Variation in female songbird state determines signal strength needed to evoke copulation

Ammon Perkes[1], Bernd Pfrommer[1], Kostas Daniilidis[1], David White[2], Marc Schmidt[1]
[1]University of Pennsylvania, Philadelphia, Pennsylvania, United States
[2]Wilfrid Laurier University, Waterloo, Ontario, Canada

Abstract

It is the female response to male signals that determines courtship success. In most songbirds, females control reproduction via the copulation solicitation display (CSD), an innate, stereotyped posture produced in direct response to male displays. Because CSD can be elicited in the absence of males by the presentation of recorded song, CSD production enables investigations into the effects of underlying signal features and behavioral state on female mating preferences. Using computer vision to quantify CSD trajectory in female brown-headed cowbirds (*Molothrus ater*), we show that both song quality and a female's internal state predict CSD production, as well as the onset latency and duration of the display. We also show that CSD can be produced in a graded fashion based on both signal strength and internal state. These results emphasize the importance of underlying receiver state in determining behavioral responses and suggest that female responsiveness acts in conjunction with male signal strength to determine the efficacy of male courtship.

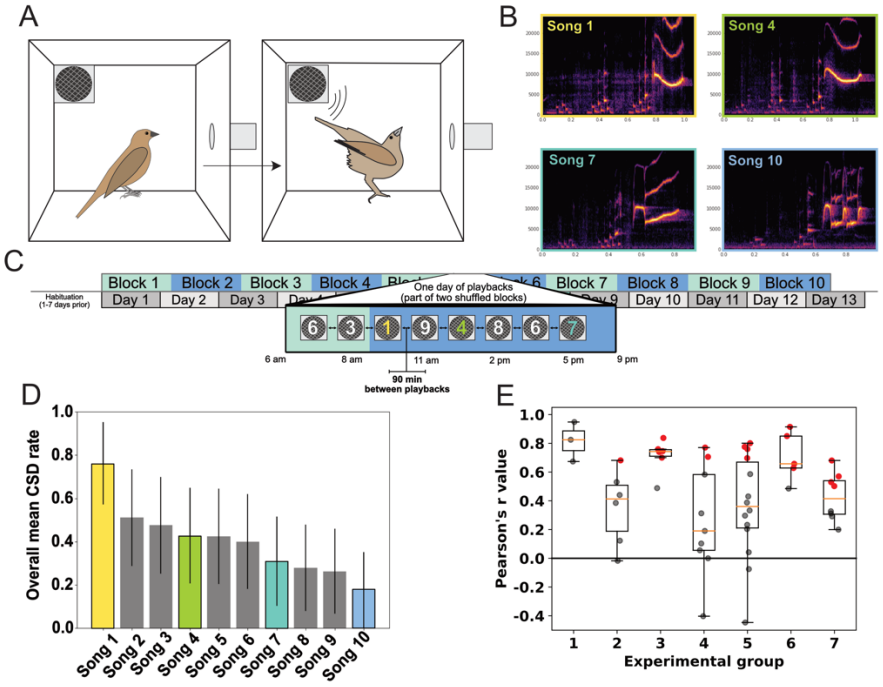
Introduction

In many species, selecting a partner for copulation is a carefully coordinated process that is mediated by courtship, typically consisting of a range of prolonged, multimodal signals by males that females in turn evaluate during mating (Andersson, 1994). While males' signals are frequently dramatic, preference is generally invisible, except as manifested by some behavioral response. The song displays of oscine songbirds (suborder, Passeri) are likely the most studied example of animal courtship and have contributed significantly to our understanding of sexual selection and signaling generally. While male song has received the most attention (Catchpole & Slater, 2008; Kroodsma & Byers, 1991), much of our understanding of song function comes from the behaviors that females produce in response to male courtship displays. The most dramatic of these behaviors is a lordosis-like pose referred to as a copulation solicitation display (CSD), in which females spread their wings, raise their tail, and arch their backs to expose their cloaca (see Figure 1a). Females across the avian phylogeny produce similar displays to invite, facilitate, and control copulation, and in most species, CSD is necessary for copulation to be successful. This behavior is strongly linked to the hormonal state of the female and CSD is never observed outside of periods of breeding, when estrogen levels increase in preparation for courtship and reproduction (Kern & King, 1972; Donna L. Maney et al., 2006; Moore, 1983). In some cases CSD is initiated by the female, sometimes for many hours before copulation occurs (Fernández & Mermoz, 2003; Nice & Marler, 1958; Sheldon & Burke, 1994), but in most species studied CSD occurs as a direct and rapid response to male

displays and is immediately followed by copulation attempts (Catchpole et al., 1986; King & West, 1977; Searcy & Marler, 1981).

Although normally limited to the context of mating behavior, in some species it is possible to elicit CSD under controlled conditions using only playback of male song, allowing the stimulus to be isolated from the many variables surrounding social interactions (King & West, 1977; Kreutzer et al., 1992; Searcy & Marler, 1981). In these experimental contexts, CSD remains tied to hormonal state: females generally need natural lighting schedules and, in some species, exogenous estrogen to become sufficiently receptive to song. During playbacks, some songs evoke CSD more frequently than others (Figure 1a-d), providing a metric of song potency, and much of our understanding of the significance of song features comes from observing CSD under these conditions (King & West, 1983a; Searcy, 1984; Spitler-Nabors & Baker, 1987; Vallet & Kreutzer, 1995). While song playbacks can elicit CSD in multiple species, brown-headed cowbirds (*Molothrus ater*) readily produce CSD to song playback without requiring hormonal manipulation (King & West, 1983b), providing a powerful system for studying CSD and female preferences (Peer et al., 2017; West et al., 1981; White, 2010).

FIGURE 1



The CSD assay reveals underlying agreement of song potency across birds. A. Cartoon showing a female in isolation producing CSD in response to male song playback. **B.** Exemplar spectrograms from four of the male song used for playback. Each example is color coded to match song potency levels shown in (d). Songs were recorded previously from adult males in aviaries. **C**. Overview of the experimental paradigm showing randomized presentation of blocks of 10 songs. Each song was played 90 minutes apart to prevent habituation **D**. Song potency, defined as the mean proportion of playbacks that elicit CSD for each song, averaged across females (n=8). Vertical lines depict standard error of the mean. Song numbers match exemplars in (b). **E**. Correlation between individual and group song potency scores across all 7 experimental groups. Box and whisker plots show quartile ranges and median (red line). Red dots denote correlations where Pearson's r is significantly different from 0 (at p<.05). Random horizontal jitter was added for easier visualization of points.

Despite the importance of CSD as a tool for understanding song and its centrality to reproductive behavior across bird taxa, we lack a basic understanding of the principles which govern CSD. While it is known that song rankings observed in playbacks reflect actual song preferences (Adrian L. O'Loghlen & Rothstein, 1995; West et al., 1981), it is unknown whether preferred songs produce CSD that are more robust than those produced to lower potency songs, or if CSDs can be grouped into distinct subtypes. Unlike song, which can be described in terms of acoustic features with millisecond precision, CSD has typically been described qualitatively, and we know little about the timing and natural variation of CSD or how these features may correlate with song preference.

Given how little is known about the behavior, it is no surprise that the neural mechanisms that govern CSD production are largely unknown. The selectivity of CSD behavior has been linked to regions involved in song production (Brenowitz, 1991; Maguire et al., 2013), but it is unknown whether these regions directly modulate CSD production, or whether they shape the formation of song preferences, indirectly affecting CSD. Indeed, many questions remain regarding the neural and behavioral mechanisms that control CSD production, particularly at short timescales. For example, while CSD is readily evoked by song playback in isolation, under natural conditions females hear hundreds of songs a day without producing CSD, after which they may produce a CSD to a single song from a male. So far, we lack even a conceptual framework that explains these aspects of CSD responses.

Just as early birdsong researchers established the foundational understanding of song production (Catchpole & Slater, 2008), we set out to identify the first principles of female preferences by carefully quantifying CSD and proposing a conceptual framework of CSD production. Here we provide novel experimental and computational tools to study CSD. Using a large dataset spanning a decade of experiments from different groups of birds, we show that females broadly agree on the songs that are most likely to elicit a CSD, but there is significant variability in the responses both within and across groups. This suggests that a female's response threshold interacts with male signal potency to determine CSD production. We also show that CSD is not a simple, "all-or-none" behavioral event, but rather a graded-behavior whose overall characteristics—including latency, duration, and intensity—are mediated by both signal potency and receiver state. Our findings suggest the hypothesis that short-term variation in female state can profoundly affect copulation success, and that in many species, manipulation of this state, either by the male or the female, could act as a target for optimizing copulation and ultimately reproductive output.

Results

1. Signal strength and underlying state predict CSD production

1.1 CSD response bias to male song is consistent within and across populations

To understand the regulation of CSD behavior, we assembled a large dataset that includes the responses of 70 birds from 7 different playback experiments spanning 10 years and captured at two different sites; 5 of these experiments used the same set of recorded songs, allowing for comparison across groups. All animals were naïve to both the presentation

paradigm and the songs presented, and no birds were repeated across groups (see methods for full details of group composition). As CSD is naturally evoked by song, birds received no prior training, and CSDs were not reinforced with food or other external rewards. The proportion of playbacks eliciting CSD within an experimental group defined a bird's potency (with respect to that group).

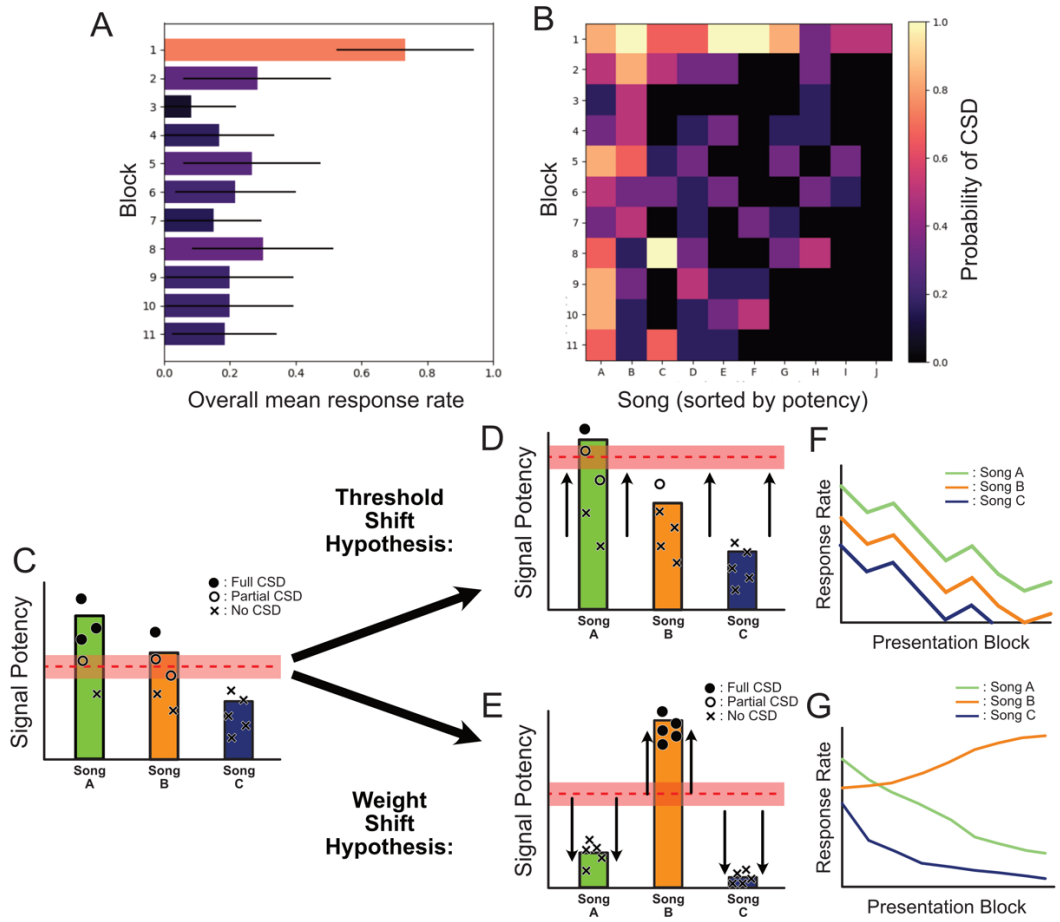
Within 6 of 7 experimental groups, some songs evoked significantly more CSD responses than others (at $p<.005$, repeated measures ANOVA, Figure S1a). We tested for agreement within groups by evaluating how the set of songs' individual song potency scores (computed for each individual in the group) correlated with the songs' overall potency (i.e., the mean of the rest of the birds in the group). We excluded 19 birds (out of 70) who responded less than 5% of the time, for a total of 51 birds across 7 groups (with mean group size of 10, ranging from 3 to 18). In 46 of these 51 birds, the mean individual responses to songs were positively correlated with their respective group's average responses (mean Pearson's $r = 0.45 +- 0.32$), and this correlation was significant (at $p<.05$) in the case of 20 out of 51 birds (Figure 1e).

In the five groups of birds that were exposed to the same set of songs, mean song response rates were strongly correlated across groups (Figure S1b) Across the 10 pairwise comparisons, mean group correlation was .72 (ranging from .47 to .89), and 4 of 5 groups correlated significantly (at $p<.05$) with at least one other group (Figure S1b). Taken together, these findings show that females display clear bias toward certain songs, and that the attractiveness of songs is maintained across groups. While past experiments have demonstrated consistent preferences (King, West, et al., 2003), this work reveals their breadth and persistence, with responses that were significantly correlated even between groups captured in locations separated by 30 miles and 10 years.

1.2 Signal potency and internal state drive CSD generation

Behavioral responses are influenced by signal potency, but they are also subject to an animal's behavioral motivation to respond. In our playback experiments, we found that female response rates decreased over the course of the experiment (as shown by a negative correlation between response rate and successive playback block number) (Figure 2a, Pearson's $r=-0.18$, $p<0.0001$,CI=-0.132:-0.210,n=3861). Furthermore, there was a significant correlation between the progression of the experiment (i.e., the number blocks of song that already been presented) and the potency of songs necessary to elicit CSD (Figure 2b, Pearson's $r=0.43$, $p<0.0001$,CI=-0.429:0.547,n=977). These effects persisted even with the removal of the first 2 blocks (Supplemental Statistics), suggesting it is not an artifact of the initial playbacks. At first glance, this pattern seems consistent with song preference developing throughout the playback by pruning away less-preferred songs. However, this pattern could also be explained by a global shift in response threshold leading to a steady sparsening of CSD responsiveness.

FIGURE 2



The decreasing CSD response rate over time could be explained either by a change in response threshold or a change to song preference. Top Row: Changes in CSD response rates with repeated playbacks. A. CSD response rate changes significantly over the course of the experiment ($p<.0001$, Repeated Measures ANOVA for 11 block groups with 10 song averages per group, $F=4.893$, $df=99$), with a dramatic and significant decrease in response rate after the first block (Tukey HSD for pairwise comparisons, $p<.05$). Here we show a representative experiment from 8 birds. Birds which responded less than 5% of the time were excluded, leaving n=6 birds. Block response rate is averaged across birds, with black lines representing the standard error of mean. **B.** Heatmap representing mean block response rate of each song (across the same 6 birds). When divided by song, it is apparent that in later blocks birds responded only to the most potent songs. This reduction in responsiveness to song could represent a change in song selectivity–pruning responses to low potency songs–or an overall decrease in responsiveness. **Bottom row: Modeling CSD response to song. C.** We conceptualize CSD production as a signal drive that must overcome some threshold in order to stimulate CSD. Any song exceeding the threshold (red horizontal bar) would generate CSD. Signals above this threshold may generate longer CSD, while signals at the margin might elicit partial CSD. Within this framework, change of CSD responsiveness could be explained either by **(D)** uniform changes in the response threshold (shifting threshold hypothesis), or **(E)** the updating of individual song weights (song weight hypothesis). Under a shifting threshold hypothesis **(F),** we would expect responses to all songs to shift in parallel. Under a song weight hypothesis **(G),** we would predict that song responses vary independently across time, with some increasing and others decreasing. C-G depict hypothetical data points, but this framework provides alternative, testable predictions (see Fig 3).

In order to explain these results and guide our experiments, we proposed a conceptual model of CSD production (Figure 2c-f), in which the CSD response is a function of the perceived song potency relative to a female's response threshold. In our model, apparent selectivity of the response emerges either by a changing of song weights, or through an increasingly strict

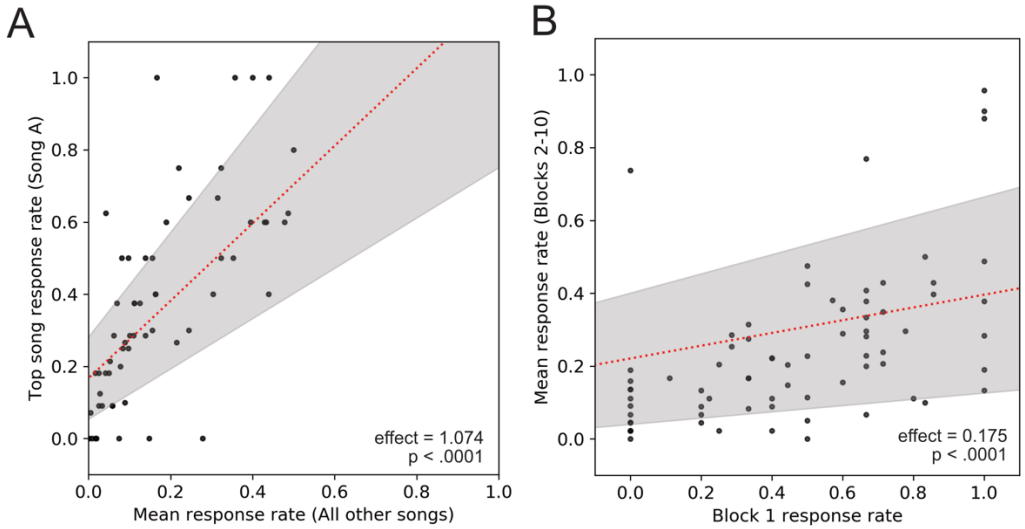
response threshold such that only high-potency songs can evoke a response. Because we were unable to directly measure response threshold or behavioral state, we used the mean block response rate (i.e., the proportion of playbacks from the same block which elicit CSD) as a proxy measure for response threshold.

1.3 Selectivity of CSD responses remain constant throughout song blocks

If changes to response selectivity are the result of a shifting response threshold, we would predict female responsiveness to the highest potency song to decrease proportionately with all other songs. Using linear mixed models to account for bird and song groups, we found a clear effect between the response rate to the most potent song and the response rate to all other songs within that presentation block (Figure 3a, effect=1.074, CI:0.267:1.450, $p<0.001$,n=90), indicating that female CSD responsiveness to all songs varies in parallel. This suggests that a shifting threshold, rather than an updated weighting of songs, drives the apparent increase in selectivity.

To further test whether song selectivity remains constant throughout the entire experiment, we measured whether female responses to songs in the first block predicted female responses to the same song in later blocks (i.e., blocks 5 and on). We observed a significant positive effect of a song's initial potency score (averaged across the group within the first block) on its mean potency during later blocks (Figure 3b, effect=0.175,CI:0.086:0.264 $p<0.001$,n=72), indicating that response selectivity represents a preexisting bias toward the most potent songs, rather than a relative preference that forms over the course of playback presentations with respect to the other songs in the block. In other words, song selectivity remains stable over time even as response threshold varies.

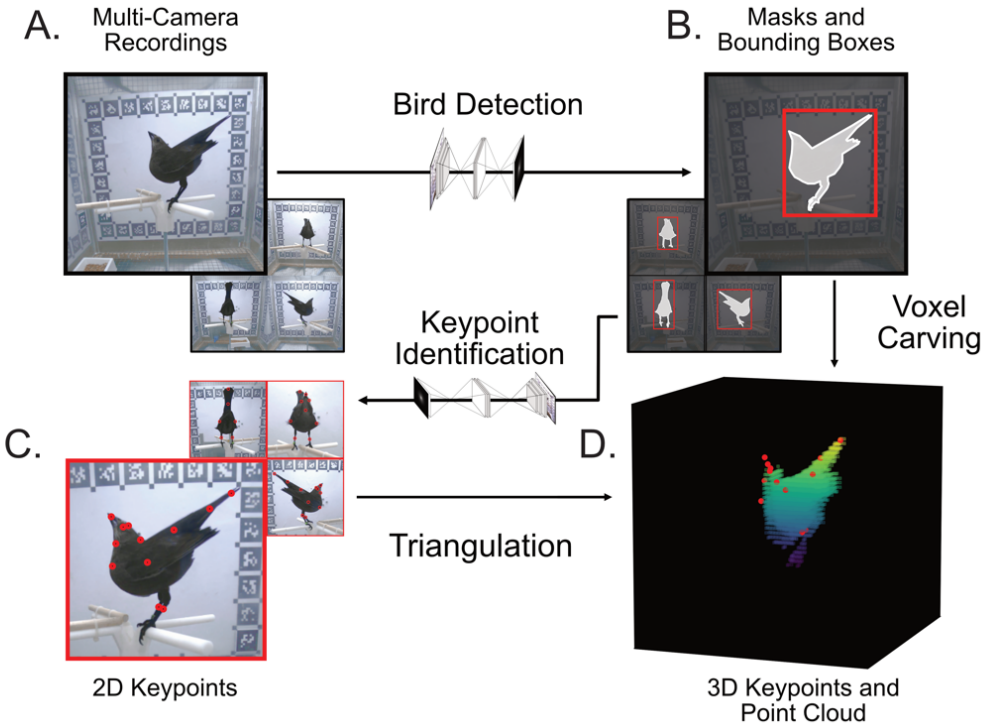
FIGURE 3



Changes in selectivity are driven by changes in baseline responsiveness. A. Scatter plot showing the top song response rate as a function of the mean response rate. Each point represents a single bird (n=64). To account for pseudoreplication of song sets, we used linear mixed models with song set and experimental group as random effects. The estimated effect (1.074) is shown in red, with the gray shade showing 95% confidence interval. The effect suggest that top song response rate decreases in parallel with overall responsiveness. **B.** A scatter plot showing mean response rate as a function of the first block response rate. Each point represents one song from one experiment averaged across the experimental group of birds (n=70). As in A, we use linear mixed models with song set and experimental group as random effects, the estimated effect (0.175) is shown in red with 95% confidence interval in gray. Response rate in the first block predicts responses in all other blocks. These data are consistent with an overall change in threshold over the course of the experiment.

As a final test for an effect of song exposure, we tested whether female CSD reaction time changes over the course of playback presentations. Reaction time often decreases in evaluation tasks as subjects gain familiarity with the stimuli (Gold & Shadlen, 2007; Mruczek & Sheinberg, 2007). Each song stimulus consisted of 2-3 introductory note clusters followed by a terminal whistle, the entire song around 1 second in duration (1.111 +/- 0.115 s). For a subset of birds (n=10), we were able to calculate latency of response. The median latency to initiate CSD was 0.668 s (+- 0.249) from song onset implying that CSD responses are initiated prior to the completion of song and typically during the second introductory note cluster. We did not find any decrease in response latency throughout the experiment (effect=0.007, CI:-0.001,0.015,p<0.094,n=203), and the mean latency observed in the first block was shorter than the length of song (mean first block latency = 0.823 +-0.319 s, while song mean length was 1.11 +- 0.110 s). These findings also show that even when hearing song for the first time, females commit to producing a CSD early in the song–during the introductory note clusters–and therefore do not rely on information from the terminal whistle.

FIGURE 4



Using computer vision to quantify bird pose in 3D. A. We use a system of four calibrated cameras to capture images of the female cowbird from multiple angles. **B.** In each frame, we first identify the mask (white) and the bounding box (red) of the bird using a neural network, maskrcnn_resnet50 (He et al, 2017), pretrained on the COCO dataset (Lin et al, 2014). We finetuned the network on a dataset of 276 masks of birds from frames of playback experiments. This outputs cropped images of the bird. **C.** Crops are then used as the input for a second neural network (Pose resnet, Sun et al, 2019) to identify 14 keypoints within the image (shown in red). This network was finetuned using hand labeled keypoints from the same dataset of 276 birds. **D.** Keypoints are triangulated across the 4 views to estimate their position in 3-dimensional space. In parallel, we use a silhouette-based approach to generate a 3D point-cloud for visualization.

2. Factors influencing CSD timing, shape, and duration

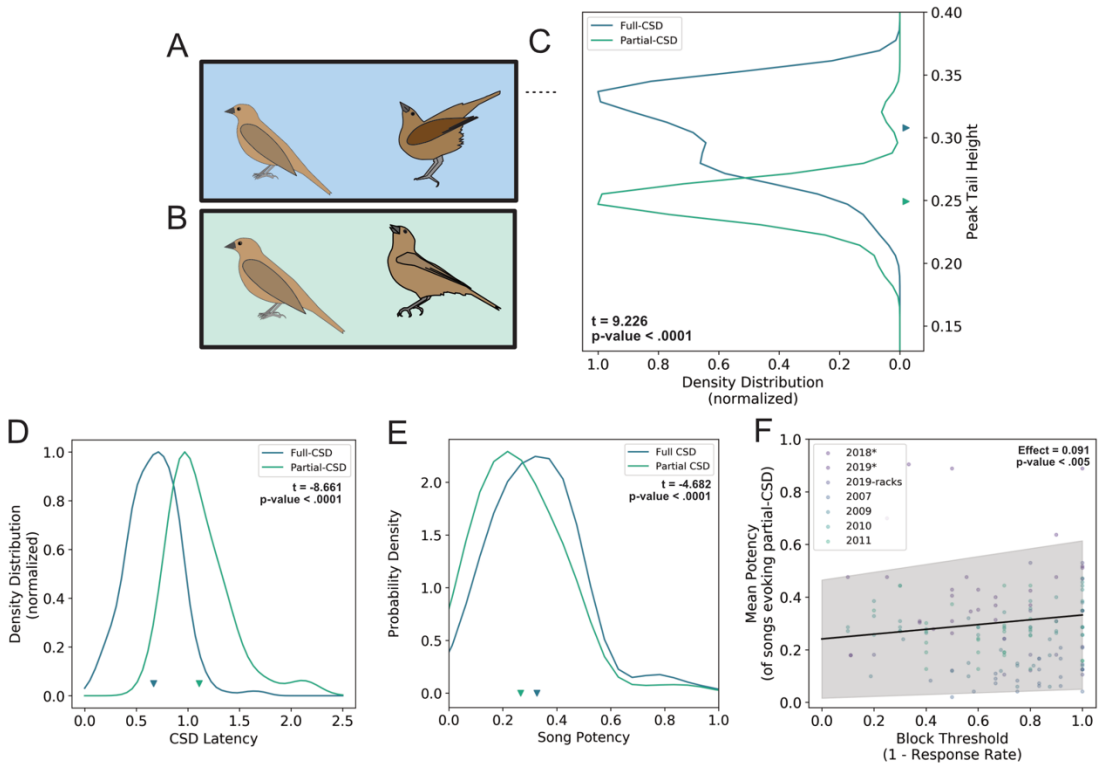
2.1 Song potency predicts intensity of CSD response.

As we have defined it, the CSD response is a clear behavioral event in which the female arches her back and lifts her tail above her body (thereby exposing the cloaca) (Supplemental Video 1). However, females often produce various levels of reduced displays, with only slight arching of the back, tilting of the head and/or spreading of the wings, and without full elevation of the tail or exposure of the cloaca. We used a neural network-based approach to reconstruct the three-dimensional pose of the bird in each frame of the video, which allowed us to track the trajectory of each body part with high temporal precision (Figure 4) to quantify CSD. These displays were visually distinct (Figure 5a-b; Supplemental Video 2) and statistically separable from standard "full-CSD" based on tail height (Figure 5c) and latency (t= -8.661, p<.0001,n=237;28, Figure 5d). Frequently these "partial-CSDs" are grouped together with full-CSD (e.g. A. L. O'Loghlen & Beecher, 1997, 1999; Ratcliffe & Weisman, 1987), although they have been separated as well (D. L. Maney et al., 2003; Adrian L. O'Loghlen & Rothstein, 2002). In our case, we distinguished between full- and partial-CSD, but we hypothesized that partial-CSDs are a version of CSD that occurs when signal drive only marginally exceeds the CSD response threshold (see Figure 2). In other words, we would expect partial-CSD to be produced by songs with a potency just sufficient to elicit a response, but not enough to elicit a full-CSD.

To test this hypothesis, we first analyzed the distribution of the potency scores of songs presented for all the trials which elicited a) CSD, b) partial-CSD, or c) no CSD. Song potency is defined as the mean proportion of playbacks that elicited CSD for a given song within an experimental group. Consistent with the hypothesis that partial-CSD is a marginal-response, we found that the mean song potency of all trials eliciting partial-CSD (mean= 0.265 +-.160,n=237) was significantly lower than those eliciting full-CSD (mean=0.326 +- 0.163,n=482; t=-4.682, p<0.0001, , Figure 5e), but significantly higher than the mean song potency of playbacks that failed to elicit CSD (mean=0.213 +- 0.143,n=855; t=4.836, p<0.0001).

Because the threshold for eliciting a CSD varies throughout the experiment (see Figure 1), we would expect the mean potency of trials that elicit partial-CSD within a given block to shift as well, reflecting a change in the current marginal potency (i.e., the signal potency that just meets the threshold) (see Figure 2). Consistent with this prediction, we found a significant effect between a given block's response threshold (measured as 1 – mean response rate) and the mean potency of songs that elicited partial-CSD within that block (Figure 5f, effect=0.091,CI=0.034:0.149,n=150, p<0.005).

FIGURE 5



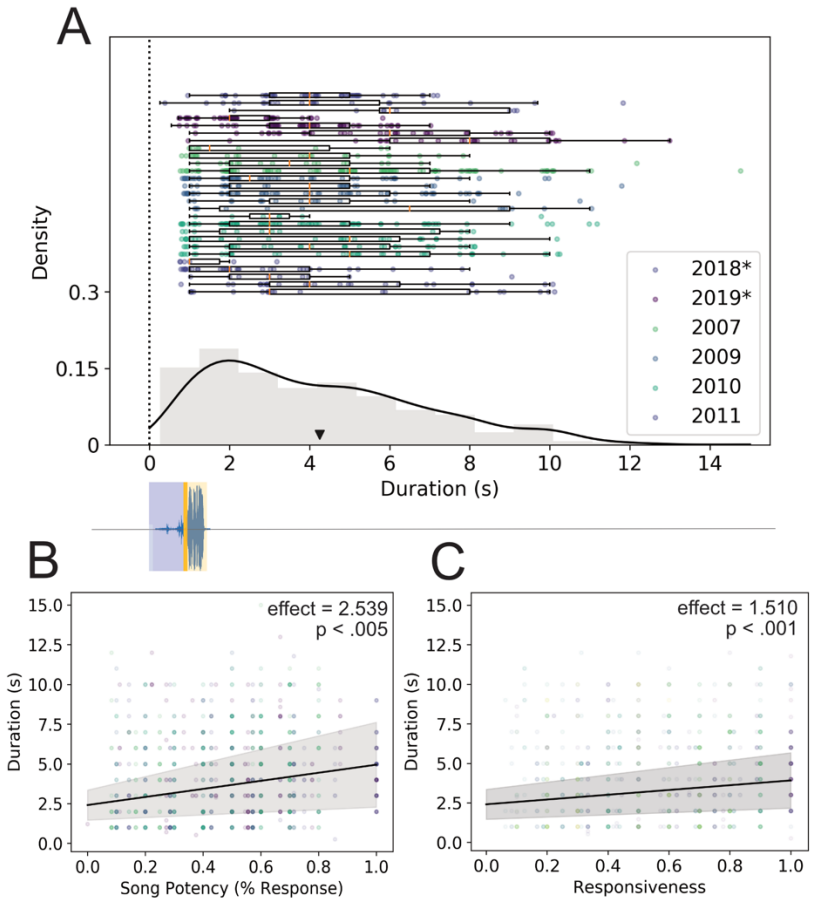
Partial-CSD is a reduced version of CSD. A. Schematic depicting full-CSD, characterized by wing spread and arching of the back, with the tail raised above the level of the body. **B.** Schematic depicting an example partial-CSD. In contrast to full-CSD, females only present a slightly arched back and lowered wing, and the tail never is fully raised above the level of the body. Full and partial-CSD appear related in appearance but clearly distinct. **C.** Density distributions of the tail heights across all full-CSD (blue, n=153) and partial-CSD (green, n=23). Arrows show the mean tail height for each behavior. Full- and partial-CSD heights were (by definition) significantly different (t = 9.226, p<0.001,n=137;26), although overlap existed due to variation in individual bird size and perching height. **D.** The density distribution for latency of full-CSD and partial-CSD, showing that full-CSD occurs significantly faster than partial-CSD, although there is substantial overlap between the two. Density distributions modes for panels (c) and (d) were scaled to have each maximum at 1 for ease of comparison. **E.** Females produce partial-CSD more frequently to lower potency songs than they do for full-CSD. This plot shows the density distribution of the songs in trials which elicited full- or partial-CSD. The potency of song trials eliciting full-CSD was significantly higher than those which elicited partial-CSD. All density distributions were smoothed using a gaussian kernel using Scott's rule (Scott, 1992). **F.** The set of songs which elicit partial-CSD varies with response threshold. The scatter plot showing the mean potency of songs eliciting partial-CSD as a function of block threshold. Each point represents a single block from a single bird (n=208). Colors denote birds from different experimental groups. The black line shows a significant effect of block response rate on the songs which elicit partial-CSD in a given block, based on a linear mixed effect model with group, song, and bird identity as random effects, with the grey shading showing the 95% confidence interval.

2.2 <u>CSD duration correlates with song potency and female behavioral state</u>

In a variety of behaviors studied, response duration varies with stimulus intensity (Cattell, 1886; Turesson et al., 2009). Consistent with CSD being a graded response, we hypothesized that signal potency should also correlate with posture duration. For six playback experiments (n=832 postures from 27 birds), we quantified posture duration as the time (to the nearest 1s) from the initial vertical motion of the tail to the moment the tail returned below the

level of the body. Across all birds, mean CSD duration was 4.25 (+/- 2.67) seconds. To account for the multiple non-independent experimental variables, including the significant individual variation between birds (ANOVA, F=5.996,p<.0001), we used a linear mixed model to test for the effect of song potency and response threshold on CSD duration. We found a significant effect between song potency and CSD duration, with higher potency songs stimulating postures of longer duration (Figure 6b, e=2.960, CI: 0.649-5.271, p<.017). We also found an effect of the overall response rate on duration (Figure 6c, e=1.083, CI: 0.282-1.884, p<.008) suggesting that birds also produce longer-duration CSD when their response threshold is lower.

FIGURE 6

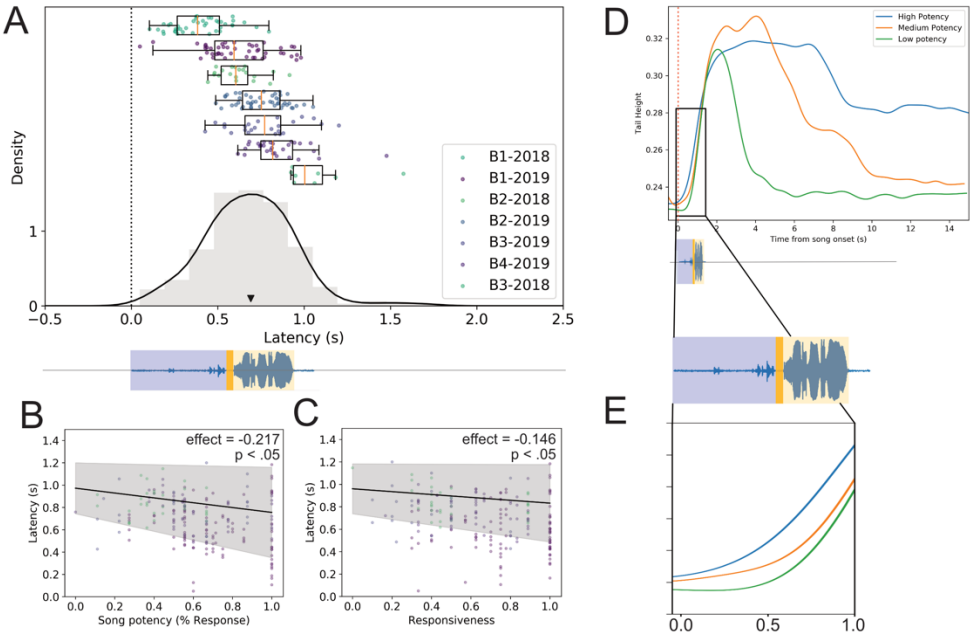


Song potency and underlying behavioral state both significantly determine CSD duration. A. The distribution of duration for 29 birds across 6 experimental groups for which we could calculate duration (n=821 postures). Each row represents an individual bird with each point representing an individual posture. Random horizontal jitter was added for ease of visualization. Box and whisker plots show the distribution of the data, and colors denote experimental groups. The black triangle shows the mean duration of CSD. The groups marked with asterisks (2018*, 2019*) were recorded in small, isolated conditions for the purposes of computer vision. The amplitude waveform of a typical cowbird song is shown below the graph. Introductory notes are highlighted in blue and the terminal whistle in yellow. **B.** Scatter plot showing CSD duration as a function of song potency. The black line represents the estimated effect of song potency on CSD duration using a linear mixed model with bird, block, song, and experimental group as random effects and song potency and block response rate as fixed effects. **C**. A similar plot of CSD duration as a function of female responsiveness. For both b and c, grey shading shows the 95% confidence interval predicted by the model.

2.3 CSD latency correlates with song potency and female behavioral state

Similar to duration, in most behavioral tasks, response latencies vary with both signal strength (Blaxter & Hoss, 1981; Laming, 1968) and signal ambiguity (Gold & Shadlen, 2007). We tested whether signals that most reliably elicit CSD responses also produce responses with the shortest latencies. Using a playback chamber equipped with multiple cameras and precise audio-video calibration at 50 Hz frame rate, we were able to calculate the onset latency for 222 CSDs (from 7 birds, Figure 7a). Latency was defined as the delay between song onset and the moment the tail feathers began to rise. Mean latency was 0.691 s (+- 0.293 std), and there was significant individual variation between birds (Anova, F=24.107 p<.0005). Using linear mixed models to account for song and bird identity, we found a significant effect of song potency on CSD latency (Figure 7b, e=-0.217, CI: -0.394:-0.040, p<.05) with higher potency songs eliciting the shortest onset responses. Defining response threshold by the mean responsiveness within a given presentation block, our analysis also revealed a significant effect of response threshold on latency (Figure 7c, e=-0.146, CI: -0.289:-0.003, p<.05) with CSD latencies being shorter in trials when females had a lower response threshold. To summarize our timing results, we found that certain songs evoke more CSD, and these high potency songs produce shorter latency, longer duration CSD (Figure 7d-e), and that highly responsive birds produce shorter-latency, longer duration CSD, separate from the effect of song potency.

Figure 7



Commitment to produce a CSD occurs early in the song and onset latency correlates with both song potency and underlying behavioral state. As in Figure 6, **A**. depicts the distribution of CSD latencies for 203 postures from 7 individual birds for which we could calculate precise latency. The black triangle shows the mean latency. Scatter plots, overlayed with box and whisker plots, show individual bird distributions. Below, an example waveform shows the timing of cowbird song, and its division into introductory notes (blue) and whistle (yellow). Note that CSD onset occurs before song termination suggesting that commitment to produce a display occurs early in the song during the production of the relatively quiet introductory note clusters. **B.** A scatter plot showing CSD latency as a function of song potency. Each point represents one CSD. **continued:**

Continued from figure 7: The black line represents the estimated effect of song potency on CSD latency, based on a linear mixed model with song potency and response rate as fixed effects and bird, block, song, and experimental group as random effects. **C.** CSD latency decreases as a function of female responsiveness. Scatter plots show observed CSD latencies, each point representing a single CSD, with the black line is the estimated effect using the same model in b. For b and c, grey bars show 95% confidence intervals. Bird colors are consistent for a, b, and c. **D.** Three exemplar postures from a single bird demonstrate the effect of song potency on CSD duration and latency (cutout). CSD duration was longer for higher potency songs. There was no effect of potency on peak height (Figure S3d). **E.** Zooming in on the 1 second following song onset, the latency of CSD is shorter for higher potency songs (Figure 7b), although there was no effect of potency on tail velocity (Figure S3b).

2.4 CSD shape varies significantly with behavioral state but not with song potency

CSD has generally been defined as a single behavioral class. The observation, however, that song potency and response threshold influence CSD onset and duration, as well as the probability of partial displays, suggest that CSD is in fact a graded response. Full-CSD could contain further gradations, and these could even represent different behavioral types that encode meaningful information. To investigate this possibility, we used our computer vision approach to track the precise trajectory of CSD in 3 dimensions in a select group of birds (n = 142 postures from 6 birds) allowing us to quantify the position and velocity of each body part at each moment in time (see Figure 4). To define the magnitude of CSD, we used maximum recorded tail velocity and the peak tail height. We found that song potency had no effect on max velocity (Figure S2b, effect=0.399,CI: -0.76:–1.563) and only a weak, non-significant effect on peak height (Figure S2d, e=0.056, -0.003:0.115, p<.07). On the other hand, female response threshold had a significant effect on max velocity (Figure S2a, e=1.477, CI: 0.487:2.468, p<.005) but no effect on peak height (Figure S2c, e=0.012, CI: -0.029:0.054).

Because there are many ways, besides tail height and velocity, that CSD could vary, we wanted an unbiased approach to characterize CSD that was not limited to the metrics we selected. We used a quantitative representation of CSD in which each frame was described by a 91-dimensional vector of the pairwise distances between key points (Figure S3b) to train a random-forest classifier. To prevent bias introduced by variation in bird shape, we trained a distinct classifier for each bird using a training set of labeled playbacks (80% of trials). Using a withheld set of labeled playbacks (20% of trials), we tested whether the classifier could predict the song potency or female response threshold during the trial based on the CSD evoked. We found no evidence of discrete CSD types based on song potency, with the classifier only performing better than chance for 1 of 7 birds (Figure S2f). In contrast, our classifier performed significantly better than chance in 4 of 7 birds when trained to predict response threshold. This effect was strongest in the last few seconds of the CSD (Figure S2e).

In summary, this suggests that while songs with higher potency do evoke shorter latency and longer duration CSD, there were no detectable differences in the CSD themselves. The postures of more responsive females, on the other hand, were distinct based on measured velocity and could be classified based on the trajectory of CSD offset.

2.5 Female CSD responses are not composed of distinct subcomponents

Many complex behaviors can be divided into a sequence of discrete and simpler subcomponents (D. J. Anderson & Perona, 2014; Berman et al., 2014). One clear example is the song of many songbirds, which is made up of individual syllables produced in a highly stereotyped manner (Baker & Boylan, 1995; T. J. Gardner et al., 2005). The specific acoustic and temporal properties of these subcomponents, which can include both the syllables as well as the gaps that separate them, can vary significantly based on social context (Glaze & Troyer, 2006). In some cases, syllable transition probabilities can also vary with social context (Sakata et

al., 2008). Given the role of CSD in sexual signaling, and that the same neural circuit that controls singing has been shown to regulate aspects of CSD selectivity (Brenowitz, 1991; Perkes et al., 2019), we wanted to test the possibility that CSD posture might also represent a sequence of behavioral subcomponents, whose order and/or temporal features within the sequence might reflect some aspect of behavioral state or song potency. For example, while the peak of CSD posture appears consistent, with the tail elevated and wings spread to expose the cloaca, we wondered whether early elements of the posture (e.g., early head tilt or wing motion) could be thought of as separate subcomponents that are independently regulated.

We used trajectories of all individual keypoints to describe CSD sequencing (Figure S3a), and the 91-dimensional vector of keypoint distances (Figure S3b) to produce a behavioral space of posture and CSD trajectories (Figure S3c-d). To assess CSD sequencing, we calculated the onset of motion for each individual keypoint (defined as the first moment following song that a particular keypoint deviated more than 2 std from baseline, Figure S3a). While the mean time of onset of head movement was slightly earlier than the leg or wing (Figure S3e, ANOVA,n=160, F=3.08 p<.01, Post-hoc Tukey HSD, p<0.05, Figure S3h), we found no significant bias for CSD initiation for any body part (χ^2=0.05 p=0.999). Nor did we observe any evidence for stereotyped sequencing within CSD itself (Chi-Squared test p>0.99 for all body parts transitions, Figure S3f).

As a final test for possible sub-components of CSD, we characterized the behavioral density of every observed pose (Figure S3d). In brief, any pose that appeared frequently (e.g., perching) should have high density, while one that was rare or highly variable (e.g., initiating CSD while hanging upside-down from the ceiling, Supplemental Video 5) should have low density. As CSD occurs, periods of high density would suggest common gestures across CSD, or recurring patterns within CSD. Here we observed a consistent spike in pose-space density only at one timepoint, (Figure S3h) corresponding with the average moment birds reached the peak of posture. This further suggests that CSD exists as a single goal posture, rather than a series of discrete gestures. In summary, we found no evidence for discrete subcomponents or sequencing of CSD, and found that the pose for CSD itself is stereotyped while the transitional motion towards it is variable.

3. Song evokes other postural responses

3.1 Song presentation evokes a suite of non-CSD postural behaviors

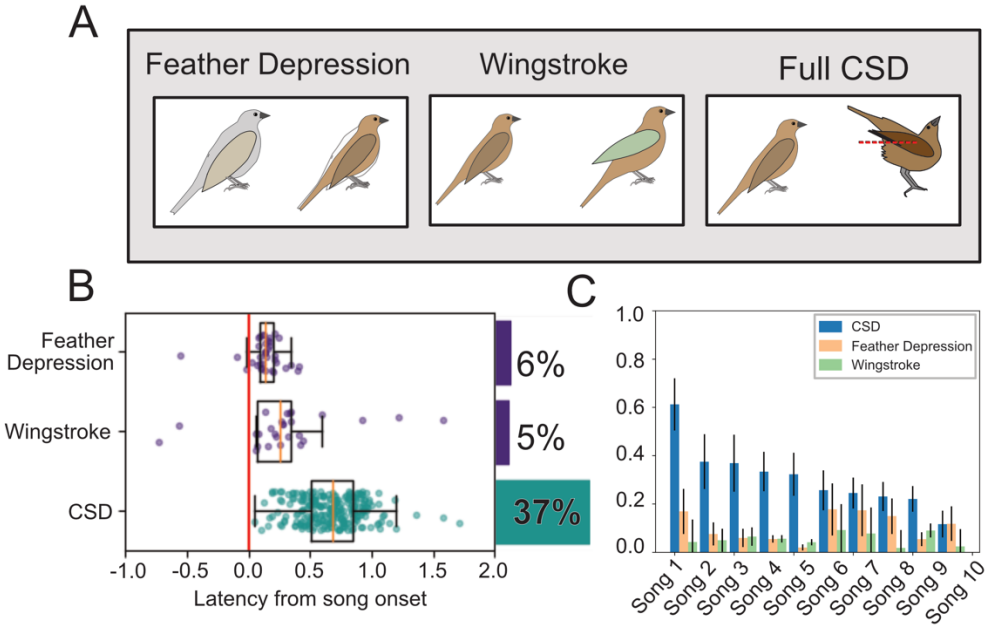
In addition to CSDs, which are rare in natural contexts, female birds respond to song with a number of postural as well as vocal responses (Riebel, 2003; West & King, 1988a). Some of these responses have been suggested to act as signals of song preference (Carouso-Peck & Goldstein, 2019; West & King, 1988a) and might therefore be involved in courtship signaling. Using our high-resolution observation boxes, we wished to assess the range of non-vocal behaviors that could be evoked by male song under these laboratory conditions. We were particularly interested if these behaviors showed the same type of song specificity as observed for CSD.

We quantified the full range of behavior produced by females (n=10) during song playbacks and quantified 8 behaviors that females consistently produced in the 5 seconds following song (i.e., in >5% of responses, across 527 total song presentations). These behaviors,

arranged by increasing frequency, were 1) leg adjustment (26 times, 3%), 2) gaze shift (38 times, 5%) 3) wingstrokes (rapid, unilateral wing motion, 41 times, 5%) 4) rapid feather depressions (47 times, 6%), 5) partial-CSD (55 times, 7%) and 6) full-CSD (280 times, 37%). Interestingly, although females often respond to male song with precisely timed chatter calls in large aviaries (H. L. Anderson et al., 2021), we did not observe vocal call responses in our recording boxes. The precise criterion used for classifying each of these behaviors is described in the methods section with examples of each behavior available as supplemental videos.

In quantifying behavioral responses and selectivity, we chose to focus particularly on wingstrokes and feather depressions (Figure 8a) because these behaviors had been previously linked to song preference in social contexts (Morris, 2008; West & King, 1988a). Feather depressions and wingstroke-like movements occasionally occurred during baseline conditions, however both wingstrokes and feather depressions were clearly evoked by song onset. We compared the occurrence of these behaviors in the 5 seconds prior to song compared to the 5 seconds following song onset. Song presentation evoked a significant increase in both wingstroke behavior (probability of occurring during vs before playback: 0.04 vs 0.009, N=548, t=3.326, p<0.001) and feather depression (0.005 vs 0.065, N=548, t =5.448, p<0.0001).

FIGURE 8



Song presentation evokes a ranges of postural displays which each have characteristic latencies and song bias. A. Schematic depictions of the 3 behaviors for which we quantify latency. Full CSD was defined by the tail being raised above the level of the body (red line) **B.** The distribution of latency for these 3 behaviors with respect to song onset (behaviors occurring more than 1 second prior to song onset are not shown). CSD latency (Mean 0.668 +- .016 SEM, n=237) was significantly slower than feather depressions (.166 +- 0.015 SEM, n=37) and wingstrokes (.379 +- 0.071 SEM, n=30) (ANOVA, F=23.6, p<0.0001, Tukey-HSD corrected pairwise comparisons at p<0.001). Behaviors occurring prior to song onset were excluded from the analysis, as were wingstrokes occurring later than 1s, to focus exclusively on song evoked behaviors. **C.** The proportion of feather depressions correlates with song potency, while wingstroke production appears unrelated. The bar plot shows the proportion of playbacks that evoked postural behaviors, sorted by CSD (i.e., song potency). Proportions are averaged across the 9 birds for which we could calculate latency, black lines show +- 1 SEM for the mean across birds.

3.2 Wingstrokes and feather depressions occur significantly faster than CSD responses

As with CSD, we quantified the latency (defined by the first frame the behavior was visible) for both wingstrokes and feather depressions. There were significant differences in latency between the 4 behaviors observed (i.e., wingstrokes, feather-depressions, full-CSD, and partial-CSD, Figure 8c, ANOVA, F=59.7, n=332 events total, p<.0001), and both feather depression (0.166 +- 0.087 s) and wingstrokes (0.379 +- 0.354 sec) showed onset latencies that were significantly shorter than CSD (Tukey-HSD adjusted p-values <0.001). Notably, these findings suggest that both wingstrokes as well as feather depression occur during the very first acoustic elements of the song whereas CSD (Figure 8c, mean latency = 0.668 +- 0.249) occurs after the bird has heard most of the introductory note clusters.

3.3 Rapid non-CSD responses such as feather depression correlate with song potency

Although the rapid latencies seemed indicative of a reflexive, startle response, we wondered whether these responses might actually be based on an evaluation of the song signal. We found no significant song bias for either wingstrokes (F=0.438) or feather depressions (F=0.533; ANOVA, p > 0.77 for both), but given our small sample size, this was not surprising. Because we had an established metric of song potency (based on CSD responses), we could ask whether the ratio of behavioral responses correlated with the known mean potency scores (Figure 8b). We found no significant correlation between feather depression and song preference (r=0.446, CI: -0.255:0.840, p<0.20) and interestingly, even though it has been shown in large aviaries that wingstrokes correlate with song preference (White et al., 2006), we did not detect any correlation between song preference and wingstroke production (Pearson's r=0.293, CI: -0.412:0.779, p = 0.411) suggesting, at least under these conditions, that wingstrokes are not strongly biased by song potency.

Discussion

Most innate survival behaviors are driven by sensory triggers. For behaviors such as escape, the nature of the sensory stimuli that drive these responses are well-characterized because the stimulus-response relationship is clearly defined (Dill, 1974; R. C. Eaton et al., 2001; Wei et al., 2021). This allows for the ability to manipulate the signal and identify the specific features that elicit the behavior. Combined with careful quantification of behavioral responses, these types of studies have been able to identify in detail how stimuli elicit the behavior and produce a detailed understanding of the neural mechanisms underlying these critical survival behaviors (R. Eaton, 1984; R. C. Eaton et al., 2001; Evans et al., 2019; Roberts, 1992). In contrast, the mechanisms underlying reproductive behaviors remain less well understood, in part because most sensory signals that trigger behaviors–including female copulatory responses–involve close, even physical, contact (Noble, 1979), and are therefore challenging to quantify and manipulate in a controlled manner. During lordosis in rodents, for example, initiation of this copulatory behavior is mediated primarily by cutaneous activation of the flanks, a stimulus that is difficult to precisely reproduce experimentally, given the diversity of sensory areas that are activated during such stimulation (Kow & Pfaff, 1975; Pfaff et al., 2008).

An additional challenge with studying reproductive behaviors is that the relevant triggers are shaped by female preferences, and as such vary significantly across species.

Songbirds offer a unique opportunity to study the fine-grained features of the signals that drive reproduction because the relevant behaviors can be elicited in a laboratory setting without the presence of a male, using the presentation of easily-defined acoustic signals (King & West, 1977; Searcy & Marler, 1981). The copulatory response, known as the copulation solicitation display (CSD), shares many of the same features as lordosis, including postural arching of the back to both support the mounting male and expose the reproductive organ. Because the same signals that elicit CSD (i.e., male song displays) are used within the context of other courtship behavior (Kroodsma & Byers, 1991), CSD can also be used to evaluate a female's innate signal preference. In the current study, we developed a novel approach to capture pose, providing a quantitative representation of CSD, produced in response to varying song stimuli, all captured with millisecond precision. We observed that adult female cowbirds showed a consistent bias for specific songs, and these high potency songs also elicited the shortest onset latencies and longest response durations. Song bias was consistent across females and was present from the first presentation of song. In addition to signal strength, an animal's internal state was a strong predictor of the response to song, showing that copulatory behavior is controlled by the combination of signal potency and behavioral state.

Commitment to produce a CSD occurs early in the song

Our quantitative analyses of CSD timing showed that the latency to produce CSD was short, often beginning within 250ms of song onset, a full 750ms before song was complete. These latencies are consistent with previous work suggesting that CSD is evoked by the lower amplitude introductory note clusters that make up the first component of the song (King et al., 1986; West et al., 1979). CSD initiation was not reliant, at least in our laboratory setting, on the high amplitude terminal whistle as previous suggested (Rothstein & Fleischer, 1987). Preliminary experiments confirmed that presentation of introductory notes alone is sufficient to elicit CSD (Supplementary Video 6). These introductory notes are produced by rapid switching between two sides of the avian vocal organ, the syrinx. Because of the precision required, these syllables likely represent an honest signal of male fitness (Lahti et al., 2011; Podos, 1997). Rapid left-right switching appears to be a consistent feature of preferred syllables across songbird species (Suthers et al., 1999) and these syllables are particularly effective in evoking female copulatory responses (Suthers et al., 2012; Vallet & Kreutzer, 1995). In songbirds, successful forced copulation is rare (Low, 2004; McKinney & Evarts, 1998) so these hyper-stimulating signals may represent a form of sensory coercion. Interestingly, in social contexts female cowbirds often produce loud, broadband "chatter" vocalizations in response to male song that effectively jam the introductory notes. Our recent work shows that these vocalizations are able to prevent females from going into posture (H. L. Anderson et al., 2021), and suggest this might be an effective strategy females have evolved for maintaining reproductive autonomy.

Female cowbirds appear to have an innate agreement of song potency

Based on prior findings (West & King, 1988b) and prevailing theories of mate selection (Andersson, 1994; Real, 1990), we initially assumed that females might form a comparative ranking between the different songs presented. We therefore predicted that onset latency would decrease, and the strength of song bias would increase, as females accumulated information about song quality over repeated presentations. Surprisingly, we observed that females responded preferentially from the very first presentation, and these preferences did not change over repeated playbacks, nor did the latency to respond to song decrease. Because all females were naïve to the songs presented, and the females were tested over the course of ten years, our findings suggest that females have an innate bias towards certain song types and that females across different experimental years agree on that bias. In this way the CSD assay appears consistent with other no-choice assays, in which variation in the absolute stimulus strength drives the behavior (Beckers & Schul, 2004), rather than a paired preference task, in which the response is dependent on the opposing stimulus presented (Phelps et al., 2006).

These findings contrast with those of prior studies (West & King, 1988b, White, personal observations) that suggested female preference was based on a comparative ranking of the songs that were presented. In our experiments, females do respond exclusively to the highest potency songs in later blocks, but this appears to be the result of an overall decreased responsiveness to all songs presented to the female, including the highest potency songs. These findings can be explained by a simple threshold model where birds show an overall increase in the response threshold which consequently excludes responses to low potency songs while also decreasing the response probability to high potency songs without changing the relative signal drive.

The rapid response onset as well as the stereotyped nature of the posture indicate a tight, perhaps even reflexive, link between the song stimulus and CSD. These features, however, do not rule out the possibility of top-down control. In rodents, the copulatory posture lordosis is also driven reflexively by well-characterized circuits in the brainstem, but this response, in both mammals and birds, is modulated in a top-down manner by higher-order forebrain centers (Ball & Balthazart, 2020; Maguire et al., 2013; Perkes et al., 2019; Pfaff et al., 2008). It is certainly possible that in richer social contexts females do update their song weightings over short timescales with respect to male exposure–it is clear that female song preferences are at least partially learned (Riebel, 2003, 2009), and that these preferences remain plastic into adulthood (King, West, et al., 2003; Nagle & Kreutzer, 1997)–however as our model shows, this type of song learning is not necessary for flexible behavior. Song selectivity can emerge, even where preferences are static, merely by raising the response threshold. If the response threshold is also subject to feedback from song exposure (as shown by (Remage-Healey et al., 2012)), threshold modulation could provide a simple strategy for responding only to the highest quality songs. Each time a female is exposed to high potency song, it could raise the response threshold, increasingly the probability that lower potency songs are excluded. This mechanism produces a best-of-n ranking, without requiring the female to remember and compare every male observed. Threshold-based regulation of selectivity has been proposed previously in túngara frogs (Lynch et al., 2006), and may be a consistent mechanism of behavioral modulation, allowing innate responses to be context specific without the need for cognitive control.

Signal potency shapes latency, duration, and overall response vigor of CSD response.

CSD is an innate and highly stereotyped behavior more easily evoked after periods of sensory deprivation (King & West, 1977; Searcy & Marler, 1981). Previous literature referred to these types of behaviors as fixed action patterns (Lorenz, 1937; Schleidt, 1974). Indeed, our observations that females initiated CSD even when caught in non-upright positions (e.g., hanging from the wall or ceiling, Supplemental Videos 3 and 4) are consistent with this idea. However, our results show that CSD is a graded behavior with variation in both the timing and intensity of the display. Notably, females sometimes produced limited versions of the behavior, referred to as partial-CSDs, which occurred in 7% of trials (16% of total CSD) and could be statistically distinguished from full-CSD based on latency and tail height. A similar pattern has been observed during lordosis in rodents where females sometimes produce "marginal lordosis," with the decreased vigor of the response being qualitatively linked to weaker stimulation (Hardy & DeBold, 1972). These observations suggest that copulatory behaviors are not truly fixed, and can be shaped by both stimulus and context. These findings are consistent with a recent reinterpretation of escape behaviors, where these previously assumed highly stereotyped behaviors show quantitative differences in response characteristics based not only on context but also stimulus type and strength (Wei et al., 2021). Taken together, this supports the emerging view that survival behaviors are more malleable than previously thought, thereby allowing the organism to tailor its innate responses based on context to achieve the optimal survival outcome.

While prior studies have shown that signal potency affects copulatory behavior, to our knowledge, no prior work has been able to quantify the timing and trajectory of a copulatory response to a single stimulus with the precision afforded in our study. As with birdsong (T. Gardner et al., 2001; Long et al., 2010), the quantification of the postural aspects of reproductive behavior makes it possible to link neural activity to behavioral output, as well as identify fundamental aspects of behavioral variability. In the current study, the ability to precisely measure CSD while presenting females with a range of different stimuli offered a unique opportunity to probe the interaction of signal potency and behavioral state. Overall, potent songs drove CSDs that were both shorter latency and longer duration (see Figure 7d). Potent songs also produced more full-CSDs and fewer partial-CSDs (compared to lower potency songs), although as the response threshold increased, higher potency songs produced more partial-CSDs, and lower potency songs failed to elicit any response. This again highlights the link between signal strength and the nature of copulatory behavior, and future experiments can identify neural mechanisms that mediate this relationship.

A model for achieving sexual autonomy by modulating response threshold

In addition to signal potency, the internal state of the individual is critical in determining response characteristics for many innate survival behaviors (Evans et al., 2018; Kim et al., 2017; Wei et al., 2021). While song potency clearly had a strong effect on CSD response probability and timing, we found that the prior response rate of the female, reflecting the bird's underlying behavioral state, also predicted the probability of producing CSD to a given trial. In fact, the response rate to adjacent trials was nearly as strong a predictor of CSD as the song type presented, and like potency, this underlying response threshold was a strong independent

608 predictor of CSD latency and duration, as well as the velocity of motion when producing the
609 CSD.
610 For many behaviors, but especially reproductive behaviors, underlying responsiveness is
611 strongly affected by hormonal state. For lordosis, for example, the intensity and duration of the
612 lordosis posture increases with higher estrogen levels (Hardy & DeBold, 1971) and decreases
613 with higher levels of corticosterone (deCatanzaro et al., 1981). Prior studies also make it clear
614 that the seasonal transition to producing CSD is dependent on estrogen, (D. Maney & Pinaud,
615 2011; Donna L. Maney et al., 2006) however there is conflicting evidence as to whether the
616 level of estrogen alters CSD once females are responsive to song (Leboucher et al., 1998;
617 Moore, 1983; Searcy & Capp, 1997). In our study, we observed fluctuations in response rates
618 over times scales that were short and often unpredictable. For example, some females
619 produced CSD to every song presented in a day, while other females (or the same female the
620 following day) did not produce a single CSD, even though the same songs were presented. We
621 did not measure hormone levels in our study, and so cannot conclude which hormones (if any)
622 were responsible for this variation. Regardless of the mechanism, these observations suggest
623 that the underlying behavioral state of the female is necessary, and in some cases sufficient, to
624 evoke a copulatory response regardless of song potency. These findings may explain in part
625 why male signal quality frequently fails to predict reproductive success (King, White, et al.,
626 2003; Takahashi et al., 2008; White et al., 2010) and at best explains only a portion of the
627 variance (Andersson, 1994).
628 We have conceptualized this process as a response threshold that signal strength must
629 exceed to evoke a response, and this model has important implications for our understanding
630 of courtship behaviors. If the behavioral response threshold is highly variable, or even directly
631 malleable, male signal quality will be a poor predictor of reproductive success, as the signal
632 itself is less important than how and when the signals is used. Although studies of sexual
633 selection often focus on variation in the absolute value of male signals, evidence from a variety
634 of species suggests that courtship skill–i.e., the effective, even strategic, use of signals–is critical
635 to maximizing reproductive output (Baird et al., 2007; Mennill et al., 2002; Vinnedge & Verrell,
636 1998). By incorporating a variable response threshold into our models of courtship we can
637 better understand the selective pressures on male behavior.
638 Female behavior can also be better understood by considering a variable threshold for
639 stimulus evoked responses. Females of various species produce behaviors that apparently serve
640 to affect their own responsiveness. For example, female ringdoves self-stimulate using cooing
641 vocalizations, effectively lowering their response threshold to male sexual stimulus and
642 promoting copulation (Cheng, 1992). Female cowbirds on the other hand appear to use chatter
643 vocalizations to obstruct song and prevent CSD (Anderson et al, 2021) by reducing the strength
644 of the stimulus, rather than changing the response threshold. In both cases, females alter their
645 own sensitivity via active behavioral responses, providing a level of behavioral control on an
646 otherwise autonomic, involuntary response.
647 While the neural pathways that mediate these copulatory responses are still being
648 discovered and many questions remain (Ball & Balthazart, 2020), our findings here help to
649 contextualize prior work within the framework of a variable response threshold. Multiple
650 studies have suggested the auditory-motor forebrain region HVC is necessary for song
651 selectivity (Brenowitz, 1991; Del Negro et al., 1998; Maguire et al., 2013), showing that lesions

652 to HVC result in a loss of CSD selectivity. Based on this, we had previously inferred that HVC
653 must act immediately in response to song stimulus, inhibiting CSD to low potency songs and
654 releasing inhibition on higher potency songs (Perkes et al., 2019), however the loss of selectivity
655 observed following HVC lesions could also be explained by a decreased response threshold.
656 Rather than directly inhibiting the CSD response, HVC could modulate the effect of acoustic
657 feedback on underlying responsiveness. Experiments directly testing these hypotheses are
658 necessary to understand the role and importance of variable response thresholds in social
659 behavior, and these experiments are made possible by the approaches established here. By
660 establishing high-throughput, quantitative representations of behavior, we can take advantage
661 of the potential of CSD as a window into the mechanisms and meaning of sexual signals and
662 copulatory responses.
663

664 # Methods

665

666 **Subjects & Housing**

667 A total of 85 female cowbirds were used in these experiments. All birds were wild-
668 caught near Philadelphia (in Chester or Montgomery Counties) under appropriate state and
669 federal trapping permits. Birds were captured from 3 different sites spaced 15-30 miles apart,
670 between the years of 2007 and 2018. Outside of playback experiments, birds were housed
671 either in small flocks in aviaries (at least 1.2x1.2x2.4m)) or co-housed in smaller wire cages
672 (56cmx56cmx46cm). All housing was supplied with perches and birds had access to food (a
673 modified Bronx zoo diet for omnivorous birds mixed with seed) and water, *ad libitum*. Bird
674 subjects can be divided into three distinct groups: traditional playbacks, cage-rack playbacks,
675 and posture-analysis playbacks. In each case, the institutional Animal Care and Use Committee
676 (IACUC) at the University of Pennsylvania inspected and approved their use for their respective
677 studies.

678 **Cohoused playbacks:** 65 adult females, captured from Montgomery county between
679 the years 2007-2010 were housed in large (18.3x6.1x4m) outdoor aviaries. Aviaries contained
680 live vegetation, feeding shelters, and perches. Beginning in March, females were separated
681 from males. During playback experiments from May to June, birds were housed in pairs in
682 sound attenuation chambers (1 x 1 x 1 m). Cohousing has been shown to reduce stress and has
683 been shown to have no influence on playback responses (Eastzer et al., 1985; King & West,
684 1983b; Smith et al., 2000). After playbacks, females were returned to large aviaries.
685

686 **Cage Racks:** 8 adult females were captured in Chester County in Fall 2018 and housed in
687 an indoor aviary (measuring 2.4x1.5x2.4m), along with 9 males and 10 additional females (used
688 for a separate study). Programmed light timers (Intermatic, Model: ET70415CR) allowed us to
689 maintain the birds on natural, local light schedule throughout the year. In March, the males
690 were moved to a different room, and all 18 females were individually housed in two 9-cage
691 wire racks [a 3x3 rack of cages, each measuring 56 x 56 x 46 cm], where they remained during
692 playback experiments in May and June (described below).
693

Posture Analysis playbacks: 12 adult females, also captured from Chester County, 5 in the fall of 2017, and 7 more in fall of 2018, were housed in pairs in our vivarium (in the same 9-cage racks) from December to March, to ensure they were habituated to enclosed housing. These birds were maintained on the same natural light conditions as those tested in wire cages above. Females were never co-housed with male cowbirds in their individual cages, but males were always present in adjacent cages. In March, prior to breeding season, the males were moved to a separate room to prevent song habituation (see (West & King, 1985)). During playbacks, females were moved to plexiglass multi-camera cages (55x55x48cm) contained within larger sound attenuation chambers (76x76x76cm), set to long days (14 light:10 dark), similar to their natural lighting conditions. Following playbacks, birds were returned to wire cages.

Playback Procedures

The procedure for song playbacks to natural birds has been described previously (Eastzer et al., 1985; White et al., 2010). Females, overwintered in captivity, and isolated from males during breeding season consistently produce CSD in response to song playback, and the number of CSDs provides a potency metric for each song. Song potency is defined as the mean proportion of CSDs produced during playbacks, averaging across birds and song trials. In these experiments, females heard the playback of a single song every 90 minutes, for a total of 6-8 songs throughout the day. Speaker parameters and distance varied slightly between experimental groups, but sound intensity at the bird was maintained between 80 and 90 dB (measured via sound meter at the bird location). Songs were arranged in pseudorandom blocks, such that females heard every song in the set before repeating. Females were always naïve to the presented songs prior to playback. Playback procedures take 2-4 weeks per bird (depending on the number of songs and repetition blocks). We excluded birds that produced fewer than 5 postures or 5% of playbacks (whichever was fewer) from our analysis. Except where noted, all experiments used a single song set, recorded from aviaries in May 2008 (Maguire et al., 2013). In all cases, songs were recorded by placing 4-6 Sennheiser RF condenser microphones (Sennheiser Electronic Corporation, 1 Enterprise Drive, Old Lyme, CT, USA) in locations within the aviaries to maximize our chances of obtaining recordings from males vocalizing within 0.3 m on axis of the microphones. We recorded to a Marantz PMD670 solid state digital recorder using uncompressed WAV at 48KHz. For inclusion, the unweighted signal-to-noise ratio measured peak-to-peak had to be at least 53 ± 2 dB measured between 300 and 12 000 Hz. Specific playbacks procedures are as follows: **Traditional playbacks**: Sets of birds (in 2007 and from 2009-2011) were housed in pairs in 1m x 1m x 1m sound isolation boxes. Behavior was captured using a CCTV camera. Song playback was initiated manually, via a computer connected to an LG XCSS amplifier on Bose 161 speakers located in each sound-attenuating chamber. Presentation was initiated once birds were visible in the camera and perched. In all but two groups, we used a set of songs recorded from aviaries in Montgomery County, PA in 2008 (Maguire et al., 2013). In 2007 we used a different set of songs, also recorded in aviaries from local males (Ronald et al., 2017). In 2009 we used songs recorded under controlled conditions as part of a separate experiment (Gersick & White, 2018). **High through-put cage racks:** We developed a new approach to increase the overall numbers of playback experiments possible in a smaller space (Supplemental Video 7). In 2019, females were individually housed

in standard cage racks described above, inside an animal housing room. Females could see, but not physically interact with, adjacent birds. Two speakers (PreSonus E3.5-3.5" 2-way near field studio monitors, PreSonus audio Electronics, Inc., Baton Rouge, LA, U.S.A) mounted at the far corners of the room, 2m away from the cages. Acoustic foam was placed on walls and behind the cages to reduce echo. Behavior was captured using 4 webcams, such that each cage was visible in at least 2 webcams. A custom python script automated the playback and video capture. The script tracked bird motion (by computing the cumulative pixel difference between one frame to the next and attempted to wait until total motion was low, but would play a song regardless after 1 minute. The code is available on GitHub (github.com/aperkes/SmartBox).

Posture Analysis: In 2018 and 2019, birds were individually housed in a 56x56x48 cm plexiglass cage, stationed inside a larger wooden sound isolation chamber (76x76x76 cm, adapted from instructions from accousticsolutions.com). We used 3d printed perches bolted onto threaded metal rods to encourage the female to remain perched in the center of the cage. Sound was played over computer speakers (Dell Rev A00 2.0), measured at 80 dB at 10cm distance. Recording and playback were automated using a custom python script. We estimated the position of the bird in real time using a color discriminator to detect the bird, and voxel carving to estimate the position. By tracking the bird, we were able to wait to play the song until the bird was resting on the perch for 10s. All code for recording video is available upon request.

Behavioral Observations

In all experiments, video recordings were scored by various trained observers from the Schmidt or White labs. We noted CSD (including full or partial, described below), and also recorded duration, expect in the case of the cage-rack playbacks, in which video resolution and frame rate precluded measuring duration with confidence. During the posture-analysis experiments, human observers also manually identified a range of female responses to male song playbacks. The behaviors scored were defined as follows: **CSD** (also called "Full-CSD") characterized by arching of the back to elevate the head and tail while spreading the wings. To qualify as CSD, the tail needed to be raised above the body. CSD was defined as **"partial-CSD"** if the female initiated the characteristic wing spread and/or back arching, but did not elevate the tail above the body; **wingstroke:** characterized by rapid, unilateral motion of the wing (West & King 1988), which was not produced within 1 second of flying or hopping motion; **feather depression:** smoothing of the feathers causing a visible reduction in feather volume; **startle:** sudden body shutter, often accompanied by crouching as if readying for motion; **gaze shift:** defined by rapid head motion (at least 3 head movements in a 1 second window). For CSD, partial-CSD, wingstroke, and feather depression the latency (in ms) was defined as the first frame in which the behavior was visible. For CSD, duration was calculated as the time (in seconds) between initiation and the moment the tail sank below the body after reaching the peak (see Figure 2a). Observers in posture-analysis experiments scored behaviors without audio or song information to prevent song or song onset bias.

Camera Setup

Multi-view camera cages were equipped with four FLIR Grasshopper color cameras (FLIR Systems, Inc) able to capture behavior with a minimum of 1024x1024 resolution and less than 6ms shutter time operating in global shutter mode (Figure 9a). The focal distance was set to

the center of the cage and the adjustable aperture of the lenses were set to yield an acceptable compromise between depth of field and achievable shutter time. The shutter times were synchronized via hardware trigger, with one master camera triggering the remaining 3 cameras. Running at an acquisition rate of 50 frames per second (fps) we did not observe frames being dropped.

We captured the audio in the cage (using a Logitech webcam), and also captured the copied audio output (by splitting the audio-output and plugging it into the microphone port). The audio and video were stored with their arrival timestamps in a single data object (using the Robot Operating System, ROS (Quigley et al., 2009), and the offsets of these timestamps were carefully calibrated using a percussive reference sound (tapping a metal rod on the perch). This prevented error based on variable delay in the audio and video sampling. Using the stamped audio information, we identified the precise onset of song by cross-correlating the song recording with the audio file.

Precise calibration is required to triangulate objects across cameras. We performed intrinsic calibration (defining the intrinsic parameters of the camera) once using a printed checkerboard, using the radial-tangential distortion model with the ROS calibration toolbox [https://github.com/ros-perception/image_pipeline.git]. The extrinsic calibration (i.e. the position and orientation of the cameras) was subject to significant variation due to thermal expansion of the acrylic walls and mechanical perturbations whenever the cage was opened. For this reason, we performed full extrinsic calibration from the first image of every recording. AprilTag markers (Olson, 2011) were placed on the four walls of the cage. Their positions were carefully measured to serve as an initialization prior to a full pose graph optimization using TagSLAM software [github.com/berndpfrommer/tagslam]. All code for calibration and data gathering is available upon request.

Computer Vision Quantification

To convert the time-aligned, calibrated, multi-view recordings into quantified pose we followed a 3-step process, summarized in figure 9 as bird detection, keypoint identification, and triangulation. Bird detection: each frame passes through a mask detector, which identifies pixels that belong to a bird in an image. To perform mask detection, we used an existing deep neural network (maskrcnn_resnet50 (He et al, 2017), included in the pytorch package (https://pytorch.org/) in python. The model was pretrained from the COCO dataset, a large database of images (Lin et al., 2014), and performed reasonably well on bird detection but often failed when birds adopted CSD. To improve performance, we finetuned the model on a custom dataset we created of 276 masks of birds at various moment of CSD inside the posture-analysis cage and labeled using the website supervise.ly. We used an 80:20 split for training. The mask detector returns a bounding box and a mask for each frame (Figure 9b). Keypoint identification: Cropped images of the bounding box (padded to be 1.2x bigger to more reliably capture the tail) were input to a second neural network trained to identify bird keypoints. Here we also used an existing model designed for human pose recognition (Pose resnet by Leo Xiao https://github.com/leoxiaobin/deep-high-resolution-net.pytorch). This model had also been pretrained on the COCO dataset. We first retrained the output layer of the model on the large CUB dataset (Welinder et al., 2010). Since the CUB dataset lacked the tail tip, we inferred the location of the tail tip by finding the line segment that began at the back keypoint, passed

through the tail keypoint, and ended at the bounding box. We then finetuned the model on our own custom dataset of the 276 images with labeled keypoints. This approach generated accurate keypoint identification that was robust to the unusual positions adopted during copulatory displays (Figure 4c,Supplemental Video 8). Triangulation: With keypoints detected in each frame, we perform triangulation to estimate the 3d position of each point. In short, we found each pairwise triangulation point and select the location that is optimized for low reprojection error and high keypoint confidence. In parallel, we used the silhouettes from the mask detection to generate a 3d point cloud. Taken together, this provided a 3d volume surrounding a simple skeleton (Figure 9d, supp v1). All image processing code and the models used are available on GitHub (github.com/aperkes/keypoint_detection_public).

Computer Vision Analysis

To produce a simple representation of pose during CSD, we used the height of the tail tip keypoint. This proved intuitive and highly tractable. From this trajectory we identified a number of inflection points within the course of CSD (See Figure 7a). To reduce the effect of noise from mis-identified frames, all keypoint trajectories were smoothed using a 1d gaussian kernel with alpha=5. Baseline: we calculate the mean baseline for the tail height in the 1s prior to song onset; latency: defined by the first moment of deviation (> 5 std) from baseline; max velocity: we use a 2s window following onset and find the point of maximum velocity; stabilization: defined as the first time, following the max velocity, with a non-positive velocity. In other words, it was the first time the tail stopped rising; stable height: we recorded the height at the point of stabilization; refraction: first we calculated the distance between stable height and baseline, and defined refraction as the first point, following peak height, that the tail dropped below 66% of the peak height. The time between max-velocity and refraction provided an automated metric of duration, although for the analysis we opted to use our hand labeled duration (as well as latency) scores in order to maximize accuracy.

Binary Classifiers

In order to assess what contextual information is encoded in CSD, we trained classifiers using information from all keypoints. We used the random forest classifier from the scikit-learn package in python. The training algorithm attempts to weight the categories equally to avoid always selecting the most common label (for example in the case of song potency, potent songs produce more postures, so are over represented in the dataset). Here, the pose was represented at each time point by a 91-dimensional vector of all pairwise distances. This provided a useful representation which is robust to noise from a single keypoint and invariant to position and rotation. We used 2 different approaches to assess information content. We first trained a classifier on the entire posture (in a 4s window following song onset). We then trained a classifier at each time point, providing a trajectory of classification accuracy over time. In each case, we randomly sampled 80% of the data for a training set and used the remainder to score accuracy, bootstrapping 10 times to get a distribution of performance. To assess classifier performance, we calculated chance performance by training the same model on shuffled labels. This allowed us to test whether the classifier was significantly better than what would be obtained by guessing based on any remaining label count bias. We optimized our parameters to maximize our CSD vs non-CSD accuracy, then used the same parameters to

compare specific contexts (i.e., song potency and response rate). All postural analysis code is available on GitHub (github.com/aperkes/PostureAnalysis).

Pose Analysis

To test for the existence of distinct gestures within posture, we first calculated the velocity of each keypoint in each dimension, calculated as $V_t=x_t - x_{t-1}$, then took the normal of the 3 dimensions (sqrt(vx^2 + vy^2 + vz^2)). This provides a positive value giving the magnitude of the velocity for any given time point (regardless of the direction). We calculated the mean velocity at baseline within a 1 second window prior to song onset, accounting for both background motion and computer noise. We then identified spikes in the velocity, periods where the velocity increases beyond 3 standard deviations of baseline (Supplemental Figure 3a). We treated the first moment above baseline as the onset of a distinct event, from which we could calculate the distribution of motion for individual parts and transition probabilities. For keypoints which were rigidly attached (e.g. the beak, eyes, and crown) we averaged the onset time to treat them as a single part (e.g. the head). This helped to reduce the parameter space when evaluating transition probabilities. Since there were occasional ties (we performed these analyses at a sampling resolution of 20 Hz), we added a random jitter to break ties, repeating the analysis 50 times and taking the median result to compensate for the randomness.

To produce a lower dimensional representation of the full pose information we performed principal component analysis on the 91-dimensional vector described above (Supp Figure 3c). To limit the effect of individual bird proportions driving variability, we normalized the pairwise distances, scaling them by the mean across all birds (i.e. for each bird, each pairwise measurement was multiplied by the across bird mean divided by the within bird mean). We selected the first 11 dimensions, which accounted for 88% of the variance (we chose 11 as it was the point after which each additional principal component accounted for less than 1% of the total variation). The first two principal components accounted for 60% of the variation and provided visually intuitive representations of the pose trajectory of CSD (Figure S3c). It also provided a way to measure density within this pose space (Figure S3e). We used a gaussian kernel to estimate density across the pose space. To capture the average trajectory density, we performed 1000 iterations and calculated the density at each time point at a randomly sampled trajectory.

Statistics

To properly account for the non-independence of song and bird, we used linear mixed models with song potency and responsiveness as fixed effects, and bird, song, block, and aviary as random intercepts, and other similar models, which are described in the results. All analyses were performed in python 3.7 using standard available analysis packages. Most statistical tests were performed using SciPy (Virtanen et al., 2020) or statsmodels (Seabold & Perktold, 2010) , except mixed regression, which was performed using Pymer4.0 (Jolly, 2018). All plots were generated using matplotlib . Regression plots show the estimated effect and 95% confidence intervals based on linear mixed effects models. “CI” in the text describes the 95% confidence interval. Description statistics are written as mean +/- standard deviation unless otherwise noted. The full output of all statistical tests is available in Supplemental Statistics.

Data Availability

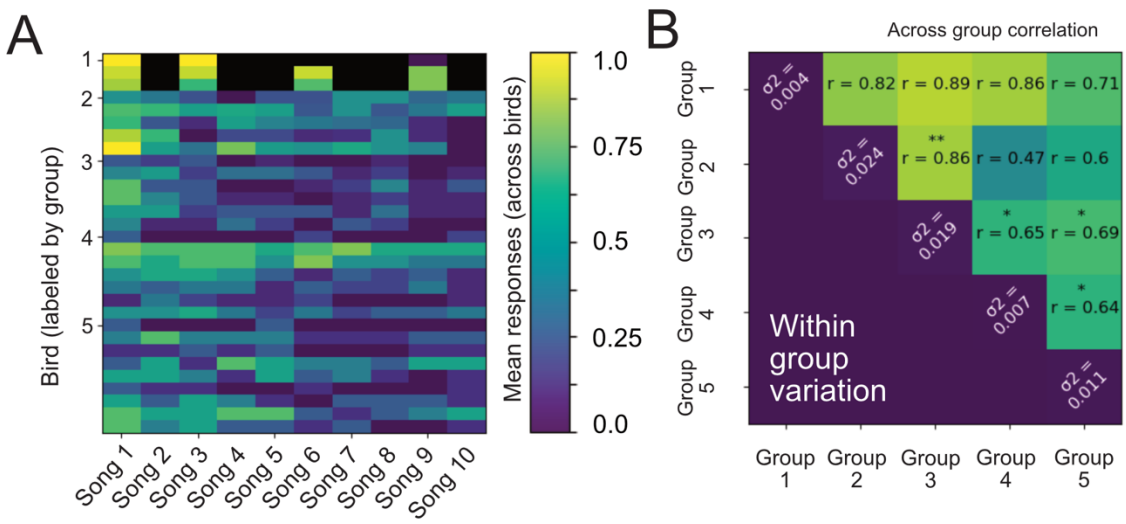
Example videos, along with all meta information, human annotations, and reconstructed postural trajectories are archived on Dryad. All analysis code is available on Dryad, as well as in an interactive notebook on GitHub (github.com/aperkes/PostureAnalysis). The full dataset of videos is available by request.

Acknowledgments: We wish to acknowledge Nikos Kolotouros for his key assistance in implementing the computer vision approach, Jakub Jarmula for tirelessly constructing an experimental apparatus, H. Luke Anderson, Chloe Perkes, and Marc Badger for their invaluable comments on the approach and manuscript, and the cooperative cowbirds who made this work possible. 3D printed perches were courtesy of the University of Pennsylvania Libraries' Biomedical Library.

Competing Interests: No competing interests declared

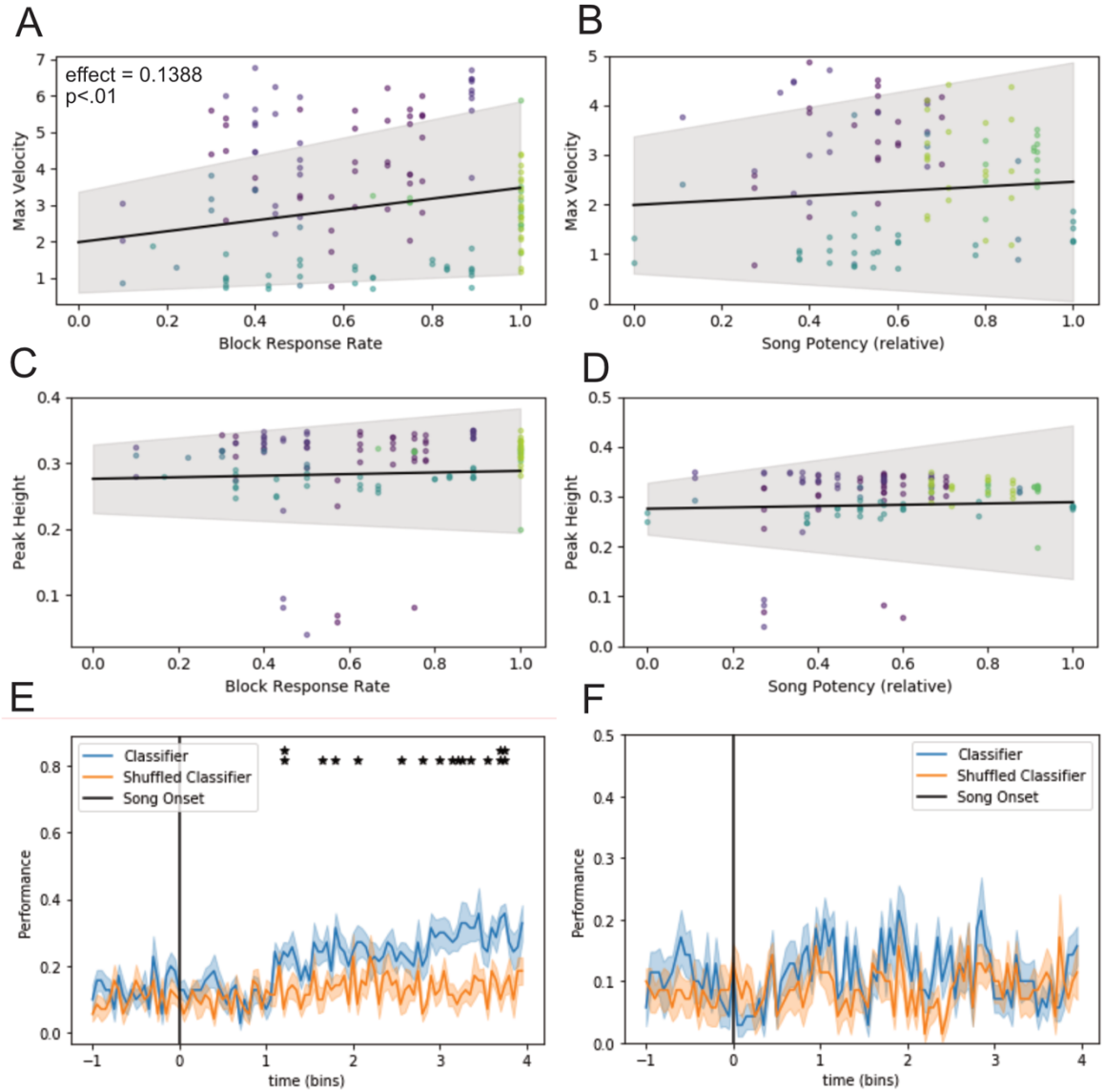
Supplemental Figures

Figure S1



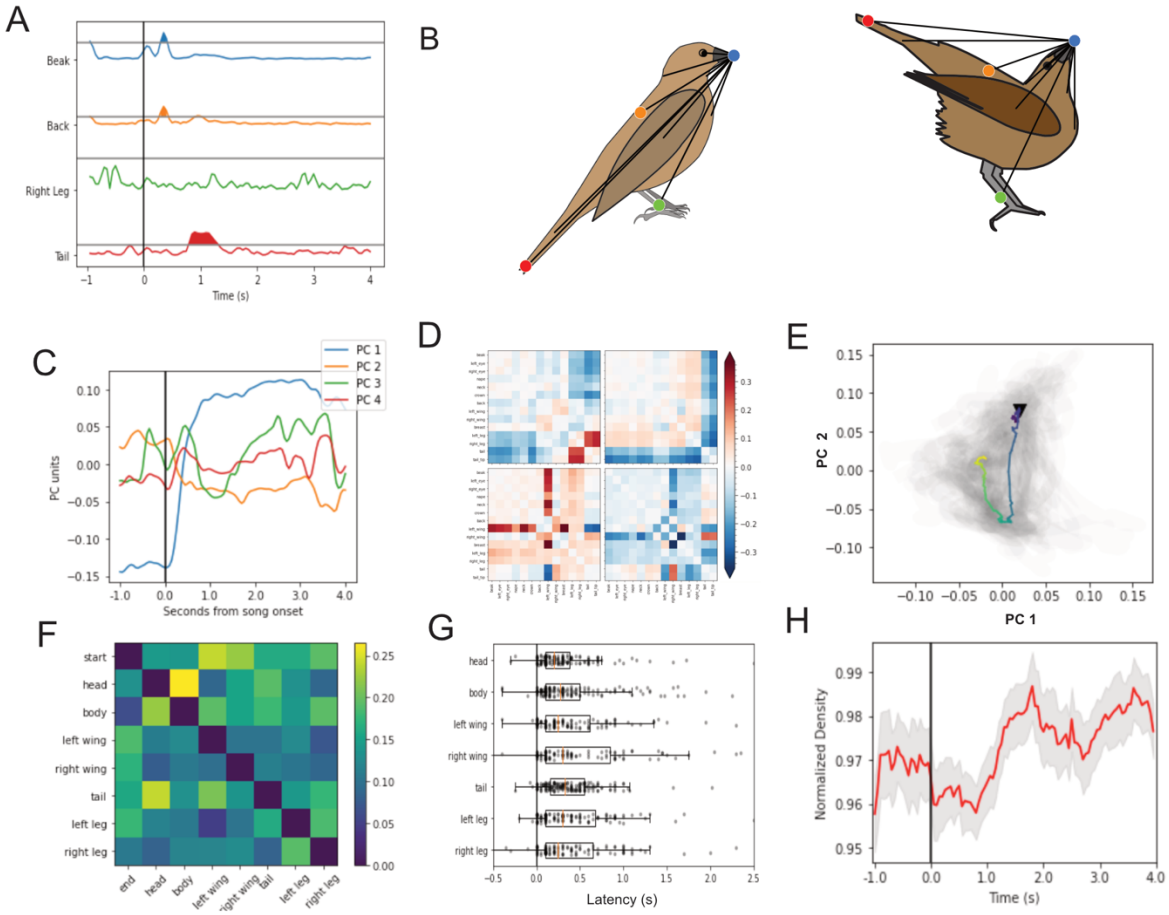
Females display consistent bias toward songs. Multiple females hear the same sets of songs in a given experiment and we used the same set of songs across multiple years of experiments. **A.** The response rate of 30 females, sorted by groups, to the same set of songs, sorted by potency. The first 3 birds were only exposed to 4 of the songs. **B**. Pairwise correlations between group potency scores for the same 5 groups of birds. Asterisks mark correlations which were significantly at * = p<0.05, ** = p<0.01. The diagonal shows the within-group variance, representing the level of consistency of bias within groups.

Figure S2



Posture shape is a function of underlying behavioral state, but there is no significant effect from song potency. A-D. scatterplots show max velocity and peak height as a function of block response rate and song potency. In each plot, every point represents a posture, and the colors depict individual birds (colors match colors from Figure 7). Max velocity was the peak velocity recorded as the tail lifted following song onset, while peak height is simply the maximum tail height during CSD. Both metrics were quantified using our computer vision-based quantification described in Figure 4. To account for pseudoreplication, we used linear mixed models with experimental group, bird, and block as random affects, and max velocity and peak height as fixed effects. Black lines show the estimated effect with the 95% confidence interval in gray. We found a significant effect only for A, the effect of block response rate on max velocity. indicated a positive effect. **E.** A random forest classifier was trained to predict response threshold from a single frame the quantification of posture. The classifier was trained using 80% of recorded trials (n_train=138) and tested on the remainder (n_test=34). We retrained the classifier 10 times, randomly splitting the train & test sets, in order to generate bootstrapped distributions. The blue line shows classifier performance, with (shading showing +- 1 SEM). To generate "chance" performance, we performed the same procedure but with a shuffled test set (orange bar). This ensured that our classifier was performing better than would be predicted by just guessing the mean responses. Asterisks denote time bins in which the classifier performed significantly better than chance (i.e., significantly higher than shuffled performance, T-test, df=18). *: p<.05, **: p<.01. **F.** As in E, this depicts a random forest classifier trained to predict song potency based on a CSD. This classifier did not perform significantly better than chance.

Figure S3



Computational approaches for describing CSD show no evidence of precise sequencing. To characterize posture, we tracked the trajectory of individual keypoints, as well as calculating the onset latency for each point. **A.** An exemplar posture showing the calculation of onset latency for 4 different keypoints. Latency was defined as the moment the keypoint velocity increased 3 standard deviations above the mean baseline (defined by the 1 second prior to song presentation). **B.** In order remove variation based on the location of the bird and have a representation that was robust to individual variations in keypoints, we described posture by the pairwise-distances between keypoints, shown in this cartoon. Here we show only the distances from the beak, but all pairwise distances are used to make 91-dimensional vector. **C.** We then performed a principal component analysis to observe correlations between the keypoints and have a lower dimensional representation of posture. Here we plot the first 4 principal components. **D.** Shows the weights for the first four principal components. These principal components define a new trajectory space. **E.** All postures for a single exemplar bird are plotted as the first principal components. This PCA was performed on the postures of this bird. The colored line shows a single posture progressing from t=-1 (purple) to t=4 (yellow). The black shows the onset of song. Given these representations of pose, we can test for CSD sequencing in multiple ways. **F.** Shows the transition probabilities from one point to another. None are significantly different from chance (chi-squared test). **G**. Shows the distribution of onsets for each body part (based on the measurements defined in A. For G and F, multiple, rigidly connected keypoints are averaged to a single body part (e.g., the head is composed of the beak, and eyes). Here, we observed that there was significant variation in the timing based on parts (ANOVA, F =3.08, p<0.005) with the head starting slightly earlier on average than other parts (Tukey-HSD head vs left wing and head vs tail significant at p<.05). This could be explained somewhat by a startle response occurring prior to CSD onset. **H.** In order to test for actemes, i.e., sub-gestures which compose CSD, we plotted the density as a function of time for the postural space defined in E. Actemes would manifest as periods of high density, divided by more variable, lower density periods. We detected one wide peak around the time the bird reaches the characteristic CSD pose, and a second, lower peak as the bird returns to its neutral pose. We produced a bootstrapped distribution by randomly selection from posture trajectories and iteration (n=1000 times). The Red line shows the mean density for a representative bird, and gray shading shows the standard deviation. This only suggests one stereotyped aspect of CSD, the full posture, while the transitions to and from CSD appear variable.

References

Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. In *Neuron*. https://doi.org/10.1016/j.neuron.2014.09.005

Anderson, H. L., Perkes, A., Gottfried, J. S., Davies, H. B., White, D. J., & Schmidt, M. F. (2021). Female signal jamming in a socially monogamous brood parasite. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2020.10.011

Andersson, M. (1994). Sexual Selection: In *Sexual Selection*. Princeton University Press. https://doi.org/10.2307/j.ctvs32s1x.22

Baird, T. A., Hranitz, J. M., Timanus, D. K., & Schwartz, A. M. (2007). Behavioral attributes influence annual mating success more than morphological traits in male collared lizards. *Behavioral Ecology*. https://doi.org/10.1093/beheco/arm092

Baker, M. C., & Boylan, J. T. (1995). A catalog of song syllables of indigo and lazuli buntings. *Condor*. https://doi.org/10.2307/1369541

Ball, G. F., & Balthazart, J. (2020). Sex differences and similarities in the neural circuit regulating song and other reproductive behaviors in songbirds. In *Neuroscience and Biobehavioral Reviews* (Vol. 118). https://doi.org/10.1016/j.neubiorev.2020.07.026

Beckers, O. M., & Schul, J. (2004). Phonotaxis in Hyla versicolor (Anura, Hylidae): The effect of absolute call amplitude. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*. https://doi.org/10.1007/s00359-004-0542-3

Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of the Royal Society Interface*. https://doi.org/10.1098/rsif.2014.0672

Blaxter, J. H. S., & Hoss, D. E. (1981). Startle response in herring: The effect of sound stimulus frequency, size of fish and selective interference with the acoustico-lateralis system. *Journal of the Marine Biological Association of the United Kingdom*. https://doi.org/10.1017/S0025315400023018

Brenowitz, E. A. (1991). Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science*. https://doi.org/10.1126/science.1987645

Carouso-Peck, S., & Goldstein, M. H. (2019). Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Current Biology*. https://doi.org/10.1016/j.cub.2018.12.026

Catchpole, C. K., Leisler, B., & Dittami, J. (1986). Sexual Differences in the Responses of Captive Great Reed Warblers (Acrocephalus arundinaceus) to Variation in Song Structure and Repertoire Size. *Ethology*. https://doi.org/10.1111/j.1439-0310.1986.tb01000.x

Catchpole, C. K., & Slater, P. J. (2008). Bird song: biological themes and variations. *Cambirdge University Press*. https://doi.org/https://doi.org/10.1017/CBO9780511754791

Cattell, J. M. (1886). The influence of the intensity of the stimulus on the length of the reaction time. *Brain*. https://doi.org/10.1093/brain/8.4.512

Cheng, M. F. (1992). For whom does the female dove coo? A case for the role of vocal self-stimulation. *Animal Behaviour*. https://doi.org/10.1016/S0003-3472(06)80016-3

deCatanzaro, D., Knipping, R. P., & Gorzalka, B. B. (1981). Antagonism of estrogen-induced lordosis by corticosterone in adrenalectomized-ovariectomized female rats and mice. *Pharmacology, Biochemistry and Behavior*. https://doi.org/10.1016/0091-3057(81)90019-8

Del Negro, C., Gahr, M., Leboucher, G., & Kreutzer, M. (1998). The selectivity of sexual responses to song displays: Effects of partial chemical lesion of the HVC in female canaries. *Behavioural Brain Research*, *96*(1–2). https://doi.org/10.1016/S0166-4328(98)00009-6

Dill, L. M. (1974). The escape response of the zebra danio (Brachydanio rerio) I. The stimulus for escape. *Animal Behaviour*, *22*(3). https://doi.org/10.1016/S0003-3472(74)80022-9

Eastzer, D. H., King, A. P., & West, M. J. (1985). Patterns of courtship between cowbird subspecies: evidence for positive assortment. *Animal Behaviour*. https://doi.org/10.1016/S0003-3472(85)80117-2

Eaton, R. (1984). Neural Mechanisms of Startle Behavior. In *Neural Mechanisms of Startle Behavior*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4899-2286-1

Eaton, R. C., Lee, R. K. K., & Foreman, M. B. (2001). The Mauthner cell and other identified neurons of the brainstem escape network of fish. In *Progress in Neurobiology*. https://doi.org/10.1016/S0301-0082(00)00047-2

Evans, D. A., Stempel, A. V., Vale, R., & Branco, T. (2019). Cognitive Control of Escape Behaviour. In *Trends in Cognitive Sciences* (Vol. 23, Issue 4). https://doi.org/10.1016/j.tics.2019.01.012

Evans, D. A., Stempel, A. V., Vale, R., Ruehle, S., Lefler, Y., & Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. *Nature*, *558*(7711). https://doi.org/10.1038/s41586-018-0244-6

Fernández, G. J., & Mermoz, M. E. (2003). Group copulation solicitation display among female Greater Rheas. *Wilson Bulletin*. https://doi.org/10.1676/03-060

Gardner, T., Cecchi, G., Magnasco, M., Laje, R., & Mindlin, G. B. (2001). Simple motor gestures for birdsongs. *Physical Review Letters*, *87*(20). https://doi.org/10.1103/PhysRevLett.87.208101

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and rules: The acquisition and reprogramming of a bird's learned song. *Science*. https://doi.org/10.1126/science.1108214

Gersick, A. S., & White, D. J. (2018). Male cowbirds vary the attractiveness of courtship songs with changes in the social context. *Behaviour*. https://doi.org/10.1163/1568539X-00003475

Glaze, C. M., & Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3387-05.2006

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. In *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Hardy, D. F., & DeBold, J. F. (1971). The relationship between levels of exogenous hormones and the display of lordosis by the female rat. *Hormones and Behavior*. https://doi.org/10.1016/0018-506X(71)90003-1

Hardy, D. F., & DeBold, J. F. (1972). Effects of coital stimulation upon behavior of the female rat. *Journal of Comparative and Physiological Psychology*. https://doi.org/10.1037/h0032536

Jolly, E. (2018). Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software*. https://doi.org/10.21105/joss.00862

Kern, M. D., & King, J. R. (1972). Testosterone-Induced Singing in Female White-Crowned Sparrows. *The Condor*. https://doi.org/10.2307/1366289

Kim, S. M., Su, C. Y., & Wang, J. W. (2017). Neuromodulation of Innate Behaviors in Drosophila. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev-neuro-072116-031558

King, A. P., & West, M. J. (1977). Species identification in the North American cowbird:

Appropriate responses to abnormal song. *Science*. https://doi.org/10.1126/science.841321

King, A. P., & West, M. J. (1983a). Dissecting Cowbird Song Potency: Assessing a Song's Geographic Identity and Relative Appeal. *Zeitschrift Für Tierpsychologie*. https://doi.org/10.1111/j.1439-0310.1983.tb00739.x

King, A. P., & West, M. J. (1983b). Female perception of cowbird song: A closed developmental program. *Developmental Psychobiology*. https://doi.org/10.1002/dev.420160408

King, A. P., West, M. J., & Eastzer, D. H. (1986). Female Cowbird Song Perception: Evidence for Different Developmental Programs within the same Subspecies. *Ethology*, *72*(2). https://doi.org/10.1111/j.1439-0310.1986.tb00609.x

King, A. P., West, M. J., & White, D. J. (2003). Female Cowbird Song Perception: Evidence for Plasticity of Preference. *Ethology*. https://doi.org/10.1046/j.0179-1613.2003.00924.x

King, A. P., White, D. J., & West, M. J. (2003). Female proximity stimulates development of male competition in juvenile brown-headed cowbirds, Molothrus ater. *Animal Behaviour*. https://doi.org/10.1006/anbe.2003.2280

Kow, L. -M, & Pfaff, D. W. (1975). Dorsal root recording relevant for mating reflexes in female rats: Identification of receptive fields and effects of peripheral denervation. *Journal of Neurobiology*. https://doi.org/10.1002/neu.480060107

Kreutzer, M., Vallet, E., & Doucet, S. (1992). Sexual responses of female canaries at the onset of song stimuli. *Experientia*. https://doi.org/10.1007/BF02118318

Kroodsma, D. E., & Byers, B. E. (1991). The function(s) of bird song. *Integrative and Comparative Biology*. https://doi.org/10.1093/icb/31.2.318

Lahti, D. C., Moseley, D. L., & Podos, J. (2011). A tradeoff between performance and accuracy in bird song learning. *Ethology*. https://doi.org/10.1111/j.1439-0310.2011.01930.x

Laming, D. R. J. (1968). Information theory of choice-reaction times. *Information Theory of Choicereaction Times*.

Leboucher, G., Béguin, N., Mauget, R., & Kreutzer, M. (1998). Effects of fadrozole on sexual displays and reproductive activity in the female canary. *Physiology and Behavior*, *65*(2). https://doi.org/10.1016/S0031-9384(98)00079-1

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-10602-1_48

Long, M. A., Jin, D. Z., & Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature*, *468*(7322). https://doi.org/10.1038/nature09514

Lorenz, K. (1937). Über die Bildung des Instinktbegriffes. *Die Naturwissenschaften*. https://doi.org/10.1007/BF01491749

Low, M. (2004). Female weight predicts the timing of forced copulation attempts in stitchbirds, Notiomystis cincta. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2004.01.006

Lynch, K. S., Crews, D., Ryan, M. J., & Wilczynski, W. (2006). Hormonal state influences aspects of female mate choice in the Túngara Frog (Physalaemus pustulosus). *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2005.10.001

Maguire, S. E., Schmidt, M. F., & White, D. J. (2013). Social Brains in Context: Lesions Targeted to the Song Control System in Female Cowbirds Affect Their Social Network. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0063239

Maney, D. L., MacDougall-Shackleton, E. A., MacDougall-Shackleton, S. A., Ball, G. F., & Hahn, T. P. (2003). Immediate early gene response to hearing song correlates with receptive behavior and depends on dialect in a female songbird. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*. https://doi.org/10.1007/s00359-003-0441-z

Maney, D., & Pinaud, R. (2011). Estradiol-dependent modulation of auditory processing and selectivity in songbirds. In *Frontiers in Neuroendocrinology* (Vol. 32, Issue 3). https://doi.org/10.1016/j.yfrne.2010.12.002

Maney, Donna L., Cho, E., & Goode, C. T. (2006). Estrogen-dependent selectivity of genomic responses to birdsong. *European Journal of Neuroscience*. https://doi.org/10.1111/j.1460-9568.2006.04673.x

McKinney, F., & Evarts, S. (1998). Sexual Coercion in Waterfowl and Other Birds. *Ornithological Monographs*. https://doi.org/10.2307/40166723

Mennill, D. J., Ratcliffe, L. M., & Boag, P. T. (2002). Female eavesdropping on male song contests in songbirds. *Science*. https://doi.org/10.1126/science.296.5569.873

Moore, M. C. (1983). Effect of female sexual displays on the endocrine physiology and behaviour of male white-crowned sparrows, Zonotrichia leucophrys. *Journal of Zoology*. https://doi.org/10.1111/j.1469-7998.1983.tb02085.x

Morris, D. (2008). The Feather Postures of Birds and the Problem of the Origin of Social Signals. *Behaviour*. https://doi.org/10.1163/156853956x00264

Mruczek, R. E. B., & Sheinberg, D. L. (2007). Context familiarity enhances target processing by inferior temporal cortex neurons. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2106-07.2007

Nagle, L., & Kreutzer, M. L. (1997). Adult female domesticated canaries can modify their song preferences. *Canadian Journal of Zoology*. https://doi.org/10.1139/z97-759

Nice, M. M., & Marler, P. (1958). Behavior of the Chaffinch, Fringilla coelebs. *Bird-Banding*. https://doi.org/10.2307/4510692

Noble, R. G. (1979). The sexual responses of the female hamster: A descriptive analysis. *Physiology and Behavior*, *23*(6). https://doi.org/10.1016/0031-9384(79)90288-9

O'Loghlen, A. L., & Beecher, M. D. (1997). Sexual preferences for mate song types in female song sparrows. *Animal Behaviour*. https://doi.org/10.1006/anbe.1996.0348

O'Loghlen, A. L., & Beecher, M. D. (1999). Mate, neighbour and stranger songs: A female song sparrow perspective. *Animal Behaviour*. https://doi.org/10.1006/anbe.1999.1125

O'Loghlen, Adrian L., & Rothstein, S. I. (1995). Culturally correct song dialects are correlated with male age and female song preferences in wild populations of brown-headed cowbirds. *Behavioral Ecology and Sociobiology*. https://doi.org/10.1007/BF00165834

O'Loghlen, Adrian L., & Rothstein, S. I. (2002). East and west coast female brown-headed cowbirds agree: Both categories of male song are sexy. *Animal Behaviour*. https://doi.org/10.1006/anbe.2002.3081

Olson, E. (2011). AprilTag: A robust and flexible visual fiducial system. *Proceedings - IEEE International Conference on Robotics and Automation*. https://doi.org/10.1109/ICRA.2011.5979561

Peer, B. D., Rivers, J. W., Merrill, L., Robinson, S. K., & Rothstein, S. I. (2017). *The Brown-Headed Cowbird: A Model Species for Testing Novel Research Questions in Animal Ecology,*

Evolution, and Behavior. https://doi.org/10.1007/978-3-319-73138-4_9

Perkes, A., White, D., Wild, J. M., & Schmidt, M. (2019). Female Songbirds: The unsung drivers of courtship behavior and its neural substrates. *Behavioural Processes*, *163*(December 2017), 60–70. https://doi.org/10.1016/j.beproc.2017.12.004

Pfaff, D. W., Kow, L. M., Loose, M. D., & Flanagan-Cato, L. M. (2008). Reverse engineering the lordosis behavior circuit. In *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2008.03.012

Phelps, Rand, & Ryan. (2006). A Cognitive Framework for Mate Choice and Species Recognition. *The American Naturalist*. https://doi.org/10.2307/3491245

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*, *51*(2). https://doi.org/10.1111/j.1558-5646.1997.tb02441.x

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., & Ng, A. Y. (2009). ROS: an open-source Robot Operating System. *ICRA Workshop on Open Source Software*.

Ratcliffe, L., & Weisman, R. (1987). Phrase order recognition by brown-headed cowbirds. *Animal Behaviour*. https://doi.org/10.1016/S0003-3472(87)80188-4

Real, L. (1990). Search theory and mate choice. I. Models of single-sex discrimination. *American Naturalist*. https://doi.org/10.1086/285103

Remage-Healey, L., Dong, S. M., Chao, A., & Schlinger, B. A. (2012). Sex-specific, rapid neuroestrogen fluctuations and neurophysiological actions in the songbird auditory forebrain. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00749.2011

Riebel, K. (2003). The “Mute” Sex Revisited: Vocal Production and Perception Learning in Female Songbirds. *Advances in the Study of Behavior*. https://doi.org/10.1016/S0065-3454(03)33002-5

Riebel, K. (2009). Chapter 6 Song and Female Mate Choice in Zebra Finches: A Review. In *Advances in the Study of Behavior*. https://doi.org/10.1016/S0065-3454(09)40006-8

Roberts, B. L. (1992). Neural mechanisms underlying escape behaviour in fishes. In *Reviews in Fish Biology and Fisheries* (Vol. 2, Issue 3). https://doi.org/10.1007/BF00045039

Ronald, K. L., Zeng, R., White, D. J., Fernández-Juricic, E., & Lucas, J. R. (2017). What makes a multimodal signal attractive? A preference function approach. *Behavioral Ecology*. https://doi.org/10.1093/beheco/arx015

Rothstein, S. I., & Fleischer, R. C. (1987). Vocal Dialects and Their Possible Relation to Honest Status Signalling in the Brown-Headed Cowbird. *The Condor*, *89*(1). https://doi.org/10.2307/1368756

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social modulation of sequence and syllable variability in adult birdsong. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.01296.2007

Schleidt, W. M. (1974). How “Fixed” is the Fixed Action Pattern? *Zeitschrift Für Tierpsychologie*. https://doi.org/10.1111/j.1439-0310.1974.tb02131.x

Scott, D. W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization (Wiley Series in Probability and Statistics). In *Wiley Series in Probability and Mathematical Statistics Applied Probability and Statistics Section* (Vol. 156).

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*.

https://doi.org/10.25080/majora-92bf1922-011
Searcy, W. A. (1984). Song repertoire size and female preferences in song sparrows. *Behavioral Ecology and Sociobiology*. https://doi.org/10.1007/BF00299499
Searcy, W. A., & Capp, M. S. (1997). Estradiol dosage and the solicitation display assay in red-winged blackbirds. *Condor*, *99*(3). https://doi.org/10.2307/1370497
Searcy, W. A., & Marler, P. (1981). A test for responsiveness to song structure and programming in female sparrows. *Science*. https://doi.org/10.1126/science.213.4510.926
Sheldon, B. C., & Burke, T. (1994). Copulation behavior and paternity in the chaffinch. *Behavioral Ecology and Sociobiology*. https://doi.org/10.1007/BF00164185
Smith, V. A., King, A. P., & West, M. J. (2000). A role of her own: Female cowbirds, Molothrus ater, influence the development and outcome of song learning. *Animal Behaviour*, *60*(5). https://doi.org/10.1006/anbe.2000.1531
Spitler-Nabors, K. J., & Baker, M. C. (1987). Sexual display response of female white-crowned sparrows to normal, isolate and modified conspecific songs. *Animal Behaviour*. https://doi.org/10.1016/S0003-3472(87)80261-0
Suthers, R. A., Goller, F., & Pytte, C. (1999). The neuromuscular control of birdsong. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.1999.0444
Suthers, R. A., Vallet, E., & Kreutzer, M. (2012). Bilateral coordination and the motor basis of female preference for sexual signals in canary song. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4755630
Takahashi, M., Arita, H., Hiraiwa-Hasegawa, M., & Hasegawa, T. (2008). Peahens do not prefer peacocks with more elaborate trains. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2007.10.004
Turesson, H., Satta, A., & Domenici, P. (2009). Preparing for escape: Anti-predator posture and fast-start performance in gobies. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.032953
Vallet, E., & Kreutzer, M. (1995). Female canaries are sexually responsive to special song phrases. *Animal Behaviour*. https://doi.org/10.1016/0003-3472(95)90082-9
Vinnedge, B., & Verrell, P. (1998). Variance in male mating success and female choice for persuasive courtship displays. *Animal Behaviour*. https://doi.org/10.1006/anbe.1998.0776
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3). https://doi.org/10.1038/s41592-019-0686-2
Wei, D., Talwar, V., & Lin, D. (2021). Neural circuits of social behaviors: innate yet flexible. *Neuron*. https://doi.org/10.1016/j.neuron.2021.02.012
Welinder, P., Branson, S., Mita, T., Wah, C., & Schroff, F. (2010). Caltech-ucsd Birds 200. *Caltech-UCSD Technical Report*.
West, M. J., & King, A. P. (1985). Social Guidance of Vocal Learning by Female Cowbirds: Validating its Functional Significance. *Zeitschrift Für Tierpsychologie*, *70*(3). https://doi.org/10.1111/j.1439-0310.1985.tb00514.x
West, M. J., & King, A. P. (1988a). Female visual displays affect the development of male song in

the cowbird. *Nature*. https://doi.org/10.1038/334244a0

West, M. J., & King, A. P. (1988b). Vocalizations of juvenile cowbirds (Molothrus ater ater) evoke copulatory responses from females. *Developmental Psychobiology*. https://doi.org/10.1002/dev.420210605

West, M. J., King, A. P., & Eastzer, D. H. (1981). The Cowbird: Reflections on Development from an Unlikely Source. *American Scientist*.

West, M. J., King, A. P., Eastzer, D. H., & Staddon, J. E. (1979). A bioassay of isolate cowbird song. *Journal of Comparative and Physiological Psychology*, *93*(1). https://doi.org/10.1037/h0077577

White, D. J. (2010). A social ethological perspective applied to care of and research on songbirds. *ILAR Journal*. https://doi.org/10.1093/ilar.51.4.387

White, D. J., Gersick, A. S., Freed-Brown, G., & Snyder-Mackler, N. (2010). The ontogeny of social skills: experimental increases in social complexity enhance reproductive success in adult cowbirds. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2009.11.014

White, D. J., Gros-Louis, J., King, A. P., & West, M. J. (2006). A method to measure the development of song preferences in female cowbirds, Molothrus ater. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2006.01.008