J Ornithol (2011) 152:591–601
DOI 10.1007/s10336-010-0618-5

ORIGINAL ARTICLE

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

Nana Hesler · Roger Mundry · Torben Dabelsteen

Received: 18 April 2010 / Revised: 27 October 2010 / Accepted: 4 November 2010 / Published online: 4 December 2010
© Dt. Ornithologen-Gesellschaft e.V. 2010

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

Keywords Intra-sexual selection · Bird song · Song complexity · Playback experiment · *Turdus merula*

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

Communicated by T. Friedl.

N. Hesler (✉) · T. Dabelsteen
University of Copenhagen, Universitetsparken 15,
2100 Copenhagen, Denmark
e-mail: nhesler@bio.ku.dk

R. Mundry
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, 04103 Leipzig, Germany

Introduction

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

In classic speaker replacement experiments, territorial males were removed and speakers playing back conspecific song were installed in their territories. In both Great Tits

(*Parus major,* Krebs et al. 1978) and Red-winged Blackbirds (*Agelaius phoeniceus*; Yasukawa, 1981), playback containing a repertoire of different song types was more efficient in keeping out intruders than playback comprised of a single song type. However, a common problem of these studies is the fact that the single song type playbacks contained single songs from one individual while the repertoire playbacks consisted of several song types from several different individuals. Assuming that individuality may be encoded in song (Weary and Krebs 1992; but see also Gentner et al. 2000) and can be decoded by a receiver, the test birds of these experiments may have perceived the repertoire playback as song originating from multiple birds rather than one bird with a large repertoire. As a quality signal, repertoire size could be used by males to estimate other males' fighting ability in actual encounters. Only a few studies have tested the direct response to different repertoire sizes, leading to ambiguous results. In a two-speaker experiment, male territorial Field Sparrows *Spizella pusilla* approached closer to the speaker playing larger repertoires (Searcy 1983), and male Song Sparrows *Melospiza melodia* reacted more persistently to playback of larger repertoires (Stoddard et al. 1988). However, in other experiments—even one conducted with Song Sparrows—the reaction of male birds did not vary in accordance with different repertoire sizes (Song Sparrow, Searcy 1983; Sedge Warbler *Acrocephalus schoenobaenus*, Catchpole 1989). In species with discontinuous song (i.e. song which is structured into discrete songs and pauses), repertoire size is usually measured as the number of different song types (a song type always contains the same element types, which are the shortest song unit). In the Chaffinch *Fringilla coelebs*, a species with a small song type repertoire, within-song complexity is meaningful to both females and males (Leitão et al. 2006). Also, Balsby and Dabelsteen (2001) found that male Whitethroats *Sylvia communis* did not respond to differences in repertoire size, but song length, and suggested that in this species the evolution of repertoire size is not driven by male competition.

These inconsistent results may have several reasons. First, a territory owner's response towards an intruder may not only depend on the intruder's fighting ability but also on the territory owner's fighting ability relative to that of the intruder (Searcy and Beecher 2009). Therefore, a bird with a small repertoire may react highly aggressively towards a small repertoire intruder, but choose a different strategy when a large repertoire male is intruding, who would probably be a superior rival in a fight. A solution to this problem is to take the test birds' own repertoire sizes into consideration when evaluating their response to playbacks of different repertoire sizes, as was by done Balsby and Dabelsteen (2001). Second, different songbird species may use a diversity of singing styles, ranging from continuous to discontinuous singing, with repertoire sizes at different levels of song organisation (song vs element). To date, most studies on song complexity concern species with either small song type repertoires or no song types at all [for example: Sedge Warbler 30–55 syllables, Great Reed Warbler *Acrocephalus arundinaceus* 10–20 syllables (Catchpole 1980); Great Tit 2–5 song types, Red-winged Blackbird 2–6 song types, Song Sparrow 5–11 song types (Beecher et al. 2000); but also Common Nightingale *Luscinia megarhynchos* 150–230 song types (Kipper et al. 2004)]. However, it is possible that a more clear effect of repertoire size can be seen in species with large repertoires.

Therefore, we chose to investigate the Common Blackbird *Turdus merula* (from now on called 'Blackbird'), which has a large repertoire and a complex singing style without fixed song types. Full song is organised into single songs (Strophen) of about 2 s, separated by silent intervals of ca. 3 s. A typical Blackbird's full song starts with a long-ranging whistle part consisting of 3–5 low-frequency elements (1.5–3 kHz) and ends with a twitter part comprising a more variable number of elements with a broader frequency range and on average higher frequency (ca. 10 kHz) and a considerably lower sound pressure level (SPL; Todt 1981; Dabelsteen 1984; Fig. 1). Entire songs are only occasionally repeated, but there are 'motifs', that is, fixed sequences of elements, which are always sung in the same order (Dabelsteen 1984, 2005; Rasmussen and Dabelsteen 2002). These motifs are combined in a flexible way within songs, and the fact that they repeatedly occur in different songs of a male gives them some resemblance with the song types of other species. Thus, in terms of song organisation, Blackbirds are between a discontinuous singing style with song type repertoires and a continuous singing style without any song types. Previous studies have shown that there are large inter-individual differences in motif repertoire size and that song is used in both intra- and inter-sexual contexts (Rasmussen and Dabelsteen 2002). Furthermore, in another study, we found correlations between body size and repertoire size in the Blackbird, indicating that repertoire size is a potential signal of male quality in this species (Hesler 2010).

In this study, we investigate the importance of repertoire sizes in the Blackbird in an intra-sexual context. In a playback experiment, we broadcast song with different element repertoire sizes (small vs large) to territorial males, and compared their responses, taking the test birds' own repertoire sizes into account. If males use repertoire size to estimate a rival's fighting ability, we expect test subjects to react differently to the different repertoire sizes, presumably with the strength of reaction increasing as the repertoire size broadcast increases. Furthermore, we predicted that the response may depend on the bird's own repertoire size, i.e. its own quality (Balsby and Dabelsteen 2001;

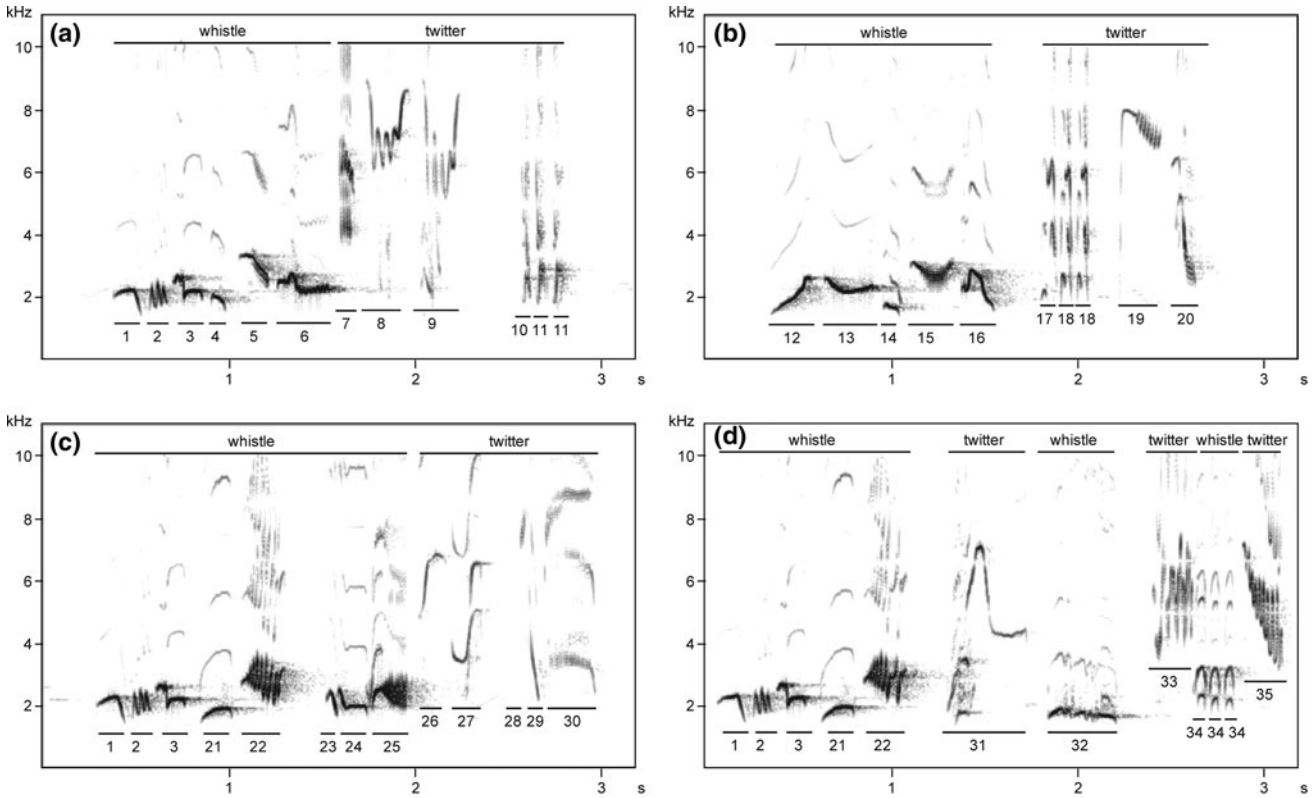


Fig. 1 **a–d** Spectrograms of four consecutive Blackbird songs with indication of whistle/twitter parts and single elements. *Numbers* indicate different element types. Note that some of the element types of song (**a**) occur also in the song (**c**, **d**)

Searcy and Beecher 2009). For example, a bird with a large repertoire size might react more strongly to playback of large repertoires, because they expect a stronger opponent. On the other hand, birds with small repertoires may react less strongly to playback of large repertoires, as they would expect an opponent who is stronger than themselves.

Methods

Study population

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

Song recordings and repertoire analyses

All recordings used were taken from colour-ringed males during dawn chorus ca. 1 h before sunrise (0340–0630 hours) in the early breeding seasons (8 April–10 June) of 2006–2008. All recorded males were in their third calendar year or older. These birds were distinguishable from second-calendar year males by a moult border in the greater coverts (Svensson 1992).

We used a directional microphone (Sennheiser K6/ME67) and a Marantz PMD 670 solid state recorder with a sample rate of 22,050 Hz (2006) or 44,100 Hz (2007, 2008) to record songs. Recordings in 44,100 Hz were resampled to 22,050 Hz. Avisoft-SASLab Pro 4.4–4.52 was used for all sound analyses. Recordings were filtered (Butterworth, high pass 1.5 kHz, low pass 10 kHz) and cut into single songs. Printed spectrograms were used for repertoire analyses (settings: FFT 512, window Hamming, bandwith 56 Hz, resolution 43 Hz, frame size 100%, overlap 75%, resolution 5.8 ms; graphical scale on prints: 1 s = 36 mm, 1 kHz = 6 mm). All repertoire analyses were carried out by the same person by visual inspection of the spectrograms (N.H.). We classified all elements in the recordings according to their overall appearance on the spectrogram (frequency and temporal characteristics). An element was defined as a single continuous trace in time without visible sound free pauses, i.e. pauses longer than ca. 0.015 s (Fig. 1). Intra-individual pause durations between two element types can be highly variable. If elements appeared at least once with a distinguishable pause, they were always considered as two elements, even if they

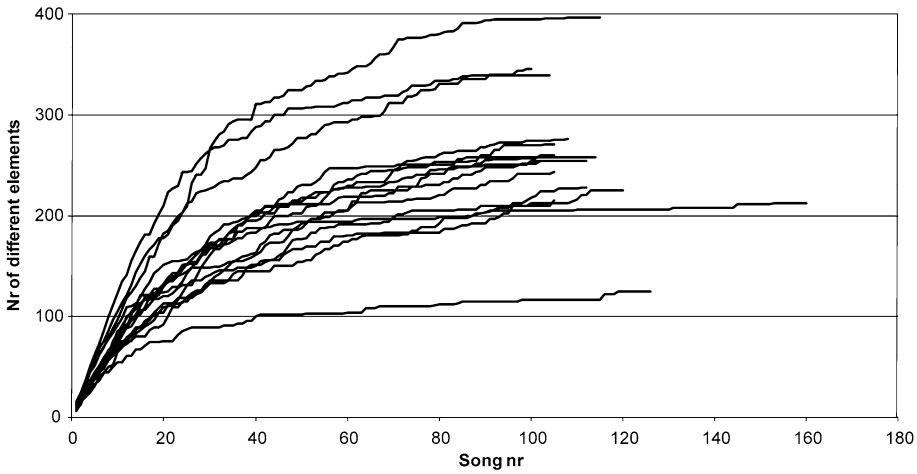


Fig. 2 Element repertoire sizes of test subjects. The number of different element types is depicted against the number of analysed songs

sometimes appeared as one continuous element. Repertoire sizes were determined by counting the number of different elements in 100 consecutive songs, including both whistle and twitter elements. We reasoned that it should be possible to adequately assess inter-individual differences on the basis of 100 songs, even if this does not always reflect a Blackbird's full repertoire (Fig. 2).

Construction of test files for playback

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

To create a large repertoire for playback, we first analysed a recording of $\geq$100 songs from a stimulus bird and then selected a sequence of 50 songs with the largest possible element repertoire. This was achieved by deleting songs from the sequence which contained only or mostly element types also present in other songs. The repertoire size in the derived sequence determined the repertoire size of the corresponding small repertoire size to be broadcast (ca. half the large repertoire size). Then, we constructed the playback sequence with the small repertoire, using the same procedure as above, except that this time songs were chosen such that they added the smallest available amount of new elements types to the sequence. These procedures resulted in large repertoire playbacks with repertoires slightly higher than in the original recording and small repertoire playbacks with repertoires considerably smaller. On average, the large repertoires had 237 different element types (range 190–276), and small repertoires 120 element types (range 85–152). Playback files were matched in such a way that the large repertoire was ca. twice as large as the small one. Recordings from birds with an originally small repertoire size were not used, since we would then create unnaturally small repertoires.

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

We matched the two recordings used for the same test bird in song duration and twitter part duration since these characteristics may influence response strength (Dabelsteen and Pedersen 1990; Balsby and Dabelsteen 2001). Also, after manipulation, songs in the large and small repertoire playback did not differ systematically in song length (mean of means $\pm$ SD; large playback repertoire: 2.69 s $\pm$ 0.28, small playback repertoire: 2.66 s $\pm$ 0.48; Wilcoxon's

matched pairs signed ranks test of the average song durations in the large versus small repertoire files from the same stimulus bird: $T^+ = 18$, $n = 8$, $P = 1$; and of average song durations of the two files which were played to the same test subject: $T^+ = 18$, $n = 8$, $P = 1$). The number of different element types per song and the ratio between whistle and twitter element types did not differ significantly after manipulation either (comparing songs in the small and large repertoire files from the same stimulus bird; number of element types: $T^+ = 30$, $n = 8$, $P = 0.11$; ratio of whistle to twitter element types: $T^+ = 25$, $n = 8$, $P = 0.38$).

Playback procedure

We conducted the playback experiments in spring 2008 (6 April–6 May, after dawn chorus and before noon) with 16 territorial males who were colour-ringed and at least in their third calendar year. Most were paired with a female which was in the nest building phase and very likely in her fertile phase. We chose this critical stage of the breeding cycle because we obtained very weak reactions during pilot experiments earlier in the season in 2007 in the same population. Each test bird was tested twice (small and large repertoire playback tests), on consecutive days (except for one bird that was tested with 3 days between tests). The order of treatments was balanced. We played only recordings from birds that had their territories in a different area of the island than the test bird throughout their lifetime. On the same day, we did not test birds in earshot of each other, and neither did we use playbacks from same stimuli birds for test subjects in earshot of each other. The test files were played back using a CD player (Sony CD D-NE330) and a Nagra DSM speaker. The output level was calibrated to maximum values of ca. 65 dB(A) at 10 m (Brühl and Kjær 2206, fast setting), which is the natural level of Blackbird full song (Dabelsteen 1981). We placed the speaker in the test subject's territory. Due to the often very small and adjacent territories on Helgoland, neighbouring males often interfered during the experiments; therefore, we placed the speaker centrally in the territory, i.e. between ca. 10–20 m from the territory border (as most territories were in the residential areas of the island, we were somewhat restricted in positioning the speakers). The speaker was placed at the same position for both experiments, put on a suitable object in a height between ca. 20 and 150 cm. Most birds on the island are very well habituated to humans and used to human activity in their territories, so test birds rarely seemed disturbed when we set up the speaker. However, we waited at least 3 min before starting the experiments. We started playback while the test bird was in sight.

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

Song and other vocalisations were also recorded, but occurred too rarely to provide data for analysis. Only twitter song was elicited frequently by playback. This is a different singing style which resembles the twitter part in full song and is used in aggressive interactions (Dabelsteen et al. 1998; Klump et al. 2010). However, due to the low amplitude of this vocalisation and the relatively high background noise level during experiments, it was not possible to quantify it reliably. When a test bird was interacting with another bird, we continued the playback but did not use behaviours during this interaction as data—five test subjects had interactions with females, one subject had an interaction with a male. We attempted to increase the sample size by testing six further males, but these data could not be used due to frequent interruption by neighbouring males.

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

Statistical analyses

We analysed data separately for the period during the playback and the 2 min afterwards. During playback, we used ten response variables: rates (n/s) of the behaviours *flight*, *tail up*, *bow*, latency and mean timing (s) of the behaviours *flight* and *tail up*. 'Mean timing' reflects the timing of the behaviours. It was calculated as the mean of all times of occurrence of the respective behaviour (in seconds after start of playback) during the observed period. We used three distance measurements: mean and minimum distance to the speaker (m) and distance slope, which is the slope of a regression between distance and time. The latter was used in order to estimate the spatial behaviour relative

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

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

We tested the influence of the playback on the subjects' behaviours using Generalized Linear Mixed Models (GLMM; Baayen 2008a). We included as fixed effects (1) stimulus repertoire size, (2) stimulus order, (3) experiment number (first or second experiment), and (4) the mean duration of the respective playback songs. We also included the test subject's own repertoire size and its interaction with the stimulus repertoire size as an additional fixed effect. Since the repertoire size of 1 of the 16 test subjects was unknown, we ran the model once including this effect and once excluding it (but achieving a larger sample size). As random effects, we initially included the ID of the tested individual and the individual from which the broadcast songs were recorded. However, we removed the latter since it appeared to be insignificant (all likelihood ratio tests except for one $P > 0.5$, one $P = 0.20$).

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

Several of the fixed effects were considered to be 'control variables' (i.e. stimulus order, experiment number,

Table 1 Results of the PCA for data obtained during and after playback

	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/s)	Sqrt	**−0.82**	0.04	0.10	–	–	–
Minimum distance (m)	–	**0.80**	−0.36	0.05	–	**0.96**	−0.09
Rate of flight (1/s)	–	−0.06	**0.84**	0.08	Sqrt	−0.10	**0.89**
Rate of tail up (1/s)	–	−0.18	**0.75**	−0.19	Sqrt	0.13	**0.88**
Latency flight (s)	Log	−0.05	**−0.68**	0.55	–	–	–
Mean timing tail up (s)	–	−0.05	0.37	**0.78**	–	–	–
Mean timing flight (s)	–	0.25	−0.31	**0.76**	–	–	–
Latency tail up (s)	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

The Kaiser–Meyer–Olkin measure of sampling adequacy (K–M–O) and Bartlett's test of sphericity (Bartlett's χ^2, *df* and *P*) justified both PCAs. Response variables which contributed most to the respective components (PC1, PC2, PC3) are in bold. After playback the behaviours *bow, flight* and *tail up* occurred too rarely to include them in the analysis.

Transf. Variables were transformed to achieve normality or uniformity if needed according to visual inspection, *log* log transformed, *sqrt* square root transformed

mean duration of the respective playback songs). We were not interested in their statistical significance per se but simply wanted to control for potential effects of these variables. If these were non-significant ($P > 0.25$), we removed them from the model prior to testing for the effects of the subject's own repertoire size and its interaction with the stimulus repertoire size.

We tested for the significance of the random effects using likelihood ratio tests (Dobson 2002; Bolker et al. 2009) based on models fit using Maximum Likelihood (rather than restricted Maximum Likelihood). To test the significance of fixed effects, we used Monte-Carlo Markov-Chain (MCMC) analysis (Baayen 2008a; Bolker et al. 2009). We checked for homogeneous variance and approximate normal distribution of residuals by visually inspecting residuals plotted against predicted values. We also checked for interrelations between predictor variables using Variance Inflation Factors (VIF). We also calculated VIFs for dummy coded stimulus number (Field 2005). The largest VIF was 1.1, ruling out collinearity issues. To achieve reliable tests of their interaction, we *z*-transformed repertoire size of the test subjects and playback, respectively, to a mean of zero and a standard deviation of one (Aiken and West 1992).

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

Results

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

The repertoire size of the playback files had no significant influence on the test birds' responses, nor was there an interaction between the test birds' and playback repertoire sizes. Also, none of our 'control variables' (the mean duration of the respective playback songs, stimulus order, number of experiment, or the identity of the stimulus bird and the test subject itself) had a recognizable effect on any of the responses. This was the case both when we did not take the subjects' own repertoire sizes into consideration (Table 2) and when we did (Table 3).

Discussion

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

In many other studies, very similar methods have been used in playback experiments with songbirds and even Blackbirds (e.g. Dabelsteen 1982; Dabelsteen and Pedersen 1985, 1990; Stoddard et al. 1988; Balsby and Dabelsteen

Table 2 GLMM results of the playback experiments

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

Given are *P* values (*MCMC*) and 95% confidence interval limits (*lower CI, upper CI*) of final models *without considering the subjects' own repertoire sizes* (one fitted for each response variable). The predictor variable, repertoire size of playback stimulus, did not have a significant effect on any of the response variables. Indicated *P* values are without correction for multiple testing. For details of the PC factors, see Table 1

Test Response variables obtained during playback, *after* response variables obtained in the 120 s after playback, *n* sample size

Table 3 GLMM results of the playback experiments

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

Given are *P* values (*MCMC*) and 95% confidence interval limits (*lower CI, upper CI*) from final models *including the subjects' own repertoire sizes* (one fitted for each response variable). None of the predictor variables had a significant influence on any of the response variables. For details of the PC factors, see Table 1. Since *after PC 2* was involved in a marginally non-significant interaction, the values for the involved variables *repertoire size of stimulus* and *repertoire size of test subject* are not indicated. Indicated *P* values are without correction for multiple testing

Test/after Response variables obtained during/in the 120 s after the experiment, *n* sample size

2001; Briefer et al. 2008) and our test subjects showed strong responses to our playbacks. The rates of the recorded behavioural patterns were comparable to those observed by Dabelsteen (1982) in a playback experiment with Blackbirds. We therefore think that the general design which we used should be suitable to detect differences in response if there were any.

A crucial question is whether the test birds were able to assess the differences in repertoire sizes in our experiments at all. We presented the stimuli over a relatively short time and the repertoire sizes in the 50 songs of our playback were much smaller than a normal full repertoire of a Blackbird, in which new element types can occur even after 300 consecutive songs (unpublished data). However, differences in repertoire sizes of the two stimuli were very clear, and if it is indeed necessary to hear the full repertoire of a rival in order to assess his quality, it would not be feasible to use repertoire sizes for quality assessment in short encounters. In our experiments, the test birds had to hear at least ca. 100 s of playback (corresponding to ca. 20 songs) in order to have the opportunity to detect differences in repertoire sizes. In most cases, the test birds started responding earlier, meaning that their first response cannot have been based on the repertoire size broadcast. However, during the rest of the playback and in the period afterwards, they could have adjusted their response if the perceived threat was dependent on repertoire size, as described by Stoddard et al. (1988) for Song Sparrows. Nevertheless, even if the test birds were able to distinguish between the two repertoire sizes, the magnitude of response was similar during the broadcasting of both repertoire sizes.

This finding may be due to a ceiling effect. In our experiment, the test birds were confronted with the simulation of an unknown intruder singing full song from within their territory—usually during a period in which the subjects' females were likely to be fertile (before/during the egg laying period). The strong immediate responses indicate that the test birds perceived the playback as a strong threat. We chose this design as we experienced very weak reactions in pilot experiments, which were carried out in 2007 but later in the season and thus in a less critical stage of the breeding cycle. To the territory holder, such an intrusion may represent the risk of losing parts of his territory or his female engaging in extra-pair copulations with the unknown intruder. Furthermore, the reaction of a male to an intruder could also influence the mate choice of observing females (Otter et al. 1999; Mennill et al. 2002). Therefore, this is possibly a situation in which the subjects always react persistently with full strength, without losing time in assessing the repertoire size broadcast. In such situations, birds may rather rely on quality signals which can be immediately evaluated. In other territory intrusion experiments with Blackbirds, males responded more strongly to song playback with longer twitter part duration (Dabelsteen and Pedersen 1990), and in a different

experiment, they reacted more strongly to a model bird with a yellow or orange beak than to models with brown beaks, which are typical for yearling males (Bright and Waas 2002). Unlike repertoire size, both twitter part duration and beak colour can be assessed immediately.

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

Our negative result does of course not rule out the possibility that males evaluate repertoire size in other situations. For example, a territory owner might listen to neighbours singing outside the territory and use their repertoire sizes for quality assessment, which could influence decisions about future interactions like investment in territorial fights or in mate guarding.

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

It is also possible that not only repertoire size per se is the selected song trait but also how it is used and organised: even birds with same element repertoire sizes could differ in sequential organisation or frequency distribution of their elements, or their ability of repertoire matching (Beecher et al. 2000). It could also be that special element types, which are energetically demanding to produce, are the selected trait. In Canaries *Serinus canaria*, it has been found that females prefer 'sexy syllables' which are energetically demanding to produce (Vallet and Kreutzer 1995; Leitner et al. 2006). Other performance-dependent song traits have been identified, which were linked to male quality and/or preferred by females, as syllable type or trill consistency (Tropical Mockingbird *Mimus gilvus*; Botero et al. 2009; Banded Wren *Thryothorus pleurostictus*; de Kort et al. 2009) and the performance of rapid, broadband trills (Swamp Sparrow *Melospiza georgiana*; Ballentine 2009), or song traits which depend on conditions during early development, as syntax accuracy (Zebra Finch *Taeniopygia guttata*; Brumm et al. 2009; Holveck et al. 2008). In our design, we created both large and small playback files from the same recording, which was always from an originally large repertoire bird, since we could not enlarge repertoire sizes, but only reduce them. We did this in order to avoid the influence of other song parameters correlated with repertoire size. However, this also means that song traits other than repertoire size were present (or absent) in both test files. If these effects were clearly stronger than repertoire size, responses towards different stimulus birds would have differed, which was not the case. However, if both repertoire size and other intrinsic song traits have an effect at the same time, an indifferent response could result. An experimental design using both large and small repertoire from the same stimulus bird for the same test bird would avoid this problem. On the other hand, in the second experiment the test bird may then recognise the identity of the stimulus bird (Weary and Krebs 1992) from the first experiment and base its response partly on that. Therefore, we decided to use stimuli from different individuals for one test bird.

Possibly the two distinct parts of the Blackbird's song, the whistle and the twitter, have different functions, as they also have different acoustic properties (the whistle part has a much further transmission range than the twitter part and is more omni-directional; Larsen and Dabelsteen 1990; Dabelsteen et al. 1993). This means that they reach different parts of the habitat. By using the habitat in different ways, female and male receivers may thus not be equally exposed to the whistle and twitter part (Dabelsteen and Pedersen 1988) so that inter- and intra-sexual selection may act differently on these parts. This could be tested by

manipulating only one of the two song parts in a playback experiment (Leitão and Riebel 2003).

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

Acknowledgments We thank T. Sacher, T. Coppack and F. Bairlein for collaboration and O. Hüppop with the team of the Vogelwarte Helgoland and C. Horn for assistance during the field work. We also thank I. Teschke for language improvement and two anonymous reviewers for their helpful comments. The study was funded by a combined PhD fellowship from NaFöG (Promotionsstipendium nach dem Nachwuchsförderungsgesetz des Landes Berlin, Germany), SNAK (Graduate School on Sense organs, Nerve systems, Behaviour and Communication, funded by the Danish Research Agency) and the University of Copenhagen to N.H. and a framework grant from the Danish Research Council for Nature and Universe (no. 272-07-0477) to T.D. This study meets the guidelines for the treatment of animals in behavioural research and teaching of the Association for the Study of Animal Behaviour and complies with laws in Germany and Denmark.

References

Aiken LS, West SG (1992) Multiple regression: testing and interpreting interactions. Sage, Newbury Park

Baayen RH (2008a) Analyzing linguistic data: a practical introduction to statistics using R. Cambridge University Press, Cambridge

Baayen RH (2008b) LanguageR: data sets and functions with "analyzing linguistic data: a practical introduction to statistics". R package version 0.953

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

Balsby TJS (2000) The function of song in whitethroats *Sylvia communis*. Bioacoustics 11:17–30

Balsby TJS, Dabelsteen T (2001) The meaning of song repertoire size and song length to male whitethroats *Sylvia communis*. Behav Process 56:75–84

Bates D, Maechler M, Dai B (2008) lme4: linear mixed-effects models using S4 classes. R package version 0.999375-31

Beecher MD, Campbell SE, Nordby JC (2000) Territory tenure in song sparrows is related to song sharing with neighbours, but not to repertoire size. Anim Behav 59:29–37

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135

Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL (2009) Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Anim Behav 77:701–706

Briefer E, Aubin T, Lehongre K, Rybak F (2008) How to identify dear enemies: the group signature in the complex song of the skylark *Alauda arvensis*. J Exp Biol 211:317–326

Bright A, Waas JR (2002) Effects of bill pigmentation and UV reflectance during territory establishment in blackbirds. Anim Behav 64:207–213

Brumm H, Zollinger SA, Slater PJB (2009) Developmental stress affects song learning but not song complexity and vocal amplitude in zebra finches. Behav Ecol Sociobiol 63:1387–1395

Buchanan KL, Catchpole CK, Lewis JW, Lodge A (1999) Song as an indicator of parasitism in the sedge warbler. Anim Behav 57:307–314

Catchpole CK (1980) Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*. Behaviour 74:149–165

Catchpole CK (1989) Responses of male sedge warblers to playback of different repertoire sizes. Anim Behav 37:1046–1047

Catchpole CK, Slater PJB (2008) Bird song—biological themes and variations, 2nd edn. Cambridge University Press, Cambridge

Collins SA (2004) Vocal fighting and flirting: the functions of birdsong. In: Marler P, Slabbekoorn H (eds) Nature's music. Elsevier Academic Press, London, pp 39–79

Dabelsteen T (1981) The sound pressure level in the dawn song of the blackbird *Turdus merula* and a method for adjusting the level in experimental song to the level in natural song. Z Tierpsychol 56:137–149

Dabelsteen T (1982) Variation in the response of freeliving blackbirds *Turdus merula* to playback of song: I. effect of continuous stimulation and predictability of the response. Z Tierpsychol 58:311–328

Dabelsteen T (1984) An analysis of the full song of the blackbird *Turdus merula* with respect to message coding and adaptations for acoustic communication. Ornis Scandinavica 15:227–239

Dabelsteen T (2005) Public, private or anonymous? Facilitating and countering eavesdropping. In: McGregor PK (ed) Animal communication networks. Cambridge University Press, Cambridge, pp 38–62

Dabelsteen T, Pedersen SB (1985) Correspondence between messages in the full song of the blackbird *Turdus merula* and meanings to territorial males, as inferred from responses to computerized modifications of natural song. Z Tierpsychol 69:149–165

Dabelsteen T, Pedersen SB (1988) Do female blackbirds, *Turdus merula*, decode song in the same way as males? Anim Behav 36:1858–1859

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

Dabelsteen T, Larsen ON, Pedersen SB (1993) Habitat-induced degradation of sound signals: quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song. J Acoust Soc Am 93:2206–2220

Dabelsteen T, McGregor PK, Lampe HM, Langmore NE, Holland J (1998) Quiet song in song birds: an overlooked phenomenon. Bioacoustics 9:89–105

de Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL (2009) Trill consistency is an age-related assessment signal in banded wrens. Proc R Soc Lond B 276:2315–2321

Dobson AJ (2002) An introduction to generalized linear models. Chapman & Hall/CRC, Boca Raton

Eens M, Pinxten R, Verheyen RF (1993) Function of the song and song repertoire in the european starling (*Sturnus vulgaris*)—an aviary experiment. Behaviour 125:51–66

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

Gentner TQ, Hulse SH, Bentley GE, Ball GF (2000) Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. J Neurobiol 42:117–133

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

Hiebert SM, Stoddard PK, Arcese P (1989) Repertoire size, territory acquisition and reproductive success in the song sparrow. Anim Behav 37:266–273

Holveck MJ, de Castro ACV, Lachlan RF, Ten Cate C, Riebel K (2008) Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. Behav Ecol 19:1267–1281

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

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

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

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

Larsen ON, Dabelsteen T (1990) Directionality of blackbird vocalization. Implications for vocal communication and its further study. Ornis Scandinavica 21:37–45

Leitão A, Riebel K (2003) Are good ornaments bad armaments? Male chaffinch perception of songs with varying flourish length. Anim Behav 66:161–167

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

Leitner S, Marshall RC, Leisler B, Catchpole CK (2006) Male song quality, egg size and offspring sex in captive canaries (*Serinus canaria*). Ethology 112:554–563

McGregor PK (1992) Quantifying responses to playback: one, many, or composite multivariate measures? In: McGregor PK (ed) Playback and studies of animal communication. Plenum Press, New York and London, pp 79–96

Mennill DJ, Ratcliffe LM, Boag PT (2002) Female eavesdropping on male song contests in songbirds. Science 296:873

Mountjoy DJ, Lemon RE (1996) Female choice for complex song in the European starling: a field experiment. Behav Ecol Sociobiol 38:65–71

Otter K, McGregor PK, Terry AMR, Burford FRL, Peake TM, Dabelsteen T (1999) Do female great tits (*Parus major*) assess males by eavesdropping? A field study using interactive song playback. Proc R Soc Lond B 266:1305–1309

R Development Core Team. (2008) R: a language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria

Rasmussen R, Dabelsteen T (2002) Song repertoires and repertoire sharing in a local group of blackbirds. Bioacoustics 13:63–76

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

Searcy WA (1983) Response to multiple song types in male song sparrows and field sparrows. Anim Behav 31:948–949

Searcy WA, Beecher MD (2009) Song as an aggressive signal in songbirds. Anim Behav 78:1281–1292

Stoddard PK, Beecher MD, Willis MS (1988) Response of territorial male song sparrows to song types and variations. Behav Ecol Sociobiol 22:125–130

Svensson L (1992) Identification guide to European passerines. Svensson, Stockholm

Todt D (1981) On functions of vocal matching: effect of counter-replies on song post choice and singing. Z Tierpsychol 57:73–93

Vallet E, Kreutzer M (1995) Female canaries are sexually responsive to special song phrases. Anim Behav 49:1603–1610

Weary DM, Krebs JR (1992) Great tits classify songs by individual voice characteristics. Anim Behav 43:283–287

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