Title: Effects of experimentally sustained elevated testosterone on incubation behaviour and reproductive success in female great tits (*Parus major*)

Authors: Berber de Jong[a,b,c], Luc Lens[b], Seyed Mehdi Amininasab[c,d], Kees van Oers[e],Veerle M. Darras[f], Marcel Eens[g], Rianne Pinxten[g,h], Jan Komdeur[c] & Ton G.G. Groothuis[a]

Affiliations:

a. Behavioural Biology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, P.O. Box 11103, 9700 CC, Groningen, The Netherlands

b. Terrestrial Ecology Unit, Department Biology, Ghent University, BE-9000, Ghent, Belgium.

c. Behavioural and Physiological Ecology, Groningen Institute for ELIFES, University of Groningen, P.O. Box 11103, 9700 CC, Groningen, The Netherlands

d. Department of Environmental Science, Behbahan Khatam Alanbia University of Technology, Iran

e. Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 50, 6700 AB, Wageningen, The Netherlands

f. Laboratory of Comparative Endocrinology, K.U. Leuven, Naamsestraat 61, B-3000 Leuven, Belgium

g. Behavioural Ecology & Ecophysiology Research Group, Department of Biology, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium

h. Didactica Research Group, Faculty of Social Sciences, University of Antwerp, B-2000, Antwerp, Belgium

Corresponding author:

Berber de Jong

Behavioural and Physiological Ecology

Groningen Institute for Evolutionary Life Sciences

University of Groningen,

P.O. Box 11103, 9700 CC,

Groningen, The Netherlands

Tel: +31-629478683

Email: berber.dejong@gmail.com

ABSTRACT

In many seasonally breeding birds, female and male testosterone (T) levels peak at the start of the breeding season, coinciding with pair bonding and nesting activities. Shortly after the onset of egg laying, T levels slowly decline to baseline levels in both sexes, but more rapidly so in females. During this period, T in males may still function to facilitate territorial behaviour, mate guarding and extra pair copulations, either via short lasting peaks or elevated basal levels of the hormone. In some species, however, males become insensitive to increased T after the onset of egg laying. It has been postulated that in these species bi-parental care is essential for offspring survival, as T is known to inhibit paternal care. However, only very few studies have analysed this for females. As females are heavily involved in parental care, they too might become insensitive to T after egg laying. Alternatively, because territorial defence, mate guarding and extra pair copulations are expected to be less important for females than for males, they may not have had the need to evolve a mechanism to become insensitive to T during the period of maternal care, because their natural T levels are never elevated during this part of the breeding season anyway. We tested these alternative hypotheses in female great tits (*Parus major*). Male great tits have previously been shown to be insensitive to T after egg laying with regard to nestling feeding behaviour (but not song rate). When females had started nest building, we experimentally elevated their T levels up to the nestling feeding phase, and measured incubation behaviour (only females incubate) and reproductive success. T did not significantly affect nest building or egg laying behaviour, although egg laying tended to be delayed in T females. Females with experimentally enhanced T maintained lower temperature during incubation but did not spend less time incubating. This might explain the reduced hatching success of their eggs, smaller brood size and lower number of fledglings we found in this study. As in this species T-dependent behaviour by females during the phase of parental care is not needed, the results support the hypothesis that in this species the need for selection in favour of T-insensitivity did not occur.

Keywords: *Parus major*; experimentally elevated testosterone; incubation behaviour; reproductive success; essential mating effort hypothesis

1. INTRODUCTION

Testosterone (T) concentrations in males remain important for reproductive success in many seasonally breeding birds after territory establishment and pairing, although they decline gradually after the onset of egg laying (Ketterson et al., 2005; Wingfield, 1990). T is known to facilitate mate guarding, extra-pair fertilization, secondary female acquisition, and/or territorial defence during parental care. For example, experimentally elevated T has been shown to increase singing behaviour to attract additional mates (European starling *(Sturnus vulgaris)*, De Ridder et al., 2000), courtship behaviour (house sparrow (*Passer domesticus)*, Hegner and Wingfield, 1987; red bishop *(Euplectes* orix*)*, Edler et al., 2011), extra-pair fertilization rate (dark-eyed junco (*Junco hyemalis),* Raouf et al., 1997) and attractiveness to females (dark-eyed junco, Enstrom et al., 1997) later on in the breeding season. Yet T has also been shown to suppress incubation behaviour (spotted sandpiper (*Actitis meculariа)*, Oring et al., 1989; yellow-legged gull *(Larus cachinnans),* Alonso-Alvarez, 2001) and nestling provisioning (Hegner and Wingfield, 1987) in males. Thus, elevated levels of T in males appear to moderate the trade-off between mating effort and parental effort (Adkins-Regan, 2005). In certain species, however, males do not respond to T elevation after egg laying with a reduction of parental behaviour (lapland longspur *(Calcarius lapponicus)*, Hunt et al., 1997; chestnut-collared longspur *(Calcarius ornatus)*, Lynn et al., 2002; great tit *(Parus major)*, Van Duyse et al., 2002; black-tailed gull *(Larus crassirostris)*, Kazama et al., 2011). Such variation in sensitivity of male parental behaviour to T after egg laying may be explained by the essential paternal care hypothesis, which postulates that in species where bi-parental care is essential for the survival of offspring, males, becomes insensitive to T (in terms of their parental behaviours) during the period of increased paternal care (Lynn et al., 2002; Lynn, 2008) in order to avoid the detrimental effects of the hormone.

In many avian species, female testosterone (T) levels rise, as in males, at the start of their breeding season but decline shortly after the start of egg laying (Ketterson et al., 2005). Although less well studied than in males, there is some evidence that the seasonal peak in T levels might be beneficial for females too. For example, early peak T levels are linked to female aggression (red-winged blackbird (*Agelaius phoeniceus*), Searcy, 1988; European starling, Sandell, 2007; blue tits *(Cyanistes caeruleus*), de Jong, 2013; tree swallow *(Tachycineta bicolor*), Rosvall, 2013a, b). This can help securing male care by outcompeting rivalling females (great tit, Slagsvold 1993, dunnock *(Prunella modularis*), Langmore et al., 2002; Sandell, 2007). Also, more aggressive females had higher reproductive success (dark-eyed junco, Cain and Ketterson, 2012). Females of many species also remain sensitive to elevated T levels later in the breeding season. However, in contrast to males, this prolonged sensitivity to T in females has mainly been associated with costs that could reduce reproductive success. For example, experimentally prolonged elevated T levels have been shown to delay the onset of egg laying (Searcy, 1988; dark-eyed junco , Clotfelter et al., 2004 ; zebra finch (*Teaniopygia guttata*), Rutkowska et al., 2005), decrease incubation temperature (Rosvall, 2013a,b), reduce brooding of nestlings (dark-eyed junco, O'Neal et al., 2008), and decrease the number of hatchlings and fledglings (spotless starling *(Sturnus unicolor)*, Veiga and Polo, 2008; spotless starling, Lopez-Rull and Gil, 2009; Rosvall, 2013a,b; dark-eyed junco, Gerlach and Ketterson, 2013) in various passerine birds. In other passerine species, however, prolonged experimentally elevated T levels does not affect the onset of egg laying (de Jong, 2013) or incubation

behaviour (European starling, Sandell et al., unpublished manuscript cited in Ketterson et al., 2005). Moreover, a few studies suggest that remaining sensitive to elevated T levels after egg laying might be advantageous for females. For example, prolonged elevated T levels are linked to female aggression (Searcy, 1988; Sandell, 2007; Rosvall, 2013a,b) which can secure male care by enabling females to outcompete rivals (European starling, Sandell, 1998; Langmore et al., 2002). A similar suggestion was made by Rosvall 2013a, demonstrating sensitivity to T during the period of parental care in tree swallow.

Thus, in several bird species, female parental care appears to remain sensitive to elevated T levels after egg laying. This has been explained by the "essential mating effort hypothesis", (Rosvall, 2013a, b). This hypothesis postulates that females in these species did not evolve insensitivity to T, since sensitivity to T is required for mediating other behaviours that are important for reproductive success, such as securing mates and nest sites, also during the period of parental care. Therefore, the benefits remaining sensitive to T by T dependent mating effort may outweigh the potential costs of suppression of parental care.

However, the majority of studies of sensitivity to T in females after egg laying have applied T implantations to test behavioural sensitivity. This may have induced much higher T levels than the endogenously produced low levels during that time period. It is therefore conceivable that females react to these artificially elevated levels, showing sensitivity to the hormone, because selection for becoming insensitive to the hormone was never necessary. In such species the costs of maintaining T production and remaining sensitive to it (reduction of parental care) would outweigh the benefits. This may be the case in many species in which females during the phase of parental care do not participate in nest defence or mate competition. On the other hand, aggression, sexual behaviour and parental care are influenced by partly the same brain areas, all containing androgen receptors (Rosvall 2013b). Therefore it is also possible that females only become insensitive to T in their parental behaviour, but not in aggressive and sexual behaviour. This is suggested by data on male great tits where experimentally elevated T levels increased the expression of song while not affecting parental care (Van Duyse et al., 2000).

In this study we examine the effects of sustained experimentally elevated T levels on incubation behaviour and reproductive success in female great tits (*Parus major*). The great tit is a socially monogamous species with bi-parental care, which is essential for the survival of the offspring (Bjorklund and Westman, 1986a). Among females there is competition for males that own a territory (Gosler, 1993), which might explain female-female aggression at the beginning of the breeding season (Slagvold, 1993). Only females build nest and incubate eggs, but both parents provide food to their nestlings. The great tit is one of the few species in which elevated T levels in males do not suppress paternal care measured as food provisioning rate, although it does increase song rate (Van Duyse et al., 2000).

It is currently unclear whether female great tits remain sensitive to elevated T levels later in their breeding season after egg-laying, when natural T levels have declined. By comparing reproductive behaviour and reproductive success between females treated at the start of nest building with long lasting T-implants or with empty implants (controls), we tested to what extent great tit females remain sensitive to T. Since the effectiveness and pattern of incubation might be influenced by T and affect

hatching success, we also measured nest temperature during incubation. The only other similar study that looked at experimentally elevated T levels and incubation behaviour in females was conducted on tree swallows, a species in which females need to defend mates and their nest cavities against intruders also during the phase of parental care (Rosvall 2013a). In great tits there is competition for nest cavities during the winter / pre-breeding period (Gosler, 1993) but later in the season, when birds are incubating, female-female aggression in the neighbourhood of the nest box has rarely been observed (more than 20 hours of personal observations BdJ). Also polygyny rarely occurs in this socially monogamous species (Björklund and Westman, 1986b). This suggests that T levels during the parental care period are not beneficial. During this period natural T levels are indeed low (Rost, 1990). Therefore, we hypothesised that selection against remaining sensitive to T has never been needed and that females will still respond behaviourally to artificially increased levels of this hormone. Under natural conditions female great tit probably never experienced detrimental behaviours induced by elevated T levels later in the breeding season. Therefore, it is unlikely that selection has acted on reducing sensitivity to T during this period. By experimentally elevating T level for a long period we expect to expose these behaviours that might be negatively affected by selection.

2. MATERIALS & METHODS

2.1 Study area and study species

The study was conducted in a population of great tits using nest boxes near the city of Antwerp, Belgium (51° 10'N, 4° 17'E), during the spring of 2009. The study area consists of a park area with deciduous forest containing 58 nest-boxes. Great tits in this population produced an average clutch of nine eggs (SE ± 0.44) per year. Second clutches rarely occur in this population (personal observation BdJ). Full day incubation starts after clutch completion and lasts on average thirteen days, although great tits already incubate their eggs for short periods at night before clutch completion (Gosler, 1993). From early March onwards, nest boxes were checked every other day to determine the onset of nest building. As soon as females started nest building, they were captured at night (in nest boxes) or during the day (in food-baited potter traps). At capture, all individuals were sexed (following Svensson, 1984) and banded with a unique metal ring and three colour rings for individual identification. Age (second calendar year or older) was determined based on the colour of the wing feather (following Cramp and Perrins, 1993).

When females were captured for implantation (see below), nests were checked daily to determine the extent of nest building (see paragraph below), the onset of egg laying and incubation, and the clutch size. The onset of incubation was determined when a female was found incubating her eggs or when the eggs were found uncovered and warm. After females started incubating, nests were checked every second day, while two days before expected hatching date, daily checks were resumed to determine hatching date and brood size. The number of hatchlings and brood size were determined on day 6. Nestlings that had died before day 6 were included in the number of hatchlings. When nestlings were 10 days old they were ringed with a metal ring and their body mass was measured to the nearest 0.1g using a digital balance. When nestlings were 15 days old, their body mass was measured again and their tarsus length was measured to the nearest 0.1mm using a calliper. Near the end of the nestling phase (when nestlings

were ca. 17 days of age), daily nest checks were resumed to determine the fledging date and number of fledged young.

2.2 Implantation procedure

Females were assigned randomly to a control (C) or a testosterone (T) treatment group, while nest building stage taken into account, ensuring an approximately equal number of females with the same nest building stage per treatment. Four nest building stages were distinguished: small parts of nesting material present (stage 1), a solid layer of nesting material present (stage 2), a nest cup present but not yet lined with hairs and feathers (stage 3), and a completed nest lined with hair and feathers (stage 4). When a nest box was empty it was labelled as stage 0. If the nest stage was intermediate of two stages the stage was assigned a half (e.g. a nest box had almost a solid layer but the bottom of the box was still visible it would be called stage 1.5). A total of nine C birds and 12 T birds were implanted, of which a total of seven C and seven T birds started breeding in our experiment (details see below). The average nest stage at the time of implantation (2; 1–2.5) of all the implanted birds was not significantly different between the treatments (mean nest building stage C females: 2; 1-2; mean nest building stage T females: 2; 0.5-2; Mann-Whitney U-test (1) = 48.5, Z = 0.39, P = 0.70), nor was it significantly different for the birds that actually started breeding in the population (nest building stage for birds breeding in the population: 1.75; 1-4; nest building stage for birds not breeding in the population: 2; 0-2; Mann-Whitney U-test (1) =22.5, Z = 0.26, P = 0.80). Just prior to implantation, females were weighed and their tarsus length and third primary feather (P3) were measured to the nearest 0.1 mm with a stop-ruler. There was no difference in body mass, tarsus length or P3 length between control and T groups prior to implantation (all P>0.48). All females of which the age was determined (C = 8; T = 9) were scored as second calendar year birds.

Both T and C females were implanted with a 6-mm long silastic tube (Degania silicone; inner diameter 0.762 mm, outer Diameter 1.651 mm), which was sealed at both ends with silcon glue (Dow corning). The implant was inserted subcutaneously along the left flank under local anaesthesia (Xylocaine, 10% spray). After implantation the small incision was sutured with tissue glue (1 × 0.5 ml Histoacryl, Braun, Germany). In T females, the implants were filled with 0.6 ± 0.015 mg crystalline testosterone (Fluka) over a length of 2 mm. C females received empty implants. The amount of T used in our experiment was determined based on a pilot study in which a higher dosage (0.8mg ± 0.025mg) suppressed normal breeding activities, such as egg laying and incubation behaviour, so the dosage used in our experiment was slightly lower. All females were implanted between 16th of March and the 12th of April 2009, and there was no difference in implantation date between the two treatments (independent t-test t_{19} = –1.00, P = 0.33), nor was there a difference in implantation date for those birds that started breeding in the population (Mann-Whitney U-test (1) = 24.5, Z = 0.0, P = 1.0). Two days after implantation, one C female was found dead. Additionally one C female and five T females did not breed in one of our nest boxes. After implantation, two out of seven C females and four out of seven T females moved to a different nest box.

2.3 Incubation measurements

As soon as a female was observed incubating, the temperature of her nest was measured, from which incubation temperature and incubation behaviour was calculated. We used a data logger (HOBO logger, Mulder-Hardenberg BV., The Netherlands) that registered the temperature inside the nest box via a sensor that was positioned in the middle of the nest on the first day of incubation. To place the sensor, the eggs were temporarily removed and a small hole was drilled in the bottom of the nest box. Through this hole a sensor was mounted in the middle of the nest cup. The logger was stored in a small green plastic box, which was taped to the lid of the nest box on the outside (Figure 1A). After mounting the sensors, the eggs were placed back into the nest box, around the sensor (Figure 1B). The sensor did not extend above the eggs. Drilling and placing the equipment did not take more than 10 min. The temperature was measured every 15 s for an average period of 9967 ± 688 min from the onset of incubation until eggs had hatched, and the measurement time did not significantly differ between treatments (independent t-test $t_{11} = –0.04$, $P = 0.96$). To validate whether temperature fluctuation recorded by the data logger coincided with presence or absence of the incubating female, a video camera was placed approximately 5 m away from five nest boxes to record when a female entered or left the nest box and when a male entered the nest box to feed the female. Before the start of each video recording, the nest box was checked to see if the female was on the nest. A total of 264 min of video observations were made with an average of 53 ± 4.5 min per nest, and presence and absences recorded on video were afterwards visually compared to the temperature data plotted against time per female. Gaps in incubation due to female absences (recorded via video) corresponded to a sharp decline in temperature, and for the presence of females on eggs showed the opposite pattern. Once the temperature data were validated, we developed threshold temperature values to identify presence and absence of females. A sharp decline in temperature of more than 1.3 °C/min for at least 4 min was considered as a gap in incubation. An increase in temperature with an initial slope of at least 0.2 °C/min and a maximum slope of at least 1.0 °C/min was considered the start of a bout of incubation. These criteria were used to automatically identify incubation bouts and gaps in the program Rhythm 1.0 (Cooper and Mills, 2005). Subsequently, we visually inspected these intervals with the program Raven Pro (1.4) and manually corrected obvious errors (see also Cooper and Mills 2005, for the selection procedure using Rhythm and Raven). We found in particular that the start of the recesses needed to be corrected manually (most of the time by a few minutes) because the drop in temperature was relatively slow in the beginning and only gradually became steep, perhaps because in a nest box species temperature changes are relatively slow compared to non-hole breeding species. Therefore Rhythm did not always select the complete recess period. From the identified time intervals of incubation bouts and gaps, we calculated the following parameters; (i) duration of every incubation bout (min); (ii) duration of every incubation gap (min); (iii) minimum nest temperature during a gap (°C); (iv) mean nest temperature during the day (combining incubation bouts and gaps) (°C); and (v) mean nest temperature during night (°C). The start of the night was defined from when the females started incubating in the evening for a period longer than 2 hours until there was a sharp decline in temperature, when the female exited her nest box in the morning. For the analyses of the incubation data we used data of 13 females (C=6; T=7). In the analyses we included the average time (min) a female spent off and on her nest during the day, the number of gaps and the average minimum temperature during incubation gaps, the overall mean incubation temperature during the day, the mean variation in temperature during the day, and mean night temperature inside the nest.

2.4 Natural hormone concentrations

To examine the natural profile of circulating T levels, female blood was collected just prior to implantation (n = 17) and from five C females that were recaptured during the nestling period (between the 2nd and 15th of May 2009). None of the recaptured females lost their implant. During these captures, 50 – 150 µl blood was taken within 30 min after capture by puncturing the brachial vein with a sterile needle (Terumo, 27 g × 3/4; 0.4 × 20 mm) and transferred into an Eppendorf tube using heparinized microhematocrit capillaries. The blood was stored on ice and centrifuged for 10 min at 7000 rpm within six hours after sampling. The plasma fraction was removed and stored at –20°C until analysis. Testosterone was quantified in plasma extracts by radioimmunoassay (RIA) using a commercial double antibody system purchased from MP Biomedicals (Solon, Ohio). For extraction, 500 µl of a 50/50 mixture of cyclohexane/ethylacetate was added to 50 µl plasma and the tubes were incubated for 10 min with continuous shaking. After centrifugation, the tubes were placed in a mixture of dry ice and ethanol for snap freezing, followed by transfer of the organic phase to a new tube. After thawing, samples were re-extracted following the same method. The combined supernatants were dried by vacuum centrifugation and stored at –20°C until further analysis. For testosterone measurements, the dried samples were dissolved in 25 µl steroid diluent buffer and further treated following the protocol of the RIA kit. The primary antibody used in this assay does not significantly cross-react with other androgens beside T (5α-dihydrotestosterone: 3.4%; 5α-androstane-3β, 17β-diol: 2.2%; 11-oxo-testosterone: 2%; all other steroids: <1%). Testosterone standards ranged from 0.10 ng/ml to 11.75 ng/ml, but the effective detection limit could be extended to 0.05 ng/ml owing to the concentration effect of the extraction procedure. All samples were measured in a single assay and the intra-assay coefficient of variation was 4.6 for medium/low and 9.1% for high concentrations.

2.5 Test of hormone implants

Because we did not recapture enough T females (see table 1) to examine the effects of the implants on T plasma concentrations, an additional laboratory experiment was conducted. In this experiment eight female great tits were implanted with T on the 10th of December 2012. The females were implanted in December, because during this non-breeding period female birds generally have very low T levels (reviewed by Ketterson et al., 2005). Thus, an increase in T levels would most likely be caused by the implants. The great tits were wild-caught and hand-reared and all the same age (2 years). Before and during the first 14 days of the experiment these females were housed in single-sex groups of eight individuals in free-flight, half-open aviaries (2.0 × 4.0 × 2.5 meters). After 14 days, four of the eight females were housed together with a male in separate aviaries. Birds were provided with *ad libitum* food and fresh water at all times. To measure baseline T plasma concentrations, a blood sample was taken prior to implantation within 10 min after capture. Next the females were implanted with silastic tubes filled with T. The implantation procedure and the implants used were equal to the field experiment (see for size above; average T weight 0.6 ± 0.01 mg). The birds returned to their aviary within 90 min after capture. Seven and 28 days after implantation another blood sample was taken to measure the effects of the implants on T plasma concentrations. All the blood samples were taken between 12:00 –14:00 GMT+1. Immediately after sampling the blood was centrifuged for 10 min at 6000 rpm, the plasma was removed, and stored at –20°C. After the last blood sample was taken the implants were removed under

local anaesthesia (Xylocaine, 10% spray) by making a small incision below the implant, and the incision was sutured with tissue glue (1 × 0.5 ml Histoacryl, Braun, Germany). During each of the captures the health of the females was checked by measuring their weight to the nearest 0.1 g. There was no significant weight change over the course of the experiment (One-way repeated measures ANOVA: $F_{1,14}$ = 2.26, P = 0.15).

The plasma samples were analysed in one assay using a commercial kit (Orion Diagnostica, Spectria Testosterone RIA kit, Espoo, Finland) with a sensitivity of 0.04 ng/ml testosterone and cross-reactivities of 4.5 % with DHT and 0.01 % with A4 as described in de Jong et al. (2013). In brief; plasma samples were defrosted, their volume was measured and 50 µl radio-actively labelled testosterone (Perkin Elmer Life and Analytical Science BV) was added to all samples to measure the accuracy of the extraction process (recovery). After an incubation time of 1 hour, 2.5 ml diethyl ether/petroleum benzine (70:30) was added and samples were vortexed and centrifuged. Samples were snap frozen by a mixture of ethanol and dry ice and decanted. The supernatant was dried under streaming nitrogen, the remaining pellet was again dissolved in 1 ml 70% methanol and samples were stored over night at –20ºC. The following day, samples were centrifuged, the methanol phase was decanted and the samples dried again under streaming nitrogen. The pellet was re-suspended in 200 µl PBS buffer. 30 µl of this mixture was used for measuring recoveries (average recovery rate for testosterone: 92.96 ± 0.89%). Hormone concentrations were measured using radio immuno assays (RIAs). Based on the standard curve values below the detection limit was calculated as being 0.10 ng/ml. The dilution curve ran parallel to the standard curve. The intra assay variation was 6.9 %.

2.6 Statistical analyses

All the data were checked for normality. Data that were not normally distributed (number of days spent incubating, incubation bouts and gaps, and fledging date) were transformed to approximate normality when possible (see below) or tested using non-parametric tests. Independent t-tests were used to analyse the effect of treatment on female characteristics (body mass (g), tarsus (mm) and wing length (mm), the onset of egg laying and incubation date, clutch size, and the number of hatchlings and fledglings. The effect of treatment on hatching date was analysed with a general linear model (GLM). Treatment was included as a fixed factor and the onset of egg laying was included as a covariate.

To test whether T implants had an effect on T plasma levels of captive females 7 and 28 days after implantation, a linear mixed model (MIXED) was used. The model included T plasma levels as a dependent variable, sample period (where sample period two = 7 days after implantation; sample period three = 28 days after implantation) as a fixed factor and female ID as a random factor. To correct for possible disturbance effects from capturing on T levels, time between entering the aviary and blood sampling was included as covariate. For the four birds that were housed separately during the last blood sampling, the time between capture and blood sampling was taken. The housing condition of the females when the third sample was taken had no effect on the hormone levels (independent t-test t_5 = 0.85, P =0.43) and was therefore not included in the analyses.

The number of incubation gaps, the mean duration of incubation bouts and gaps, mean minimum temperature, mean day temperature and mean variance in day temperature, and mean night

temperature were analysed with a GLM. Treatment was included as a fixed factor and clutch size was included as a covariate in all models, as females with larger clutches spent more time inside their nest boxes ($F_{1, 12} = 8.37$, $P = 0.02$). The mean incubation gap duration and time spent incubating were not normally distributed and were transformed with a $\log_{10}$ transformation. Linear regression models were used to quantify the relationship between the average incubation temperatures during the day or during the night and the proportion of hatching success (the number of hatchlings divided by clutch size) per nest, for each treatment separately (variance in hatching success was not equally distributed between treatments).

To test the effect of treatment on hatching and fledging success, and nestling survival until day 6, a generalized linear mixed model (GLIMMIX) was used, with a binary distribution and a logit function. An egg was scored as 1 when it hatched and as 0 when it did not hatch. A nestling was scored as 1 when it had survived until day 6 or fledged and 0 when it had not survived until day 6 or fledged. In the models treatment was included as a fixed factor and female ID as a random factor. Clutch size was included as a covariate to correct for initial differences in number of eggs. To test if the treatment of the mother had an effect on nestling body mass (g) at day 10 and day 15, growth rate (measured as the change in body mass between day 10 and 15 in g) and tarsus length (mm) at day 15, a linear mixed model was used, including treatment as fixed factor, nest box as a random factor (to correct for non-independence of nestlings of the same mother), date of measurement (for body mass), or hatching date (for hatchling growth) and brood size as co-variables. Tests were two tailed and differences were considered to be significant with a P-value <0.05. SAS (SAS® 9.2) was used to analyse the nestling characteristics and incubation data. All the other data were analysed using STATISTICA 7.0 (StatSoft, Inc.). Unless stated otherwise, average values are presented ± SEM while median values are presented ± range.

3. RESULTS

3.1 Hormone concentrations

The mean natural T plasma concentration of female great tits during the nest building phase was 1.02 ± 0.30 ng/ml with a range of 0.10–2.67 ng/ml. During the nestling phase, the mean natural T concentration of C females was 0.54 ± 0.07 ng/ml with a range of 0.31–0.67 ng/ml (see Fig. 2A). The average baseline T level of the captive female great tits was 1.92 ± 0.49 ng/ml (see Fig. 2B). T levels differed significantly between the three sampling periods (MIXED; $F_{2,11} = 9.42$, $P = 0.0041$, correcting for time: $F_{1,11} = 9.96$, $P = 0.009$). T levels during the first sampling period were significantly lower than during the second and third sampling period (1 vs 2: Tukey test; $P = 0.004$; 1 vs 3: Tukey test; $P = 0.012$), while the latter two periods did not differ significantly from each other (Tukey test; $P = 0.99$). The mean plasma T value after 7 and 28 days were 3.29 ± 0.39 ng/ml and 2.81 ± 0.22 ng/ml, respectively. As the T-implant induced values were around the same values as some of the baseline values before implantation, the implant did not induce supra-physiological values. Moreover, T concentrations during implantation did not return to the average baseline level, indicating that the implants worked over a long period. The higher levels in winter compared the average natural T levels during the nest building phase may be due to the fact that in this species females compete during winter and form dominance hierarchies in winter flocks (Gosler, 1993).

This was clearly expressed in our aviaries where there was obvious competition (personal observation KvO).

3.2 Breeding parameters

There was no effect of the T treatment on time it took females to complete the nest compared to those from the control group (Table 2). T females laid their eggs on average four days later than C females; however this difference was not significant. Clutch sizes of T and C females did not differ (Table 2).

3.3 Incubation behaviour

Treatment had a significant effect on incubation temperature; both the mean minimum temperature inside the nest box during a gap in incubation (Fig. 3A) and the mean day and night temperatures (Fig. 3B, C) were significantly lower in the T than in the control group (Table 3). Mean variance in day temperature was not affected by treatment, but clutch size did show a significant negative relationship with mean variance in day temperature. Females with large clutches showed less variation in day temperature, perhaps because eggs buffer each other against temperature changes. Mean day and night temperatures were positively correlated with hatching success in T females whereas mean day temperature was negatively correlated with hatching success in C females (Table 4). All other parameters measured were not significantly different between the treatments (Table 3).

3.4 Reproductive success

The incubation time in days (C = 12; 12-13; T = 13; 13-14) was not significantly different between the two treatments (Mann-Whitney U-test (1) = 7, Z = -1.71, P = 0.9). The average hatching date of nestlings (66 ± 1.34) from T females was five days later than that of offspring produced by control females (61 ± 1.41), but this effect was driven by a delay in egg laying as hatching date was not significantly different when correcting for the onset of egg laying (treatment: $F_{1,9} = 0.66$, $P = 0.45$; onset of egg laying: $F_{1,9} = 12.45$, $P = 0.006$). Hatching success was significantly lower in the T treated group (GLIMMIX; hatching success: $F_{1,109} = 4.01$, P = 0.048; clutch size; $F_{1,109} = 1.91$, P = 0.17). A total of 35 out of 52 (64%) eggs from T females hatched compared to 68 out of 70 (97%) eggs from C females. After hatching, a higher proportion of nestlings of T females did not survive until day 6 (10 out of 35) compared to those of C mothers (1 out of 68; GLIMMIX; nestling survival until day 6: $F_{1,109} = 8.14$, 0.0005; clutch size: $F_{1,109} = 0.08$, P = 0.78). Brood size at day 6 after hatching was also significantly smaller in the treatment group compared to the control group (Table 5.1). After day 6 nestling survival was almost equal in both groups, only one T nestling died after day 15. But because the number of hatchlings and the number of nestlings surviving to day 6 were lower for T females, the overall fledging success was significantly lower in the T treatment group (GLIMMIX; fledging success: $F_{1,109} = 9.08$, P = 0.003; clutch size: $F_{1,109} = 0.08$, P = 0.78). A total of 24 fledglings from six T nests fledged, compared to 67 fledglings from seven control nests, which was statistically significant (Table 2). The average fledging date of the nestlings did not differ between treatments. In summary, prolonged increased levels of testosterone significantly reduced reproductive success.

3.4 Nestling characteristics

On day 10, the average body mass of nestlings produced by T females (12.76 ± 0.45 g) was significantly lower than that of nestlings produced by C females (13.39 ± 0.22 g) correcting for brood size (MIXED; treatment: $F_{1,89}$ = 9.54, P = 0.003; brood size: F1,89 = 9.62, P =0.003). Nestlings of larger broods were on average lighter. The date at which 10 day old nestlings were measured did not have a significant effect on their weight (MIXED; date: $F_{1,88}$ = 2.57, P = 0.11). There was no difference in weight between 15 day old nestlings of T females (16.62 ± 0.27 g) and nestlings of C females (16.59 ± 0.18 g, MIXED; treatment: F1,88 = 0.45, P = 0.51; date: $F_{1,88}$ = 3.43, P = 0.07). The covariate brood size was negatively associated with the weight of 15 day old nestlings (MIXED; brood size: $F_{1,88}$ =5.57, P =0.02). Between day 10 and day 15, T nestlings (3.85 ± 0.28 g) grew significantly faster than C nestlings (3.20 ± 0.16 g, MIXED; treatment: $F_{1,88}$ = 25.11, P<0.0001; hatching date: $F_{1,88}$ = 26.44, P<0.0001; brood size: $F_{1,88}$ = 7.82 P = 0.006). Nestlings that hatched later in the breeding season and/or grew up in larger broods showed a reduced growth. Average nestling tarsus length did not differ between treatments (T nestlings: 19.61 ± 0.17 mm; C nestlings: 19.48 ± 0.11 mm, MIXED; treatment: $F_{1,88}$ = 0.14, P = 0.71) and was not affected by brood size or date of measurement (MIXED; brood size: $F_{1,88}$ = 1.21, P = 0.279; date: $F_{1,88}$ = 2.89, P = 0.09). In summary, females treated with prolonged elevated T levels had lighter chicks in the beginning of the rearing period, but these chicks caught up in body mass thereafter.

4. DISCUSSION

In some avian species, elevation of male T after the formation of a pair and a territory can still increase male reproductive success by mate guarding, territorial defence and obtaining extra pair offspring. However, the hormone can also suppress parental care. To circumvent the detrimental effects of T on male parental behaviour, certain species appear to have evolved insensitivity to T during the period of intensive parental care (the essential paternal care hypothesis: Lynn et al 2002, Lynn 2008). These species are generally species where male care is essential for offspring survival (for example the great tit, Bjorklund and Westland, 1986a). Alternatively, behaviour may remain sensitive to elevated T levels, when T-mediated behaviour is important for reproductive success even after egg laying, although this might be detrimental for other reproductive behaviours (the essential mating effort hypothesis, Rosvall 2013a,b).Although behavioural insensitivity to T after egg production has been mainly tested in males, it may be even more important for females that provide often most of parental care and suppressive effects of T on female care is well documented (see introduction). We studied this in female great tits, finding that, despite limited sample size and some multiple testing, several aspects of female parental care are still sensitive to T, resulting in poorer care with negative effects on fitness.

Our findings correspond with those of a recent study on female tree swallows (Rosvall 2013a), T treatment in these birds resulted in more aggression, poorer incubation and lower hatching success of the chicks. This was explained by the fact that female tree swallows need T to defend their nest after egg production and during chick rearing, even though this decreases offspring production. . However, in great tits, females do not need to defend territories after the start of incubation. Therefore there would be no need for T elevation as this would have detrimental effects on parental care and no clear benefits. Indeed, endogenous T levels are low after the start of egg laying in this species (Rost, 1990). Therefore, the fact that females remain sensitive to elevated T levels later in the breeding season, are in line with our hypothesis that there has been no need for strong selection in the past in favour of becoming insensitive to T in great tits.

4.1. Nest building behaviour

Many female bird species show a peak in T levels at the beginning of their breeding season, when females are building their nests. Therefore we were surprised to find that elevated T did not affect nest building behaviour. Only a few other studies have investigated the role of T in nest building behaviour and the results are inconsistent. In the European starling, T did not affect nest building (De Ridder et al., 2002), whereas in the blue tit, T females accelerated nest building (de Jong, 2013). The absence of a treatment effect in our study might be due to the fact that we only had a somewhat rough classification of nest completeness, and that we did not observe actual nest building behaviour.

4.2. Incubation behaviour

We found particularly strong negative effects of elevated T late in the breeding cycle, with T females producing significantly lower incubation temperatures than C females. So far, two other studies have investigated the effect of elevated T on incubation behaviour in female birds. Experimental elevation of T levels did not affect the total time females spent incubating in dark-eyed juncos (Clotfelter et al., 2004),

but did decrease incubation temperature in the tree swallow (Rosvall, 2013a). A likely explanation for the reduction in incubation temperature may be that T females had a less developed brood patch (Clotfelter et al., 2004), reducing the maximum temperature a female can reach during incubation (yellow-eyed penguin *(Megadyptes antipodes*) Massaro et al., 2006). Further research is required to confirm this, as this effect of T on the brood patch could be one of the functional explanations for why T levels dropped quickly when females start incubating.

4.2.1 Incubation behaviour and hatching success

The lower incubation temperature in the nest of T females was inversely correlated with the hatching probability of their eggs. Egg temperature during incubation is important for the development of the embryo (Webb, 1987) and low incubation temperature can cause mortality of the embryos before hatching (Deeming and Ferguson, 1991), decrease hatching success (blue tit, Nord and Nilsson, 2011), increase nestling mortality (domestic white leghorn chicken *(Gallus gallus)*, Evans, 1990), or decrease nestling weight (domestic white leghorn chicken, Suarez et al., 1996; tree swallow, Ardia et al., 2010). Therefore the lower hatching success of eggs of T females compared to C females might be causally explained by the low incubation temperature in the former. As a consequence, females treated with T produced fewer hatchlings and fledglings and therefore had a lower reproductive success. A reduction in hatchling and/or fledgling number due to elevated T in their mothers has been shown in other bird species too (O'Neal et al., 2008; Lopez-Rull and Gil, 2009; Veiga and Polo, 2008). Surprisingly, C females that showed higher mean incubation temperatures during the day, also showed lower hatching success. In fowl (*Gallus gallus domesticus*), very low and very high incubation temperature have both been shown to negatively affect embryonic development, indicating that there is an optimum incubation temperature for embryos to develop normally (Romanoff et al., 1938). Since in our study the nests with the lowest and the highest incubation temperature had lower hatching success, this might have been caused by a deviation from such an optimum incubation temperature in these nests. Alternatively, T implantation resulted not only in higher circulating concentration of the hormone in the female but also in her eggs, with the latter negatively affecting hatching success. Although there is evidence that female birds can independently regulate T concentrations in their circulation from those in her eggs (Japanese quail *(Coturnix japonica)*, Okuliarova et al 2011), T implantations can indeed lead to higher T concentrations in the egg (Groothuis and Schwabl 2008). However, from the many *in ovo* T injection studies there is no evidence at all that elevated yolk T negatively affects hatching success (Von Engelhardt and Groothuis 2011).

4.3. Nestling phase

At day 10, T nestlings had lower body masses compared to C nestlings. There are three possible (mutually not exclusive) explanations why nestling weight was affected by the T treatment. First, eggs laid by T females may have been lighter and therefore produced lighter nestlings. Great tit nestlings that hatch from lighter eggs grow more slowly during the early period of the hatching phase, but do catch up before fledging (Schifferli, 1973). However, effects of T treatment of females on egg mass are ambiguous; studies have either found no effect of T on egg mass (Clotfelter et al., 2004; Lopez-Rull and

Gil, 2009), or an increase in egg mass (Rutkowska et al., 2005). A pilot study on the effects of T on egg mass in great tits, however, did show a decrease in egg mass (R. Pinxten, unpubl. data).

Second, a lower brooding temperature of the mother might have affected the weight of the nestlings. A previous study in dark-eyed juncos showed that females treated with T indeed showed reduced brooding of young (O'Neal et al., 2008). Brooding behaviour of the mother is very important for nestlings because thermoregulation of altricial young is not yet fully developed in the early nestling stage (Dunn, 1975). Thus in our study female brooding may have resulted in chicks having to spent more energy on thermoregulation and less on energy for growth and development, resulting in lower weight. This is consistent with the fact that chick survival was lower of T nestlings during the first 6 days compared to the period of the nestling phase after day 6. After 6 days, nestlings can regulate their own temperature and therefore are less dependent on the brooding of the mother.

Third, nestling body mass may have been affected by differential food provision rates of T mothers. In the spotless starling, T treatment of females reduced female feeding rate (Veiga and Polo, 2008), whereas in dark-eyed juncos there was no difference in feeding rate between T and C females (O'Neal et al., 2008). Feeding rates were not quantified in this study, but if feeding behaviour of T females was affected negatively, then this most likely only occurred during the early nestling period since nestling weight did not differ anymore on day 15. The fact that the chicks from T treated mothers caught up in weight after day 10 may be caused by increased maternal provision. This has been found in female dark-eyed junco's in which those with natural high T levels brooded less but provisioned nestling more frequently (Cain and Ketterson, 2013). Alternatively, elevated T concentrations in the egg might have induced higher growth rates, as found in several avian studies (reviewed by Von Engelhardt and Groothuis 2011).

5. Conclusions

Overall, our results indicate that great tit females are behaviourally sensitive to elevated T levels during the period of maternal care, at the cost of incubation behaviour and hatching success of their offspring. Since in this period natural T levels are normally low we suggest that Darwinian selection for T insensitivity in great tits has not been strong because it was never needed. One potential restriction in our study was that we only measured parental care and not aggressive behaviour, in contrast to an earlier study on female tree swallows (Rosvall 2013a), which also found behavioural sensitivity to T after egg laying. However, females of our study species have less need to show aggression after egg laying. We hypothesise that in species in which females do need T induced aggressive behaviour during the phase of maternal care, the suppressive cost of T on this care are avoided either by much more moderate levels of T than induced by implantation studies, or by becoming T insensitive in maternal care but not in aggressive behaviour. That such behaviour-specific regulation can occur has been demonstrated in males in several species. This includes males of our study species, in which T implantations induce higher song rates but not at the cost of reduced parental care (Van Duyse et al. 2002).

Funding

BDJ was supported by a doctoral grant from the Flemish Agency for Innovation by Science and Technology.

Acknowledgments

We thank Peter Scheys and Ann Geens for field assistance, and Bonnie de Vries and Lut Noterdaeme for lab assistance. We thank Lewis Spuring for comments on the manuscript. The study was conducted in full compliance with Belgian and Dutch laws and regulations.

REFERENCES

Adkins-Regan, E., 2005. Hormones and animal social behavior. Princeton University Press, Princeton

Alonso-Alvarez, C., 2001. Effects of testosterone implants on pair behaviour during incubation in the Yellow-legged Gull Larus cachinnans. Journal of Avian Biology 32, 326–332.

Ardia, D.R., Perez, J.H., Clotfelter, E.D., 2010. Experimental cooling during incubation leads to reduced innate immunity and body condition in nestling tree swallows. Philosophical Transactions of the Royal Society B-Biological Sciences 277, 1881–1888.

Bjorklund, M., Westman, B., 1986a. Mate-guarding in the great git - tactics of a territorial forest-living species. Ornis Scandinavica 17, 99–105.

Björklund, M., Westman, B., 1986b. Adaptive advances of monogamy in the great tit (*Parus major*): an experimental test of the polygyny threshold model. Animal Behaviour 34, 1436-1440.

Cain, K.E., Ketterson, E.D., 2012. Competitive females are successful females; phenotype, mechanism, and selection in a common songbird. Behavioral Ecology and Sociobiology 66, 241–252.

Cain, K. E., Ketterson, E.D., 2013. Individual variation in testosterone and parental care in a female songbird; the dark-eyed junco (*Junco Hyemalis*)." Hormones and Behavior 64, 685–92.

Clotfelter, E.D., O'Neal, D.M., Gaudioso, J.M., Casto, J.M., Parker-Renga, I.M., Snajdr, E.A., Duffy, D.L., Nolan, V., Ketterson, E.D., 2004. Consequences of elevating plasma testosterone in females of a socially monogamous songbird: evidence of constraints on male evolution? Hormones and Behavior 46, 171–178.

Cramp, S., Perrins, C.M., 1993. *Parus major* Great tit, in: Handbook of the birds of europe, the Middle East and North Africa. The birds of the western palearctic. Oxford University Press, Oxford, pp. 255–281.

De Jong, B., 2013, Testosterone a female hormone – Testing the function and evolution of testosterone in female birds. University of Groningen.

De Ridder, E., Pinxten, R., Eens, M., 2000. Experimental evidence of a testosterone-induced shift from paternal to mating behaviour in a facultatively polygynous songbird. Behavioral Ecology and Sociobiology 49, 24–30.

De Ridder, E., Pinxten, R., Mees, V., Eens, M., 2002. Short- and long-term effects of male-like concentrations of testosterone on female European starlings (*Sturnus vulgaris*). Auk 119, 487–497.

Deeming, D.C., Ferguson, M.W.J., 1991. Egg Incubation: Its effects on embryonic development in birds and reptiles. Cambridge University Press, Cambridge.

Dunn, E., 1975. Timing of endothermy in development of altrical birds. Condor 77, 288–293.

578 Edler, R., Goymann, W., Schwabl, I., Friedl, T.W.P., 2011. Experimentally elevated testosterone levels
579 enhance courtship behaviour and territoriality but depress acquired immune response in Red Bishops
580 *Euplectes orix*. Ibis 153, 46–58.

581 Enstrom, D.A., Ketterson, E.D., Nolan, Jr, V., 1997. Testosterone and mate choice in the dark-eyed junco.
582 Animal Behaviour 54, 1135–1146.

583 Evans, R.M., 1990. Effects of low incubation temperatures during the pipped egg stage on hatchability
584 and hatching times in domestic chickens and ring-billed gulls. Canadian Journal of Zoology 68, 836–840.

585 Gerlach, N.M., Ketterson, E.D., 2013. Experimental elevation of testosterone lowers fitness in female
586 dark-eyed juncos. Hormones and Behavior 63, 782–90.

587 Groothuis, T.G.G., Schwabl, H., 2008. Hormone-mediated maternal effects in birds: Mechanisms matter
588 but what do we know of them? Philosophical Transactions of the Royal Society B-Biological Sciences 363,
589 1647–61.

590 Gosler, A., 1993. The great tit. Hamlyn, London.

591 Hegner, R.E., Wingfield, J.C., 1987. Effects of experimental manipulation of testosterone levels on
592 parental investment and breeding success in male house sparrows. Auk 104, 462–469.

593 Hunt, K.E., Hahn, T.P., Wingfield, J.C., 1997. Testosterone implants increase song but not aggression in
594 male lapland longspurs. Animal Behaviour 54, 1177–92.

595 Kazama, K., Sakamoto, K.Q., Niizuma, Y., Watanuki, Y., 2011. Testosterone and breeding behavior in male
596 black-tailed gulls: An implant experiment. Ornithological Science 10, 13–19.

597 Ketterson, E.D., Nolan, V., Sandell, M., 2005. Testosterone in females: Mediator of adaptive traits,
598 constraint on sexual dimorphism, or both? The American Naturalist 166, S85–S98.

599 Langmore, N.E., Cockrem, J.F., Candy, E.J., 2002. Competition for male reproductive investment elevates
600 testosterone levels in female dunnocks, *Prunella modularis*. Proceedings of the Royal Society of London
601 Series B-Biological Sciences 269, 2473–2478.

602 Lopez-Rull, I., Gil, D., 2009. Elevated testosterone levels affect female breeding success and yolk
603 androgen deposition in a passerine bird. Behavioural Processes 82, 312–318.

604 Lynn, S.E., 2008. Behavioral insensitivity to testosterone: Why and how does testosterone alter paternal
605 and aggressive behavior in some avian species but not others? General and Comparative Endocrinology
606 157, 233–240.

607 Lynn, S.E., Hayward, L.S., Benowitz-Fredericks, Z.M., Wingfield, J.C., 2002. Behavioural insensitivity to
608 supplementary testosterone during the parental phase in the chestnut-collared long-spur, - *Calcarius*
609 *ornatus*. Animal Behaviour 63, 795–803.

610 Nord, A., Nilsson J., 2011. Incubation temperature affects growth and energy metabolism in blue tit
611 nestlings. American Naturalist 178, 639–51.

612 Massaro, M., Davis, L.S., Davidson, R.S., 2006. Plasticity of brood patch development and its influence on
613 incubation periods in the yellow-eyed penguin *Megadyptes antipodes*: an experimental approach.
614 Journal of Avian Biology 37, 497–506.

615 Okuliarova, M., Groothuis, T.G.G., Skrobánek, P., Zeman, M., 2011. Experimental evidence for genetic
616 heritability of maternal hormone transfer to offspring. The American Naturalist 177, 824–34.

617 O'Neal, D.M., Reichard, D.G., Pavilis, K., Ketterson, E.D., 2008. Experimentally-elevated testosterone,
618 dark-eyed Junco (*Junco hyemalis*). Hormones and Behavior 54, 571–578.

619 Oring, L.W., Fivizzani, A.J., Elhalawani, M.E., 1989. Testosterone-induced inhibition of incubation in the
620 spotted sandpiper (*Actitis-Mecularia*). Hormones and Behavior 23, 412–423.

621 Raouf, S.A., Parker, P.G., Ketterson, E.D., Nolan, V., Ziegenfus, C., 1997. Testosterone affects
622 reproductive success by influencing extra-pair fertilizations in male dark-eyed juncos (Aves: *Junco*
623 *hyemalis*). Proceedings of the Royal Society of London Series B-Biological Sciences 264, 1599–1603.

624 Romanoff, A.L., Smith, L.L., Sullivan, R.A., 1938. Biochemistry and biophysics of the developing hen's egg.
625 III. Influence of temperature. Cornell University Agricultural Experimental Station Memoranda 216, 1–42.

626 Rost, R., 1990. Hormones and behavior - a Joint examination of studies on seasonal-variation in song
627 production and plasma-levels of testosterone in the great tit *Parus major*. Journal Fur Ornithologie 131,
628 403–411.

629 Rosvall, K.A., 2013a. Life history trade-offs and behavioral sensitivity to testosterone: An experimental
630 test when female aggression and maternal care co-occur. PLoS ONE 8, e54120.

631 Rosvall, K.A., 2013b. Proximate perspectives on the evolution of female aggression: good for the gander,
632 good for the goose? Proc. Trans R. Soc B 368: 20130083

633 Rutkowska, J., Cichon, M., Puerta, M., Gil, D., 2005. Negative effects of elevated testosterone on female
634 fecundity in zebra finches. Hormones and Behavior 47, 585–591.

635 Sandell, M.I., 2007. Exogenous testosterone increases female aggression in the European starling
636 (*Sturnus vulgaris*). Behavioral Ecology and Sociobiology 62, 255–262.

637 Schifferli, L., 1973. The effect of egg weight on the subsequent growth of nestling great tits *Parus major*.
638 Ibis 115, 549–558.

639 Searcy, W.A., 1988. Do female red-winged blackbirds limit their own breeding densities? Ecology 69, 85–
640 95.

641 Slagvold, T., 1993. Female-female aggression and monogamy in great tits *Parus Major*. Ornis
642 Scandinavica 24, 155-158.

643 Suarez, M.E., Wilson, H.R., McPherson, B.N., Mather, F.B., Wilcox, C.J., 1996. Low temperature effects on
644 embryonic development and hatch time. Poultry Science 75, 924–932.

645 Svensson, L., 1984. Identification Guide to European Passerines. Stockholm.

646 Van Duyse, E., Pinxten, R., Eens, M., 2000. Does testosterone affect the trade-off between investment in
647 sexual/territorial behaviour and parental care in male great tits? Behaviour 137, 1503–1515.

648 Van Duyse, E., Pinxten, R., Eens, M., 2002. Effects of testosterone on song, aggression, and nestling
649 feeding behavior in male great tits, *Parus major*. Hormones and Behavior 41, 178–186.

650 Veiga, J.P., Polo, V., 2008. Fitness consequences of increased testosterone levels in female spotless
651 starlings. American Naturalist 172, 42–53.

652 von Engelhardt, N., Groothuis, T.G.G., 2011. Maternal hormones in avian Eggs. In hormones and
653 reproduction of vertebrates. Vol. volume 1. Academic Press.

654 Webb, D.R., 1987. Thermal tolerance of avian embryos: A review. The Condor 89, 874.

655 Wingfield, J.C., 1990. Hormonal-control of territorial behavior in Birds - field and laboratory
656 investigations. Progress in Comparative Endocrinology 342, 697–703.

FIGURE LEGENDS

Figure 1. A) Nest box with box attached to it. Inside the box was the data logger that recorded incubation temperature. **B**) Great tit clutch with a sensor placed in between the eggs (see arrow).

Figure 2. A) Natural testosterone levels during the breeding season in female great tits. Open circles left of the dotted line are individual pre-breeding T levels. Open circles to the right of the dotted line are T levels during the nestling phase. The square is the average T level pre-breeding and the closed circle is the average T level during the nestling phase (mean ± SEM). **B)** The effect of T implants on plasma testosterone of captive female great tits. Blood samples were taken prior to implantation (baseline sample) and seven and 28 days after implantation. Each symbol represents the T-levels of one individual female.

Figure 3. (A) The mean minimum incubation temperature during the recess of incubation. (B) Mean incubation temperature during the day (including progresses and recesses of incubation). (C) Mean incubation temperature during the night. Light grey bars are the control females, dark grey bars are the testosterone females. Letters indicate a significant difference with $P < 0.05$. See table 2 for specific P-values. Means ± SEM are presented.

673 **Table 1**. Sample sizes used for the statistical analyses.

	Testosterone	Control	Total
Females implanted	12	9	21
Breeding in the population after implantation	7	7	14
Baseline T levels in the field	7	3	10[1]
T levels after implantation	4	5[1]	9
All breeding parameters until hatching	7	7	14[2]
Incubation measurements	6	7	13[3]
All parameters after hatching	7	6	13[3]
Captive birds	**Baseline**	**Day 7**	**Day 28**
T levels[4]	8	7	7

674 **[1]** Of the total of 17 females of which blood was sampled prior to implantation to measured natural T plasma
675 concentrations seven individuals from the early breeding period were excluded because an insufficient amount of
676 blood was collected for the hormone analyses. Also, two of the seven control females from the nestling period
677 were excluded for that reason.
678 [2] For the analyses of the different breeding parameters seven C and seven T nests were included until the onset of
679 egg laying.
680 [3]During the incubation phase, one nest box of a T female was lost due to vandalism. Therefore seven C and six T
681 birds were included in the analyses of the nestling phase.
682 [4] One female had lost her implant after implantation, therefore only seven of the 8 females were included for the
683 analyses of the second and third sample.

684

685 **Table 2**. Summary of the overall treatment effects on different breeding parameters, nestling
686 characteristics and reproductive output of female great tits. Data are presented as mean ± SEM or as
687 median (quartile range).

	Testosterone		**Control**		**Test**	
	Mean/ Median	**SEM/ Quartile range**	**Mean/ Median**	**SEM/ Quartile range**	**T/U**	**P**
Nest building time (days)	9.83	2.24	13.86	2.31	1.24[2]	0.24
Onset of egg laying[1]	43	1.20	39.14	1.61	1.92[2]	0.08
Clutch size	8.71	0.52	10.00	0.82	1.33[2]	0.21
Number of hatchlings	**5.83**	**1.38**	**9.71**	**0.87**	**2.46[2]**	**0.03**
Brood size at day 6	**4.17**	**1.42**	**9.57**	**0.75**	**3.44[2]**	**0.005**
Number of fledglings per nest	**4.00**	**1.37**	**9.57**	**0.75**	**3.72[2]**	**0.003**
Fledging date[1]	84	82-87	83	74-84	-1.23[3]	0.22

688 [1]The onset of egg-laying, hatching date and fledging date were scored in March days, where 1st of March
689 is 1.
690 [2]Independent t-test.
691 [3]Mann-Whitney U-test.

692

693

694 **Table 3.** Summary of the overall treatment effects on incubation behaviour of female great tits. Data are
695 presented as mean ± SEM or as median (quartile range).

	Testosterone		Control		Test
	Mean/ Median	**SEM/ Quartile range**	**Mean/ Median**	**SEM/ Quartile range**	**GLM**
On-bout time	20.00	18.31-26.45	21.13	20.62-23.43	Treatment: F1,12, = 0.16, P = 0.70; Clutch size: F1,12 = 3.92, P = 0.08
Recess time	6.44	5.23-10.30	6.39	4.43-19.17	Treatment: F1, 12, = 0.12, P = 0.74; Clutch size: P > 0.45
Nr of recesses	185.14	33.52	181.33	21.46	Treatment: F1, 12, = 0.93, P = 0.36; Clutch size: P > 0.80
Minimum temperature during a recess	**23.37**	**0.69**	**25.88**	**0.33**	**Treatment.: F1, 12, = 9.67, P = 0.01; Clutch size: P > 0.20**
Day time temperature	**28.62**	**0.79**	**31.23**	**0.44**	**Treatment.: F1, 12, = 7.59, P = 0.02; Clutch size: P > 0.70**
Night time temperature	**31.13**	**0.94**	**33.92**	**0.58**	**Treatment.: F1, 12, = 5.93, P = 0.03; Clutch size: P > 0.90**

696
697
698

699 **Table 4.** Summary of the regression analyses of the relationship per treatment between incubation
700 temperature during the day/ night and hatching success.

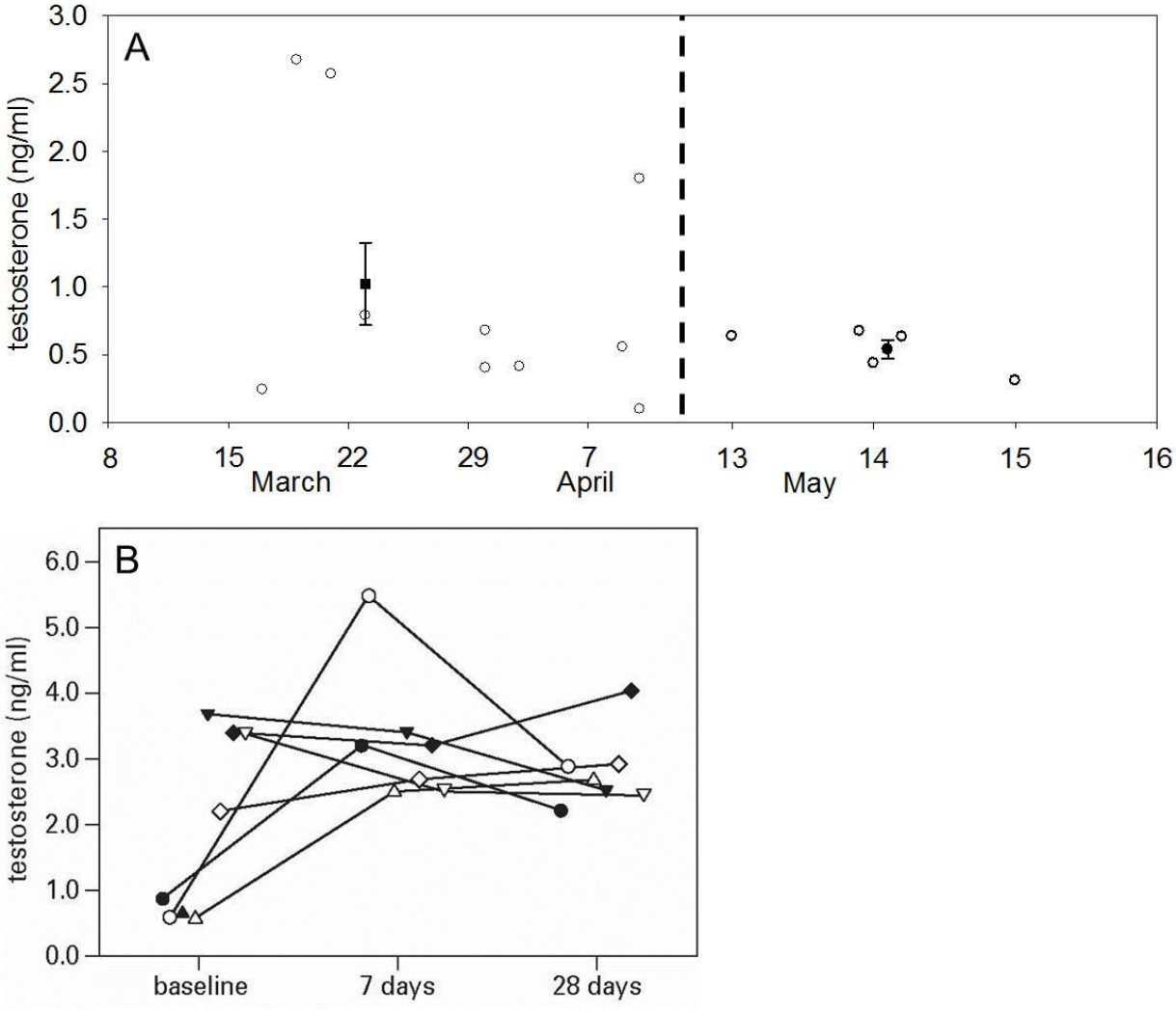
	Testosterone				**Control**			
	Intercept	Slope	R^2	*P*	Intercept	Slope	R^2	*P*
Day temperature	-6.36	0.25	0.66	**0.048**	5.90	-0.14	0.80	**0.02**
Night temperature	-6.34	0.23	0.67	**0.046**	4.59	-0.09	0.59	0.08

701

702
703

704
705

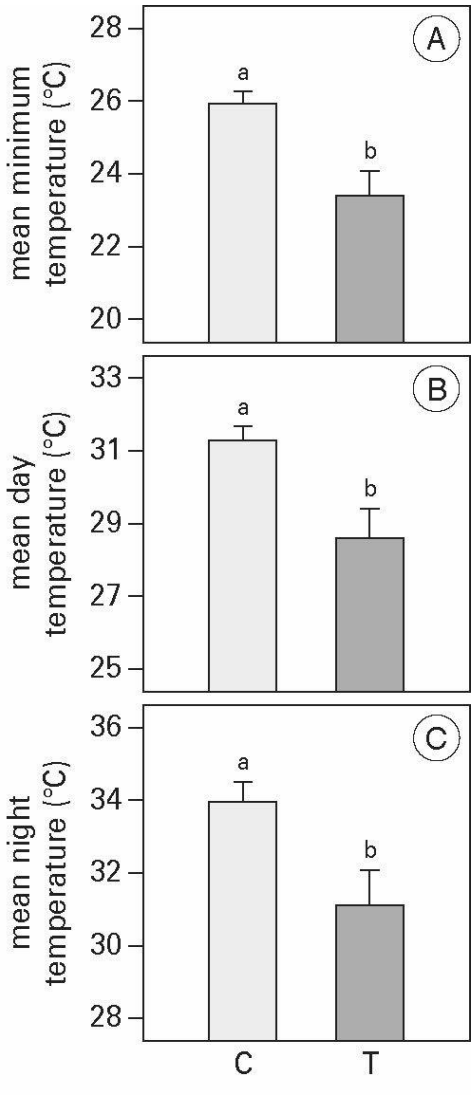
Figure 1.



709

710 Figure 2A and 2B.

711



712

713 Figure 3.

714