measure relevant for our flycatcher studies. For missing hourly data points (0.03% of data), we used an exponentially weighted moving average to replace the missing temperature values.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of captivity data

Overall, we compared data from 11 birds from 6 families in 1981 (5 females and 6 males), and 22 birds from 8 families in 2002 (11 females and 11 males). In spring, data were missing for one 2002 bird, and for autumn migratory restlessness, data were missing for three 2002 birds.

Annual cycle timing traits

We compiled data on the timing of moults, migratory restlessness, body mass, and reproductive activation (Main text, Figure 1). The timing of migratory restlessness was quantified from nightly activity profiles as described above. The timing of body mass changes was also quantified using changepoint analysis [42] to determine the date at which a bird shifted from high (winter) to low (spring) body mass states. Moult timing traits were dates of start and end. For the body and flight feather moults, we defined start as the first date on which a given moult was recorded, and end as the last date of recording this moult. We quantified variation in reproductive timing with weighted averages, weighting each measuring date by gonad size on that date. Thus, birds that showed enlarged gonads earlier in the season were assigned an earlier date, and vice versa. We did not include the declining phase of the reproductive cycle.

Because we had season-specific predictions, we analyzed timing traits in seasonal blocks. In autumn, our measures included only the end date of post-juvenile body moult and the end date of autumn migratory restlessness. We did not use the start dates because on several occasions these events may have started before data collection began. In winter, we used the start dates of winter moult of body plumage and flight feathers, and the start date of the winter drop in mass. Finally, in spring, we examined the end dates of winter body plumage and flight feather moult, the start date of spring migratory restlessness, and the weighted mean date of gonadal activation as described above.

Model construction and evaluation

We used linear mixed-effects models (lme4 package in R [52]) to test for a difference between cohorts in timing traits during autumn, winter, and spring. Because our hypotheses were structured by season and all of our predictors (traits) were in the same units (days), we first derived seasonal timing indices by averaging across seasonal traits for each individual. We thus obtained autumn, winter, and spring mean timings for each bird. We could not compare seasonal means for individuals missing data in any trait in a season, so we excluded individuals with missing data. We retained 30 birds in autumn (11 from 1981, 19 from 2002) and 28 in winter (10 from 1981, 18 from 2002). In spring, we had 23 individuals with complete data (8 from 1981, 15 from 2002); an additional 5 did not show any spring migratory restlessness or were not monitored. Therefore, we calculated two versions of the spring index, one including migratory restlessness but fewer (23) birds, and another version that excluded migratory restlessness but included 28 birds (10 from 1981, 18 from 2002). Both versions produced highly similar results in our analysis. The spring index without migratory restlessness, including the five additional individuals that were missing data, averaged 8.5 days earlier in 2002 (95% CI -17 to -0.49), compared to 9.3 days earlier (95% CI -16 to -2.9, $\chi^2_1 = 7.3$, $p = 0.007$) based on the 23 individuals with complete data (see Main text).

The response variables were the seasonal timing indices. We included a random intercept of brood ID (sibgroup) to account for any similarities in timing due to genetic similarities among siblings. The fixed effects were cohort (1981 or 2002), sex, a cohort × sex interaction, and hatch date (to account for any effect of the timing of hatching on subsequent annual cycle timing). To maximize the precision of our estimates given a small sample size, we removed non-cohort fixed effects if they were weakly supported ($p > 0.15$). We report effect sizes, 95% confidence intervals, and likelihood ratio test P values for remaining fixed effects. If there was evidence for a cohort × sex interaction, we report separate effects for males and females.

After testing seasonal indices, we repeated the above procedure for each individual timing trait and present effect sizes and confidence intervals for the effect of cohort on these traits. Our goal here was to better understand the drivers of seasonal differences while fully utilizing all data.

Analysis of field data

We tested for change in laydate (d_{lay}) with linear models. For nests where hatchdate (d_{hatch}) but not laydate was recorded, we estimated laydate with the following formula:

$$d_{lay} = d_{hatch} - (N_{egg} - 1) - 12$$

where N_{egg} is the number of eggs in the complete clutch. The constant 12 reflects the local incubation period.

We used the R package *climwin* [36] to identify the absolute spring time window (“climate window”) in which mean ambient temperature at the breeding site most closely predicted breeding phenology (Figure S4). We searched all climate windows of one week or longer in duration, up to 90 days before the last recorded laydate in our dataset (4 June). Searching a large number of climate windows increases the likelihood of a false positive result. Therefore, we used the *randwin* function to create 100 randomized datasets