the repertoire size of the corresponding small

180 repertoire size to be broadcast (ca. half the large repertoire size). Then we constructed the
181 playback sequence with the small repertoire, using the same procedure as above, except
182 that this time songs were chosen such that they added the smallest available amount of
183 new elements types to the sequence. These procedures resulted in large repertoire
184 playbacks with repertoires slightly higher than in the original recording and small
185 repertoire playbacks with repertoires considerably smaller. On average, the large
186 repertoires had 237 different element types (range 190-276), and small repertoires 120
187 element types (range 85-152). Playback files were matched in such a way that the large
188 repertoire was ca. twice as large as the small one. Recordings from birds with an
189 originally small repertoire size were not used, since we would then create unnaturally
190 small repertoires.

191 Since the natural order of songs was jumbled to a higher degree in the small than
192 large playback files, we randomized the order of songs in both the large and small
193 repertoire playback files. We however avoided direct repetitions of songs, as this occurs
194 rarely in natural song. The 50 songs of a file were played back with a delivery cadence of
195 5 sec, resulting in a total duration of 250 sec. The maximum volume of each song (peak
196 amplitude) was normalized to 90% of the maximal possible amplitude of the system, and
197 the pauses between songs were filled with background noise from the same recordings.

198 We matched the two recordings used for the same test bird in song duration and
199 twitter part duration since these characteristics may influence response strength
200 (Dabelsteen & Pedersen 1990; Balsby & Dabelsteen 2001). Also after manipulation,
201 songs in the large and small repertoire playback did not differ systematically in song
202 length (mean of means \pm sd; large playback repertoire: 2.69 sec \pm 0.28, small playback

203 repertoire: $2.66 \text{ sec} \pm 0.48$; Wilcoxon's matched pairs signed ranks test of the average
204 song durations in the large vs. small repertoire files from the same stimulus bird: $T^+ = 18$,
205 $N = 8$, $P = 1$; and of average song durations of the two files which were played to the
206 same test subject: $T^+ = 18$, $N = 8$, $P = 1$). The number of different element types per song
207 and the ratio between whistle and twitter element types did not differ significantly after
208 manipulation either (comparing songs in the small and large repertoire files from the
209 same stimulus bird; number of element types: $T^+ = 30$, $N = 8$, $P = 0.11$; ratio of whistle to
210 twitter element types: $T^+ = 25$, $N = 8$, $P = 0.38$).

211

212 Playback procedure

213 We conducted the playback experiments in spring 2008 (6. April - 6. May, after
214 dawn chorus and before noon) with 16 territorial males who were colour-ringed and at
215 least in their 3. calendar year. Most were paired with a female which was in the nest
216 building phase and very likely in her fertile phase. We chose this critical stage of the
217 breeding cycle, because we obtained very weak reactions during pilot experiments earlier
218 in the season in 2007 in the same population. Each test bird was tested twice (small and
219 large repertoire playback tests), on consecutive days (except for one bird that was tested
220 with three days between tests). The order of treatments was balanced. We played only
221 recordings from birds that had their territories in a different area of the island than the test
222 bird throughout their lifetime. On the same day, we did not test birds in earshot of each
223 other; neither did we use playbacks from same stimuli birds for test subjects in earshot of
224 each other. The test files were played back using a CD player (Sony CD D-NE330) and a
225 Nagra DSM speaker. The output level was calibrated to maximum values of ca. 65 dB(A)

226 at 10 m (Brühl & Kjær 2006, fast setting), which is the natural level of blackbird full
227 song (Dabelsteen 1981). We placed the speaker in the test subject's territory. Due to the
228 often very small and adjacent territories on Helgoland, neighbouring males often
229 interfered during the experiments; therefore we placed the speaker centrally in the
230 territory, i.e. between ca. 10-20 m from the territory border (as most territories were in
231 the residential areas of the island we were somewhat restricted in positioning the
232 speakers). The speaker was placed at the same position for both experiments, put on a
233 suitable object in a height between ca. 20 and 150 cm. Most birds on the island are very
234 well habituated to humans and used to human activity in their territories, so test birds
235 rarely seemed disturbed when we set up the speaker. However, we waited at least 3 min
236 before starting the experiments. We started playback while the test bird was in sight.

237 We recorded the following behaviours during the 250 sec of the playback and 120
238 sec afterwards: *Flight*, *bow posture* (neck stretched forward, beak pointing downwards,
239 rump feathers typically fluffed, tail fanned), *tail up posture* (tail going upwards and down
240 again in a fast movement), and *distance to speaker* (on speaker, 0 m to < 1 m, 1 m to < 5
241 m, 5 m to < 10 m, 10 m or more; for a more detailed description of the postures, see
242 Dabelsteen 1982, Dabelsteen & Pedersen 1985).

243 Song and other vocalisations were also recorded, but occurred too rarely to
244 provide data for analysis. Only twitter song was elicited frequently by playback. This is a
245 different singing style which resembles the twitter part in full song and is used in
246 aggressive interactions (Dabelsteen et al. 1998, Klump et al. 2010). However, due to the
247 low amplitude of this vocalisation and the relatively high background noise level during
248 experiments it was not possible to quantify it reliably. When a test bird was interacting

249 with another bird, we continued the playback but did not use behaviours during this
250 interaction as data. (5 test subjects had interactions with females, one subject had an
251 interaction with a male). We attempted to increase the sample size by testing 6 further
252 males, but these data could not be used due to frequent interruption by neighbouring
253 males.

254 We were able to record full song from 15 of the 16 test subjects. We made the
255 recordings during dawn chorus in the same breeding season. They had repertoire sizes of
256 257 ± 70 (mean \pm SD) element types with a large range (116-395 element types) in the
257 first 100 songs (Fig. 2). We did not know the subjects' repertoire sizes prior to the
258 experiments. Retrospective comparison of these with the playback repertoire sizes
259 showed that in sequences of 50 songs, the repertoire size of 10 test subjects was between
260 the sizes of the small and the large repertoire broadcast, 4 had repertoires larger than the
261 large playback and only one had a repertoire that was smaller than the small playback
262 with which the respective bird was tested.

263

264 Statistical analyses

265 We analysed data separately for the period during the playback and the two minutes
266 afterwards. During playback we used ten response variables: Rates (n/sec) of the
267 behaviours *flight*, *tail up*, *bow*, latency and mean timing (sec) of the behaviours *flight* and
268 *tail up*. 'Mean timing' reflects the timing of the behaviours. It was calculated as the mean
269 of all times of occurrence of the respective behaviour (in seconds after start of playback)
270 during the observed period. We used three distance measurements: mean and minimum
271 distance to the speaker (m) and distance slope, which is the slope of a regression between

272 distance and time. The latter was used in order to estimate the spatial behaviour relative
273 to time (i.e., whether the bird was approaching or moving away from the speaker over
274 time). For the period after playback we used only the rates of the behaviours *flight* and
275 *tail up* as well as the same distance measurements as before. Low latencies, high rates
276 with long occurrences and short distances indicate a strong response during and after
277 playback (e.g. Dabelsteen 1982).

278 To consolidate the response variables, we conducted two Principal Component
279 Analyses (PCA), one for response variables obtained during and one for those obtained
280 after the playback. Since distance slope was only weakly correlated to the other variables
281 (largest absolute correlation: during playback 0.299, after 0.236) we excluded it from the
282 PCAs. Large correlations between some of the variables, the Kaiser-Meyer-Olkin
283 measure of sampling adequacy, as well as Bartlett's test of sphericity (McGregor 1992)
284 indicated that both PCAs were justified (Table 1). We used Varimax-Rotation in order to
285 achieve easy interpretable factors. The PCAs extracted three principal components with
286 an Eigenvalue > 1 for the variables collected during and two for the variables collected
287 after playback (Table 1). In the following analyses we used factor scores derived from the
288 PCAs as well as distance slopes as response variables.

289 We tested the influence of the playback on the subjects' behaviours using
290 Generalized Linear Mixed Models (GLMM; Baayen 2008a). We included as fixed effects
291 1) stimulus repertoire size, 2) stimulus order, 3) experiment number (first or second
292 experiment), 4) the mean duration of the respective playback songs. We also included the
293 test subject's own repertoire size and its interaction with the stimulus repertoire size as an
294 additional fixed effect. Since the repertoire size of one of the 16 test subjects was

295 unknown, we ran the model once including this effect and once excluding it (but
296 achieving a larger sample size). As random effects we initially included the ID of the
297 tested individual and the individual from which the broadcast songs were recorded.
298 However, we removed the latter since it appeared to be insignificant (all likelihood ratio
299 tests except for one $P > 0.5$, one $P = 0.20$).

300 We ran the GLMM once for each of the seven response variables (during
301 playback: PC1, PC2, PC3, distance slope; after playback: PC1, PC2, distance slope).
302 Since some individuals left immediately after the playback ended, the sample size for the
303 period *after* was smaller. Thus, we used four different combinations of experimental
304 periods and fixed effects (during playback: 32/30 experiments; after playback: 29/27
305 experiments excluding/including the subject's repertoire size) and ran the model
306 altogether 4 x 7 times.

307 Several of the fixed effects were considered to be 'control variables' (i.e. stimulus
308 order, experiment number, mean duration of the respective playback songs). We were not
309 interested in their statistical significance per se but simply wanted to control for potential
310 effects of these variables. If these were non-significant ($P > 0.25$), we removed them
311 from the model prior to testing for the effects of the subject's own repertoire size and its
312 interaction with the stimulus repertoire size.

313 We tested for the significance of the random effects using likelihood ratio tests
314 (Dobson 2002; Bolker et al. 2009) based on models fit using Maximum Likelihood
315 (rather than restricted Maximum Likelihood). To test the significance of fixed effects we
316 used Monte-Carlo Markov-Chain (MCMC) analysis (Baayen 2008a, Bolker et al. 2009).
317 We checked for homogeneous variance and approximate normal distribution of residuals

318 by visually inspecting residuals plotted against predicted values. We also checked for
319 interrelations between predictor variables using Variance Inflation Factors (VIF). We
320 also calculated VIFs for dummy coded stimulus number (Field 2005). The largest VIF
321 was 1.1, ruling out collinearity issues. To achieve reliable tests of their interaction, we z-
322 transformed repertoire size of the test subjects and playback, respectively, to a mean of
323 zero and a standard deviation of one (Aiken & West 1992).

324 PCAs were calculated using SPSS 15.0.0 for Windows. We calculated GLMMs
325 using the function 'lmer' provided by the package 'lme4' (version 0.999375-31; Bates et
326 al. 2008) for R (version 2.9.1; R Development Core Team 2008). We derived MCMC P-
327 values using the function 'pvals.fnc' from the R-package 'languageR' (version 0.953,
328 Baayen 2008b). Variance inflation factors were run using scripts written for R by R. M.

329

330 Results

331 When we started playback, most of the test subjects reacted immediately by stopping
332 their current behaviour and/or starting to fly around. In most cases they reacted strongly
333 and aggressively, flying low over the speaker and/or landing on it while uttering
334 continuous twitter song. During playback, test birds performed 3.13 ± 1.68 (mean \pm sd)
335 *tail up postures* per minute, 2.23 ± 1.26 *flights* per minute and 1.75 ± 2.56 *bow postures*
336 per minute with a latency of 25 ± 22.45 sec until the first occurrence of *tail up posture*
337 and a latency of 32 ± 52.28 sec until the first *flight*.

338 The repertoire size of the playback files had no significant influence on the test
339 birds' responses, nor was there an interaction between the test birds' and playback
340 repertoire sizes. Also, none of our 'control variables' (the mean duration of the respective

341 playback songs, stimulus order, number of experiment, or the identity of the stimulus bird
342 and the test subject itself) had a recognizable effect on any of the responses. This was the
343 case both when we did not take the subjects' own repertoire sizes into consideration
344 (Table 2) and when we did (Table 3).

345

346 Discussion

347 In our playback experiment blackbirds did not respond differently towards playbacks of
348 different repertoire sizes. This was also the case when we took the test birds' own
349 repertoire sizes into account. Thus, with our experiment we found no evidence supporting
350 the hypothesis that male blackbirds use other males' repertoire sizes to estimate a rival's
351 quality. Considering the body of literature implying that both inter- and intra-sexual
352 selection are driving forces in the evolution of repertoires (for an overview, see Collins
353 2004), this seems to be a surprising result (but see Catchpole 1989).

354 In many other studies very similar methods have been used in playback
355 experiments with songbirds and even blackbirds (e.g. Dabelsteen 1982; Dabelsteen &
356 Pedersen 1985, 1990; Stoddard 1988; Balsby & Dabelsteen 2001; Briefer et al. 2008) and
357 our test subjects showed strong responses to our playbacks. The rates of the recorded
358 behavioural patterns were comparable to those observed by Dabelsteen (1982) in a
359 playback experiment with blackbirds. We therefore think that the general design which
360 we used should be suitable to detect differences in response if there were any.

361 A crucial question is whether the test birds were able to assess the differences in
362 repertoire sizes in our experiments at all. We presented the stimuli over a relatively short
363 time and the repertoire sizes in the 50 songs of our playback were much smaller than a

364 normal full repertoire of a blackbird, in which new element types can occur even after
365 300 consecutive songs (unpubl. data). However, differences in repertoire sizes of the two
366 stimuli were very clear and if it is indeed necessary to hear the full repertoire of a rival in
367 order to assess his quality, it would not be feasible to use repertoire sizes for quality
368 assessment in short encounters. In our experiments, the test birds had to hear at least ca.
369 100 sec of playback (corresponding to ca. 20 songs) in order to have the opportunity to
370 detect differences in repertoire sizes. In most cases the test birds started responding
371 earlier, meaning that their first response cannot have been based on the repertoire size
372 broadcast. However, during the rest of the playback and in the period afterwards they
373 could have adjusted their response if the perceived threat was dependent on repertoire
374 size, as described by Stoddard et al. (1988) for song sparrows. Nevertheless, even if the
375 test birds were able to distinguish between the two repertoire sizes, the magnitude of
376 response was similar during the broadcasting of both repertoire sizes.

377 This finding may be due to a ceiling effect. In our experiment, the test birds were
378 confronted with the simulation of an unknown intruder singing full song from within
379 their territory – usually during a period in which the subjects' females were likely to be
380 fertile (before/during the egg laying period). The strong immediate responses indicate
381 that the test birds perceived the playback as a strong threat. We chose this design as we
382 experienced very weak reactions in pilot experiments, which were carried out in 2007 but
383 later in the season and thus in a less critical stage of the breeding cycle. To the territory
384 holder, such an intrusion may represent the risk of losing parts of his territory or his
385 female engaging in extra-pair copulations with the unknown intruder. Furthermore, the
386 reaction of a male to an intruder could also influence the mate choice of observing

387 females (Otter et al. 1999; Mennill et al. 2002). Therefore, this is possibly a situation in
388 which the subjects always react persistently with full strength, without losing time for
389 assessing the repertoire size broadcast. In such situations, birds may rather rely on quality
390 signals which can be immediately evaluated. In other territory intrusion experiments with
391 blackbirds, males responded more strongly to song playback with longer twitter part
392 duration (Dabelsteen & Pedersen 1990), and in a different experiment, they reacted more
393 strongly to a model bird with yellow or orange beak than to models with brown beaks,
394 which are typical for yearling males (Bright & Waas 2002). Unlike repertoire size, both
395 twitter part duration and beak colour can immediately be assessed.

396 We also considered a playback design using two speakers simultaneously, so that
397 the test bird would have to decide between the stimuli even though both impose a strong
398 threat, as used by Leitão & Riebel (2003) or de Kort et al. (2009). However, the
399 assessment of large repertoires seems to be a rather complex task requiring to listen for
400 some time. In our experiments, the differences in repertoire sizes were not detectable
401 before listening to at least 20 songs and we wanted to avoid that in a two speaker design,
402 the test birds would decide for one speaker earlier during playback. In the mentioned
403 studies song traits were compared which can be immediately assessed (Leitão & Riebel
404 2003: songs with vs. without the ‘terminal flourish’; de Kort et al. 2009: consistency of
405 trill notes within one song).

406 Our negative result does of course not rule out the possibility that males evaluate
407 repertoire size in other situations. For example, a territory owner might listen to
408 neighbours singing outside the territory and use their repertoire sizes for quality

409 assessment, which could influence decisions about future interactions like investment in
410 territorial fights or in mate guarding.

411 The large inter-individual differences in the test birds' repertoire sizes found in
412 this and previous studies (Rasmussen & Dabelsteen 2002) and our findings in another
413 study, that larger males have larger repertoire sizes (Hesler 2010), imply that repertoire
414 size could serve as a quality signal in blackbirds. It is possible that repertoire size is
415 linked to male quality as in several other species (e.g. Hiebert et al. 1989; Mountjoy &
416 Lemon 1996; Buchanan et al. 1999; Balsby 2000; Kipper et al. 2006) and used for quality
417 assessment, but only in inter-sexual contexts. This idea complements findings that sedge
418 warbler females react differently to playback of different repertoire sizes while males do
419 not (Catchpole 1989), and that in an experiment with starlings *Sturnus vulgaris*, territorial
420 males sang with a larger repertoire in the presence of a female than in the presence of a
421 male (Eens et al. 1993). Catchpole (1989) suggested that in the sedge warbler inter-sexual
422 selection has resulted in more complex songs and intra-sexual selection in simpler and
423 shorter songs. However, if females are able to use song characteristics for male quality
424 assessment, it would seem advantageous for males to do so as well. That repertoires in
425 general do not play a role in intra-sexual contexts in blackbirds seems unlikely, as they
426 also have a distinct singing style, the 'aggressive twitter' (Dabelsteen et al. 1998), which
427 is used exclusively in intra-sexual contexts during close range aggressive interactions
428 with males and has an even larger repertoire than normal full song (Klump et al. 2010).

429 It is also possible that not only repertoire size per se is the selected song trait, but
430 also how it is used and organised: even birds with same element repertoire sizes could
431 differ in sequential organisation or frequency distribution of their elements, or their

432 ability of repertoire matching (Beecher et al. 2000). It could also be that special element
433 types, which are energetically demanding to produce, are the selected trait. In canaries
434 *Serinus canaria* it has been found that females prefer ‘sexy syllables’ which are
435 energetically demanding to produce (Vallet & Kreuzer 1995; Leitner et al. 2006). Other
436 performance dependent song traits have been identified, which were linked to male
437 quality and/or preferred by females, as syllable type or trill consistency (tropical
438 mockingbird *Mimus gilvus*, Botero et al. 2009; banded wren *Thryothorus pleurostictus*,
439 de Kort et al. 2009) and the performance of rapid, broadband trills (swamp sparrow
440 *Melospiza georgiana*, Ballentine 2009), or song traits which depend on conditions during
441 early development, as syntax accuracy (zebra finch *Taeniopygia guttata*, Brumm et al.
442 2009; Holveck et al. 2008). In our design, we created both large and small playback files
443 from the same recording, which was always from an originally large repertoire bird, since
444 we could not enlarge repertoire sizes, but only reduce them. We did this in order to avoid
445 the influence of other song parameters correlated with repertoire size. However, this also
446 means that song traits other than repertoire size were present (or absent) in both test files.
447 If these effects were clearly stronger than repertoire size, responses towards different
448 stimulus birds would have differed, which was not the case. However, if both repertoire
449 size and other intrinsic song traits have an effect at the same time, an indifferent response
450 could result. An experimental design using both large and small repertoire from the same
451 stimulus bird for the same test bird would avoid this problem. On the other hand, in the
452 second experiment the test bird may then recognise the identity of the stimulus bird
453 (Weary & Krebs 1992) from the first experiment and base its response partly on that.
454 Therefore we decided to use stimuli from different individuals for one test bird.

455 Possibly the two distinct parts of the blackbird's song, the whistle and the twitter,
456 have different functions, as they also have different acoustic properties (the whistle part
457 has a much further transmission range than the twitter part and is more omni-directional,
458 Larsen & Dabelsteen 1990; Dabelsteen et al. 1993). This means that they reach different
459 parts of the habitat. By using the habitat in different ways, female and male receivers may
460 thus not be equally exposed to the whistle and twitter part (Dabelsteen & Pedersen 1988)
461 so that inter- and intra-sexual selection may act differently on these parts. This could be
462 tested by manipulating only one of the two song parts in a playback experiment (Leitão &
463 Riebel 2003).

464 In conclusion, we found no evidence for the hypothesis, that male blackbirds use
465 repertoire sizes to assess a rival's fighting ability. However, the interpretation of negative
466 results is difficult. Possibly a ceiling effect lead to indifferent responses to the different
467 repertoire sizes. Also, it is possible that males use repertoire sizes for quality assessment
468 in other contexts than territory intrusion, such as listening to song outside their territory.

469

470 Acknowledgements

471 We thank T. Sacher, T. Coppack and F. Bairlein for collaboration and O. Hüppop with
472 the team of the Vogelwarte Helgoland and C. Horn for assistance during the field work.
473 We also thank I. Teschke for language improvement and two anonymous reviewers for
474 their helpful comments. The study was funded by a combined PhD fellowship from
475 NaFöG (Promotionsstipendium nach dem Nachwuchsförderungsgesetz des Landes
476 Berlin, Germany), SNAK (Graduate School on Sense organs, Nerve systems, Behaviour
477 and Communication, funded by the Danish Research Agency) and the University of

478 Copenhagen to NH and a framework grant from the Danish Research Council for Nature
479 and Universe (no.272-07-0477) to TD. This study meets the guidelines for the treatment
480 of animals in behavioural research and teaching of the Association for the Study of
481 Animal Behaviour and complies with laws in Germany and Denmark.

482

483 References

484 Aiken, L S & West, S G 1992 Multiple Regression: Testing and Interpreting Interactions
485 Newbury Park, Sage.

486 Baayen, R H 2008a languageR: Data sets and functions with "Analyzing Linguistic Data:
487 A practical introduction to statistics".

488 Baayen, R H 2008b Analyzing Linguistic Data: A practical introduction to statistics using
489 R. University Press, Cambridge.

490 Ballentine, B (2009) The ability to perform physically challenging songs predicts age and
491 size in male swamp sparrows, *Melospiza georgiana*. Anim Behav 77, 973-978.

492 Balsby, T J S (2000) The function of song in whitethroats *Sylvia communis*. Bioacoustics
493 11, 17-30.

494 Balsby, T J S and Dabelsteen, T (2001) The meaning of song repertoire size and song
495 length to male whitethroats *Sylvia communis*. Behavioural Processes 56, 75-84.

496 Bates, D, Maechler, M and Dai, B (2008) lme4: Linear mixed-effects models using S4
497 classes.

498 Beecher, M D Campbell, S E, and Nordby, J C (2000) Territory tenure in song sparrows
499 is related to song sharing with neighbours, but not to repertoire size. Anim Behav
500 59, 29-37.

501 Bolker, B M, Brooks, M E, Clark, C J, Geange, S W, Poulsen, J R, Stevens, M H, and
502 White, J S (2009) Generalized linear mixed models: a practical guide for ecology
503 and evolution. *Trends in Ecology & Evolution* 24, 127-135.

504 Botero, C A, Rossman, R J, Caro, L M, Stenzler, L M, Lovette, I J, de Kort, S R, and
505 Vehrencamp, S L (2009) Syllable type consistency is related to age, social status
506 and reproductive success in the tropical mockingbird. *Anim Behav* 77, 701-706.

507 Briefer, E, Aubin, T, Lehongre, K, and Rybak, F (2008) How to identify dear enemies:
508 the group signature in the complex song of the skylark *Alauda arvensis*. *Journal*
509 *of Experimental Biology* 211, 317-326.

510 Bright, A and Waas, J R (2002) Effects of bill pigmentation and UV reflectance during
511 territory establishment in blackbirds. *Anim Behav* 64, 207-213.

512 Brumm, H, Zollinger, S A, and Slater, P J B (2009) Developmental stress affects song
513 learning but not song complexity and vocal amplitude in zebra finches. *Behav*
514 *Ecol Sociobiol* 63, 1387-1395.

515 Buchanan, K L, Catchpole, C K, Lewis, J W, and Lodge, A (1999) Song as an indicator
516 of parasitism in the sedge warbler. *Anim Behav* 57, 307-314.

517 Catchpole, C K (1980) Sexual selection and the evolution of complex songs among
518 European warblers of the genus *Acrocephalus*. *Behaviour* 74, 149-165.

519 Catchpole, C K (1989) Responses of male sedge warblers to playback of different
520 repertoire sizes. *Anim Behav* 37, 1046-1047.

521 Catchpole, C K and Slater, P J B 2008 *Bird song - biological themes and variations*, 2
522 edn: Cambridge University Press.

523 Collins, S A (2004) Vocal fighting and flirting: the functions of birdsong. In Marler P,
524 and Slabbekoorn H (eds) *Nature's Music*, Elsevier Academic Press, London, pp.
525 39-79.

526 Dabelsteen, T (1981) The Sound Pressure Level in the Dawn Song of the Blackbird
527 *Turdus merula* and a Method for Adjusting the Level in Experimental Song to the
528 Level in Natural Song. *Z Tierpsychol* 56, 137-149.

529 Dabelsteen, T (1982) Variation in the Response of Freelifving Blackbirds *Turdus merula*
530 to Playback of Song: I.Effect of Continuous Stimulation and Predictability of the
531 Response. *Z Tierpsychol* 58, 311-328.

532 Dabelsteen, T (1984) An analysis of the full song of the Blackbird *Turdus merula* with
533 respect to message coding and adaptations for acoustic communication. *Ornis*
534 *Scandinavica* 15, 227-239.

535 Dabelsteen, T and Pedersen, S B (1985) Correspondence between Messages in the Full
536 Song of the Blackbird *Turdus merula* and Meanings to Territorial Males, as
537 Inferred from Responses to Computerized Modifications of Natural Song. *Z*
538 *Tierpsychol* 69, 149-165.

539 Dabelsteen, T and Pedersen, S B (1988) Do Female Blackbirds, *Turdus merula*, Decode
540 Song in the Same Way as Males? *Anim Behav* 36, 1858-1859.

541 Dabelsteen, T and Pedersen, S B (1990) Song and information about aggressive
542 responses of blackbirds, *Turdus merula*: evidence from interactive playback
543 experiments with territory owners. *Anim Behav* 40, 1158-1168.

544 Dabelsteen, T, Larsen, O N, and Pedersen, S B (1993) Habitat-induced degradation of
545 sound signals: Quantifying the effects of communication sounds and bird location

546 on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song.
547 Journal of the Acoustic Society of America 93, 2206-2220.

548 Dabelsteen, T, McGregor, P K, Lampe, H M, Langmore, N E, and Holland, J (1998)
549 Quiet song in song birds: an overlooked phenomenon. Bioacoustics 9, 89-105.

550 Dabelsteen, T (2005) Public, private or anonymous? Facilitating and countering
551 eavesdropping. In McGregor P K (ed) Animal Communication Networks,
552 Cambridge University Press, Cambridge, pp. 38-62.

553 de Kort, S R, Eldermire, E R B, Valderrama, S, Botero, C A, and Vehrencamp, S L 2009
554 Trill consistency is an age-related assessment signal in banded wrens. Proc Roy
555 Soc London Ser B 276, 2315-2321.

556 Dobson, A J (2002) An introduction to generalized linear models. Chapman & Hall/CRC,
557 Boca Raton,FL.

558 Eens, M, Pinxten, R, and Verheyen, R F (1993) Function of the Song and Song
559 Repertoire in the European Starling (*Sturnus vulgaris*) - An Aviary Experiment.
560 Behaviour 125, 51-66.

561 Field, A (2005) Discovering statistics using SPSS. Sage Publications, London.

562 Gentner, T Q, Hulse, S H, Bentley, G E, and Ball, G F (2000) Individual vocal
563 recognition and the effect of partial lesions to HVC on discrimination, learning,
564 and categorization of conspecific song in adult songbirds. J Neurobiol 42, 117-
565 133.

566 Hesler N (2010) Song complexity in common blackbirds – an honest signal of male
567 quality? Dissertation, University of Copenhagen.

568 Hiebert, S M, Stoddard, P K, and Arcese, P (1989) Repertoire size, territory acquisition
569 and reproductive success in the song sparrow. *Anim Behav* 37, 266-273.

570 Holveck, M J, de Castro, A C V, Lachlan, R F, Ten Cate, C, and Riebel, K (2008)
571 Accuracy of song syntax learning and singing consistency signal early condition
572 in zebra finches. *Behav Ecol* 19, 1267-1281.

573 Kipper, S, Mundry, R, Hultsch, H, and Todt, D (2004) Long-term persistence of song
574 performance rules in nightingales (*Luscinia megarhynchos*): a longitudinal field
575 study on repertoire size and composition. *Behaviour* 141, 371-390.

576 Kipper, S, Mundry, R, Sommer, C, Hultsch, H, and Todt, D (2006) Song repertoire size is
577 correlated with body measures and arrival date in common nightingales, *Luscinia*
578 *megarhynchos*. *Anim Behav* 71, 211-217.

579 Klump, B, Hesler, N, and Dabelsteen, T (2010) Quiet song in common blackbirds -
580 repertoire size and use in an aggressive context. Conference abstract, Bioacoustics
581 in press.

582 Krebs, J, Ashcroft, R, and Webber, M (1978) Song repertoires and territory defence in
583 the great tit. *Nature* 271, 539-542.

584 Larsen, O N and Dabelsteen, T (1990) Directionality of Blackbird vocalization.
585 Implications for vocal communication and its further study. *Ornis Scandinavica*
586 21, 37-45.

587 Leitão, A and Riebel, K 2003 Are good ornaments bad armaments? Male chaffinch
588 perception of songs with varying flourish length. *Anim Behav* 66, 161-167.

589 Leitão, A, Ten Cate, C, and Riebel, K 2006 Within-song complexity in a songbird is
590 meaningful to both male and female receivers. *Anim Behav* 71, 1289-1296.

591 Leitner, S, Marshall, R C, Leisler, B, and Catchpole, C K (2006) Male song quality, egg
592 size and offspring sex in captive canaries (*Serinus canaria*). *Ethology* 112, 554-
593 563.

594 McGregor, P K (1992) Quantifying responses to playback: one, many, or composite
595 multivariate measures? In McGregor P K (ed) *Playback and Studies of Animal*
596 *Communication*, Plenum Press, New York and London, 79-96.

597 Mennill, D J, Ratcliffe, L M, and Boag, P T (2002) Female Eavesdropping on Male Song
598 Contests in Songbirds. *Science* 296, 873.

599 Mountjoy, D J and Lemon, R E (1996) Female choice for complex song in the European
600 starling: a field experiment. *Behav Ecol Sociobiol* 38, 65-71.

601 Otter, K, McGregor, P K, Terry, A M R, Burford, F R L, Peake, T M, and Dabelsteen, T
602 (1999) Do female great tits (*Parus major*) assess males by eavesdropping? A field
603 study using interactive song playback. *Proc Roy Soc London Ser B* 266, 1305-
604 1309.

605 R Development Core Team. (2008) R: A language and environment for statistical
606 computing. R Foundation for Statistical Computing. Vienna, Austria.

607 Rasmussen, R and Dabelsteen, T (2002) Song repertoires and repertoire sharing in a local
608 group of Blackbirds. *Bioacoustics* 13, 63-76.

609 Sacher, T, Engler, J, Gorschewski, A, Gottschling, M, Hesler, N, Bairlein, F, and
610 Coppack, T (2006) Die Helgoländer Amselpopulation: ein Modell für
611 Populationsgenetik und Zugbiologie. *Ornithologischer Jahresbericht Helgoland*
612 16, 76-84.

613 Searcy, W A (1983) Response to multiple song types in male song sparrows and field
614 sparrows. Anim Behav 31, 948-949.

615 Searcy, W A and Beecher M D (2009) Song as an aggressive signal in songbirds. Anim
616 Behav 78, 1281-1292.

617 Stoddard, P K, Beecher, M D, and Willis, M S (1988) Response of territorial male song
618 sparrows to song types and variations. Behav Ecol Sociobiol 22, 125-130.

619 Svensson, L (1992) Identification Guide to European Passerines. Svensson, Stockholm.

620 Todt, D (1981) On Functions of Vocal Matching: Effect of Counter-replies on Song Post
621 Choice and Singing. Z Tierpsychol 57, 73-93.

622 Vallet, E and Kreutzer, M (1995) Female Canaries Are Sexually Responsive to Special
623 Song Phrases. Anim Behav 49, 1603-1610.

624 Weary, D M and Krebs, J R (1992) Great tits classify songs by individual voice
625 characteristics. Anim Behav 43, 283-287.

626 Yasukawa, K (1981) Song repertoires in the red-winged blackbird (*Agelaius phoeniceus*):
627 A test of the Beau Geste hypothesis. Anim Behav 29, 114-125.

628

629

630 Tables

631 **Table 1** Results of the PCA for data obtained during and after playback. The Kaiser-
632 Meyer-Olkin measure of sampling adequacy (K-M-O) and Bartlett's test of sphericity
633 (Bartlett's χ^2 , df and P) justified both PCAs. Response variables which contributed most
634 to the respective components (PC1, PC2, PC3) are in bold. After playback the behaviours
635 *bow*, *flight* and *tail up* occurred too rarely to include them in the analysis. Transf.:
636 Variables were transformed to achieve normality or uniformity if needed according to
637 visual inspection; log: log transformed, sqrt: square root transformed.

	during playback				after playback		
K-M-O	0.58				0.49		
Bartlett's χ^2	114.2				48.6		
Bartlett's df	30				6		
Bartlett's P	<0.001				<0.001		
variable	transf.	PC1	PC2	PC3	transf.	PC1	PC2
mean distance [m]	-	0.94	0.07	0.08	-	0.96	0.12
rate of bow [1/sec]	sqrt	-0.82	0.04	0.10	-	-	-
minimum distance [m]	-	0.80	-0.36	0.05	-	0.96	-0.09
rate of flight [1/sec]	-	-0.06	0.84	0.08	sqrt	-0.10	0.89
rate of tail up [1/sec]	-	-0.18	0.75	-0.19	sqrt	0.13	0.88
latency flight [sec]	log	-0.05	-0.68	0.55	-	-	-
mean timing tail up [sec]	-	-0.05	0.37	0.78	-	-	-
mean timing flight [sec]	-	0.25	-0.31	0.76	-	-	-
latency tail up [sec]	log	-0.12	-0.23	0.69	-	-	-
Eigenvalue		2.95	2.10	1.44		1.88	1.59
% variance explained		32.76	23.37	15.98		47.02	39.69

638

639 **Table 2** GLMM results of the playback experiments. Given are P-values (MCMC) and
640 95% confidence interval limits (lower CI, upper CI) of final models *without considering*
641 *the subjects' own repertoire sizes* (one fitted for each response variable) . The predictor
642 variable, repertoire size of playback stimulus, did not have a significant effect on any of
643 the response variables. Test: response variables obtained during playback; after: response
644 variables obtained in the 120 sec after playback. N: sample size. Indicated P-values are
645 without correction for multiple testing. For details of the PC factors see Table 1.

Response variable	Predictor variable	lower CI	upper CI	p MCMC	N
test distance slope	repertoire size of stimulus	-0.321	0.473	0.733	32
test PC 1	repertoire size of stimulus	-0.479	0.234	0.552	32
test PC 2	repertoire size of stimulus	-0.311	0.379	0.840	32
test PC 3	repertoire size of stimulus	-0.515	0.203	0.371	32
after distance slope	repertoire size of stimulus	-0.836	0.936	0.906	29
after PC 1	repertoire size of stimulus	-0.491	0.248	0.471	29
after PC 2	repertoire size of stimulus	-0.575	0.206	0.382	29

646

647 **Table 3** GLMM results of the playback experiments. Given are P-values (MCMC) and
648 95% confidence interval limits (lower CI, upper CI) from final models *including the*
649 *subjects' own repertoire sizes* (one fitted for each response variable). None of the
650 predictor variables had a significant influence on any of the response variables.
651 Test/after: response variables obtained during/in the 120 sec after the experiment. N:
652 sample size. For details of the PCs see Table 1. Since *after PC 2* was involved in a
653 marginally non-significant interaction, the values for the involved variables *repertoire*
654 *size of stimulus* and *repertoire size of test subject* are not indicated. Indicated P-values are
655 without correction for multiple testing.

Response variable	Predictor variable	lower CI	upper CI	p MCMC	N
test distance slope	repertoire size of stimulus	-0.457	0.320	0.734	30
test distance slope	repertoire size of test subject	-0.628	0.159	0.231	30
test PC 1	repertoire size of stimulus	-0.479	0.280	0.577	30
test PC 1	repertoire size of test subject	-0.487	0.315	0.659	30
test PC 2	repertoire size of stimulus	-0.414	0.306	0.798	30
test PC 2	repertoire size of test subject	-0.590	0.188	0.297	30
test PC 3	repertoire size of stimulus	-0.421	0.289	0.752	30
test PC 3	repertoire size of test subject	-0.085	0.642	0.140	30
after distance slope	repertoire size of stimulus	-0.769	0.999	0.697	27
after distance slope	repertoire size of test subject	-0.230	1.585	0.151	27
after PC 1	repertoire size of stimulus	-0.443	0.318	0.738	27
after PC 1	repertoire size of test subject	-0.708	0.102	0.146	27
after PC 2	interaction between repertoire size of stimulus and test subject	-0.906	0.049	0.080	27

656

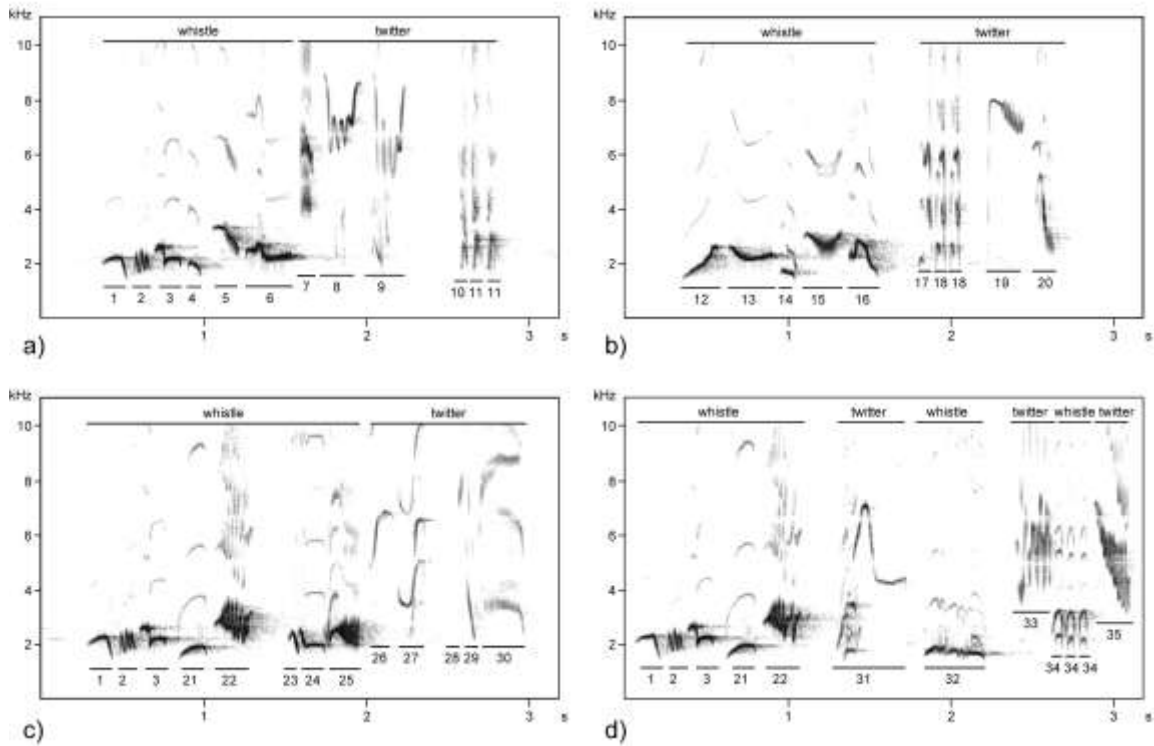
657

658 Figure captions

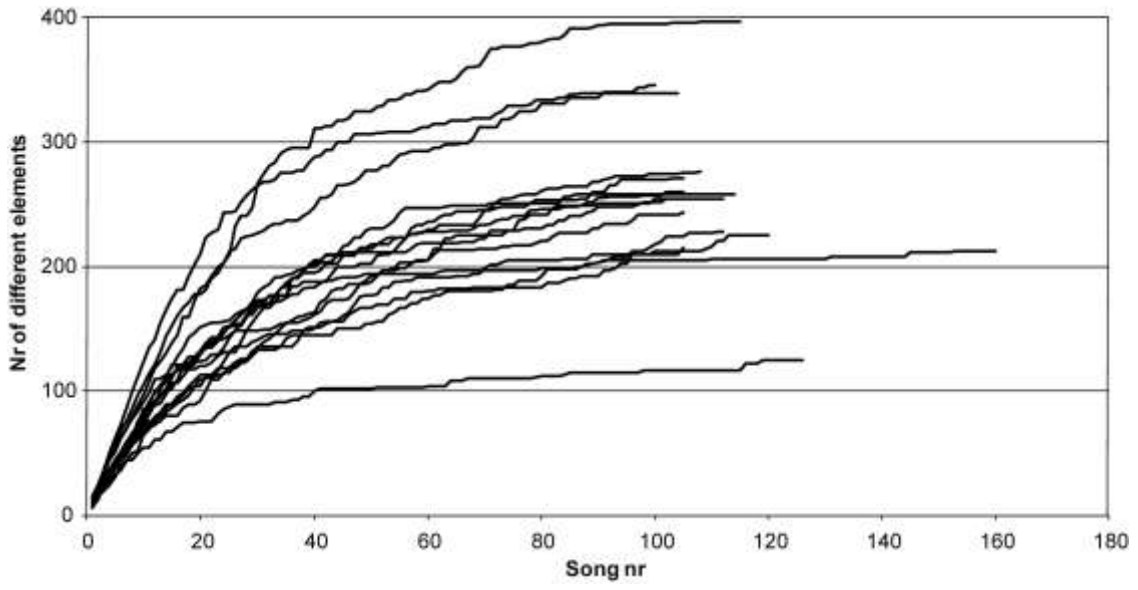
659 **Fig. 1** Spectrograms of four consecutive blackbird songs with indication of
660 whistle/twitter parts and single elements. Numbers indicate different element types. Note
661 that some of the element types of song 1a) occur also in the song 1c) and d).

662

663 **Fig. 2** Element repertoire sizes of test subjects. The number of different element types is
664 depicted against the number of analysed songs.



665



666