Supplementary Materials & Methods

Supplementary Figures and Legends:

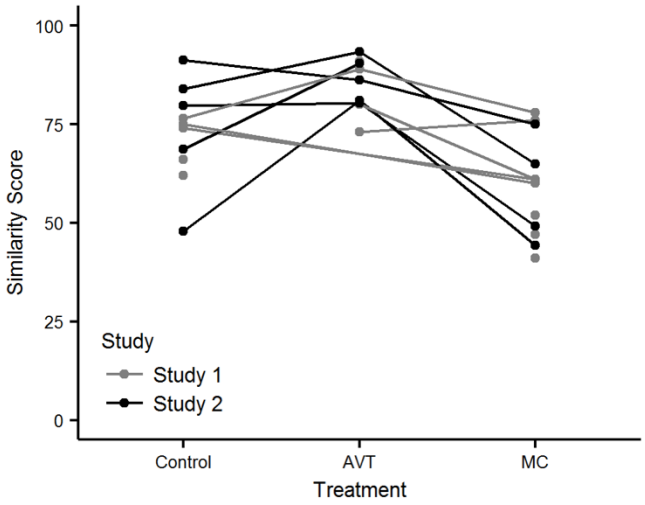

Figure S1: Individual similarity scores at day 90 for both Study 1 and Study 2. Lines connect siblings within the same family. The similarity score for subjects in Study 1 is from a single song recording, whereas the mean similarity score is shown for subjects in Study 2.

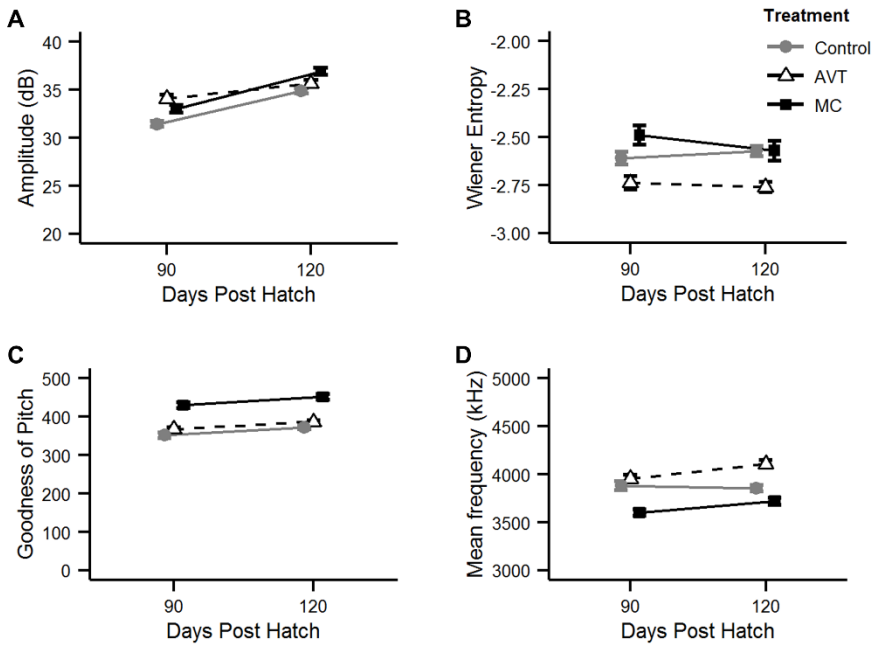

Figure S2: Acoustic features of song recorded on 90 and 120 days post-hatch. Mean ± SE of A) Amplitude (dB) of the song motif. All groups became louder between 90 dph and 120 dph (t = 10.1, $p < 0.0001$). AVT birds were louder on 90 dph (t = 3.0, p = 0.01), but decreased less in loudness between 90 dph and 120 dph than both MC and Control males (Control-AVT: $X^2(1)$ = 14.4, p = 0.0003; AVT-MC: $X^2(1)$ = 16.2, p = 0.0002). B) Wiener entropy as measured using Sound Analysis Pro (SAP). Wiener entropy slightly but significantly increased in the Control group between 90 dph and 120 dph (t = 2.5, p = 0.02), but it decreased in both the AVT and MC males (AVT: t = -2.4, p = 0.03; MC: t = -3.2, p = 0.004). C) Goodness of pitch. The goodness of pitch actually increased between 90 dph and 120 dph (t = 5.4, p = 0.0002). However, MC birds had a higher goodness of pitch than Control males (t = 2.6, p = 0.02). D) Mean frequency (kHz). MC males had a lower mean frequency compared to Controls (t = -2.8, p = 0.009). Additionally, both AVT and MC males, but not Controls, increased in mean frequency between 90 dph and 120 dph (AVT: t = 2.7, p = 0.01; MC: t = 2.2, p = 0.04). Control subjects are depicted with circles and a solid gray line, AVT with triangles and dashed black line, and MC as squares and solid black line.

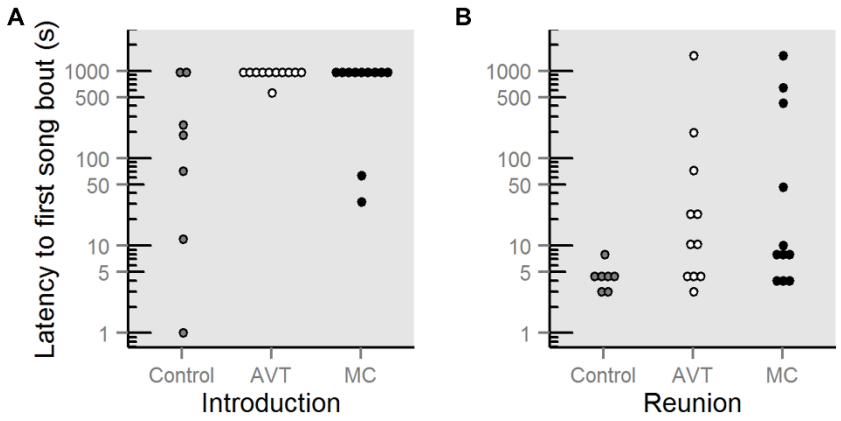

Figure S3: Latency to sing when introduced to and later reunited with a female partner. (A) Dot plot of the latency (s) until the male's first song bout during introduction to a novel female conspecific ("partner"). AVT and MC males took significantly longer to sing than Controls when first introduced to a novel female (Kruskal-Wallis test: $X^2(2) = 9.26$, $p = 0.01$; AVT-Control: $p = 0.005$, MC-Control: $p = 0.05$; AVT-MC: $p = 0.5$) (B) Latency (s) during reunion with the female partner following a one-hour separation after being housed with that female for one week. AVT and MC males also took significantly longer to sing upon reunion (Kruskal-Wallis test: $X^2(2) = 7.1$, $p = 0.03$; AVT-Control: $p = 0.02$, MC-Control: $p = 0.02$; AVT-MC: $p = 0.9$). Latency to first song bout (s) is shown on a $\log_{10}$-scale**.**

	Duration(s)				*Amplitude (dB)*				*Pitch*			
Predictors	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
Intercept	737.249	88.109	8.367	**0.000**	31.169	0.987	31.585	**0.000**	1186.953	93.270	12.726	**0.000**
DPH (120)	-52.210	16.470	-3.170	**0.005**	3.715	0.366	10.142	**0.000**	27.173	12.717	2.137	**0.048**
Treatment (AVT)	-104.705	124.546	-0.841	0.410	3.187	1.079	2.954	**0.011**	124.776	94.786	1.316	0.207
Treatment (MC)	90.143	130.548	0.690	0.497	1.801	1.141	1.578	0.139	-197.378	100.791	-1.958	0.068
Treatment (AVT)* DPH(120)	51.562	23.296	2.213	**0.038**	-1.968	0.518	-3.798	**0.002**				
Treatment (MC)*DPH(120)	-12.670	23.676	-0.535	0.598	0.152	0.527	0.288	0.778				

	Mean Frequency (kHz)				*Peak Frequency (kHz)*				*Goodness of Pitch*			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
Intercept	3861.651	63.142	61.158	**0.00E+00**	3839.487	70.819	54.216	**0.00E+00**	352.743	23.606	14.943	**1.50E-08**
DPH (120)	-5.069	38.757	-0.131	0.897	-11.629	43.921	-0.265	0.793	19.905	3.700	5.380	**0.000**
Treatment (AVT)	93.177	88.839	1.049	0.303	105.937	99.628	1.063	0.297	22.449	25.784	0.871	0.403
Treatment (MC)	-258.672	92.597	-2.794	**0.009**	-278.558	103.830	-2.683	**0.012**	71.680	27.366	2.619	**0.024**
Treatment (AVT)* DPH(120)	145.944	54.795	2.663	**0.013**	179.942	62.094	2.898	**0.007**				
Treatment (MC)*DPH(120)	121.286	55.746	2.176	**0.038**	127.605	63.173	2.020	0.053				

	Wiener Entropy				*Frequency Modulation*				*Amplitude Modulation*			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
Intercept	-2.667	0.134	-19.883	**2.15E-14**	39.109	1.529	25.571	**1.67E-11**	-0.010	0.001	-8.882	**1.07E-10**
DPH (120)	0.094	0.038	2.459	**0.023**	1.075	0.368	2.923	**0.013**	0.002	0.001	3.077	**0.004**
Treatment (AVT)	-0.060	0.156	-0.385	0.704	2.095	1.439	1.455	0.172	-0.001	0.002	-0.862	0.394
Treatment (MC)	0.163	0.165	0.989	0.335	1.652	1.530	1.079	0.302	-0.002	0.002	-1.036	0.307
Treatment (AVT)* DPH(120)	-0.130	0.054	-2.408	**0.026**	-1.502	0.520	-2.888	**0.014**	-0.003	0.001	-3.012	**0.005**
Treatment (MC)*DPH(120)	-0.177	0.055	-3.227	**0.004**	-2.769	0.529	-5.237	**2.39E-04**	-0.001	0.001	-1.278	0.209

Table S1: Linear mixed model (LMM) results for acoustic features of song. Summary of the linear mixed models testing for an interaction effect between treatment and day (90 dph vs 120 dph) on individual acoustic features (Duration ($X^2(2) = 8.3$, $p = 0.02$), Amplitude ($X^2(2) = 19.9$, $p < 0.0001$), Pitch ($X^2(2) = 7.8$, $p = 0.02$), Mean Frequency ($X^2(2) = 8.0$, $p = 0.02$), Peak Frequency ($X^2(2) = 8.7$, $p = 0.01$), Goodness of Pitch ($X^2(2) = 5.7$, $p = 0.06$), Wiener Entropy ($X^2(2) = 11.1$, $p = 0.004$), Frequency Modulation ($X^2(2) = 26.4$, $p < 0.0001$), and Amplitude Modulation($X^2(2) = 9.0$, $p = 0.01$)) as dependent variables. The fixed effects are Treatment, Day Post-Hatch (dph), and the interactions. Individual ID nested within Family ID was included as a random effect. The LMM models were selected based on model comparisons using likelihood ratio tests. To test the significance of each parameter within the models, we used the Kenward-Roger approximation to get approximate degrees of freedom and the t-distribution (SE = standard error, bold numbers indicate significance, * refers to an interaction term).

Supplementary Experimental Procedures:

Study 1 Breeding Conditions

Seventy-two unpaired adult males and females bred in large aviaries (1.2 x 0.9 x 0.6 m) and male chicks hatched within 40 days were used as subjects. Starting on 2 days post-hatch (dph) through 8 dph, following genetic sexing, subjects were injected intracranially with 2μL of either 1) 10ng of AVT; 2) 50ng of Manning Compound (MC), a potent V1a and mild OT receptor antagonist; or 3) 0.9% isotonic saline vehicle control [1,2]. Intracranial (IC) injections were performed using a sterile 31G stainless steel insulin syringe [3]. Both AVT and Manning Compound act at multiple nonapeptide receptor subtypes in zebra finches, including the VT4 (V1aR), VT3 (OT-like), and V2 receptors [4,5].

Chicks within the same clutch were randomly assigned to different treatment groups, such that treatment was unrelated to hatching order. One Control male was found dead on 38 dph and was subsequently excluded from the experiment. The number of males that completed treatment and survived until adulthood are as follows: AVT (n = 11); Manning Compound (n = 11); Control (n = 7). After 39.8 ± 5.4 dph, subjects were removed from their natal aviary and housed in same-sex aviaries in a separate room from the parents. The only interactions between subjects and their parents after moving into same sex housing occurred during the weekly 15 minute four-way affiliative preference tests which took place from 30 dph to 86 dph [3].

Study 2 Breeding Conditions

We used cross-fostering on the second day post-hatch to create a within-family design. The genetic sex of the subject birds was determined on the day of hatching and chicks were then cross fostered on 2 dph to create families with three male subjects (one per treatment group) and one non-subject female sibling. All experimental subjects and female siblings were raised by foster parents until 60 dph. The experiment was conducted using four temporally separate family cohorts (n = 7 families), each with a total clutch size of four. However, two families in the same aviary were excluded from analyses because one of the male parents was highly aggressive towards his own offspring (two of which died), his female partner (who also died), and the other juveniles within the aviary, resulting in a highly abnormal developmental environment for all surviving offspring within the aviary. Additionally, one subject was excluded from the study because of incorrect genetic sexing. Any additional females were cross-fostered into nests in a non-experimental aviary to be raised for future studies. Thus, the final number of subjects at the end the experiment totaled 14, consisting of AVT (n = 5), Manning Compound (n = 4), and saline control (n = 5) subjects.

For both Study 1 and Study 2, birds were kept on a 14/10h light/dark cycle and were provided with seed, cuttlebone, water, and grit *ad libitum* throughout the study, with supplemental hard-boiled eggs weekly during the egg laying period. Additionally, each aviary was equipped with a nest box and nest-building material (coconut fiber), allowing the parent pairs to construct nests and breed. Nest boxes were checked daily in the morning to record the number of eggs and chicks. On the day of hatching, subjects were genotyped to determine the genetic sex using DNA extracted from feather follicle tissue[3]. Chicks were marked with non-toxic markers for individual identification. In both studies, all researchers were blinded to treatment until after the completion of the experiment. All procedures were developed with veterinary supervision and approved by Cornell University's Institutional Animal Care and Use Committee.

Study 1 Song Recordings

After reaching sexual maturity, subjects were randomly assigned an unmanipulated, sexually-naïve, and unpaired female pair partner. In order to obtain high-quality recordings of the male subjects' songs, all introductions between the subjects and their pair partner were performed on 90 dph in a room with no other birds, in a small aviary (57 x 32 x 42 cm) enclosed by sound attenuating foam. The subject was first placed in the cage, followed by the partner. Behavior (song latency, number of song bouts) was scored for the 15 min following introduction, though pair was sometimes left in the cage for 20-45 min if a male did not sing during the first 15 min. After the introductions, the pair was moved into a small pair aviary (.57 x .32 x .42 m or .61 x .36 x .43 m) in a colony room. The pairing aviaries were arranged such that they were visually, but not acoustically, isolated from other pairs in the room. Subjects were then housed with the partner for a total of seven days. High-quality songs were recorded from social fathers and other adult males in the breeding aviaries using a similar method for comparisons to juvenile songs. All songs were recorded using a highly-directional cardioid microphone (Sennheiser ME66).

As a result of profound effects of treatment on singing, a large number of males did not sing (AVT: n = 10/11; MC: n = 9/11; and Control: n = 2/7) during this introduction. Thus, we attempted to record songs from these males in the colony room over the course of the following week. Recordings were obtained either by an observer in the colony room or from recordings of reunion with the partner following a 1hr separation. These song recordings were processed

to remove extraneous noise prior to analysis. In the final analysis, we obtained high-quality song recordings from n = 4 AVT males, n = 8 MC males and n = 6 Control males, which was a subset of all males in the experiment.

In addition, for all males in the experiment (AVT: n = 11; MC: n = 11; and Control: n = 7), we recorded the latency to the first song bout during the first introduction to the female partner. We also recorded both the latency to sing and the number of song bouts in the reunion after 1 hr separation from the female partner, with whom they had been housed in a small pair aviary for seven days.

Study 2 Song Recordings

Songs were recorded from subjects between 50 and 120 dph. To obtain song recordings at each time point, subject males were removed from their home aviaries and individually isolated in sound attenuation chambers overnight. The chambers were kept in a separate room and were constructed from coolers (0.94 x 0.38 x 0.38 m) internally lined with sound attenuating foam. Individual transport cages (0.46 x 0.23 x 0.25 m) along with a microphone were placed in the chambers. Keeping the juveniles in the room overnight allowed them to become accustomed to the chamber and social isolation increased subsequent motivation to sing. Each chamber was equipped with overhead white LED lights, and subjects were provided with seed, cuttlebone, and water *ad libitum* while in the chambers. The next day immediately following lights on, a female was introduced to the cage with the subject. Song was recorded every three days from 50 to 60 dph, every ten days from 60 dph until 90 dph, and on 120 dph for 1 hr each day. The female used to elicit song was the same on 90 and 120 dph for each male subject, to control for differing motivation to sing to different females. Male subjects were kept in the sound attenuation chambers for as long as necessary to record multiple song bouts, even when that required several hours of recording.

Song Recording Analysis

In Study 1, one motif was cropped at random from subjects' recordings. In Study 2, ten motifs were cropped at random from juvenile songs recorded at 90 dph and at 120 dph. Motifs with background noise, female calls, or cage noise were excluded from the sample. Introductory notes were identified and excluded from analysis. A total of four core motifs were cropped out at random from the songs recorded from the focal juvenile's father and the best core motif was chosen based on the least amount of background noise or cage noise. Sound files were saved as uncompressed digital audio to a hard disk with a sampling rate of 44.1 KHz.

As a measure of song learning success, the recordings were used to analyze juvenile song match to paternal song using Sound Analysis Pro 2.0 (SAP) [6]. SAP measures song similarity between juvenile and paternal song by splitting up the songs into syllables, defined as discrete sound units bounded by silent intervals. For each tutor-juvenile pair of songs, SAP calculates the probability that the goodness of match between the songs would have occurred by chance [7]. Our analysis focused primarily on the scores of song similarity, accuracy, and sequential match percentage. Percent song similarity is defined as the percentage of tutor sounds included in the juvenile's crystallized song. Tutor-pupil pairs typically have similarity scores between 65 and 95, whereas random pairs typically have scores ranging between 20 and 45 [6]. Song accuracy is the average local similarity per millisecond across the crystallized song. Sequential match is calculated by comparing song tempo and rhythm between the tutor song and the juvenile's crystallized song [6]. In SAP, the similarity, accuracy, and sequential match were determined using the Explore & Score feature segmentation tool (entropy: -9.5; FFT data window: 9.27 ms; contour threshold: 10; frequency range: 11025 Hz; advance window: 1.00 ms). The motifs of each juvenile's song were compared to the respective father motif, using the same father motif for juvenile day 90 song and day 120 song analyses.

AVT has also been shown to affect acoustic features of vocalizations in other vertebrate species, so we tested whether the treatment impacted the spectral features of the song which have good articulatory correlates. For the Study 2 subject songs, we used Sound Analysis Pro 2.0 (SAP) to analyze each subject's song for the following acoustic features: duration (s), amplitude (dB), pitch, mean frequency (kHz), peak frequency (kHz), Goodness of Pitch, Wiener Entropy, amplitude modulation, and frequency modulation.

In order to investigate individual song variation between motif renditions, and to determine whether inter-group differences in similarity scores were primarily driven by variation in tutor imitation fidelity at the level of the individual syllable or the full motif, we compared all syllables from the chosen tutor motif with all syllables of the ten randomly chosen motifs of each juvenile's crystallized song at 120 dph. For each syllable pair, percent similarity and accuracy were computed. Matching syllables were identified based on syllable pairs with the highest similarity scores. Syllables for which SAP calculated a similarity score of zero were not included in analysis. Average percent similarity and accuracy of the ten renditions of each syllable and their standard deviations were computed for each juvenile-tutor syllable pair, which was used to calculate overall syllable similarity between pupils and tutors. We then determined the overall percentage of syllables learned from the tutor based on the number of syllables produced by the juvenile which had a >40% similarity score with the equivalent tutor syllable.

Social Isolation Tests

The day after subjects from Study 1 only were first observed having fledged from the nest, we assessed subjects' responses to isolation from their family and subsequent reunion with the male parent. The social isolation tests were performed in a testing apparatus (60 x 41 x 36 cm) in a separate room from the breeding cage. Two aviaries of paired adults were in the room but were behind a curtain to provide ambient colony noise. After one minute of acclimation, we recorded behavior in isolation for 9 min total. Next, the male parent was placed in the aviary with the fledgling for five additional minutes. The video was scored for the number of perch hops, saccadic head movements, and long tonal calls performed by the subject per minute.

Four-way Affiliative Preference Tests

We assessed the Study 1 subjects' preference for being proximal to parents or unfamiliar conspecifics weekly from day 30 to 86 in four-choice proximity tests with two males, two females, the parent pair, or no conspecifics as the four stimulus choices, similar to [8]. For testing, the subject was removed from its aviary and placed alone in a plus-shaped testing cage (61 x 61 x 41 cm) in a separate testing room, which was flanked on three sides by cages containing pairs of stimulus birds. The three stimulus cages (61 x 36 x 45 cm) were positioned next to the subject's cage. One stimulus cage contained the subject's parents, one contained two unfamiliar adult females, and one contained two unfamiliar adult males. Subjects were allowed to acclimate in the apparatus for one minute prior to recording. Tests were 15 minutes long and were videotaped from behind a blind with no human in the room. The testing cage contained three stimulus zones of proximity (ZOPs). The remainder of the cage (the center portion and the zone nearest to the video camera which was not proximal to any birds) was considered a neutral (non-proximity) zone. The same pool of 20 males and 20 females was used as conspecific stimuli in a random order for each subject. The stimuli were unfamiliar to the subject at the time of presentation, with the location of each stimulus set varied randomly. The total time that the subject's head was in each of the three proximity zones was recorded. Proximity is a valid indicator of family and sexual and pairing interest in this species, because these relationships are marked by close physical proximity [9]. The testing period, with nine weekly tests, covered the majority of the juvenile period, allowing us to measure changes in affiliative preferences across juvenile development.

All tests were recorded with a Canon Vixia HFM31 HD camera. Digital videos were coded by trained assistants who were blind to treatment. In addition, all researchers were blinded to treatment throughout the experiment, until after data collection and video coding was complete.

Statistical Analyses

All statistical analyses were performed with R software (R Development Core Team 2007). We used random slope linear mixed models (LMM) to test the effect of the treatment on the acoustic features of the song and measures of similarity to the social father. We used the *lmer* function of the lme4 package [10] which allowed us to define multiple distinct random factors. In these models, Treatment (and song recording date in Study 2) were specified as fixed factors. The interaction effect considered was Treatment*Day (in Study 2 only). Random factors were individual ID (18 levels in Study 1 and 17 levels in Study 2), nested within Family ID (13 levels in Study 1 and 6 levels in Study 2). To test the effect of the treatment on the syllable-by-syllable similarity and accuracy scores, we used a random slope LMMs. Random factors were individual ID (17 levels), nested within Family ID (6 levels). To test the effect of treatment on the number of number of notes copied from the tutor, we computed a measure of the proportion of notes copied by dividing the number of comparable notes by the total number of notes in the tutor's song. We then performed a random slope LMM similar to above with Family ID as a random factor.

To perform model comparisons for the LMM models, we used likelihood ratio tests to compare the full model to a reduced null model with only the factor of interest removed using the *anova* function to perform a chi-square test. To test the significance of each fixed effect within a model, we used the Kenward-Roger approximation to get approximate degrees of freedom and the t-distribution to get p-values (Kenward-Roger in the pbkrtest package) [11]. In addition, we performed post hoc tests on the interaction terms using the *testInteractions* function in the phia package [12]. We also calculated marginal and conditional R^2 to measure effect size for the model using the *r.squaredGLMM* function in the MuMIn package [13–15].

In order to test for treatment effects on whether or not males sang when introduced to a female, we used a chi-square test. To test the effect of IC injections on song latency (which were not normally distributed), we used a non-parametric Kruskal-Wallis test, followed by a pair-wise Wilcoxon test to perform planned comparisons between the different treatment groups.

Supplementary References

1. Goodson, J. L., Lindberg, L. & Johnson, P. 2004 Effects of central vasotocin and mesotocin manipulations on social behavior in male and female zebra finches. *Horm. Behav.* **45**, 136–143. (doi:10.1016/j.yhbeh.2003.08.006)

2. Manning, M. et al. 1989 Solid-phase synthesis of 16 potent (selective and nonselective) in vivo antagonists of oxytocin. *J. Med. Chem.* **32**, 382–391. (doi:10.1021/jm00122a016)

3. Baran, N. M., Sklar, N. C. & Adkins-Regan, E. 2016 Developmental effects of vasotocin and nonapeptide receptors on early social attachment and affiliative behavior in the zebra finch. *Horm. Behav.* **78**, 20–31. (doi:10.1016/j.yhbeh.2015.10.005)

4. Leung, C. H., Goode, C. T., Young, L. J. & Maney, D. L. 2009 Neural distribution of nonapeptide binding sites in two species of songbird. *J. Comp. Neurol.* **513**, 197–208. (doi:10.1002/cne.21947)

5. Manning, M. et al. 2012 Oxytocin and vasopressin agonists and antagonists as research tools and potential therapeutics. *J. Neuroendocrinol.* **24**, 609–628. (doi:10.1111/j.1365-2826.2012.02303.x)

6. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. 2000 A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176. (doi:10.1006/anbe.1999.1416)

7. Tchernichovski, O. & Mitra, P. 2002 Towards quantification of vocal imitation in the zebra finch. *J. Comp. Physiol. A* **188**, 867–878. (doi:10.1007/s00359-002-0352-4)

8. Adkins-Regan, E. & Leung, C. H. 2006 Sex steroids modulate changes in social and sexual preference during juvenile development in zebra finches. *Horm. Behav.* **50**, 772–778. (doi:10.1016/j.yhbeh.2006.07.003)

9. Clayton, N. S. 1990 Assortative Mating in Zebra Finch Subspecies, *Taeniopygia guttata guttata* and *T. g. castanotis*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **330**, 351–370. (doi:10.1098/rstb.1990.0205)

10. Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H. & Dai, B. 2014 *lme4: Linear mixed-effects models using Eigen and S4*. [cited 2015 Jan. 21].

11. Højsgaard, U. H. S. 2014 *pbkrtest: Parametric bootstrap and Kenward-Roger-based methods for mixed model comparison*. [cited 2015 Jun. 1].

12. Rosario-Martinez, H. D., Fox, J. & Team, R. C. 2015 *phia: Post-Hoc Interaction Analysis*. [cited 2015 Jun. 1].

13. Bartoń, K. 2015 *MuMIn: Multi-Model Inference*.

14. Johnson, P. C. D. 2014 Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods Ecol. Evol.* **5**, 944–946. (doi:10.1111/2041-210X.12225)

15. Nakagawa, S. & Schielzeth, H. 2013 A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142. (doi:10.1111/j.2041-210x.2012.00261.x)