To construct standardized visuals (e.g. Fig. 1*B,C*; Movies S2, S3) of the area of influence during periods of illumination, we cast radar resolution cells of the 0.5° elevation sweep to a regular spatial grid (i.e. raster image, ≈0.002° x 0.002°) using an equidistant cylindrical projection. We used maximum values of reflectivity and those nearest the radar for radial velocity when two or more resolution cells occupied a cell. We used the mean value in each cell for periods with and without illumination for aggregate plotting.

Acoustic Analysis. Because of the high intensity of calling activity at the site, in which many calls overlapped in time and frequency, and to minimize effects of different microphones, we used the amplitude in the 6-9 kHz frequency band to derive an index of calling activity. We applied a 10^{th}-order Butterworth band-pass filter with corner frequencies 6 kHz and 9 kHz to the dataset (see Fig. S6). We then calculated mean amplitude values for the 6-9 kHz frequency band for consecutive one-minute non-overlapping windows. Finally, we normalized the resulting time series to obtain a relative calling activity index, hereafter "normalized amplitude." A normalized amplitude of 1 represents the maximum observed calling activity.

To estimate numbers of calls from normalized amplitude, we manually counted flight calls from spectrograms (Hann Window, FFT size 512, overlap 87.5%, 375 Hz grid spacing; (3)) in 40 one-minute periods during the night of 11 September 2015. We randomly selected these periods during the night, while ensuring that there was equal representation from each quartile of the normalized amplitude distribution. Normalized amplitude was an excellent predictor of vocal activity ($R^2 = 0.90$, $P < 0.0001$; Fig. S11), demonstrating that it is an appropriate measure of vocal activity from flight calls. In this linear model, we forced the regression through the origin to avoid the impossible scenario of negative flight calls (i.e. there should be zero normalized amplitude when there are zero calls, but this assumes no interfering noises, which was not always the case).

In order to directly compare acoustic and radar observations with linear models, we downsampled acoustic observations to the frequency with which radar observations were gathered. We achieved this by simply selecting the nearest one-minute calling sample for each radar observation, provided that it occurred within three minutes of the radar observation.

Visual Observations. Visual observations represented, to the best of observers' abilities, estimates of numbers, species, and flight behaviors of birds. AF used Zeiss and Kowa optics (10 x 50 binoculars and 20-60 zoom x 85 spotting scope, respectively, in 2008, 2010, and 2012-2015) and Swarovski optics (12 x 50 binoculars and 30-70 x 95 spotting scope in 2016). These observations are archived as specified in the Methods. See Movie S1.

Hypothetically, decreases in average radial velocities observed by radar for nocturnally migrating birds during periods of illumination could mean either that birds' mean flight speeds slowed as they passed the installation, or that individual birds maintained flight speeds but, because many birds started circling, appeared to decrease in average speed relative to the radar station. We used visual observations to determine which of these scenarios was occurring.

Statistical Analyses. We used generalized additive models (R package mgcv (4)) to quantify the effects of illumination on birds' behaviors. We tested the categorical factors of light (on/off) and year on four metrics: standardized peak density; the total number of birds present within 0.5 km of the installation; the radial velocities of birds above the installation; and the number of flight calls recorded beneath the site. We looked separately at 0.5° and 1.5° radar sweeps. Because the light shutdown procedures took several minutes to complete, and to allow birds time to respond to the change of treatment, we excluded data points within 5 minutes of an on/off transition. In addition to the categorical factors listed above, we included two smooth terms (thin plate regression splines with basis dimension chosen automatically): 1) time of night and 2) mean bird density between 2-20 km away, fitted separately for each year. These terms accounted for any overall variation in densities and behavior through each night unrelated to local light pollution (e.g. due to weather factors and regular circadian patterns; see (5-7)) and additionally served to account for autocorrelation. Importantly, in our model of vocal activity, we also included the peak bird density above the installation (as measured by radar) as a continuous predictor to account for variation in calling explained simply by the number of birds present. For each metric, we compared models with three possible combinations of categorical factors: light alone; light and year; and light and year with an interaction. We evaluated these models with the Akaike Information Criterion (AIC) and selected the model with the lowest AIC score. However, if the model with the lowest AIC score was within 1 AIC unit of a model with fewer parameters, we used the more parsimonious model. We checked the distribution of model residuals and applied data transformations when necessary. Initially, the residuals for the models of standardized peak density, total number of birds, and number of flights calls were