Animal Behaviour 81 (2011) 731–739

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Circulating testosterone levels do not affect exploration in house sparrows: observational and experimental tests

Ariane Mutzel [a,*], Bart Kempenaers [a], Silke Laucht [a], Niels J. Dingemanse [a], James Dale [a,b,1]

[a] Department Behavioural Ecology & Evolutionary Genetics, Max Planck Institute for Ornithology
[b] Institute of Natural Sciences, Massey University

ARTICLE INFO

Article history:
Received 9 September 2010
Initial acceptance 21 October 2010
Final acceptance 3 January 2011
Available online 4 February 2011
MS. number: 10-00603

Keywords:
exploration
exploratory behaviour
house sparrow
novel environment
Passer domesticus
personality
testosterone

Despite rapidly accumulating evidence for the existence of consistent individual differences in suites of correlated behaviours (i.e. 'animal personalities'), little is known about proximate mechanisms causing such variation. Individual variation in circulating levels of testosterone (T) is often hypothesized to underpin personality traits such as aggressiveness and exploratory behaviour. Here we provide a comprehensive test of this hypothesis. We quantified variation in exploratory behaviour of a novel environment in a captive population of wild-caught male house sparrows, *Passer domesticus*. We then investigated the relationship between the observed behaviours and circulating levels of T, using two approaches. First, we tested whether measures of exploratory behaviour correlated with (1) point-sampled plasma T levels and (2) T-dependent ornamentation (bill coloration) in 114 males. Neither direct nor indirect estimates of individual variation in T levels were correlated with the assayed behaviours. Second, we experimentally increased plasma T levels of 21 males with T implants, using 21 placebo-implanted males as a control group. Experimentally induced between-individual variation in T levels did not increase the amount of between-individual variation in exploratory behaviour. Our results thus strongly suggest that, in house sparrows, between-individual variation in circulating levels of T cannot serve as a causal explanation for the existence of individual variation in exploratory behaviour.

Animals constantly have to respond to changes in physical and social components of the environment. Whereas behavioural plasticity has long been assumed to represent an adaptive response to such environmental changes (Piersma & Drent 2003; Dall et al. 2004), accumulating evidence suggests that individuals do not show the full range of behavioural trait values present in their population (Réale & Dingemanse 2010). In other words, individuals often differ consistently in their behaviour over a range of environmental contexts (Dall et al. 2004; Sih et al. 2004; Réale et al. 2007). Such consistent individual differences in behaviour are referred to as 'animal personalities' (Dall et al. 2004; Réale et al. 2007; Dingemanse et al. 2010) and personality traits that are correlated are commonly referred to as 'behavioural syndromes' (Sih et al. 2004; Bell 2007; Réale et al. 2007). Despite abundant research focused on animal personalities, surprisingly little is known about the potential role of androgens as proximate mechanisms that underpin consistent individual variation in behaviour.

Consistent individual variation in behaviour might be linked proximately to individual differences in circulating hormone levels (Sih et al. 2004; Bell 2007; Réale et al. 2007). Baseline testosterone (T) levels in particular have been suggested to be a hormonal marker for individual differences in personality (Sellers et al. 2007). For instance, it has been shown that T influences the personality traits aggressiveness, boldness and activity (e.g. Wingfield et al. 1987; Koolhaas et al. 1999; Lynn et al. 2000), which jointly constitute a behavioural syndrome in birds (e.g. Verbeek et al. 1996; Drent & Marchetti 1999; Van Oers et al. 2004) and various other taxa (Réale et al. 2007). In addition, earlier observational and experimental studies indicated that T might be linked to behavioural persistency (Andrew 1972; Rogers 1974; Young & Rogers 1978). This trait is likely to be associated with the slow/fast personality type, as slow-exploring birds seem to show greater behavioural flexibility, whereas fast-exploring birds tend to stick to routine-like behaviour (Marchetti & Drent 2000). Finally, seasonal fluctuations in T levels

* Correspondence: A. Mutzel, Department Behavioural Ecology & Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner-Strasse, 82319 Seewiesen, Germany.
E-mail address: mutzel@orn.mpg.de (A. Mutzel).
[1] J. Dale is at the Institute of Natural Sciences, Massey University, Private Bag 102904, North Shore City 0745, Auckland, New Zealand.

0003-3472/$38.00
doi:10.1016/j.anbehav.2011.01.001

(Hegner & Wingfield 1986) could potentially explain seasonal differences in exploration within individuals because both T levels and the tendency to explore seem to peak in spring (Dingemanse et al. 2002; Mettke-Hofmann 2007; Quinn et al. 2009).

Since hormones typically act on several physiological mechanisms simultaneously, they could thereby mediate suites of correlated behavioural traits (Ketterson & Nolan 1999). Shared common hormonal mechanisms could thus also explain the occurrence of behavioural syndromes (Bell 2007). Even though numerous studies have shown that differences in personality traits are in part genetically determined (reviewed in Van Oers et al. 2005; Penke et al. 2007), relatively few studies have explicitly investigated hormonal pathways of androgens by which consistent between-individual differences in exploratory behaviour, a key avian personality trait, might come about (but see King 2002).

In this study, we used house sparrows, *Passer domesticus*, to evaluate the relationship between exploratory behaviour (known to be part of a behavioural syndrome in birds) and circulating levels of T within and between individuals. We asked whether (experimentally induced) consistent individual variation in T affects exploratory behaviour (i.e. causes consistent individual variation in behaviour). We therefore quantified exploration of a novel environment in a captive population of male house sparrows. Such behaviour has frequently been used as a standard measure to quantify variation in a key 'avian personality' trait (see Dingemanse et al. 2002) and has been shown to correlate with numerous other behavioural traits, such as boldness, aggression and dominance (Dingemanse & De Goede 2004 and references therein).

First, we used observational data to test whether variation in exploratory behaviour correlated with two separate indexes of T levels: (1) a direct point estimate of circulating plasma T levels from a blood sample and (2) bill coloration. The degree of melanization of the bill has been shown to be a reliable method for estimating a recent 'running average' of circulating T levels (Keck 1933; Witschi 1936; Pfeiffer et al. 1944; Laucht et al. 2010), and is highly repeatable between consecutive years in the birds used for this study (Laucht et al. 2010). Second, we experimentally manipulated T levels of individual males with two distinct objectives. (1) We evaluated whether changes in behaviour within individuals were underpinned by changes in T levels, by comparing within-individual changes in behavioural traits before and after the T implantation. (2) We tested whether experimentally induced individual variation in T levels affected the amount of consistent individual variation in exploratory behaviour, by comparing a standardized index of variation between individuals (repeatability) with and without controlling for treatment effects.

METHODS

Animals and Housing

In December–January 2005–2006, about 1 year prior to the first experiments, we caught 136 male house sparrows with mist nets set up in barns around rural areas of Bavaria, Germany. The birds were transported to the institute by car, in dark compartments ($12 \times 12 \times 12$ cm) of a wooden box, and released into aviaries within 30–180 min of capture. They were housed in groups of 5–10 in adjacent semioutdoor aviaries (1.2×2.0 m and 4.0 m high) with one of the short sides enclosed only by chicken wire. All aviaries were fitted with two long perches crossing the aviaries (1.2 m), two wooden nestboxes, natural beech and spruce branches and sawdust. The birds were held under ambient outdoor temperatures and natural daylight conditions. They were fed ad libitum with a commercial seed mix ('Waldvogelfutter', RKW Süd, Universal Kraftfutterwerk, Kehl, Germany), sunflower seeds, crushed corn and wheat, pellets and laying mash and were given unrestricted access to drinking and bathing water and sand. We allowed the birds to breed every year, which they readily did. However, as the sex ratio was highly male biased (wild males were preferentially caught), not all males obtained a mate. The birds were caught and kept in captivity under licence from the government of Upper Bavaria. After we finished our experiments, the birds were kept for further studies.

Indexes of Male Testosterone Levels

Blood plasma testosterone

We collected 150–200 µl of blood from each individual by puncturing the brachial vein. All samples were obtained between 0900 and 1430 hours. To reduce variation in plasma T levels caused by variation in handling stress, we bled all individuals within 15 min after entering the housing aviaries for catching. Laucht et al. (2010) has shown for the same data set that the time between the start of catching birds and blood sampling does not affect plasma T levels, implying that effects of handling stress did not bias the data. After centrifuging the blood, we extracted the plasma and froze it at 80 °C. Plasma T levels were measured at the endocrine laboratory of the Institute for Zoo and Wildlife Research, Berlin, using the enzyme immunoassay described by Roelants et al. (2002). The interassay coefficient of variation (CV) was 12.3% and the intra-assay CV was 9.0%; for further details see Laucht et al. (2010).

Measured T levels were tested for cross-season repeatability as our observational study was conducted in the run-up to the breeding season (spring), whereas our experimental study was conducted in the nonbreeding season (autumn). Individual variation in spring plasma T levels was correlated, albeit weakly, with variation observed in the autumn (Pearson correlation: $r_{40} = 0.35$, $P = 0.02$; Appendix Fig. A1).

Bill coloration

In house sparrows, bill coloration varies from pale horn to jet-black and this variation can be quantified using 'brightness' measured on digitized photographs (Laucht et al. 2010). We took three standardized pictures of each bird the day before each behavioural trial. The focal bird was held laterally in front of a photographic grey card and photographed with a Canon Power Shot S2 IS using the flash and standardized settings. We used digital photograph processing software written in open source R version 2.8.0 (R Development Core Team 2008) to measure colour parameters on photographs. We measured grey card brightness (on the Hue, Saturation, Brightness colour scale) on each photo as the mean value of three different randomly chosen spots around the bill, and bill brightness as the mean of three different randomly chosen spots on the lower bill. We then averaged scores across the three photos. To compensate for slight variation in light conditions between pictures, we standardized bill brightness values with the mean brightness value of all the grey card brightness measures. Bill brightness was measured independently by two observers (A.M. and J.D.) and showed a high interobserver repeatability (Pearson correlation: $r_{134} = 0.95$, $P < 0.0001$). We used the mean brightness value of both scorers for further analysis.

Novel Environment Test

Following recommendations by Réale et al. (2007), we modified the standard novel environment test described by Verbeek

et al. (1994) for great tits, *Parus major*, so that it was suitable for house sparrows. The experimental room was an outdoor aviary (1.7 × 2.0 m and 3.5 m high) with side walls made of wire mesh and sackcloth, rear wall and ceiling made of wire mesh with a semitransparent plastic cover and the ground covered with sawdust and hay. The front consisted of a wooden wall, a one-way screen (45 × 45 cm) through which observations were made, and a sliding door through which the focal bird was released into the observation room. We equipped the aviary with nine objects, providing 10 different positions: an artificial tree, a nest-shaped bag on the wall, a nestbox (=two positions), a food bowl, a perch, a tunnel-shaped bag on the ceiling, a hanging tree, a mirror and a shelf (for a more detailed description see Appendix Fig. A2). We scored the position of the focal bird in the aviary for 30 min after introduction (using an event recorder; The Observer 5.0.31, Noldus Information Technology, Wageningen, The Netherlands). Based on these observations, we measured (1) exploratory behaviour, the total number of objects visited within 30 min, and (2) activity, the total number of hops (including only movements when the bird moved 10 cm or more) and flights within and between objects and positions. Exploratory behaviour did not correlate with activity (Spearman correlation: $r_S = -0.08$, $N = 114$, $P = 0.39$) and these two behavioural measures were therefore regarded as independent behavioural axes. We further recorded weather conditions and temperature for each trial.

Observational Study

In February–March 2007, 114 of the 136 captured birds were exposed to a first round of behavioural trials (between 0830 and 1600 hours). Blood samples for measuring T and pictures of the bill were collected for all the birds ($N = 136$) during the middle of the test period (early March 2007; Laucht et al. 2010). Each bird from a single housing aviary was captured with a hand net, weighed (±0.05 g measurement accuracy), and photographed on the afternoon before each round of behavioural trials, and then transferred to a cage (1.2 × 0.4 m and 0.4 m high) stocked with ad libitum food and water. The following day, we placed each focal bird in a darkened box (45 × 35 cm and 20 cm high) connected to the experimental room via a sliding door for a period of 5 min, enabling the bird to recover from handling stress. We started the trial by opening the sliding door and releasing the bird into the aviary. The focal bird was visually but not acoustically isolated from other group members throughout the trial.

Our observational study was conducted during the onset of the breeding season (in Germany, the average laying date for house sparrows is mid-April, but nest sites are already defended from January onwards; Glutz von Blotzheim 1997), because there is substantial individual variation in plasma T levels at this time of year (Kempenaers et al. 2008; Laucht et al. 2010), implying that any interindividual covariance with other traits (here: exploratory behaviour and activity) would also be estimable.

Approximately 9 months after the first round of trials (November–December 2007), we repeated the same trials described above with 48 randomly chosen (from the initial 114) birds to determine whether exploratory behaviour and activity each varied consistently between individuals over time, where repeatability was calculated following Lessells & Boag (1987).

Experimental Study

In November–December 2008, we performed another round of behavioural trials on a random subsample of 42 of the 114 previously tested birds. Half of the birds were assigned to a T treatment group, the other half to a control (C) group. Two weeks before the study, we separated the birds according to treatment and housed them in constant groups of five to seven birds throughout the experimental procedure. We implanted 21 birds subcutaneously with T pellets designed to imitate the upper range of plasma T levels during the breeding season in free-living house sparrows (approximately 5–6 ng/ml; Hegner & Wingfield 1986) with a release time of 21 days (Cat. No. A-151, 1.5 mg/pellet, IRA, Sarasota, Florida, U.S.A.). We implanted the other 21 birds with placebo pellets. We used the following implantation procedure. We caught all birds from one aviary to minimize disturbance and placed them individually in dark wooden boxes until the implantation took place. While holding the bird, we made a small incision (about 3 mm) in the upper layer of the skin on the back, inserted the pellet with tweezers and closed the incision with special glue (VetGlu, B.Braun Vet Care GmbH, Tuttlingen, Germany). Following implantation, birds were released immediately into their housing aviary. The maximum time between catching and releasing was 30 min. Our study was conducted in the nonbreeding season (autumn), because at this time of year plasma T levels are well below the peak values (Laucht et al. 2010), implying that experimental manipulation of plasma T levels would produce detectable increases within the natural range (i.e. ceiling effects might have occurred if the manipulation had been done during the breeding season instead).

Each of the 42 birds was behaviourally tested three times at intervals of 4 weeks using the same procedure as described above: once 3 weeks before implanting (trial 1), once 1 week after implanting (trial 2) and once 5 weeks after implanting (trial 3). In November–December 2008, we randomly chose three birds from the T treatment group and three birds from the C group within each test day, and assayed them in a random order to control for potential effects of time of day and temperature. We carried out the trials blindly, that is, without knowing to which treatment group the focal bird belonged. We took blood samples of every bird immediately after each of the three trials.

Statistics

Observational study

We used R version 2.8.0 (R Development Core Team 2008) for statistical analysis unless stated otherwise. T values were log transformed to achieve a normal distribution. Non-normally distributed data were Box–Cox transformed (number of objects, activity and bill brightness) and all data were standardized (i.e. expressed in standard deviation units) prior to analysis as recommended by Rasbash et al. (2004). We used general linear models (GLMs) to assess the relationship between the number of objects and activity and a number of response variables, where we used backward elimination of nonsignificant terms ($P \geq 0.05$) as a model selection criterion (Crawley 1993). We used the chi-square-distributed Wald statistic to calculate significance of predictor variables. The relationship between plasma T level and ornamentation measures and the relationship between T estimates and exploratory behaviour and activity were analysed with Pearson and Spearman rank correlations, respectively.

Experimental study

For statistical analysis of treatment effects, we used MLwiN version 2.0 (Rasbash et al. 2004). Data were Box–Cox transformed when necessary (number of objects and activity) to meet the distributional assumptions (normality) of multivariate linear mixed models (LMMs). T values were log transformed, and all data were standardized (expressed in standard deviation units) prior to analysis as recommended by Rasbash et al. (2004). A multivariate

LMM with individual as random effect and with normal errors was used to test simultaneously for manipulation effects on activity, exploratory behaviour, bill brightness and plasma T while taking into account the covariance between these four variables between and within individuals. Trial (before, during, after), treatment group (control versus T) and their interaction were fitted as fixed effects (all categorical variables). We used the chi-square-distributed Wald statistic to calculate significance of all fixed and random effects. Post hoc comparisons were performed to investigate further whether T treatment had a significant effect on plasma T levels. Repeatability was calculated from the multivariate LMM as the proportion of between-individual variation relative to the sum of the between- and within-individual variance for each trait (Rasbash et al. 2004) and standard errors were calculated following Becker (1984).

Ethical Note

Captive conditions

The birds used for this study were held in captivity for about a year prior to the study. We are confident that captivity did not have any adverse effect on the wellbeing of the birds because (1) the mean body mass of the captive birds (29.19 ± 1.72 g) was within the natural range for house sparrows in southern Germany (range 27–35.5 g, mean 29.9 g; Lowther & Cink 2006); (2) the annual mortality rate of our captive birds was between 0.05 and 0.09, which is much lower than mortality rates in free-living house sparrows (0.4–0.5; Senar & Copete 1995); and (3) our population readily breeds in the captive conditions provided, indicating conditions are generally favourable.

Testosterone implantation surgery

The implantation procedure took about 3 min per bird. The perforation of the rather loose upper layer of the skin did not cause bleeding. Moreover, none of the birds showed any sign of pain during the procedure (e.g. twitching) and all birds started feeding within 15 min after release (A. Mutzel, personal observation). The procedure applied is nearly identical to the implantation of passive integrated transponder (PIT) tags. Fitting birds with PIT tags has been shown to have no adverse effect on the survival and recruitment of nestling and adult great tits (Nicolaus et al. 2008). We did not use anaesthesia as this can have harmful effects on birds, including a high risk of death (Ludders 1998). In this study, experimental treatment did not appear to have any long-term adverse effect on the birds, because the mortality rate of the experimental birds in 2009 was slightly lower than that of nonexperimental birds (0.07 versus 0.09). The implantation was approved by the government of Upper Bavaria and carried out in accordance with the German animal protection law.

Table 1
Sources of variation in exploratory behaviour and activity in male house sparrows ($N = 114$)

	Exploratory behaviour		Activity	
	Wald χ^2_1	P	Wald χ^2_1	P
Date	0.02	0.89	0.18	0.68
Time of day	11.69	**<0.001**	0.83	0.36
Weather condition	2.68	0.10	8.52	**0.03**
Temperature	0.25	0.62	4.84	**0.03**
Bird mass	0.53	0.47	5.18	**0.02**

The results are from univariate GLMs with exploratory behaviour or activity as response variable, and date, time of day (in minutes after sunrise), weather conditions (sun/clouds/precipitation), temperature (in °C) and bird mass fitted as predictor variables. We used backward elimination of nonsignificant terms; Wald chi-square values given are for the inclusion of the variable in the final model. Significant P values are indicated in bold.

Table 2
Spearman rank correlations for exploratory behaviour, and activity, with circulating plasma T level and bill brightness (bill) in male house sparrows ($N = 114$)

	Exploratory behaviour		Activity	
	r_S	P	r_S	P
Plasma T	−0.02	0.81	0.04	0.67
Bill	0.00	0.97	−0.04	0.64

RESULTS

Observational Study

As expected, bill brightness at the time of blood sampling was negatively correlated with plasma T levels (Pearson correlation: $r_{134} = 0.25$, $P = 0.003$): birds with a paler bill had on average lower plasma T levels than birds with darker bills. Exploratory behaviour was significantly affected by time of day for first tests (Table 1). Date, temperature, weather conditions and body mass did not correlate with exploratory behaviour, and were therefore not included in the final model. Plasma T levels and bill brightness on

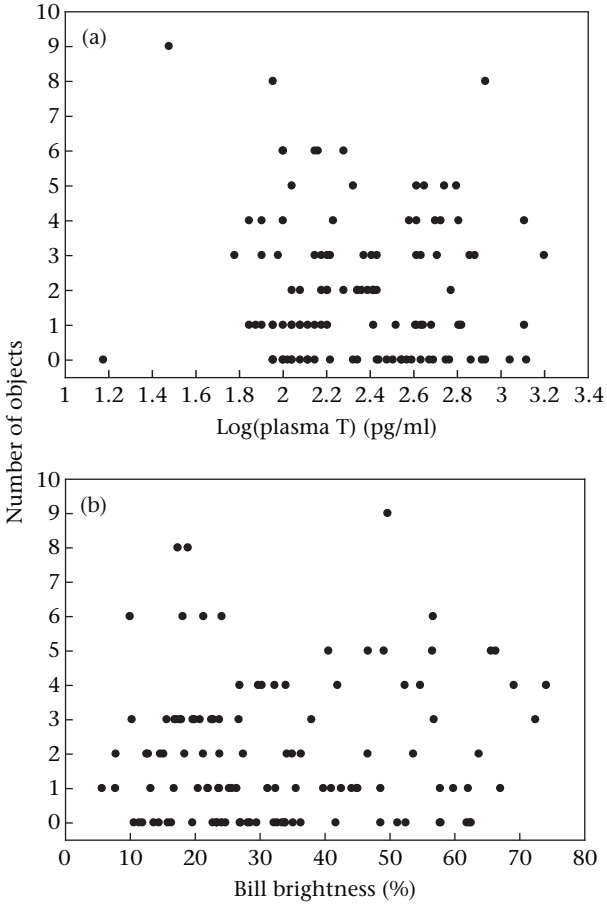

Figure 1. Correlation between exploratory behaviour (actual number of objects explored within 30 min) and two estimates of individual T levels in male house sparrows. (a) Plasma T level ($r_S = -0.02$) and (b) bill brightness ($r_S = 0.00$) as a percentage.

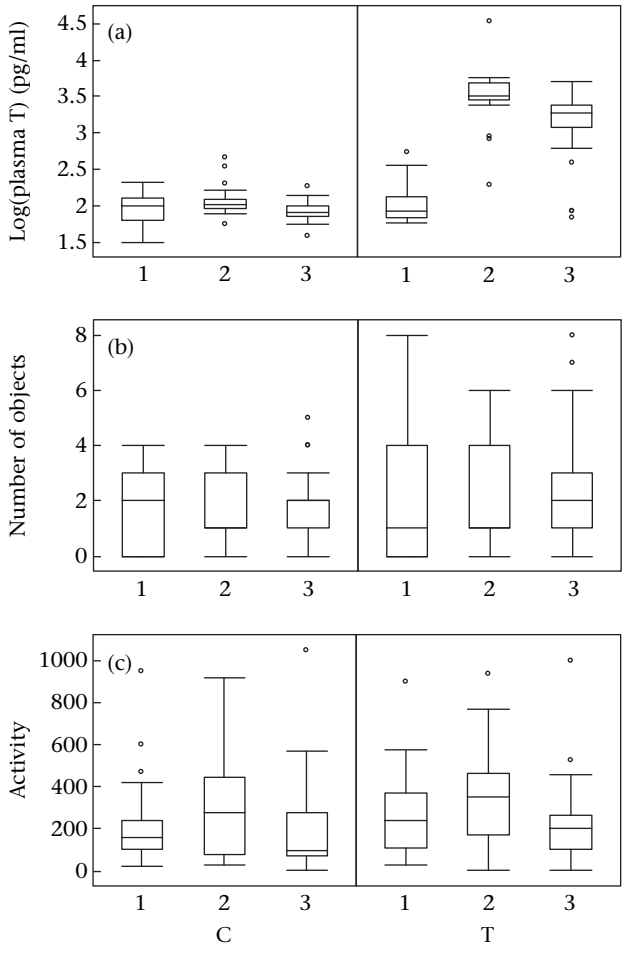

Figure 2. Plasma T levels and behaviour in the control (C) group and the testosterone (T) treatment group before (trial 1), 1 week after (trial 2) and 5 weeks after (trial 3) the implantation in male house sparrows. (a) Measured plasma T levels, (b) exploratory behaviour (actual number of objects visited) and (c) activity (total number of hops and flights). Box plots show the median and the interquartile range from the 25th to the 75th percentile. Whiskers indicate the 10th to the 90th percentile. Dots show outliers.

the day before the behavioural assay did not correlate significantly with exploratory behaviour (Table 2, Fig. 1). Activity was influenced by temperature, weather conditions and body mass before the experiment, but not by date or time of day (Table 1). Activity levels were higher on relatively rainy and cool days. Heavier birds were also more active than relatively light birds. Neither plasma T nor bill brightness was correlated with activity (Table 2). The behaviour in a novel environment varied consistently between the 48 individuals that were retested 9 months after their initial test. Repeatability was 0.45 ± 0.12 ($F_{47,48} = 2.62$, $P < 0.001$) for exploratory behaviour and 0.48 ± 0.11 ($F_{47,48} = 2.82$, $P < 0.001$) for activity.

Experimental Study

To investigate whether T treatment affected plasma T levels and bill brightness, we examined the interaction between trial and treatment group (the actual T treatment took place during trial 2) and changes in these traits with trial. As expected, we found highly significant effects of T implantation on plasma T levels (interaction trial*treatment: $\chi^2_2 = 117.87$, $P < 0.001$) and bill brightness (interaction trial*treatment: $\chi^2_2 = 70.85$, $P < 0.001$), implying that within-individual variation in T causally influenced the expression of these phenotypic traits. To explore the nature of the treatment effect, we performed post hoc comparisons. Within the C group, plasma T levels did not differ between trials (Fig. 2a, Appendix Table A1). Within the T treatment group, we found a significant effect of trial on plasma T levels. The highest levels were recorded directly following T manipulation (trial 2) and were still elevated in trial 3 compared to T levels at the beginning of the experiment (Fig. 2a, Appendix Table A1). Furthermore, plasma T levels differed significantly between the C and T groups after the experimental manipulation (in trials 2 and 3) but not before, implying that we succeeded in our aim to induce between-individual variation in T levels (see also Table 3). Birds from the T treatment group showed significantly higher plasma T levels than individuals from the C group ($P < 0.001$; Fig. 2a, Appendix Table A1). Plasma T levels of the birds from the C group ranged between 0.03 and 0.45 ng/ml (mean 0.11 ± 0.06 ng/ml). In the T treatment group plasma T levels were 0.06–0.55 ng/ml (mean 0.13 ± 0.12 ng/ml) in trial 1, 0.19–33.66 ng/ml (mean 4.72 ± 6.79 ng/ml) in trial 2 and 0.07–5.07 ng/ml (mean 1.92 ± 1.39 ng/ml) in trial 3. Excluding the unusually high plasma T value of one bird in trial 2 (33.66 ng/ml) did not change the results of the study.

To investigate whether T treatment had an effect on exploratory behaviour and activity, we examined the interaction of trial and treatment group (see above). Contrary to our prediction, neither exploratory behaviour ($\chi^2_2 = 0.35$, $P = 0.84$; Fig. 2b) nor activity ($\chi^2_2 = 0.3$, $P = 0.86$; Fig. 2c) was affected by T treatment. Next, we ran univariate LMMs to assess whether T treatment relates to individual differences in behaviour. Even though activity and exploratory behaviour were both highly repeatable, there was no significant change in repeatability of these behaviours before and after controlling for trial, treatment group and their interaction

Table 3
Variance components, repeatability values and significance for assayed phenotypic traits

	Controlling for fixed effects?	Variance component±SE		Repeatability		
		Between-individual variance	Within-individual variance	r±SE	χ^2_1	P
Activity	Yes	0.53±0.15	0.42±0.06	0.56±0.08	37.78	**<0.001**
	No	0.53±0.15	0.47±0.07	0.53±0.09	32.93	**<0.001**
Exploration	Yes	0.64±0.17	0.34±0.05	0.66±0.07	54.58	**<0.001**
	No	0.65±0.17	0.35±0.05	0.65±0.07	53.44	**<0.001**
Plasma T	Yes	0.03±0.02	0.18±0.03	0.14±0.10	2.31	0.129
	No	0.22±0.11	0.77±0.12	0.22±0.10	5.57	**0.018**
Bill	Yes	0.33±0.09	0.28±0.04	0.55±0.09	35.62	**<0.001**
	No	0.46±0.14	0.54±0.08	0.46±0.09	24.47	**<0.001**

Results are from univariate general mixed models with individual as random effect and activity, exploratory behaviour, plasma T level and bill brightness (bill) as dependent variables ($N = 42$ male house sparrows). In models where we controlled for fixed effects, the fixed effects included in the model are trial (before, during, after), treatment group (T and C) and the interaction between trial and treatment group. Significant P values are indicated in bold.

(Table 3): only 1% of the observed between-individual variance in activity and 0.1% of the variance in number of objects examined could be attributed to trial, treatment group and their interaction. These findings thus confirm experimentally that between-individual variation in T levels is not linked to between-individual variation in exploratory behaviour.

DISCUSSION

T Levels and Exploratory Behaviour

This study showed that individual differences exist in exploratory behaviour and activity in male house sparrows, but that these differences cannot be attributed to variation in circulating levels of the hormone T. Plasma T titre was not correlated with these two personality traits within our observational data set. Moreover, within individuals, experimentally elevated plasma T levels caused no changes (i.e. within-individual plasticity) in either exploratory behaviour or activity. Finally, experimentally induced differences in T levels did not affect the amount of individual variation in these behaviours.

In agreement with these findings, we did not detect any significant relationship between a reliable short-term proxy of T level (bill brightness) and the two personality traits. Importantly, we confirmed that bill brightness correlated significantly with circulating plasma T levels in our descriptive data set, and that implantation with T caused the bill of male sparrows to darken considerably (Witschi 1936; Pfeiffer et al. 1944). Bill coloration thus reliably predicts a short-term (i.e. 3–4 weeks) 'running average' of T levels. Our failure to find associations between this proxy for T and the two behaviours underlines the absence of T-mediated variation in exploratory behaviour.

Few other studies have investigated the relationship between T and consistent between-individual differences in behavioural traits. In agreement with our results, an observational study by Sellers et al. (2007) on humans also failed to detect any correlation between levels of circulating T and openness to experience, a human personality trait considered to be the equivalent of exploration behaviour in nonhuman animals (Gosling & John 1999). Similarly, in male greylag geese, *Anser anser*, T did not covary with another component of avian personality (aggressiveness; Kralj-Fišer et al. 2010). The results from other studies imply that the behavioural effects of circulating levels of T are either inconsistent or context specific. Some studies on domesticated birds have provided evidence for a causal effect of circulating plasma T levels on certain behavioural traits. For instance, domestic chickens, *Gallus gallus domesticus*, with experimentally increased circulating levels of T showed increased persistence in search for a particular type of food and pecked more often on one particular square before moving on to the next square (Andrew 1972; Rogers 1974; Young & Rogers 1978). However, in another study on domestic chickens, Archer (1973) failed to detect an effect of T on the number of squares explored during a novel environment test in a nonforaging context. However, T-treated males showed shorter latencies to move in a novel environment than controls. In contrast, King (2002) found no relationship between T and the latency to move after transfer to a novel environment in garter snakes, *Thamnophis sirtalis*, but showed instead that defensive behaviour was influenced by circulating levels of T. Furthermore, Jones & Andrew (1992) found T-related differences between male and female domestic chickens in the response to novelty. However, this effect seemed to be context dependent, with males being more explorative in a novel environment but less explorative in a foraging context. Other studies found that T is only related to dominance and aggression when dominance status is uncertain but not in socially stable situations (Ruiz-de-la-Torre & Manteca 1999; Collias et al. 2002). The potential effects of T on exploration and activity might therefore not be general, but rather depend on the context in which the behaviour is shown. For instance, it is possible that T mediates the expression of certain behaviours within a social context, but that it is less important in nonsocial situations, for example in a novel environment. This context dependency of the link between T and behaviour has recently been suggested by Koolhaas et al. (2010) for the relationship between T and aggression. To test this idea, it would be interesting to investigate the relationship between circulating plasma T levels and a number of different personality traits while manipulating context (i.e. social versus nonsocial).

Cross-season Repeatability of T

In many vertebrate species, plasma T levels show pronounced seasonal variation within individuals, with peak values occurring during the breeding season (Kempenaers et al. 2008). Whereas such seasonal changes in T levels have received considerable attention at the within-individual level, fairly little is known about (1) patterns of individual variation (e.g. repeatability) in plasma levels, (2) whether individual variation exists only in certain seasons, and (3) whether any differences between individuals in plasma T levels are maintained across seasons (i.e. is there cross-season repeatability?; Kempenaers et al. 2008; Laucht et al. 2010). Such information is valuable because it would facilitate further interpretation of our findings, as our study was carried out in two different seasons. We discuss here two scenarios. First, there is no cross-season repeatability of plasma T. This scenario would imply any individual variation observed within a season is not proximately linked with variation observed in other seasons (i.e. spring and autumn levels of T should be regarded as different 'traits') and that we tested different questions with our observational and experimental studies. Additionally, our experimental study could not be used to test whether any link between plasma T and personality traits documented in our observational study was causal. Second, there is cross-season repeatability of plasma T. This scenario would imply that individual variation observed within a season is proximately linked with variation observed in other seasons (i.e. spring and autumn levels of T measured the same trait). We tested for a cross-season repeatability of plasma T and found that there was a statistically significant but low (Pearson: $r = 0.35$) correlation. This result implies that our experimental study can be used to evaluate the causal link between plasma T and exploratory behaviour observed in other seasons (however, also see Laucht et al. 2010).

Alternative Pathways by which T could Affect Personality

Although our observational study did not provide any supportive evidence in favour of a relationship between circulating plasma T levels and exploratory behaviour and activity, there are several alternative pathways by which T could still mediate between-individual differences in these personality traits (Ketterson & Nolan 1992). For instance, there might be genetically determined variation in T receptor density, affinity and specificity that does not necessarily have to be linked with circulating levels of T (Adkins-Regan 2005; Ball & Balthazart 2008). Moreover, it is known that heritable variation in plasma-binding proteins can have substantial effects on the availability of a certain hormone (Dufty et al. 2002; Ball & Balthazart 2008). Potential effects of T on behaviour acting via between-individual differences in such alternative pathways might simply be obscured when one measures

circulating plasma T levels only. However, under such scenarios we would expect changes in behaviour following experimental elevations of T, which we did not find. Therefore it seems rather unlikely that such alternative hormonal mechanisms would affect variation in exploratory behaviour.

Nevertheless, we cannot completely exclude the possibility of a link between T and exploratory behaviour. For instance, phenotypic expression of T receptors could match phenotypic expression of circulating hormone levels. In this case endogenous levels of T might already be sufficient to saturate the receptors and experimentally increasing T levels would not have any additional effect on behaviour. It might also be that T receptor sensitivity is lower in the nonbreeding season than in the breeding season (Canoine et al. 2007). In this case, we do not expect to find a strong effect of T treatment on behaviour. However, this scenario is rather unlikely, because T treatment had a strong effect on bill coloration, a trait that has been suggested to be a good indicator of quality and/or behavioural strategies (Laucht et al. 2010).

T could also (indirectly) affect behaviour through an alternative pathway involving the stress hormone corticosterone. There is good evidence that corticosterone responses are linked with certain behavioural types (reviewed in Cockrem 2007). Moreover, numerous studies have shown that corticosterone and T titres tend to covary (e.g. Wilson et al. 1979; Siegel 1980; but see also Hau et al. 2010). Therefore, observed links between T and behaviour in certain contexts may be caused by a direct effect of T on the behaviour or an indirect effect via corticosterone. More studies are needed to investigate potential direct and indirect links between corticosterone and T and behaviour.

This study did not provide any evidence in support of a relationship between circulating plasma T levels and exploratory behaviour. We should underscore that we focused here on activational effects of T on behaviour. T could still be involved in forming this personality trait through its organizational effects, that is, by causing irreversible differences in behavioural phenotype during early development of an individual (Daisley et al. 2005; Groothuis et al. 2005). Therefore, another fruitful next step would be to investigate the relationship between maternal T deposited in eggs and personality traits.

Conclusions

We showed that male house sparrows differ consistently in exploratory behaviour and activity in a novel environment. This variation was not related to circulating levels of plasma T. Our results thus challenge the assumption of a T-based mechanism causing variation in these personality traits. Furthermore, our results suggest that T is not part of potential proximate mechanisms that cause behavioural syndromes, which include exploratory behaviour or activity. However, T might still be involved in creating individual differences in other personality traits, such as boldness or aggressiveness or influence activity and exploration within other, for example social, contexts.

Acknowledgments

We thank Erich Koch, Agnes Türk and Christian Völk, Andrea Wittenzellner and Rebecca Wutzler for technical support and Magnus Rudorfer, Edith Bodendorfer, Annemarie Grötsch and Frances Weigl for animal care. We are grateful to Wolfgang Forstmeier, Alain Jacot, Holger Schielzeth, and Mihai Valcu for discussions and to Claudia Mettke-Hofmann and three anonymous referees for comments on the manuscript. The study was financed by the Max Planck Society.

References

Adkins-Regan, E. 2005. *Hormones and Animal Social Behavior*. Princeton, New Jersey: Princeton University Press.

Andrew, R. J. 1972. Recognition processes and behavior with special reference to effects of testosterone on persistence. In: *Advances in the Study of Behavior* (Ed. by D. S. Lehrmann, R. A. Hinde & E. Shaw), pp. 175–208. New York: Academic Press.

Archer, J. 1973. Influence of testosterone on chick behavior in novel environments. *Behavioral Biology*, **8I**, 93–108.

Ball, G. F. & Balthazart, J. 2008. Individual variation and the endocrine regulation of behaviour and physiology in birds: a cellular/molecular perspective. *Philosophical Transactions of the Royal Society B*, **363**, 1699–1710.

Becker, W. A. 1984. *A Manual of Quantitative Genetics*. Washington: Academic Enterprises.

Bell, A. M. 2007. Future directions in behavioural syndromes research. *Proceedings of the Royal Society B*, **274**, 755–761.

Canoine, V., Fusani, L., Schlinger, B. & Hau, M. 2007. Low sex steroids, high steroid receptors: increasing the sensitivity of the nonreproductive brain. *Developmental Neurobiology*, **67**, 57–67.

Cockrem, J. F. 2007. Stress, corticosterone responses and avian personalities. *Journal of Ornithology*, **148**, S169–S178.

Collias, N. E., Barfield, R. J. & Tarvyd, E. S. 2002. Testosterone versus psychological castration in the expression of dominance, territoriality and breeding behavior by male village weavers (*Ploceus cucullatus*). *Behaviour*, **139**, 801–824.

Crawley, M. J. 1993. *GLIM for Ecologists*. Oxford: Blackwell Scientific.

Daisley, J. N., Bromundt, V., Möstl, E. & Kotrschal, K. 2005. Enhanced yolk testosterone influences behavioral phenotype independent of sex in Japanese quail chicks *Coturnix japonica*. *Hormones and Behavior*, **47**, 185–194.

Dall, S. R. X., Houston, A. I. & McNamara, J. M. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, **7**, 734–739.

Dingemanse, N. J. & De Goede, P. 2004. The relation between dominance and exploratory behavior is context-dependent in wild great tits. *Behavioral Ecology*, **15**, 1023–1030.

Dingemanse, N. J., Both, C., Drent, P. J., Van Oers, K. & Van Noordwijk, A. J. 2002. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, **64**, 929–938.

Dingemanse, N. J., Kazem, A. J. N., Reale, D. & Wright, J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–89.

Drent, P. J. & Marchetti, C. 1999. Individuality, exploration and foraging in hand raised juvenile great tits. In: *Proceedings of the 22nd International Ornithology Congress, Durban* (Ed. by N. J. Adams & R. H. Slotow), pp. 896–914. Johannesburg: BirdLife South Africa.

Dufty, A. M., Clobert, J. & Møller, A. P. 2002. Hormones, developmental plasticity and adaptation. *Trends in Ecology & Evolution*, **17**, 190–196.

Glutz von Blotzheim, U. N. 1997. *Handbuch der Vögel Mitteleuropas*. Wiesbaden: AULA-Verlag GmbH.

Gosling, S. D. & John, O. P. 1999. Personality dimensions in nonhuman animals: a cross-species review. *Current Directions in Psychological Science*, **8**, 69–75.

Groothuis, T. G. G., Müller, W., von Engelhardt, N., Carere, C. & Eising, C. 2005. Maternal hormones as a tool to adjust offspring phenotype in avian species. *Neuroscience and Biobehavioral Reviews*, **29**, 329–352.

Hau, M., Ricklefs, R. E., Wikelski, M., Lee, K. A. & Brawn, J. D. 2010. Corticosterone, testosterone and life-history strategies of birds. *Proceedings of the Royal Society B*, **277**, 3203–3212.

Hegner, R. E. & Wingfield, J. C. 1986. Behavioral and endocrine correlates of multiple brooding in the semicolonial house sparrow, *Passer domesticus* .1. Males. *Hormones and Behavior*, **20**, 294–312.

Jones, R. B. & Andrew, R. J. 1992. Responses of adult domestic cocks and capons to novel and alarming stimuli. *Behavioural Processes*, **26**, 189–199.

Keck, W. N. 1933. Control of the bill color of the male English sparrow by injection of male hormone. *Proceedings of the Society for Experimental Biology and Medicine*, **30**, 1140–1141.

Kempenaers, B., Peters, A. & Foerster, K. 2008. Sources of individual variation in plasma testosterone levels. *Philosophical Transactions of the Royal Society B*, **363**, 1711–1723.

Ketterson, E. D. & Nolan, V. 1992. Hormones and life histories: an integrative approach. *American Naturalist*, **140**, S33–S62.

Ketterson, E. D. & Nolan, V. 1999. Adaptation, exaptation, and constraint: a hormonal perspective. *American Naturalist*, **154**, S4–S25.

King, R. B. 2002. Family, sex and testosterone effects on garter snake behaviour. *Animal Behaviour*, **64**, 345–359.

Koolhaas, J. M., Korte, S. M., De Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W. & Blokhuis, H. J. 1999. Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience and Biobehavioral Reviews*, **23**, 925–935.

Koolhaas, J. M., de Boer, S. F., Coppens, C. M. & Buwalda, B. 2010. Neuroendocrinology of coping styles: towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology*, **31**, 307–321.

Kralj-Fišer, S., Weiss, B. M. & Kotrschal, K. 2010. Behavioural and physiological correlates of personality in greylag geese (*Anser anser*). *Journal of Ethology*, **28**, 363–370.

Laucht, S., Kempenaers, B. & Dale, J. 2010. Bill color, not badge size, indicates testosterone-related information in house sparrows. *Behavioral Ecology and Sociobiology*, **64**, 1461–1471.
Lessells, C. M. & Boag, P. T. 1987. Unrepeatable repeatabilities: a common mistake. *Auk*, **104**, 116–121.
Lowther, P. E. & Cink, C. L. 2006. House sparrow (*Passer domesticus*). In: *The Birds of North America Online* (Ed. by A. Poole). Ithaca: Cornell Laboratory of Ornithology. http://bna.birds.cornell.edu/bna/species/012.
Ludders, J. W. 1998. Respiratory physiology of birds: considerations for anesthetic management. *Seminars in Avian and Exotic Pet Medicine*, **7**, 3–9.
Lynn, S. E., Houtman, A. M., Weathers, W. W., Ketterson, E. D. & Nolan, V. 2000. Testosterone increases activity but not daily energy expenditure in captive male dark-eyed juncos, *Junco hyemalis*. *Animal Behaviour*, **60**, 581–587.
Marchetti, C. & Drent, P. J. 2000. Individual differences in the use of social information in foraging by captive great tits. *Animal Behaviour*, **60**, 131–140.
Mettke-Hofmann, C. 2007. Object exploration of garden and Sardinian warblers peaks in spring. *Ethology*, **113**, 174–182.
Nicolaus, M., Bouwman, K. M. & Dingemanse, N. J. 2008. Effect of PIT tags on the survival and recruitment of great tits *Parus major*. *Ardea*, **96**, 286–292.
Penke, L., Denissen, J. J. A. & Miller, G. F. 2007. The evolutionary genetics of personality. *European Journal of Personality*, **21**, 549–587.
Pfeiffer, C. A., Hooker, C. W. & Kirschbaum, A. 1944. Deposition of pigment in the sparrow's bill in response to direct applications as a specific and quantitative test for androgen. *Endocrinology*, **34**, 389–399.
Piersma, T. & Drent, J. 2003. Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology & Evolution*, **18**, 228–233.
Quinn, J. L., Patrick, S. C., Bouwhuis, S., Wilkin, T. A. & Sheldon, B. C. 2009. Heterogeneous selection on a heritable temperament trait in a variable environment. *Journal of Animal Ecology*, **78**, 1203–1215.
R Development Core Team 2008. *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing. http://www.R-project.org.
Rasbash, J., Steele, F., Browne, W. & Prosser, B. 2004. *A User's Guide to MLwinN Version 2.0*. London: Centre for Multilevel Modelling, Institute for Education, University of London.
Réale, D. & Dingemanse, N. J. 2010. Personality and individual social specialisation. In: *Social Behaviour: Genes, Ecology and Evolution* (Ed. by T. Szekely, A. Moore & J. Komdeur), pp. 417–441. Cambridge: Cambridge University Press.
Réale, D., Reader, S. M., Sol, D., McDougall, P. T. & Dingemanse, N. J. 2007. Integrating animal temperament within ecology and evolution. *Biological Reviews*, **82**, 291–318.
Roelants, H., Schneider, F., Goritz, F., Streich, J. & Blottner, S. 2002. Seasonal changes of spermatogonial proliferation in roe deer, demonstrated by flow cytometric analysis of c-kit receptor, in relation to follicle-stimulating hormone, luteinizing hormone, and testosterone. *Biology of Reproduction*, **66**, 305–312.
Rogers, L. J. 1974. Persistence and search influenced by natural levels of androgens in young and adult chickens. *Physiology & Behavior*, **12**, 197–204.
Ruiz-de-la-Torre, J. L. & Manteca, X. 1999. Effects of testosterone on aggressive behaviour after social mixing in male lambs. *Physiology & Behavior*, **68**, 109–113.
Sellers, J. G., Mehl, M. R. & Josephs, R. A. 2007. Hormones and personality: testosterone as a marker of individual differences. *Journal of Research in Personality*, **41**, 126–138.
Senar, J. C. & Copete, J. L. 1995. Mediterranean house sparrows (*Passer domesticus*) are not used to freezing temperatures: an analysis of survival rates. *Journal of Applied Statistics*, **22**, 1069–1074.
Siegel, H. S. 1980. Physiological stress in birds. *Bioscience*, **30**, 529–534.
Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. 2004. Behavioral syndromes: an integrative overview. *Quarterly Review of Biology*, **79**, 241–277.
Van Oers, K., Drent, P. J., De Goede, P. & van Noordwijk, A. J. 2004. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society B*, **271**, 65–73.
Van Oers, K., De Jong, G., Van Noordwijk, A. J., Kempenaers, B. & Drent, P. J. 2005. Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour*, **142**, 1185–1206.
Verbeek, M. E. M., Drent, P. J. & Wiepkema, P. R. 1994. Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour*, **48**, 1113–1121.
Verbeek, M. E. M., Boon, A. & Drent, P. J. 1996. Exploration, aggressive behavior and dominance in pair-wise confrontations of juvenile male great tits. *Behaviour*, **133**, 945–963.
Wilson, E. K., Rogler, J. C. & Erb, R. E. 1979. Effect of sexual experience, location, malnutrition, and repeated sampling on concentrations of testosterone in blood-plasma of *Gallus domesticus* roosters. *Poultry Science*, **58**, 178–186.
Wingfield, J. C., Ball, G. F., Dufty, A. M., Hegner, R. E. & Ramenofsky, M. 1987. Testosterone and aggression in birds. *American Scientist*, **75**, 602–608.
Witschi, E. 1936. The bill of the sparrow as an indicator for the male sex hormone. I. Sensitivity. *Journal of Experimental Zoology*, **67**, 484–486.
Young, C. E. & Rogers, L. J. 1978. Effects of steroidal hormones on sexual, attack, and search behavior in isolated male chick. *Hormones and Behavior*, **10**, 107–117.

APPENDIX

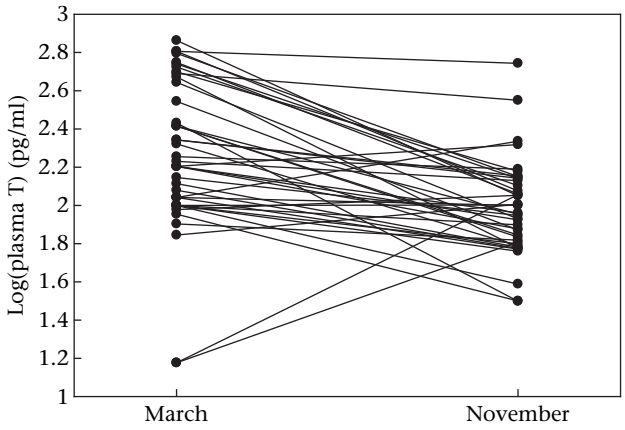

Figure A1. Plasma T levels in March and November 2007 of the individuals used for the experimental study ($N = 42$). The black lines connect plasma T measurements of the same individual at the two seasons. Parallel lines indicate cross-season repeatability of T levels.

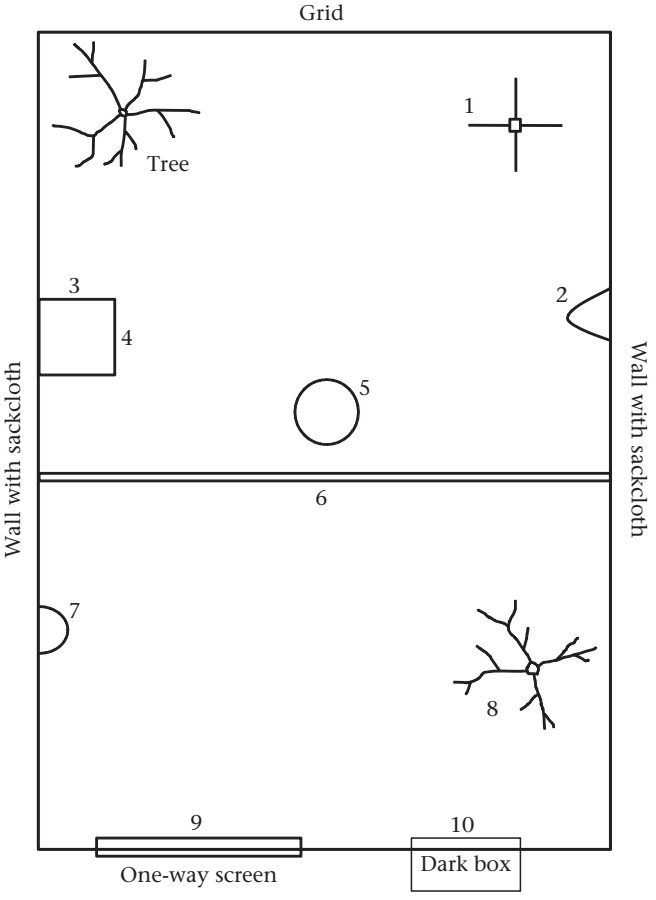

Figure A2. Schematic overview of the observation room. (1) Artificial tree (made of wood with a trunk of 4 × 4 cm, height 1.5 m and four cylindrical branches of 20 cm), (2) bag on wall (nest shaped, made of sackcloth), (3) + (4) nestbox (roof and inside), (5) food bowl with five sunflower seeds (on the ground, diameter 30 cm), (6) perch (crossing the aviary, length 1.7 m), (7) bag on ceiling (tunnel shaped, made of sackcloth), (8) hanging tree (branch hanging from the ceiling, 50 cm in length), (9) mirror (one-way screen through which observations were made), (10) shelf (in front of the sliding door).

Table A1
Results from a post hoc comparison of a linear mixed model with individual as random effect and trial, treatment group and their interaction as fixed effects

	χ^2	*df*	*P*	β±SE
Control group				
Trial 2 vs 1				
Activity	3.058	1	0.080	0.348±0.199
No. of objects	0.054	1	0.816	0.042±0.179
Bill brightness	6.129	1	**0.013**	0.401±0.162
Plasma T	1.608	1	0.205	0.168±0.033
Joint	10.43	4	**0.034**	
Trial 3 vs 1				
Activity	0.358	1	0.550	−0.119±0.199
No. of objects	0.464	1	0.496	0.122±0.179
Bill brightness	15.49	1	**<0.001**	0.637±0.162
Plasma T	0.025	1	0.874	−0.021±0.133
Joint	17.42	4	**0.002**	
Trial 3 vs 2				
Activity	5.508	1	**0.019**	−0.466±0.199
No. of objects	0.201	1	0.654	0.080±0.179
Bill brightness	2.133	1	0.144	0.236±0.162
Plasma T	2.035	1	0.154	−0.189±0.133
Joint	11.77	4	**0.019**	
Testosterone treatment group				
Trial 2 vs 1				
Activity	1.011	1	0.315	0.200±0.199
No. of objects	0.830	1	0.362	0.163±0.179
Bill brightness	14.11	1	**<0.001**	−0.608±0.162
Plasma T	248.7	1	**<0.001**	2.090±0.133
Joint	255.2	4	**<0.001**	
Trial 3 vs 1				
Activity	1.342	1	0.247	−0.230±0.199
No. of objects	2.095	1	0.148	0.259±0.179
Bill brightness	63.41	1	**<0.001**	1.289±0.162
Plasma T	131.4	1	**<0.001**	1.519±0.133
Joint	186.5	4	**<0.001**	
Trial 3 vs 2				
Activity	4.683	1	**0.030**	−0.430±0.199
No. of objects	0.288	1	0.592	0.096±0.179
Bill brightness	17.70	1	**<0.001**	−0.681±0.162
Plasma T	18.55	1	**<0.001**	−0.571±0.133
Joint	44.34	4	**<0.001**	
Within trials				
Trial 1				
Activity	0.805	1	0.370	0.278±0.300
No. of objects	0.083	1	0.773	0.088±0.305
Bill brightness	0.062	1	0.803	0.060±0.241
Plasma T	0.367	1	0.545	0.087±0.143
Joint	1.602	4	0.808	
Trial 2				
Activity	0.187	1	0.665	0.130±0.300
No. of objects	0.470	1	0.493	0.209±0.305
Bill brightness	15.52	1	**<0.001**	−0.948±0.241
Plasma T	197.5	1	**<0.001**	2.009±0.143
Joint	212.4	4	**<0.001**	
Trial 3				
Activity	0.307	1	0.580	0.166±0.300
No. of objects	0.544	1	0.461	0.225±0.305
Bill brightness	60.07	1	**<0.001**	−1.865±0.241
Plasma T	129.5	1	**<0.001**	1.627±0.143
Joint	185.2	4	**<0.001**	

Trial differences within the control group, trial differences within the testosterone treatment group and treatment group differences within trials are shown. Significant *P* values are indicated in bold.