The effects of rearing conditions on sexual traits and preferences in zebra finches

Holveck, M.J.

Citation

Holveck, M. J. (2008, February 28). *The effects of rearing conditions on sexual traits and preferences in zebra finches*. Retrieved from https://hdl.handle.net/1887/12621

Chapter 2

Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts

Marie-Jeanne Holveck and Katharina Riebel

In *Animal Behaviour*, 2007, 74, 297-309

Chapter 2

ABSTRACT

Male mating signals are often multidimensional, potentially providing multiple messages to females. However, the relative importance of different signal dimensions and their context dependency are poorly understood. Even in a well-studied species such as the zebra finch, *Taeniopygia guttata*, an important avian model for the study of mate choice, there is little consensus on the relative weighting of visual versus acoustic signals in mate choice. We therefore tested the consistency and repeatability of female mating preferences across different test contexts, presenting male song only or full courtship displays. We concurrently conducted a detailed analysis of male song characteristics and morphological traits. Females' individual preferences were consistent across three commonly used binary test paradigms (operant and phonotaxis tests with songs and association tests with live males). Preference direction was thus independent of test contexts. Preference strength was repeatable only between the operant and live male tests, possibly because these two tests allowed active interaction with songs or males whereas exposure to songs in the phonotaxis test was passive. The song structure parameters that predicted female preferences best were context independent and also predicted male morphology. We conclude from the combined results that song structure (in addition to song rate or absolute output as previously suggested) does contain sufficient information on the singer for female mate choice. We suggest that the earlier focus on song rate rather than song content might partly account for the differences between studies in the importance attributed to acoustic versus visual signals.

KEYWORDS: context-repeatable mating preference, male choice test, mate choice, multiple signals, operant test, phonotaxis test, song structure, *Taeniopygia guttata*, zebra finch.

Despite intensive research on female mate choice and the evolution of secondary sexual traits over the past few decades (Andersson, 1994), surprisingly little is known about within-population variation in female preference (Jennions and Petrie, 1997; Widemo and Saether, 1999). Females' mating decisions are often based on multidimensional signals providing a wide range of messages and involving different sensory modalities (e.g. acoustic and visual) propagated on different temporal and spatial scales (Candolin, 2003; Hebets and Papaj, 2005). Two types of not mutually exclusive functional hypotheses have been suggested to account for multiple ornaments and mating signals: they could provide (1) back-up messages or (2) multiple messages on different aspects of male quality. Within-population variation in female mating preferences (i.e. the response to sample stimuli) can arise through variation in condition or context and/or genetic, cultural or phenotypic compatibility (Jennions and Petrie, 1997; Qvarnström, 2001; Widemo and Saether, 1999). If females within a population differ in whether they predominantly look for direct or indirect benefits (e.g. resource-holding potential versus genetic benefits), they may pay attention to different signals to choose the most suitable male (Candolin, 2003). Differences in female mating preferences are also likely to be influenced by social factors such as the intensity of male-male or female-female competition (Jennions and Petrie, 1997; Widemo and Saether, 1999). For instance, mate density may affect female choosiness (i.e. the time and effort the female is prepared to invest in finding and assessing mates), the cost of sampling and sampling strategies (i.e. the decision rule adopted in mate assessment).

Both the multiple message and back-up signal scenarios might lead to context dependency of the weighting of particular signals (Candolin, 2003; Wagner, 1998), a poorly understood issue (Candolin, 2003; Jennions and Petrie, 1997). It has rightly been pointed out that different test methods might inadvertently lead to context-specific weighting of different signals. For example, they may allow different levels of interaction between males and females (Waas and Wordsworth, 1999), which may influence which traits females pay more attention to. Hence, some of the documented variation in female preferences might not arise from differences between females or populations but might be an artefact arising from the wide range of different methods used to measure female mating preferences (Wagner, 1998). One of our aims in this study was to examine the impact of the test method on estimated female preferences in zebra finches, *Taeniopygia guttata*.

In songbirds, one of the most intensively studied taxa in current studies of mate choice behaviour, there is ample evidence that male song is important for female choice (Catchpole and Slater, 1995; Searcy and Yasukawa, 1996). However, the relative weighting of song versus other phenotypic features of males such as morphological traits and display intensity remains poorly understood (Collins et al., 1994; Patricelli et al., 2003). Furthermore, surprisingly little is known about which structural features of a song determine its attractiveness to females, and how such features might relate to male quality. The current literature on mating preference in the zebra finch, a species that has been intensively studied experimentally in the laboratory, illustrates rather well how difficult it is to test the relative importance in female mate choice of acoustic and visual signals provided by male courtship displays such as song rate and beak colour (Collins, 1994; Collins and ten Cate, 1996; Forstmeier and Birkhead, 2004; ten Cate and Mug, 1984; Zann, 1996). Zebra finches therefore provide a good model with which to compare the consistency and the repeatability of female mating preferences as well as to examine what exact attributes make songs attractive to females and what the preferred song features say about the singer. To this end we tested females across three different commonly used test paradigms involving single or multiple sensory modalities. An operant test with song as reinforcer (Houx and ten Cate, 1999; Leadbeater et al., 2005; Riebel, 2000; Riebel et al., 2002) and a phonotaxis test both tested preferences for the acoustic signal alone (Clayton, 1988; Miller, 1979a; Miller, 1979b; Neubauer, 1999). A spatial association test with a choice between two live males presented a test situation where the acoustic signal was combined with additional static and dynamic visual and behavioural signals. The latter is by far the most commonly used mate choice test type (reviewed in Forstmeier and Birkhead, 2004). We assessed both preference consistency (identical direction) and repeatability (identical strength) of the within-individual preference for a specific stimulus in different test contexts. Subsequently, we compared the outcome of an in-depth song analysis with female song preferences and male morphological traits to test whether structural song parameters can predict female preferences on the one hand and male quality on the other.

If multiple signals act as multiple messages then we expect more consistency and repeatability in female preferences when comparing the two tests involving song only (operant and phonotaxis tests) than when comparing tests involving single versus multiple sensory modalities (i.e. phonotaxis or operant versus live male tests). However, if multiple signals act as back-up signals, females should not differ in their preference direction between the tests involving song only and the live male tests.

SUBJECTS AND REARING CONDITIONS

We used wild-morph domesticated zebra finches (N = 35, 17 males and 18 females) from an outbred breeding colony at Leiden University, The Netherlands. The birds were kept on a 13.5:10.5 h light:dark schedule (lights on between 0700 and 2030 hours CET) at 20-22ºC and 35-50% humidity. Birds had ad libitum access to a commercial tropical seed mixture enriched with GistoCal mineral and vitamin powder (Beaphar B.V., Raalte, The Netherlands), drinking water and cuttlebone. This basic diet was supplemented three times a week with 3-4 g of 'egg food' (Witte Molen, B.V., Meeuwen, The Netherlands) per bird, twice a week with branches of millet and once a week with germinated tropical seeds. All subjects had been reared by their parents in standard laboratory cages (80 x 40 cm and 40 cm high) equipped with a nestbox. At 65 days (i.e. after the peak of the sensitive period for song learning, Slater et al., 1988), they were moved to be housed in single-sex groups with eight to nine individuals per cage (100 x 60 cm and 60 cm high). All subjects were about 2 years old (20 ± 5 months, N = 35) and had no breeding experience when the experiments started. Subjects that encountered each other in preference tests were unfamiliar to each other and had a coefficient of relatedness less than 0.125.

PREFERENCE CONSISTENCY AND REPEATABILITY

Methods

Stimulus preparation

Following Sossinka and Böhner (1980), we call an individual's specific syllable sequence 'the motif'. A 'song' consists of a series of introductory syllables followed by several repetitions of the motif (range 1-10 for nondirected songs). For the preparation of the stimulus songs, we recorded nondirected songs of 17 males. Males were placed singly in a cage (70 x 30 cm and 45 cm high) on a wooden shelf (100 x 55 cm) at a height of 120 cm in a sound attenuation chamber (100 x 200 cm and 220 cm high). Songs were recorded at a distance of 75 cm from the cage (Sennheiser MKH40 microphone, Wedemark, Germany and Sony TCD5 Pro II cassette recorder, Tokyo, Japan). Songs were digitized (25 000-Hz sample rate) using Signal/Rts software

(Engineering Design, Belmont, MA, U.S.A.) and a lowpass filter (cutoff frequency 10 000 Hz; Frequency Devices 900C/9L8B, Haverhill, MA, U.S.A.). We chose one natural song per male and digitally deleted those introductory syllables at the beginning that were additional to the number of introductory syllables that occurred as part of the motif within songs. All chosen songs had four motifs. Using the Praat sound analysis software (version 4.2.07 forWindows, freely available from http://www.praat.org) songs were highpass filtered at 500 Hz (smoothing = 100 Hz) to remove low-frequency background noise. Amplitudes were root mean-square equalized (peak digitally scaled to 1 with Praat software).We formed 18 unique stimulus dyads without combining the same two males twice. However, we assigned each male two (*N* = 15) or three times (*N* = 2) to a different stimulus dyad by matching song duration as much as possible (average difference in song duration between dyads ± SD = 0.41 ± 0.3 s).

Preference tests

Each of the 18 females was tested with one of the 18 unique stimulus dyads in the operant test, the phonotaxis test and the live male test. The test order was fully balanced with regard to test type. With a total of 18 females and six possible combinations, there were three females per possible test order. At least 7 days of rest separated two consecutive tests (time between tests 1 and 2 ± SD = 10.1 ± 3.8 days; time between tests 2 and 3 = 13.7 ± 6.8 days).

The experimental set-up of the operant test has been described in detail elsewhere (Houx and ten Cate, 1999; Riebel, 2000; Riebel and Smallegange, 2003; Riebel et al., 2002). Briefly, there were two different songs associated with two different response keys (Fig. 1a). Pecking either of the two keys triggered a playback of a song via a loudspeaker (Blaupunkt CB 4500, Hildesheim, Germany) with a maximum amplitude of 70 dB at 30 cm from the speaker (re 20 µPa; CEL-231 sound level meter, fast response F and low range A LO settings, Lucas CEL Instruments Ltd, Hitchin, Herts, U.K.). Experimental cages were placed singly in sound attenuation chambers (100 x 200 cm and 220 cm high) on a shelf (100 x 55 cm) at a height of 70 cm. Observation was possible through a one-way mirror. A custom-built minicomputer with an Oki MSM6388 (Tokyo, Japan) sound chip controlled the playback, automatically swapped the stimuli between the two keys each night and kept a data log. The tested females were transferred to the operant cages between 0830 and 0930 hours. The red

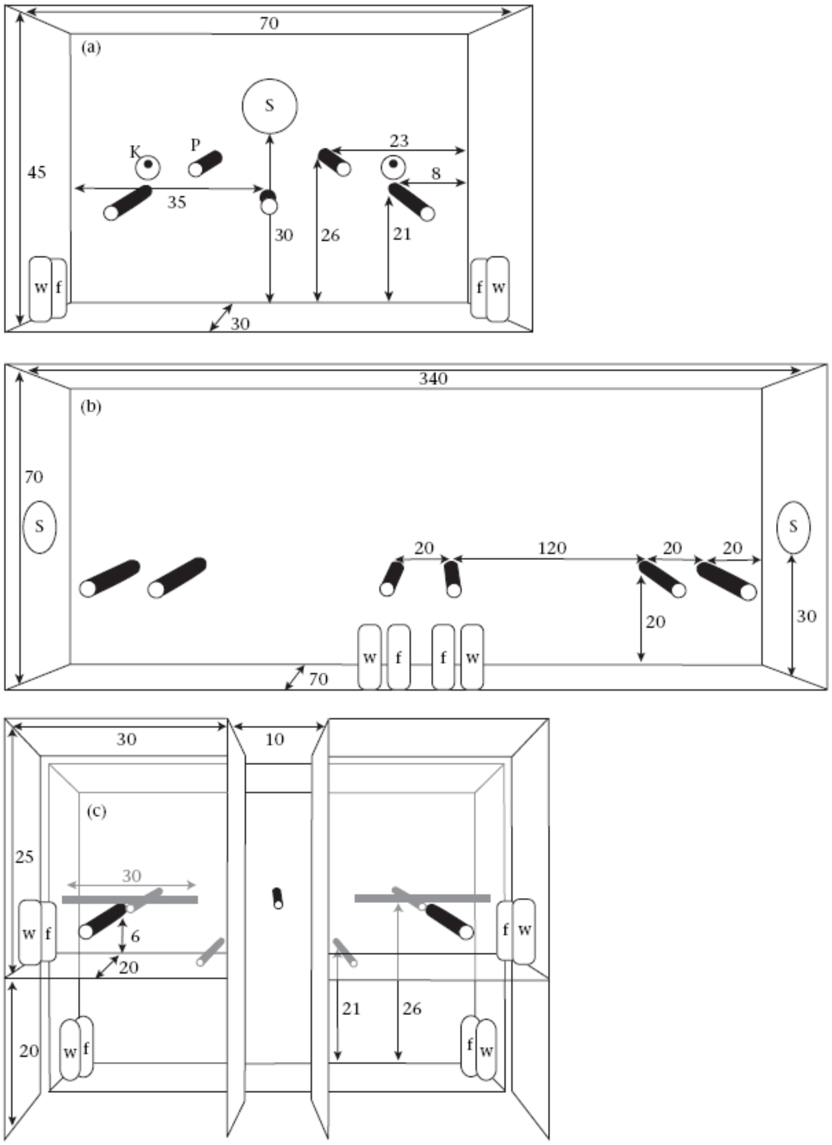


Figure 1. Schematic front views of the test designs used for (a) operant, (b) phonotaxis and (c) live male tests; (c) shows two side cages in front of a central test cage with the same dimensions as in the cage used for the operant test (except for perch height). Cages were made from plywood but with the long front side made from wire mesh. S: speaker; K: pecking keys; P: perch; w: water; f: food. All measures in cm.

LED lights of the keys were switched on only during the days of the training phase. Four of 18 females learned to peck the keys by autoshaping after they accidentally pecked the keys. Those that had not started to do so after 2 days underwent a training procedure (one or two daily sessions of about 20 min each for 1-8 days, $\overline{X}$ ± SE = 3.6 ± 2.3 days, N =13, one female did not learn to perform the operant task). During training, we first drew the attention of the birds to the keys by flashing the LED lights. We rewarded birds with playback (same songs as stimulus songs) when they perched near the key, then when exploring the plastic disk surrounding the key, until they finally started to peck the keys proper. Once a female pecked both keys regularly on the same day after the training session, the preference test started the next day and lasted 2 full days. As before, the stimuli were swapped between the two sides daily but the red LED lights were now switched off.

For the phonotaxis test, a long test cage (340 x 70 cm and 70 cm high; Fig. 1b) was placed at the far end of a rectangular experimental room. The size of the cage allowed the female to approach or evade the stimulus. All sides of the cage were solid plywood and padded with foam for sound attenuation, except the long front side which had wire mesh. Loudspeakers were attached to each end wall behind a central opening. The cage had six perches. Perching on the two outermost perches was counted as an approach (i.e. within 40 cm of the loudspeaker); the remainder of the cage including the floor was defined as neutral. Parallel to the long side of the test cage at a distance of 170 cm, a dark plastic partition with a small central opening (20 x 5 cm) divided the room and hid the observer and the playback equipment from the cage. We edited stimuli with Cool edit 2000 software version 1.0 for Windows (Syntrillium Software Corporation, Scottsdale, AZ, U.S.A.). The playback stimuli consisted of a 1-min sequence of four repetitions of the same song with 5-s silent intersong intervals. This song rate is within the naturally observed range for zebra finches (Sossinka and Böhner, 1980; Zann, 1996). The two songs assigned to a stimulus dyad were edited as one 2-min stereo file with one stimulus on the left channel (for the first minute) and the other one on the right (for the second minute). This allowed continuous alternating playback of the two stimuli via the two loudspeakers (same speaker and settings as in the operant test) during the 14-min test (see below) using the loop mode of the CD-player (Venturer DM8802-00, Venturer Electronics Inc, Markham, Ontario, Canada, U.K. and JVC AXR562BK stereo amplifier, JVC, Yokohama, Japan) We randomly assigned both the cage side for the first playback and the first stimulus used at the beginning of each test.

For the live male test, the test cage holding the female was situated in one of the sound attenuation chambers also used for the operant tests. Stimulus males were placed into two small cages in front of the long front side of the test cage (Fig. 1c) so that they could not see each other. The tested female could perch near either male or avoid associating with them altogether when perching in the centre or on the floor of her cage. The two small cages with the stimulus males could be obstructed from the female's view by a solid grey plastic panel that we could pull up by a string from outside the experimental chamber.

In both phonotaxis and live male tests, we moved the birds to their respective test cages to acclimatize the day before the preference test. We conducted test the following day between 0830 and 1230 hours. Each test started with a baseline observation period (range 15-17 min) without stimulus exposure. At this stage, the plastic panels were still lowered between side cages and test cage in the live male test (Fig. 1c). This was followed by 14 min with the plastic panels raised in the live male tests or with 14 alternating song playbacks (i.e. seven per stimulus) in the phonotaxis test. After a second stimulus-free period (range 15-17 min; lowered screen in the live male test) during which the stimuli were reversed between sides, another 14 min of stimulus exposure followed. Every 14 min, stimulus exposure started when the tested female was in the centre of the cage. Females had to visit at least one of the two stimuli in each 14-min test session and had to spend at least 10% of the total test time in front of either stimulus for the data to be included in the analyses. Of 18 females, 17 fulfilled these criteria.

The study was approved by the local Ethical Committee (Dierexperimentencommissie Universiteit Leiden).

Statistical analyses

We defined the preferred stimulus as the stimulus that was preferred in at least two of the three tests. Preference strength was set equal to the preference ratios (choices for the preferred stimulus divided by the total number of choices). Preference ratios were either the relative number of keypecks for the preferred song (operant test) or (for comparison with earlier studies) the relative time spent close to the preferred stimulus in the phonotaxis and live male tests. However, we were interested in seeing whether preferences in these two tests were also reflected by the number of times a female approached a specific stimulus. Thus for the phonotaxis and live male tests, we also calculated visit ratios (number of visits to the preferred stimulus divided by the total number of visits). Female activity level was defined as the absolute number of keypecks (operant test) or visits (phonotaxis and live male tests).We log-transformed activity level for statistical analyses. Visit intensity was expressed as the percentage change from the previous event, i.e. the number of switches between stimuli (N_s) in relation to the total number of visits (V) minus two to correct for the fact that the last visit of both test sessions (2 days for the operant test, two times 14 min for the phonotaxis and live male tests) could not be followed by a switch (visit intensity = $(100 \times N_s)/(V - 2)$). Visit intensity can range from 0% (only one stimulus received repeated visits within each session) to 100% (the visited stimulus changed at each visit). Visit time was calculated as the time spent near both stimuli as a percentage of the total test time (applicable only for the phonotaxis and live male tests). For statistical comparisons of the three preference tests, all proportion and percentage data (i.e. preference ratios, visit ratios, visit intensity and visit time) were normalized (calculating Z scores) to account for differences in mean and variance.

In our comparison of female preference across the three tests, we made a distinction between the consistency of preference direction (a nonparametric measure with binary scoring 0 or 1: the stimulus with the highest preference ratio was defined as preferred) and the repeatability of preference strength (quantitative measure of preference ratios). To test the consistency of preferences across the three tests we used Cochran Q test. The repeatability of preference ratios across the three tests was calculated following Lessells and Boag (1987) using a one-way ANOVA with preference ratios as the dependent variable and female identity as the between-subjects factor. The standard error of the repeatability estimate R was calculated as the square root of the sampling variance of the intraclass correlation (Becker, 1984). We also estimated repeatability between the three preference ratio sets two at a time resulting in a significance level of $\alpha = 0.025$ after a correction for multiple tests and compared them with the repeatability estimates of the other measures of female choice behaviour. Effects of test order (between-subjects factor) and type (within-subject variables) were tested on all measures of female choice behaviour with two-factor mixed ANOVAs. All statistical analyses were two tailed and calculated using SPSS statistical software, release 10.0.7 (SPSS Inc., Chicago, IL, U.S.A.); means are given ± 1 SD.

Results

Only one female did not respond in both the operant and phonotaxis tests and was excluded from further analyses. Female preferences were consistent across the three tests (Cochran *Q* test: $Q_2 = 6$, $N = 17$, exact $P = 0.11$; Table 1). Across all tests ($N = 51$, 17 females x 3 tests), there were only three instances where females changed the direction of their preferences. It was always the preference in the live male test that differed from the preferences in the operant and phonotaxis tests (i.e. the preference ratio changed from >0.5 to <0.5 for females 2, 5 and 12; Fig. 2, Table 1). Preference ratios were independent of test type and order and significantly repeatable between operant and live male tests ($R = 0.56$, Table 2).

Table 3 gives the means of all measures of female choice behaviour per test. Preference ratios were highly correlated with visit ratios both in the phonotaxis (Pearson correlation: $r_{15} = 0.71$, $P < 0.01$) and live male tests ($r_{15} = 0.75$, $P < 0.001$). None of the measures was affected by test order, except for visit ratio (calculated only for the phonotaxis and live male tests, see Methods and Table 2). Visit ratios increased with test order suggesting that a previous exposure to the stimuli (either to songs only or to males) subsequently reinforced the number of visits to the preferred stimulus; but note that for this comparison the operant test was not taken into account, which means that, depending on test order position of the operant test, comparisons had to be made between first and second, second and third or first and third test. Test type significantly affected activity levels with the highest value for the operant test (Table 2), but preference ratios were independent of activity levels in all three tests. Preference ratios were not correlated with number of keypecks (operant test: $r_{15} = -0.29$, $P = 0.26$) or number of visits (phonotaxis test: $r_{15} = -0.42$, $P = 0.10$; live male test: $r_{15} = -0.38$, $P = 0.14$).

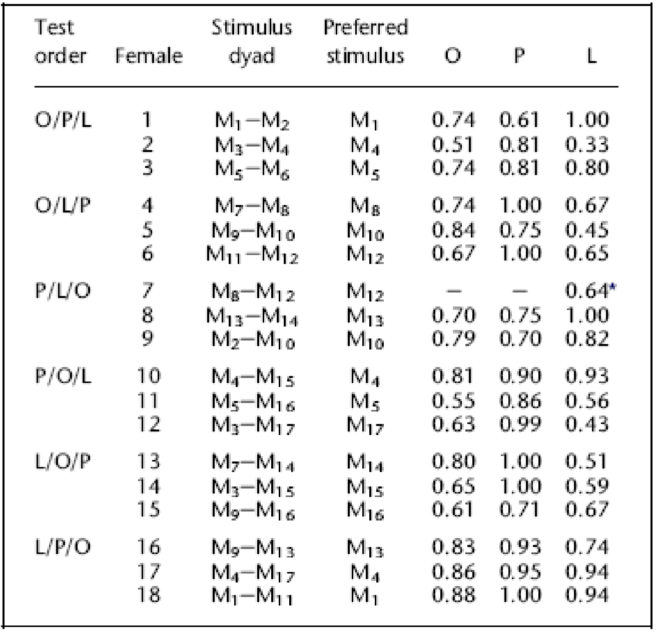
PREFERENCES AND MALE TRAITS

Methods

Song analysis

All songs were analysed blind to male identity by M.J.H. Table 4 and Fig. 3 explain the catalogue of song measures in detail. Sound density was assessed with a gating function (Signal/Rts software) that

Table 1. Preferred stimulus within dyads (i.e. stimulus preferred in at least two tests out of three) and associated preference ratios in the three test designs for the 18 tested females



Test order	Female	Stimulus dyad	Preferred stimulus	O	P	L
O/P/L	1	M_1–M_2	M_1	0.74	0.61	1.00
	2	M_3–M_4	M_4	0.51	0.81	0.33
	3	M_5–M_6	M_5	0.74	0.81	0.80
O/L/P	4	M_7–M_8	M_8	0.74	1.00	0.67
	5	M_9–M_{10}	M_{10}	0.84	0.75	0.45
	6	M_{11}–M_{12}	M_{12}	0.67	1.00	0.65
P/L/O	7	M_8–M_{12}	M_{12}	–	–	0.64*
	8	M_{13}–M_{14}	M_{13}	0.70	0.75	1.00
	9	M_2–M_{10}	M_{10}	0.79	0.70	0.82
P/O/L	10	M_4–M_{15}	M_4	0.81	0.90	0.93
	11	M_5–M_{16}	M_5	0.55	0.86	0.56
	12	M_3–M_{17}	M_{17}	0.63	0.99	0.43
L/O/P	13	M_7–M_{14}	M_{14}	0.80	1.00	0.51
	14	M_3–M_{15}	M_{15}	0.65	1.00	0.59
	15	M_9–M_{16}	M_{16}	0.61	0.71	0.67
L/P/O	16	M_9–M_{13}	M_{13}	0.83	0.93	0.74
	17	M_4–M_{17}	M_4	0.86	0.95	0.94
	18	M_1–M_{11}	M_1	0.88	1.00	0.94

O: operant test; P: phonotaxis test; L: live male test. M_1-M_{17}: 17 males used for song or live stimuli.

*This female is listed for completeness, but as the only nonresponder in two tests was excluded from the statistical analyses.

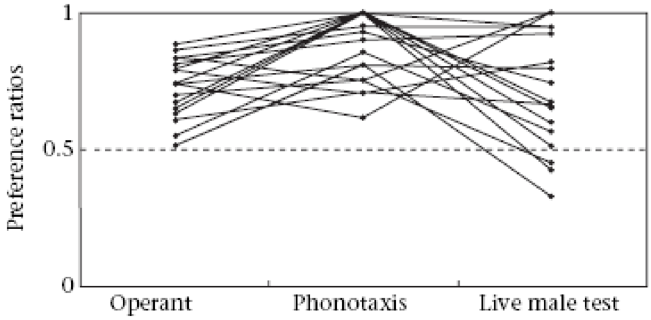
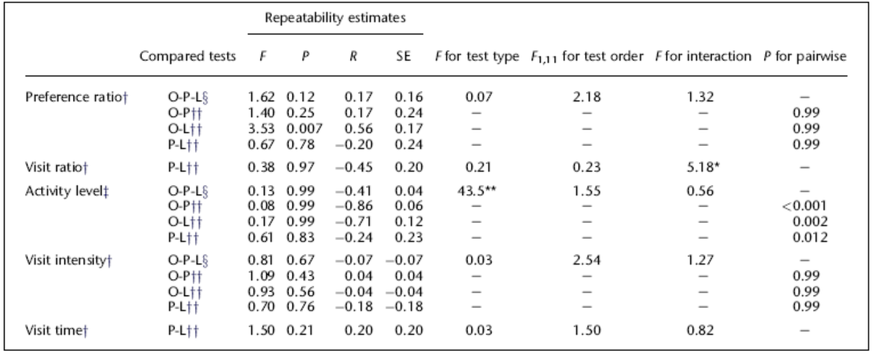


Figure 2. Preference ratios per individual for the preferred stimulus (i.e. stimulus preferred in at least two of the three tests) across the three test designs for all females (for test order see Table 1).

Table 2. Repeatability estimates and effect of test type and test order on all measures of female choice behaviour across the three tests and/or in pairwise comparisons (for test type effect with Bonferroni adjustment for multiple comparisons)



		Repeatability estimates							
	Compared tests	*F*	*P*	*R*	SE	*F* for test type	$F_{1,11}$ for test order	*F* for interaction	*P* for pairwise
Preference ratio†	O-P-L§	1.62	0.12	0.17	0.16	0.07	2.18	1.32	–
	O-P††	1.40	0.25	0.17	0.24	–	–	–	0.99
	O-L††	3.53	0.007	0.56	0.17	–	–	–	0.99
	P-L††	0.67	0.78	−0.20	0.24	–	–	–	0.99
Visit ratio†	P-L††	0.38	0.97	−0.45	0.20	0.21	0.23	5.18*	–
Activity level‡	O-P-L§	0.13	0.99	−0.41	0.04	43.5**	1.55	0.56	–
	O-P††	0.08	0.99	−0.86	0.06	–	–	–	<0.001
	O-L††	0.17	0.99	−0.71	0.12	–	–	–	0.002
	P-L††	0.61	0.83	−0.24	0.23	–	–	–	0.012
Visit intensity†	O-P-L§	0.81	0.67	−0.07	−0.07	0.03	2.54	1.27	–
	O-P††	1.09	0.43	0.04	0.04	–	–	–	0.99
	O-L††	0.93	0.56	−0.04	−0.04	–	–	–	0.99
	P-L††	0.70	0.76	−0.18	−0.18	–	–	–	0.99
Visit time†	P-L††	1.50	0.21	0.20	0.20	0.03	1.50	0.82	–

O: operant test; P: phonotaxis test; L: live male test.

*P < 0.05; **P < 0.001; two-factor mixed ANOVAs with assumed sphericity values for preference ratio, activity level and visit intensity and Greenhouse-Geisser values for visit ratio and visit time.

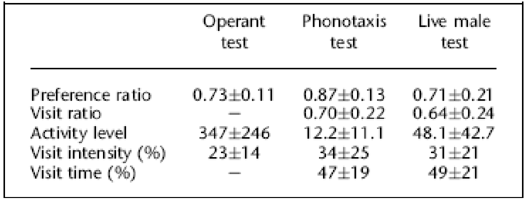
†Z score-transformed data.

‡Log-transformed data.

§$F_{16,34}$ for repeatability estimates, $F_{2,22}$ for test type and $F_{10,22}$ for the interaction 'test type x test order'.

††$F_{16,17}$ for repeatability estimates, $F_{1,11}$ for test type and $F_{5,11}$ for the interaction 'test type x test order'.

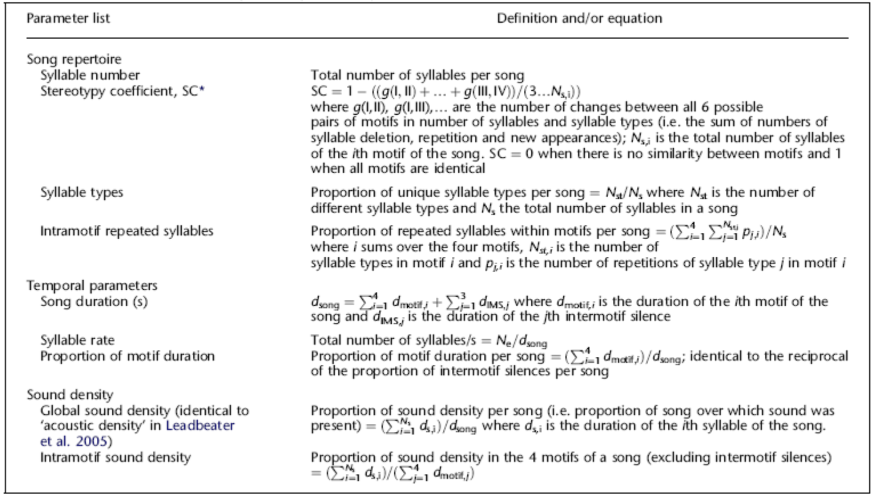
Table 3. Means ± 1 SD of all measures of female choice behaviour in the three tests ($N = 17$)



	Operant test	Phonotaxis test	Live male test
Preference ratio	0.73±0.11	0.87±0.13	0.71±0.21
Visit ratio	–	0.70±0.22	0.64±0.24
Activity level	347±246	12.2±11.1	48.1±42.7
Visit intensity (%)	23±14	34±25	31±21
Visit time (%)	–	47±19	49±21

identified all time points where the sound level exceeded a power output of 0.05 V (for at least 10 successive ms) as a sound and the rest as silence (Leadbeater et al., 2005). We used the amplitude contour of the song (Fig. 3a) as an automatic measure of the number of syllables per song (also checked by visual analysis of spectrograms). This provided an objective criterion for separating syllables by eliminating

Table 4. Definition of the nine investigated song structure parameters



Parameter list	Definition and/or equation
Song repertoire	
Syllable number	Total number of syllables per song
Stereotypy coefficient, SC*	$SC = 1 - ((g(\mathrm{I,II}) + \ldots + g(\mathrm{III,IV}))/(3 \ldots N_{s,i}))$ where $g(\mathrm{I,II})$, $g(\mathrm{I,III})$,... are the number of changes between all 6 possible pairs of motifs in number of syllables and syllable types (i.e. the sum of numbers of syllable deletion, repetition and new appearances); $N_{s,i}$ is the total number of syllables of the ith motif of the song. $SC = 0$ when there is no similarity between motifs and 1 when all motifs are identical
Syllable types	Proportion of unique syllable types per song $= N_{st}/N_s$ where N_{st} is the number of different syllable types and N_s the total number of syllables in a song
Intramotif repeated syllables	Proportion of repeated syllables within motifs per song $= (\sum_{i=1}^{4} \sum_{j=1}^{N_{st,i}} p_{j,i})/N_s$ where i sums over the four motifs, $N_{st,i}$ is the number of syllable types in motif i and $p_{j,i}$ is the number of repetitions of syllable type j in motif i
Temporal parameters	
Song duration (s)	$d_{song} = \sum_{i=1}^{4} d_{motif,i} + \sum_{j=1}^{3} d_{IMS,j}$ where $d_{motif,i}$ is the duration of the ith motif of the song and $d_{IMS,j}$ is the duration of the jth intermotif silence
Syllable rate	Total number of syllables/s $= N_e/d_{song}$
Proportion of motif duration	Proportion of motif duration per song $= (\sum_{i=1}^{4} d_{motif,i})/d_{song}$; identical to the reciprocal of the proportion of intermotif silences per song
Sound density	
Global sound density (identical to 'acoustic density' in Leadbeater et al. 2005)	Proportion of sound density per song (i.e. proportion of song over which sound was present) $= (\sum_{i=1}^{N_s} d_{s,i})/d_{song}$ where $d_{s,i}$ is the duration of the ith syllable of the song.
Intramotif sound density	Proportion of sound density in the 4 motifs of a song (excluding intermotif silences) $= (\sum_{i=1}^{N_s} d_{s,i})/(\sum_{j=1}^{4} d_{motif,j})$

See Fig. 3 for abbreviations.

*To measure the stereotypy of syllable sequencing between the four motifs of each stimulus song, we introduced a new measure we termed 'stereotypy coefficient', SC, which is slightly different but more detailed than the stereotypy score (SS) proposed by Scharff and Nottebohm (1991). When applying both formulae to our data we found the values from SS and SC to be significantly correlated with each other (Pearson correlation: $r_{15} = 0.59$, $P = 0.013$).

the substantial variation caused by human decision (Jones et al., 2001a). However, this method yielded several composite syllables that authors applying the 'sudden change in frequency' criterion would have split (Williams and Staples, 1992). We accepted this since little is known about whether zebra finches perceive composite syllables as one or a quick succession of elements (but see Franz and Goller, 2002). However, one should be aware that the average syllable/element repertoire might differ between published studies because of the criteria used and not because of differences in repertoire size between colonies (which, however, might also exist, Slater and Clayton, 1991).

Morphometry analysis

Immediately after the live male test (between 0930 and 1200 hours), we weighed each male (±0.1 g) on a Sartorius BL600 scale. The following morphometric measures (Baumel et al., 1979) were taken with callipers (±0.05 mm): tarsus length (distance from the right tibiotarsus-tarsometatarsal joint to the point of the tarsometatarsal joint at the base

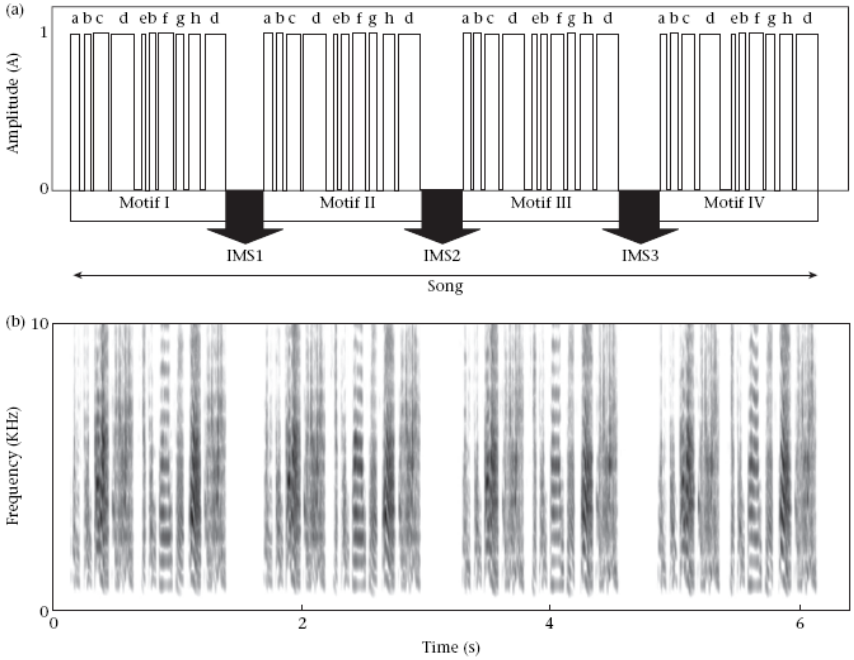


Figure 3. Example of (a) the output of the gating procedure used to measure sound density for a zebra finch song with (b) the relevant spectrogram. In this example, the song is composed of eight syllable types with a total of 10 syllables, i.e. there are repeated syllables within a motif. a, b, c, d, e, f, g, h: Syllable types. IMS: intermotif silence.

Table 5. Morphological traits of the 17 males used to test female preferences and repeatability estimates of the measures

				Repeatability estimates			
Morphological traits	Mean±SD	Range	N	$F_{16,34}$	P	R	SE
Mass (g)	16.3±3.2	13.0–24.5	17	–	–	–	–
Tarsus length (mm)	15.6±0.7	14.6–17.6	17	49.2	<0.001	0.94	0.02
Wing chord (mm)*	57.5±1.1	55.3–58.8	16	7.7	<0.001	0.69	0.11
Beak length (mm)	10.8±0.4	10.2–11.8	17	80.0	<0.001	0.96	0.02
Beak width (mm)	7.0±0.3	6.7–7.6	17	75.2	<0.001	0.96	0.02
Body condition index*	0.0±1.0	−2.1–1.6	16	–	–	–	–
Beak upper area (mm^2)	38.0±2.0	34.8–42.1	17	–	–	–	–

*$F_{15,32}$ since one male with broken primary feathers was excluded from the analyses.

of the right middle anterior toe); wing chord (from the bend of the flattened right wing (wrist) to the tip of the longest primary feather); beak length (tip of the upper mandible to the end of the culmen at its intersection with the forehead); beak width (between the two lateral sides of the upper mandible above the nostrils). All measures were taken three times, were highly repeatable (Table 5) and involved brief handling of 2-3 min per bird.

We calculated body condition index (BCI) as the standard residual index of the linear regression of body mass on structural size (Jakob et al., 1996). All morphometric measures were ln transformed to meet the homoscedasticity assumption of the regression analyses described below. The proportion of variance in ln body mass explained by the four size variables was determined by simple linear regressions (Gosler et al., 1998). We found that ln tarsus length, ln beak width, ln wing chord and ln beak length explained 67, 65, 50 and 10% of the variance in ln body mass, respectively. We then tried to improve the explained variance in ln body mass by using the factors extracted with principal components analyses (PCA; orthogonal rotation: varimax with Kaiser normalization) including different combinations of variables (with always at least tarsus length and wing chord). The first principal component (PC) of size (PC1; 74% of explained variance after rotation) based on tarsus length and wing chord explained more variance in ln body mass (80%) than any of the other structural size parameters alone. Our BCI thus refers to the standard residuals of the simple linear regression of ln body mass on the first PC based on tarsus length and wing chord (Table 5). Adding beak width to the PCA only slightly improved the explained variance in mass (82%) whereas adding beak length weakened it (74%). Beak width and length could not be entered at the same time in the PCA because they were too weakly correlated (Pearson correlation: $r_{15} = -0.05$, NS). We therefore decided to use a combined measure of beak length and width to approximate beak surface (mm^2) of the upper mandible area (beak upper area = beak length x width/2; Table 5).

Statistical analyses

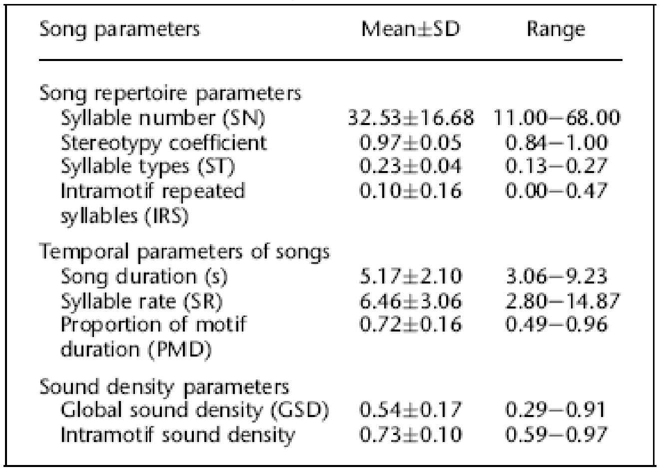
We did our analysis of song as a predictor of female preference in the two song preference tests (operant and phonotaxis tests), as female choice in the live male test could have been based on any male phenotypic trait (e.g. male morphology and display intensity, Collins and ten Cate, 1996). All 17 successfully tested females showed perfect consistency in preference direction between the operant and phonotaxis tests (Table 1). Nevertheless, test context differed and none of the measures of female choice behaviour was repeatable between these two tests (Table 2). Thus we analysed for each test separately which of the measured song structure parameters predicted relative preference strength (i.e. preference ratio). For this analysis, the use of absolute values for the song parameters was inappropriate as each female could choose between two stimuli only rather than the whole set. We therefore calculated relative differences in song parameters between the two stimuli of a dyad (value of preferred stimulus - value of nonpreferred stimulus divided by the sum of both stimuli) to test whether song structure predicted female preferences, but worked with the absolute measurement values when analysing whether variation in song parameters predicted variation in male morphological traits. To avoid multiple testing, we reduced song structure parameters with principal components analyses (orthogonal rotation: varimax with Kaiser normalization) after removing highly intercorrelated parameters (Table 6). To ensure that the scores of the principal components were uncorrelated we used the Anderson-Rubin method (Field, 2000), a necessary precondition for the subsequent stepwise linear regression analyses, which aimed to test whether the principal components obtained from the two separate PCAs could predict (1) female preference ratios (arcsine transformed to meet regression assumptions) in the operant and phonotaxis tests and (2) male morphology.

Results

Song preferences and relative differences in song structure

The PCA of the song structure parameters led to the extraction of three principal components (PCs) with eigenvalues >0.9 that we labelled 'relative performance', 'relative sound density' and 'relative proportion of different syllables', according to the relative load of the six entered parameters (Table 7). The PC1 'relative performance' accounted for 34% of the variation in female preference ratios in the operant test ($F_{1,15} = 7.7$, $R^2 = 0.34$, $P = 0.014$) and for 28% in the phonotaxis test ($F_{1,15} = 5.8$, $R^2 = 0.28$, $P = 0.03$). The preferred songs within dyads had a higher proportion of motif duration per song (i.e. lower proportion of intermotif silences) and more syllables than the less preferred songs in both preference tests (Fig. 4a). The predictive value of the PC1 'relative performance' for female preference ratios in the operant test was reinforced by the PC2 'relative sound density' when included in the stepwise regression in the model at step 2 ($F_{1,15} = 8.7$, $R^2 = 0.74$, $P = 0.004$). Thus, in the operant test, the preferred songs within dyads also had a higher intramotif sound density than the less preferred songs (Fig.

Table 6. Song structure parameters (see Table 4 for definition) of the 17 males used to test female preferences



Song parameters	Mean±SD	Range
Song repertoire parameters		
Syllable number (SN)	32.53±16.68	11.00–68.00
Stereotypy coefficient	0.97±0.05	0.84–1.00
Syllable types (ST)	0.23±0.04	0.13–0.27
Intramotif repeated syllables (IRS)	0.10±0.16	0.00–0.47
Temporal parameters of songs		
Song duration (s)	5.17±2.10	3.06–9.23
Syllable rate (SR)	6.46±3.06	2.80–14.87
Proportion of motif duration (PMD)	0.72±0.16	0.49–0.96
Sound density parameters		
Global sound density (GSD)	0.54±0.17	0.29–0.91
Intramotif sound density	0.73±0.10	0.59–0.97

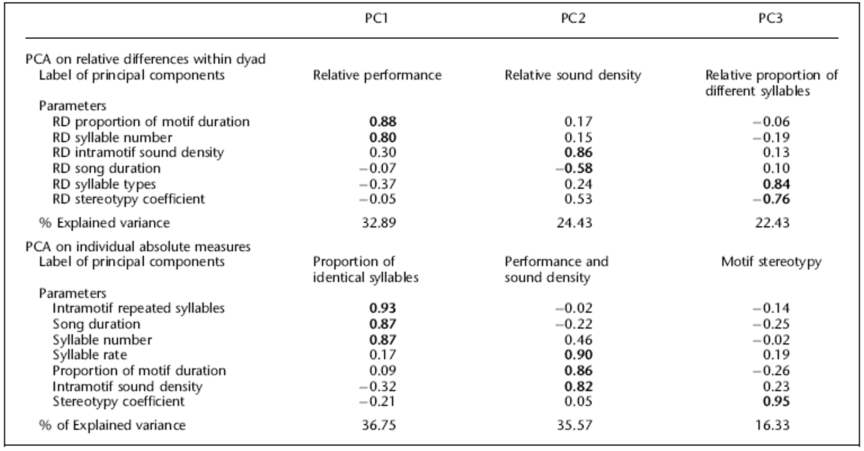
SN/SR, ST/IRS, PMD/GSD: The two parameters of each pair were highly intercorrelated (i.e. Pearson $-0.9 > r > 0.9$) and could not be entered at the same time into a principal components analysis (PCA). For the PCA testing whether relative differences in song structure parameters predicted female preference ratios, SR, IRS and GSD were not entered (SN/SR: $r_{15} = 0.98$; ST/IRS: $r_{15} = -0.96$; PMD/GSD: $r_{15} = 0.92$); for the PCA testing whether absolute measures of song structure parameters predicted male morphology, ST and GSD were not entered (ST/IRS: $r_{15} = -0.98$; PMD/GSD: $r_{15} = 0.92$; all $P < 0.001$). The parameters entered in the two independent PCAs differed because the relative difference for IRS and the absolute measure for ST could not achieve a normal distribution even after appropriate transformation. However, we are confident that it did not affect our interpretation of results since ST and IRS were highly correlated for both relative differences and absolute measures of song structure parameters (both $r_{15} < -0.96$, $P < 0.001$). All other relative differences and absolute measures of the song structure parameters were normally distributed (one-sample Kolmogorov-Smirnov test: for relative differences: all $Z < 1.18$, $N = 17$, $P > 0.12$; for absolute measures: all $Z < 1.32$, $N = 17$, $P > 0.06$).

4b).

Male morphology and song structure

This second PCA led again to the extraction of three principal components with eigenvalues >0.8, but the loading of these had changed. They were thus now labelled differently: ‘proportion of identical syllables’, ‘performance and sound density’ and ‘motif stereotypy’ (Table 7). PC2 ‘performance and sound density’ accounted for 32% of the variation in male beak length ($F_{1,15} = 7.1$, $R^2 = 0.32$, $P =$

Table 7. Rotated component matrices and percentages of explained variance (after rotation) of the two principal components analyses (PCA) of song structure parameters



	PC1	PC2	PC3
PCA on relative differences within dyad			
Label of principal components	Relative performance	Relative sound density	Relative proportion of different syllables
Parameters			
RD proportion of motif duration	**0.88**	0.17	−0.06
RD syllable number	**0.80**	0.15	−0.19
RD intramotif sound density	0.30	**0.86**	0.13
RD song duration	−0.07	**−0.58**	0.10
RD syllable types	−0.37	0.24	**0.84**
RD stereotypy coefficient	−0.05	0.53	**−0.76**
% Explained variance	32.89	24.43	22.43
PCA on individual absolute measures			
Label of principal components	Proportion of identical syllables	Performance and sound density	Motif stereotypy
Parameters			
Intramotif repeated syllables	**0.93**	−0.02	−0.14
Song duration	**0.87**	−0.22	−0.25
Syllable number	**0.87**	0.46	−0.02
Syllable rate	0.17	**0.90**	0.19
Proportion of motif duration	0.09	**0.86**	−0.26
Intramotif sound density	−0.32	**0.82**	0.23
Stereotypy coefficient	−0.21	0.05	**0.95**
% of Explained variance	36.75	35.57	16.33

Analyses were based on relative differences (RD) within dyads (principal components extracted when initial eigenvalues >0.9; rotation converged in 8 iterations) and on individual absolute measures (principal components extracted when initial eigenvalues >0.8; rotation converged in 4 iterations). Some of the parameters entered in the PCA on relative differences are different from those entered in the PCA on absolute measures because of different intercorrelations between song parameters in the two PCAs (see footnotes of Table 6). Values in bold indicate those variables that contributed most to a particular principal component.

0.018), 29% in beak upper area ($F_{1,15} = 6.0$, $R^2 = 0.29$, $P = 0.027$), 25% in mass ($F_{1,15} = 5.0$, $R^2 = 0.25$, $P = 0.040$) and 22% in tarsus length ($F_{1,15} = 4.2$, $R^2 = 0.22$, $P = 0.059$). Songs with a higher syllable rate, a higher proportion of motif duration per song (i.e. lower proportion of intermotif silences) and a higher intramotif sound density predicted larger beak length, beak upper area, mass and, almost significantly, tarsus length. The predictive value of the PC2 ‘performance and sound density’ for beak length was reinforced by the PC3 ‘motif stereotypy’ when included in the stepwise regression in the model at step 2 ($F_{1,15} = 9.3$, $R^2 = 0.57$, $P = 0.003$). The PC3 ‘motif stereotypy’ also accounted for 51% of the variation in beak width ($F_{1,15} = 15.4$, $R^2 = 0.51$, $P = 0.001$). Songs with higher stereotypy were associated with longer and larger beaks. None of the included parameters predicted wing chord or BCI.

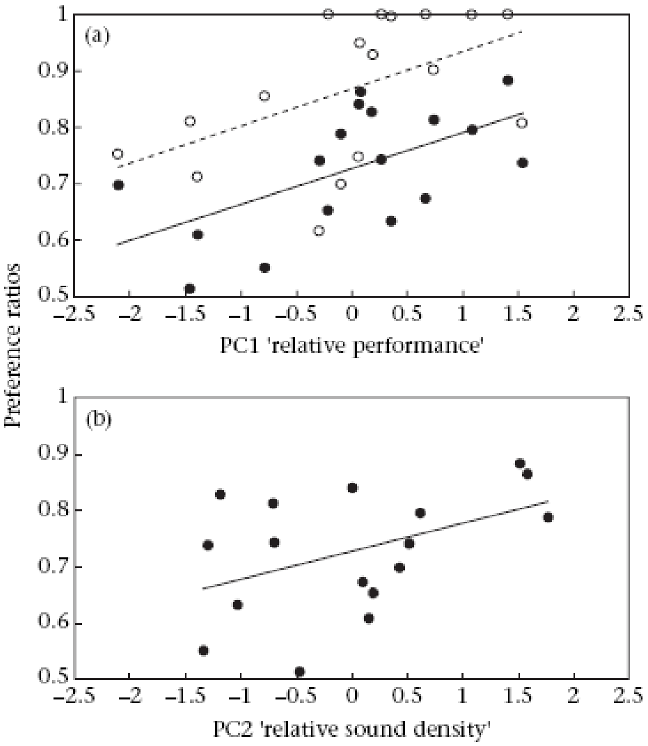


Figure 4. Relation between preference ratios and (a) the first principal component (PC1 ‘relative performance’) in the operant test (●, —) and in the phonotaxis test (○, - - -) and (b) the second principal component (PC2 ‘relative sound density’) in the operant test (linear regression lines are shown).

DISCUSSION

Females preferred the same song or the singer of this song in all three tests. As the preference was consistent independent of whether females could choose only between songs or between songs and a live male, male zebra finch song must contain sufficient information for females to judge male quality. The outcome of the multiple regression analyses further confirmed this: song characteristics predicted both variation in female preferences and male traits. The three commonly used binary choice designs not only yielded concordant results on the direction of female preference, but in two of them, the operant test and the live male test, the magnitude of relative preference strength was repeatable. This was highly surprising: females were tested only once in each test paradigm and in different orders, which left substantial potential scope for the outcome to be influenced not only by the test context but also by short-term fluctuation in female motivation. Furthermore, the operant test might be considered a highly artificial context (key pecking for song reward) whereas the live male test provided the opportunity to see, hear and interact with two live males. Why was the preference strength for males or their songs repeatable between these two tests, but not between the two tests involving song only (i.e. operant and phonotaxis tests)?

A number of aspects differed between these tests. In the much larger phonotaxis cage females had to travel further to be close to the stimuli than in the operant test and in the live male test cage. Therefore the phonotaxis test could have demanded more effort to approach the stimuli than the other two tests, which may have influenced female preference strength. Perhaps more importantly though, the operant test was an active choice design that allowed *ad libitum*, repeated sampling and active control of exposure to the stimulus by females. In the phonotaxis test, females could not actively control the playback or interact with it. The significant repeatability of preference strength between operant and live male tests further supports the notion that active operant-conditioning techniques are a highly suitable means for assessing direction and strength of female song preferences (for discussion see Riebel and Slater, 1998). Reassuringly, given the large existing body of literature on phonotaxis tests, although preference strength differed, the actual direction of females' preferences was consistent with the other two tests.

A low activity level, i.e. a low motivation in assessing stimuli (as in the phonotaxis test), and sequential rather than simultaneous presentation of stimuli may yield fewer opportunities to compare stimuli, which in turn might directly affect the expressed preference (Brooks and Endler, 2001; Rowland et al., 1995; Wagner, 1998). However, females in the phonotaxis test could hear both songs without having to perch close to the speakers. Furthermore, in all three tests, preference strength was independent of the motivation to sample stimuli (preference strength and activity level were not correlated).

Females could also have been affected by the variation in time between tests (Johnsen and Zuk, 1996; Morris et al., 2003) or learning opportunities between repeated tests (Hager and Teale, 1994) arising from the different test situations. However, we found that the measures of female choice behaviour were independent of the test order making these explanations unlikely.

Females are known to prefer males with higher song rates (Collins et al., 1994; de Kogel and Prijs, 1996; Forstmeier, 2004; Houtman, 1992), so could differences in song rate explain the differences in preference strength between the tests? In the operant test, female activity level determined the song rate. However, song rate was not measured in the live male test. Although it was obvious from observations during the tests that males sang at different rates, we do not know whether the repeatable preference strength between the operant and live male tests was associated with song rates. However, we do know that motif rate per song was identical in both playback tests but did not lead to repeatable preference strength. Furthermore, the different song rates between the three tests did not alter the direction of preference either. Therefore, song rate is unlikely to be an important factor in explaining our results, suggesting that song structure per se contains sufficient information for females to base their decisions on. Hence, song structure must either contribute disproportionately to female choice or is highly correlated with other male features relating to female choice.

Some song structure parameters did indeed predict variation in female preference strength. Females had stronger preferences for songs with lower proportions of intermotif silences and more syllables than the alternative stimulus songs (PC1 ‘relative performance’) in both operant and phonotaxis tests. This finding is consistent with the idea that female attention may depend on sound continuity in the song (Goller and Daley, 2001) and with the finding that females prefer larger syllable repertoires (Neubauer, 1999). However, there are two caveats. First, we measured syllable repertoire size and not element repertoire size for which the outcome may be different. Second, repertoire size in the zebra finch is perhaps better defined by the number of syllable or element categories (Nowicki et al., 2002a) rather than by the total number of syllables. In the operant test, preference strength was also predicted by a higher intramotif sound density (PC2 ‘relative sound density’). Sound density within motifs was not found to be a good predictor of female preferences by Leadbeater et al. (2005), but they tested females with single-motif songs whereas we used four-motif songs. Furthermore, we found that sound density reinforced the predictive value of the PC1 ‘relative performance’, not that it was a significant predictor of female preference on its own. Furthermore, females might not necessarily perceive the proportion of intermotif silences (parameter of highest load in PC1 'relative performance') and the intramotif sound density (PC2 'relative sound density') separately, as both parameters measure the amount of sound per unit of time, albeit different aspects of this.

Overall, the weighting of song structure parameters of the PC1 'relative performance' was not dependent on the test design. Preference functions are probably less affected by context than choosiness (i.e. the effort and time that an individual is prepared to invest in finding and assessing mates) and sampling strategies (i.e. the decision rule adopted in mate assessment). This suggestion is reinforced by the high repeatability of female song preferences across substantial time spans (Riebel, 2000).

The main predictors of song preference strength were also predictors of a number of male phenotypic aspects. Therefore our results suggest that some structural song parameters that are thought to signal male quality are highly correlated with other preferred traits, which supports the idea that different characteristics are back-up signals (Johnstone, 1996). Songs with higher syllable rate, a lower proportion of intermotif silences and higher intramotif sound density (i.e. PC2 'performance and sound density') predicted larger male beak length, beak upper area and mass, but not current male body condition (BC1). High syllable rates are associated with increased male attractiveness (reviewed in Gil and Gahr, 2002) and may signal male quality since fast singing patterns are hypothesized to require more vocal-respiratory coordination (Gil and Gahr, 2002; Goller and Daley, 2001) and fast syllable delivery is more demanding on the motor system (Franz and Goller, 2003). In addition, higher sound intensity songs may stimulate the female sensory system more than low-intensity ones (e.g. Basolo, 1990; Endler, 1992; Kirkpatrick and Ryan, 1991).

Our results could be taken to suggest great redundancy in male courtship display, at least over the modest time of our measurements, and an overall population-level preference for some song characteristics. However, the switching of preference by the three females in the live male test suggests that some females might rank signals differently or that no single trait might signal absolute quality. Different ornaments may indicate condition over different timescales, with some reflecting condition during development while others are more dynamic and respond to changes in condition at adulthood (e.g. Hill et al., 1999; Møller et al., 1998; Scheuber et al., 2003a; Scheuber et al., 2003b; Sorenson and Derrickson, 1994). In close-ended learners such as zebra finches, song structure may reflect how well a male fared during posthatch development and thus provide a condition-dependent signal of males' early development (Nowicki et al., 1998; Spencer et al., 2003; Spencer et al., 2005b, but see Gil et al., 2006). Such long-term effects of early experience may also be seen in other male sexually selected traits (Birkhead et al., 1999; de Kogel and Prijs, 1996; but see Blount et al., 2003). However, while song structure is likely to remain stable in close-ended learners, other sexually selected traits such as song rate and beak colour can change after a relatively short period of manipulation of food and exercise in the laboratory (Birkhead et al., 1998) or after an immune challenge (Faivre et al., 2003). Earlier studies had found the two traits to be correlated when the differences between past and present condition were smaller (Collins et al., 1994; Houtman, 1992). This could explain why contradictory results have been found on the weighting of these condition-dependent traits in mate choice (Collins, 1994; Collins and ten Cate, 1996; Forstmeier and Birkhead, 2004). This also suggests that some condition-dependent signals, because of their dynamic character, may switch between being back-up and multiple message signals (within and between individuals). Our data, together with the literature on the zebra finch, support the view that the 'multiple messages' and 'back-up signal' hypotheses should not be seen as mutually exclusive (Candolin, 2003; van Doorn and Weissing, 2004).

ACKNOWLEDGMENTS

We thank Carel ten Cate, Machteld N. Verzijden, Ben Hatchwell and two anonymous referees for their valuable comments on the manuscript. We also thank Gabriël J.L. Beckers and Jean Secondi for assistance with stimulus song preparation and Henny Koolmoes for assistance with animal care. This study and M.J.H. were funded by a research grant of the Netherlands Organization for Scientific Research (NWO/ALW) to K.R.