VOL. 194, NO. 1 THE AMERICAN NATURALIST JULY 2019

Hidden Markov Models Reveal Tactical Adjustment of Temporally Clustered Courtship Displays in Response to the Behaviors of a Robotic Female

Anna C. Perry,[1,*] Alan H. Krakauer,[2] Richard McElreath,[3] David J. Harris,[4] and Gail L. Patricelli[1,2]

1. Animal Behavior Graduate Group, University of California, Davis, California 95616; 2. Department of Ecology and Evolution, University of California, Davis, California 95616; 3. Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; 4. Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida 32611

Submitted March 4, 2018; Accepted January 29, 2019; Electronically published May 31, 2019

Online enhancements: appendixes, videos. *Dryad data:* https://dx.doi.org/10.5061/dryad.sn0c503.

ABSTRACT: We present a statistical approach—a custom-built hidden Markov model (HMM)—that is broadly applicable to the analysis of temporally clustered display events, as found in many animals, including birds, orthopterans, and anurans. This HMM can simultaneously estimate both the expected lengths of each animal's display bouts and their within-bout display rates. We highlight the HMM's ability to estimate changes in animals' display effort over time and across different social contexts, using data from male greater sage grouse (*Centrocercus urophasianus*). Male display effort was modeled across three sites in two experimental treatments (robotic female simulating interested or uninterested behavior) and in the presence or absence of live females. Across contexts, we show that sage grouse males primarily adjust their bout lengths rather than their within-bout display rates. Males' responses to female behavior were correlated with male mating success: males with more matings showed high display persistence regardless of female behavior, while males with fewer matings tended to invest selectively in females that were already showing interest in mating. Additionally, males with higher mating success responded more to the presence of a female than males with fewer matings did. We conclude with suggestions for adapting our HMM approach for use in other animal systems.

Keywords: hidden Markov model, display bouts, display rate, mate choice, courtship tactics, greater sage grouse.

Introduction

A diverse array of animals repeat displays in bouts across different contexts of animal communication, such as during contests, parent-offspring exchanges, or courtship interactions. Here, we define a display bout as a cluster of repeated display events (e.g., visual and/or auditory signals of the same type; Martin and Bateson 2007). Bout-structured display effort consists of at least two components. Setting aside variation among individual display events (e.g., variation in signal magnitude or length), an animal could adjust either (1) their display persistence, by changing the consecutive number of displays contained in each bout, or (2) their within-bout display rate, by adjusting the relative lengths of the intervals of time separating each display event. In many cases, such adjustments in display effort can increase fitness, as when males that display for longer durations or at higher rates are preferred by females (Fiske et al. 1998; Murai and Backwell 2006; Delaney et al. 2007; Byers et al. 2010). However, most males cannot display at their peak levels indefinitely; these males may need to tactically adjust their display bout behavior across different conditions (Patricelli et al. 2016). Males' display tactics may depend on their underlying state (e.g., energetic reserves), environmental context (e.g., predation risk), or social context (e.g., available partners). To answer broader questions about animals' display efficacy or costs across different conditions, behavioral ecologists therefore need to be able to characterize how animals' display bout behavior changes over time.

Given that bout-structured display effort is composed of at least two components (persistence and within-bout rate), behavioral ecologists would benefit from using statistical models that characterize both components. However, analyses that ignore bout structure are still common (e.g., analyses that count the total number of display events or that average the lengths of the intervals separating displays; Patricelli and Krakauer 2010; Milner et al. 2011). These "bout-agnostic" analyses collapse each animal's multidimensional display effort to a single metric, potentially discarding important information (Perry et al. 2017*a*). Another common approach is to classify the intervals between display events

---

[*] Corresponding author; email: ancperry@ucdavis.edu.

ORCIDs: Perry, https://orcid.org/0000-0001-8154-295X; Krakauer, https://orcid.org/0000-0001-9177-7963; Harris, https://orcid.org/0000-0003-3332-9307; Patricelli, https://orcid.org/0000-0003-3397-8390.

Am. Nat. 2019. Vol. 194, pp. 1–16. © 2019 by The University of Chicago. 0003-0147/2019/19401-58306$15.00. All rights reserved.
DOI: 10.1086/703518

relative to a "bout criterion interval" (BCI). In this approach, intervals whose length exceeds the BCI are classified as between-bout breaks (Sibly et al. 1990; Martin and Bateson 2007). Preprocessing raw interval data in this way has several drawbacks, however. For example, observers using sufficiently different BCIs could classify bouts for the same behavioral sequence differently and thus disagree about an animal's display characteristics (e.g., an animal's number of bouts, mean bout durations, or mean within-bout display rates). Moreover, use of a constant BCI value precludes biologists from investigating how the intervals separating bouts of display activity tend to vary in length across different individuals, in different contexts, or across timescales.

To address these shortcomings, we present a statistical approach—a custom-built hidden Markov model—that will be broadly useful to behavioral ecologists analyzing temporally clustered display events. Hidden Markov models (HMMs) are a statistical machine-learning technique developed in the late 1960s and early 1970s to describe time series data (as reviewed in Rabiner 1989; Seymore et al. 1999). In the animal behavior field, these models are generally used to predict animal movement and foraging patterns (e.g., MacDonald and Raubenheimer 1995; Zucchini et al. 2008; Patterson et al. 2009). The HMM that we present has been tailored to estimate the probability that a given interval between display events belongs to one of two bout-related states: either an active bout of display or a longer break between bouts. Our HMM solves the problem of bout-agnostic models by explicitly accounting for bout structure and also solves the problem of a BCI-based approach by flexibly classifying interdisplay intervals (rather than using a global threshold). This dynamic, state-based framework also gives us the ability to examine whether an animal's display behavior is changing because of differences in display persistence or rate and to describe the covariance structure between these two components. We present this model, using empirical data from an experimental study conducted in greater sage grouse (*Centrocercus urophasianus*) that examine males' display effort across different social contexts.

Greater sage grouse (hereafter "sage grouse") are an ideal field system for studying display bout behavior. Sage grouse males gather on leks (communal display grounds) during the breeding season, where they can be readily observed displaying for several hours each morning over several months (Wiley 1973*a*; Gibson and Bradbury 1985). Sage grouse males' displays, called struts, follow a stereotyped motor pattern whose duration (approx. 2 s) has relatively little intra- or intermale variation (Wiley 1973*a*). As in many chorusing insects, frogs, or other birds (Robbins 1998; Gerhardt and Huber 2002), the intervals between display events alternate between quick bursts of display activity and longer between-bout breaks. The overall quantity and rate of males' displays is positively correlated with their mating success, which is determined by female choice in this system (Hartzler 1972; Wiley 1973*a*, 1973*b*; Gibson and Bradbury 1985; Vehrencamp et al. 1989; Boyce 1990; Gibson et al. 1991; Spurrier et al. 1991, 1994; Patricelli and Krakauer 2010). However, most previous analyses have ignored the bout structure of these displays, which limits our understanding of the relative importance of males' display persistence versus rate. Thus far, no morphological or plumage traits measured in male sage grouse seem to correlate with male mating success (Gibson and Bradbury 1985), although females in one study did avoid males with experimentally applied hematomas on their air sacs (Boyce 1990).

We apply our HMM (as well as a bout-agnostic model and a BCI-based model) to a data set of interdisplay intervals collected from sage grouse males. We quantified males' display effort in the presence/absence of live females and during two experimental treatments conducted with a biomimetic female robot. This robotic female allowed us to examine how males adjust their behaviors in response to the social context of courtship, using a controlled female stimulus (see Patricelli and Krakauer 2010). Specifically, we used a robotic female capable of mimicking "interested" (upright) and "uninterested" (foraging) female behavior. Sage grouse females generally assume an upright stationary posture before soliciting a mating, whereas foraging females have a lower probability of soliciting (Wiley 1973*b*; Perry et al. 2017*b*). As these behaviors correlate with females' willingness to continue an interaction (or to copulate), a male's fitness may depend on his display tactics (i.e., on how he adjusts his display persistence and/or rate in response to his social context). We do not yet know whether these female behaviors are adaptions to elicit male responses (signals) or incidentally convey information (cues); we refer to the upright and foraging behaviors as "cues" hereafter.

We use our HMM to characterize how males' display bout behavior changes over time and across contexts, showcasing how this state-based statistical model can be used to determine how individual display components (persistence and/or rate) change across different circumstances. We also demonstrate how the individual-level varying (i.e., random) effects in our HMM can reveal correlations between display tactics and another variable of interest: mating success. More specifically, we test whether males' display tactics in the presence/absence of females on the lek and in response to interested/uninterested female cues can be used to predict the number of matings that each male achieved.

In mating systems where female choice is based on males' display performance, males that respond tactically to female cues may reduce their display costs and/or increase their signal efficacy, leading to higher reproductive success (Patricelli et al. 2016). Here, we consider two mutually exclusive hypotheses regarding the relationship between males' responses to the robotic females' behavior and their mating success. Under the "differential-social-skills" hypothesis, prospective mates prefer males that reserve longer and/or more intense display bouts for females giving positive feedback (e.g., Patricelli et al. 2002, 2004; Rodríguez et al. 2012; Sullivan-Beckers and Hebets 2011, 2014). This relationship may arise, for example, when males that are more responsive to female cues are less likely to startle prospective mates during courtship (Patricelli et al. 2002). Alternatively, females—regardless of their apparent interest in mating—might prefer males that are capable of displaying at elevated levels for an extended period, rather than males that are more responsive to female behavior. In these systems, the males that are most responsive to female cues may need to be more selective about how their effort is deployed, because they are unable to maintain elevated display levels across contexts (e.g., because of greater constraints caused by differences in their genetics, body condition, or energetic reserves). Patricelli and Krakauer (2010) refer to this kind of resource-dependent variation as the "differential-constraints" hypothesis. Under this hypothesis, males that are unable to maintain their display effort during interactions with "uninterested" females are expected to have lower mating success on average. To address these hypotheses using our HMM, we modeled males' mating success as a function of their individual-level varying slopes, which describe how each male responded to different cues from the robotic female.

Our HMM constitutes a significant improvement over the bout-agnostic model and the BCI-based model. Specifically, the HMM could estimate the effects of social and environmental context on animals' display effort better than either of the simpler models. Moreover, the bout-agnostic model struggled to explain the low display rates associated with long between-bout breaks; its residuals indicated that this model's assumptions were strongly violated. Our state-based model can be broadly applied to other study systems, since many other vertebrate and invertebrate species display in bouts (Mowles and Ord 2012).

Material and Field Methods

Lek Observations

We monitored three leks approximately 25 km southeast of Hudson, in Fremont County, Wyoming. We collected behavioral data for 11 territory-holding males on the Monument lek (MNT), 12 males on the Chugwater lek (CHG), and at least 28 unique males on the Cottontail lek (COT); our analysis from COT also includes data from 45 cases where a male could not be individually identified across different days (see "Male ID" in app. A; apps. A, B are available online). Observations were conducted daily from March 13 to May 4, 2012 (except for 2 days with poor weather), beginning at first light and continuing until no birds remained on the lek. Males' breeding activity and behavior were video-recorded with one or two high-definition cameras recording in 1080i (Sony HDR-HC1, Tokyo, and Canon Vixia HV40, New York). Field assistants recorded this footage from blinds located on an observation hill approximately 75–200 m from each lek. These videos were later cross-referenced with field notes specifying individual males' on-lek positions, reported in relation to a grid of stakes arrayed at 10-m intervals (Krakauer et al. 2009; Patricelli and Krakauer 2010). Individual males were identified through spotting scopes via unique color band combinations and tail plumage patterns (Wiley 1973*b*). This combination of data sources enabled us to identify and track individual males on each video. One trained observer per lek used these videos to record the position and time of each strut event during each experimental trial. This research was approved by the Wyoming Game and Fish Department (permit 405) and the University of California, Davis, Animal Care and Use Committee (protocol 11662).

Robotic Female and Experimental Protocol

Following Patricelli and Krakauer (2010), we used a robotic female to elicit male display behavior, ensuring that all males received a controlled female stimulus. In this study, we used a second-generation robot that can move freely across the lek on four independently rotating wheels and can pivot to face courting males (see "Specifications for Robotic Female" in app. A). The robot can also tilt forward and move its neck to simulate pecking at the ground or rotate the head from side to side to simulate looking around (see videos 1 and 2, available online). We controlled all robotic movements with radio signals from a blind positioned on the edge of the lek. The robot's electronic components were concealed beneath a composite of real skins gathered from multiple female sage grouse (casualties collected by local wildlife managers). As Patricelli and Krakauer (2010) found with the first-generation robot, we found that males responded to our second-generation robot in a similar way as they would to a typical real female. Specifically, individual males' estimated tendencies to continue a display bout (i.e., their display effort) with the robot and with real females are strongly related ($R^2 = 0.72$; $N = 96$ males; Forbey et al. 2017).

We conducted all robot experiments in the morning between 0630 and 0820 hours from April 7 to May 1, 2012. We performed two robot treatments, simulating (1) the posture and head movements of an upright female, to mimic "interested" behavior, and (2) the posture and movements of a foraging female, to mimic "uninterested" behavior (fig. 1; videos 1 and 2, respectively). Females looking back and forth from an upright body position are more likely to solicit copulations later in the same lek visit, when compared to females pecking at the ground (Perry et al. 2017*b*; see also Wiley

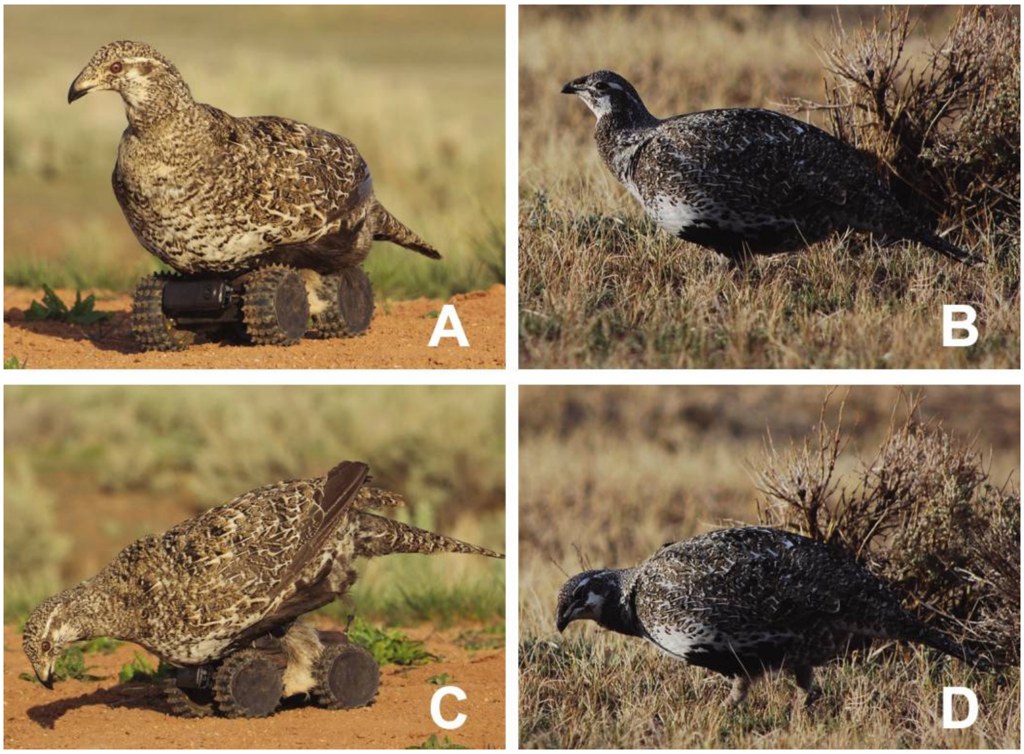


Figure 1: Comparison of the upright/interested position (*A*, *B*) and the foraging/uninterested position (*C*, *D*) in both the robotic female (*A* and *C*) and a real female sage grouse (*B* and *D*). See videos A1 and A2, available online.

1973*b*). We performed a total of 10 experimental trials across the three leks; each lek experienced both experimental treatments. We included lek ID as a fixed effect in our subsequent analyses to account for possible between-lek differences as well as potential treatment-order effects: MNT and CHG (the two smaller leks) received the "interested" treatment first, whereas COT experienced the "uninterested" treatment first. We waited at least 4 days before performing a different treatment on the same lek.

Each robot treatment was preceded by a 180-s pretrial period. Because males will spontaneously strut even when no females are present (Wiley 1973*a*), display effort during this period enabled us to measure baseline display levels for each male in our analyses. We initiated the pretrial period and performed all treatments when no real females were within 50 m of the lek, to simplify the social environment during experimental trials (Patricelli and Krakauer 2010). After each pretrial, we used small portable speakers (Auvio Universal Portable Speaker System) to play vocalizations typically associated with the arrival of a female grouse on the lek (Hartzler 1972; Patricelli and Krakauer 2010). Next, the robot emerged from an observation blind on the edge of the lek and moved to a target location between two or three males' territories, stopping approximately 5 m from these focal males (mean distance to nearest male: 4.9 m; standard deviation [SD]: 3.5 m). We selected target locations randomly with respect to male mating success, and we included all males on the lek (up to 55 m away) in our final analyses, regardless of whether they were part of the focal group. At each target location, the robot simulated behaviors associated with either the "interested" or the "uninterested" treatment for 180 s before moving to a new target location (at least 4 m away) for another 180-s sample. This movement pattern is similar to the way that real females move between males' territories on the lek (Gibson 1996). We kept the experimental treatment consistent between consecutive samples to avoid within-day treatment-order effects. The robot spent an average of 188 s (SD = 146 s; minimum [min] = 49 s; maximum [max] = 631 s) moving between locations on the lek. The robot visited up to five target locations per trial (mean = 2.7; SD = 1.4; min = 1) to reach all available groupings of males on the lek, which allowed us to collect samples of each male's display effort at various distances from the robot (mean male-to-robot distance = 17.1 m; SD = 11.0 m; min = 1.4 m; max = 54.6 m). Across all 10 experimental trials, each identified male was observed with the robot approximately five times, on average (SD = 2; min = 1; max = 10).

Natural Courtship Data

We also sampled males' display behavior during their interactions with real females. We chose time periods where there was only a single real female on the lek, analogous to our robot trials, and recorded data for males up to 50 m away from that female. We included 18 of these 180-s observation periods in our data set (8 on MNT, 6 on CHG, and 4 on COT). These observations overlapped temporally with our experiments (10 nonconsecutive days between April 5 and May 3, between 0539 and 0718 hours). We used these data to compare males' display effort during interactions with robotic versus real females (Forbey et al. 2017) and to characterize how males' display effort changed as the breeding season progressed.

Mating Success

We assessed each male's mating success by summing his total number of copulations during the 2012 breeding season (e.g., Gibson et al. 1991; Patricelli and Krakauer 2010). We did not observe any repeated copulations between the same male and female on the same morning. In sage grouse, male mating success provides an effective, if imperfect, proxy for actual male reproductive success (Semple et al. 2001; Bird et al. 2013). Although mating events observed in the field may not provide an exhaustive assessment of males' overall reproductive fitness (e.g., by potentially undersampling multiple paternity events or matings occurring off lek or at night; Bird et al. 2013), they provide us with a direct link between male display performance and female choice during the observation period.

Quantifying Male Display Effort

Two prior methods used to quantify the strut displays of male sage grouse are (1) counting the number of struts per minute (Hartzler 1972; Wiley 1973*b*; Gibson and Bradbury 1985; Spurrier et al. 1991, 1994) and (2) calculating the arithmetic or harmonic mean interstrut interval length for a set observation window (Wiley 1973*a*; Gibson and Bradbury 1985; Gibson et al. 1991; Patricelli and Krakauer 2010). Both of these methods collapse the distribution of interstrut intervals to a single number before analysis. Instead, we decided to adopt a modeling approach that uses information from each interstrut interval (i.e., the period of time between the termination of one strut and the beginning of the next; hereafter "interval"). Our data set included 8,315 fully and partially observed intervals (fig. 2). To calculate these intervals, we treated each strut display as a fixed-length event (lasting 2 s), since struts' durations have little intra- or intermale variation (Wiley 1973*a*). In the "Discussion," we suggest possible extensions to our HMM (described below) that would relax this assumption for applications in other signaling systems where the individual display events are more variable.

Statistical Methods

Statistical Model

We used a hidden Markov model (HMM) to infer males' courtship tactics from sequences of interdisplay intervals. Figure 2*B* shows the distribution of these interval lengths: a sharp peak of short intervals with an extended tail of longer intervals. Our HMM represents these two types of intervals as different "hidden" states (Rabiner 1989; Seymore et al. 1999), which correspond to different internal states for a male grouse: each male is either in an active bout of display (producing short intervals) or resting between bouts (which yields an extended interval). Each male's internal state is "hidden" in the sense that it cannot be directly measured and must be inferred from the male's behavior. In our sage grouse example, a 120-s interval between struts clearly corresponds to a break between more intense bouts of display, but additional context may be needed to classify a male's state during a 12-s interval.

In HMMs, each interval's predicted state (and thus its length) depends on the state assigned to the preceding interval in the sequence. This nonindependence between consecutive intervals (i.e., temporal autocorrelation) allows the model to accommodate biologically complex sequence data (e.g., after a long interval, males may be more likely to transition to an active bout of display than to strut once and return to another between-bout break). Our HMM must therefore estimate two sets of parameters: one controlling the probability that a male grouse will transition from one hidden state to another (fig. 3*A*) and the other determining the distribution of possible interval lengths that he can produce in each state (fig. 3*B*). We used the gamma distribution to describe the interval lengths generated by each state (fig. 3*B*); this distribution often describes lengths of time between events (like the more common, but less flexible, exponential distribution).

Our model defines a likelihood function (see "Joint Likelihoods for HMM and Mating Analysis" in app. A), which can be used to infer how multiple factors affect males' transition probabilities and their average interval lengths. We included the following variables as fixed effects for both sets of parameters: type of female stimulus (i.e., female absent, real, or robotic); experimental treatment (i.e., robot interested or uninterested); male proximity to female stimulus; date in breeding season; time since sunrise; and focal lek (see "Predictor Variables Included in the HMM" in app. A). We also included varying (i.e., random) intercepts to model the variance structure arising from repeated measures collected from the same individual, trial day, or sampling period

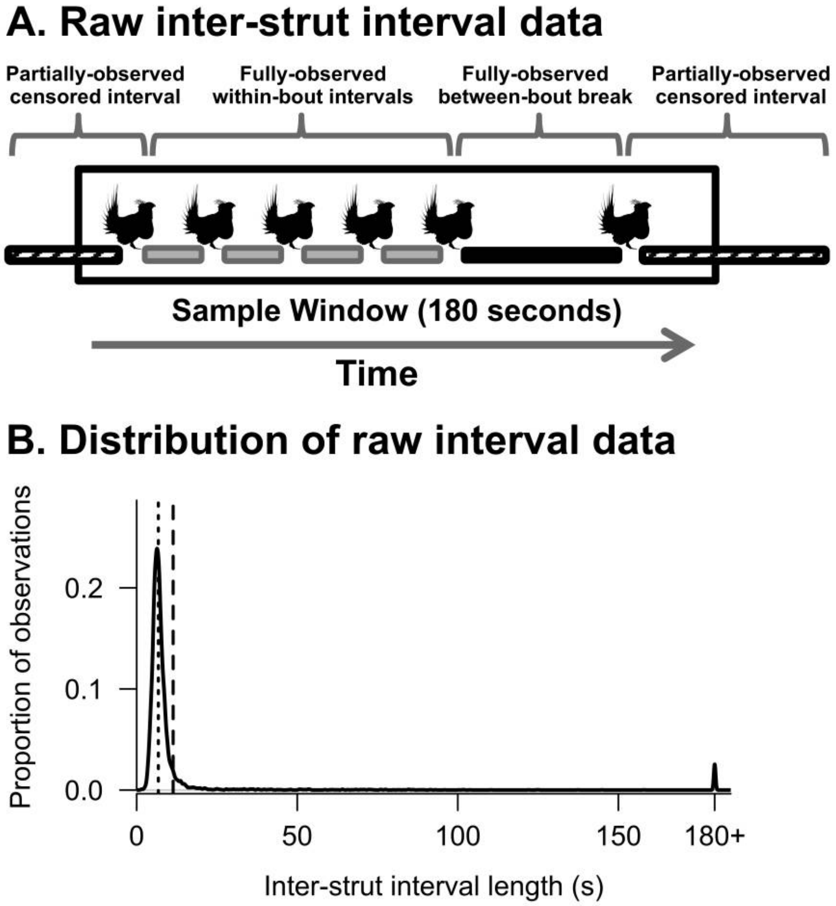


Figure 2: *A*, Hypothetical sampling window with raw interval data, where single display events (grouse silhouettes) separate both fully observed, within-bout intervals (gray bars) and longer, between-bout breaks (black bar) as well as partially observed, censored intervals (bars with dashes). *B*, Distribution of interstrut intervals observed across all sampling periods. The peak at 180 s represents specific censored intervals where the time between struts spanned an entire sampling window. The dotted vertical line shows the harmonic mean interval length at 6.75 ± 0.03 (bootstrapped SE) s (which describes only short interstrut intervals), while the dashed vertical line shows the arithmetic mean at 11.35 ± 0.28 s (approx. 90% of the raw intervals are shorter than this value). Neither mean adequately represents longer interval lengths that, although less frequent, occupy a greater proportion of a given sampling window.

(Gelman and Hill 2007; McElreath 2016). Like our fixed effects, these intercepts influence both display parameters (fig. 3). Each male, for example, has an intercept term specifying how much his display effort deviates from the typical male response in each display process. Because males display and interact with one another on leks, we could not measure each individual's display effort independently of other males' behavior. We were able to account for this nonindependence among males, in part, by including varying intercepts for each sampling window (e.g., for samples where many males simultaneously increased/decreased their effort). Finally, we used varying slopes to evaluate males' individual-level responsiveness to different social contexts (i.e., to estimate individual-level behavioral plasticity; see Dingemanse et al. 2010).

We fit two candidate HMMs: a baseline model with no treatment effect and another version with treatment. In the baseline model, males could vary only in their responsiveness to an absent, real, or robotic female, while the alternative model also included males' responsiveness to interested and uninterested robot behavior. If males differ in their responsiveness to female behavior, then the predictive accuracy of the HMM with the treatment effect should be better than that of the baseline model.

Censored Observations

Our sampling periods ended after 180 s, so we did not observe the beginning and/or ending of some intervals; these partially observed intervals are called censored observa-

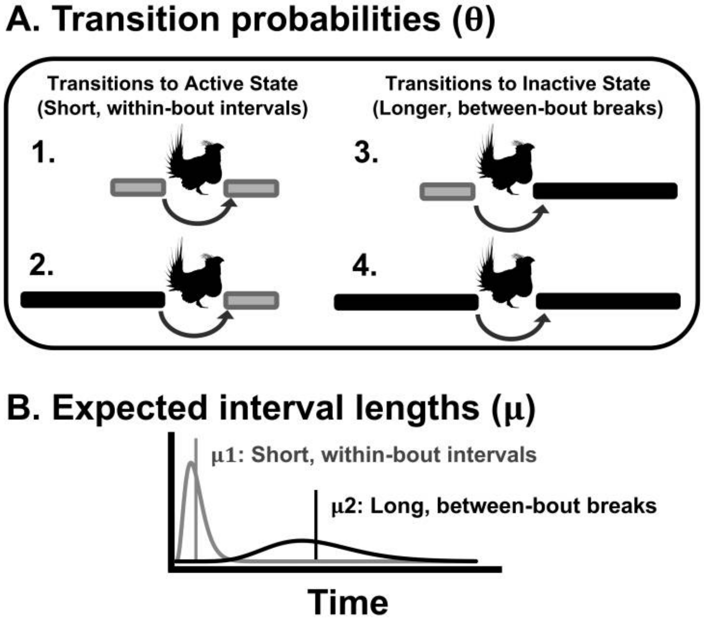


Figure 3: Conceptual diagram of the hidden Markov model (HMM), illustrating how combining the two sets of model parameters (θ and μ) contributes to a more complete understanding of each male's display effort. Here, we model male display behavior as a combination of active (within-bout) and inactive (between-bout) states. *A*, Visual representation of the probability matrix (θ) defining four possible state transitions between short, within-bout (gray boxes) and longer, between-bout (black boxes) interdisplay intervals (where each grouse silhouette represents a single display event; in our two-state model, transitions 1 and 3 or 2 and 4 sum to 1). *B*, Two hypothetical gamma distributions, showing the range of durations associated with short intervals ($\mu1$) and long intervals ($\mu2$).

tions (specifically, type I, right-censored observations; Lagakos 1979; Klein and Moeschberger 2003). Although the full durations of these censored intervals are unknown, we do have a lower bound on their possible lengths (fig. 2*A*). Proper handling of these partially observed intervals is critical for avoiding biased inference. For example, ignoring the portion of each interval that occurred outside our observation windows would lead the model to underestimate the average interval length, while discarding all partially observed intervals would disproportionately affect individuals with low display effort (i.e., nonresponse bias; Gelman and Hill 2007, p. 531–533). Our hypotheses require accurately characterizing both individuals with consistently low display effort and individuals that tactically reduce their effort in certain social contexts. To avoid biased estimates for these individuals, we used Bayesian imputation (see McElreath 2016, chap. 14) to represent our uncertainty about the range of possible durations for partially observed intervals.

Mating Success Analysis

To determine how males' display behavior relates to their mating success, we used parameter estimates from our HMM to predict the number of matings each male obtained, modeled as a Poisson-distributed variable. This Poisson generalized linear mixed model was fit simultaneously with our HMM, giving us a joint posterior distribution (see "Joint Likelihoods for HMM and Mating Analysis" in app. A). Calculating a joint posterior avoids many of the pitfalls associated with inferences based on point estimates obtained from varying effects (Hadfield et al. 2010), because the full uncertainty of each parameter is propagated through the model. This technique allows male coefficients with greater support to have a greater impact on the models' posterior and allows information to flow back from the mating model to inform our estimates of individual differences in male display behavior in the HMM (McElreath 2016, chap. 14). Both candidate models for the mating analysis allowed mating success to depend on all available male-specific intercepts and slopes from the HMM (see "Joint Likelihoods for HMM and Mating Analysis" in app. A). Since one candidate HMM included an additional predictor describing males' responsiveness to the robotic female's behavior, the corresponding mating model included this additional parameter as well.

Model Fitting and Evaluation

We implemented the joint posterior distribution for the HMM and mating analysis in the Stan programming language (Stan, ver. 2.6.0, and RStan, ver. 2.6.0; Stan Development Team 2015*a*, 2015*b*), which uses an efficient Markov chain Monte Carlo (MCMC) technique known as Hamiltonian Monte Carlo to collect samples from the posterior distribution of parameters (Monnahan et al. 2016). We calculated the HMM's likelihood, using the forward algorithm (Rabiner 1989), a recursive calculation that gives the probability of the observational sequence, by marginalizing (i.e., averaging) over all possible paths through the hidden states. We also used the R programming language to organize our raw data and analyze the model output (ver. 3.2.0; R Development Core Team 2015). For each candidate model, we generated a posterior distribution using two independent chains of 6,000 iterations each, discarding the first 2,000 "warm-up" iterations of each chain (Carpenter et al. 2017). The code and underlying data for each candidate model are deposited in the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.sn0c503 (Perry et al. 2019).

To evaluate our alternative hypotheses, we compared candidate HMM and mating models with and without the treatment effect. More specifically, we compared how well the candidate HMMs predicted each 180-s display sequence ($N = 667$) and how well the mating models predicted the mating success of each identified male ($N = 51$). We estimated each model's predictive accuracy, using the Watanabe-Akaike (or "widely applicable") information criterion (WAIC; Watanabe 2010; Gelman et al. 2014; Vehtari and Gelman 2014). The WAIC metric approximates each model's out-of-sample deviance, similar to the more familiar Akaike information criterion metric. We calculated ΔWAIC values and their standard errors to account for uncertainty in our estimates (Vehtari and Gelman 2014). Finally, we also calculated delta values for the deviance information criterion (DIC) to confirm that both model selection criteria (i.e., WAIC and DIC) produced the same qualitative results.

Additional Analyses of Males' Display Effort

We fit two additional models to our interval data to compare our HMM approach to simpler models of display effort. For the first model, we fit a linear mixed model to the natural log of males' harmonic mean interval lengths (see "Likelihood for Linear Mixed Model Predicting Males' Harmonic Mean Interval Lengths" in app. A for full details). We also calculated a bout criterion interval (BCI)—a threshold value defining the minimum length of a between-bout break—and used this BCI to classify each interval in our data set. We then fit a generalized linear mixed model that predicts the proportion of intervals in our data that were shorter or longer than the BCI value (which is analogous to the first set of parameters in our HMM predicting whether a given interval is short or long; see "Likelihood for Generalized Linear Mixed Model Using a 'Bout Criterion Interval'" in app. A). The code and underlying data for these additional analyses are deposited in the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.sn0c503 (Perry et al. 2019). See also Perry et al. (2017*a*) for a quantitative comparison of these models' performance on simulated data where display events are grouped in bouts.

Results

Below, we highlight model predictions for changes in males' display effort over time, their responses to different female stimuli, and the relationship between male responsiveness and mating success. Tables B1–B3 (tables B1–B8 are available online) summarize the model output, including mean posterior values, standard errors, and effective sample sizes for all parameters. Gelman and Rubin's (1992) "potential scale reduction factor" ($\hat{R}$) did not detect any problems with MCMC convergence ($\hat{R} < 1.01$ for all parameters).

Predictions from the HMM

The first set of HMM parameters make up a 2×2 transition matrix (fig. 3*A*) describing males' movements through the hidden states (e.g., beginning or ending a bout). Rather than describing the entire transition matrix under multiple treatments and conditions, we focus on the probability that a typical male will continue a display bout that has already begun ($p_{\text{continuing}}$; fig. 3*A.1*; see "Average Male Conditions" in app. A). We focus on this value because it determines bout length: the number of intervals per bout follows a geometric distribution whose mean is $1/(1 - p_{\text{continuing}})$, and the number of struts per bout is equal to 1 plus this value (fig. 4*A*). In general, we found that males had greater variation in their predicted bout lengths than in their predicted mean interval durations (reported in seconds), as described below.

We found that $p_{\text{continuing}}$ declined during the portion of the breeding season that we observed (i.e., after peak mating activity; fig. 4*A*). For example, the average male in the uninterested treatment continues his bout after 86.4% of his strut events (standard error [SE] = 5.2%) on day 5 of our observations (April 9), declining to 67.0% (SE = 14.3%) by day 25 (April 29; see "Average Male Conditions" in app. A). The intervals between struts also lengthened slightly over the course of the breeding season (table B4), although this had a much smaller effect on males' total strut count.

We also found that female stimulus type and treatment affected $p_{\text{continuing}}$ (fig. 4). For example, on day 25, a typical male continues bouts after 67.0% of his struts if the robot is uninterested. This increases by 9.8 percentage points (pp; SE = 7.3), on average, to 76.8% when the robot is interested and by a similar amount for a real female (11.9 pp; SE = 8.0). Finally, with no female stimulus present, $p_{\text{continuing}}$

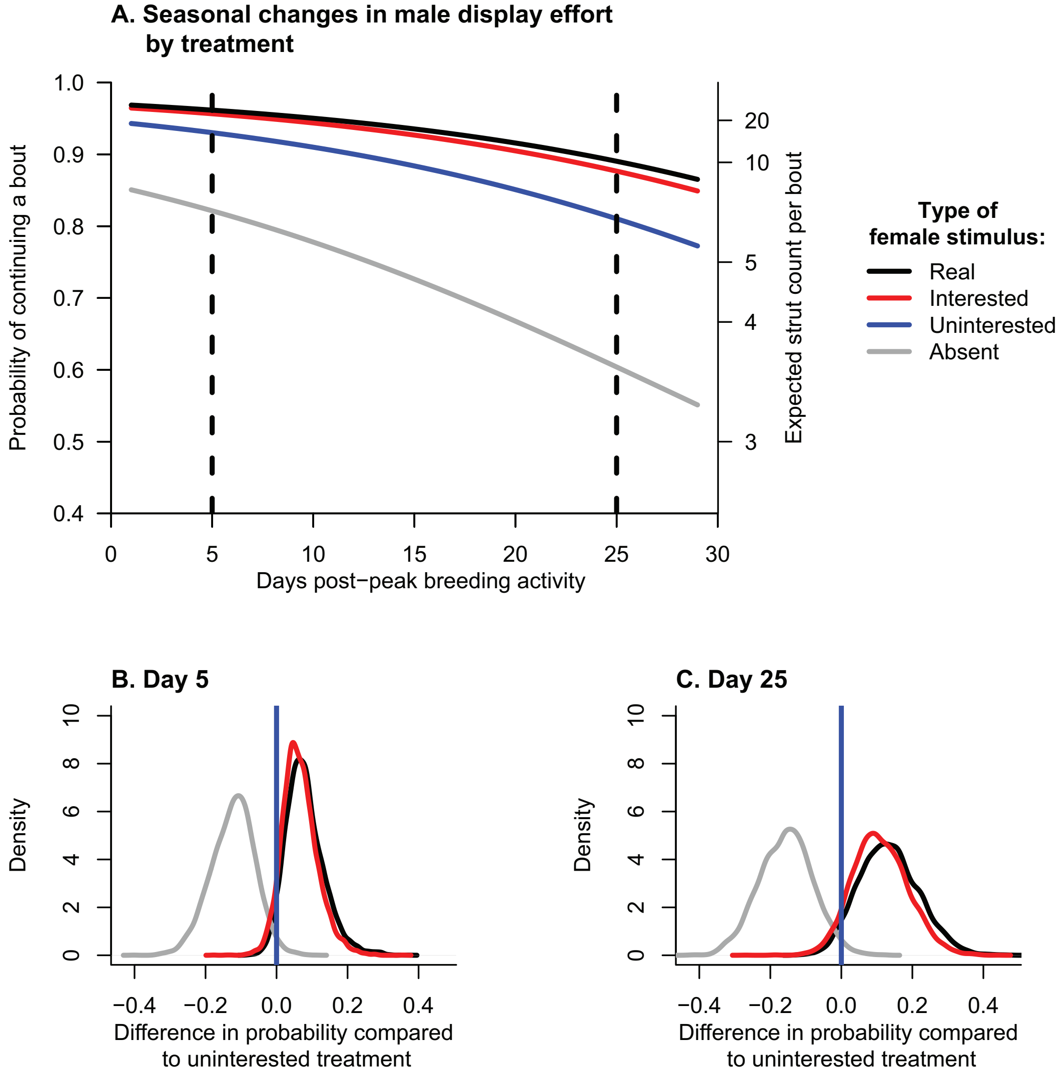


Figure 4: *A*, Mean posterior estimates for the probability that the average male will continue a bout of display with one real female (black line), with an interested robot (red line), with an uninterested robot (blue line), or in the absence of a female stimulus (gray line) at the median male-to-female distance (10.63 m, when in the presence of female stimuli) on the Monument lek. The secondary *Y*-axis translates these probabilities to expected struts per bout, using the nonlinear relationship described in the main text: $1 + 1/(1 - p_{\text{continuing}})$. *B* and *C* show posterior uncertainty about the differences among social conditions for $p_{\text{continuing}}$. The vertical line at $x = 0.0$ represents equivalency between a social context (i.e., real female, interested robot, or absence of stimuli) and the uninterested treatment. *B* shows these contrasts on day 5, and *C* shows them on day 25 (these two days are indicated by the dashed vertical lines in *A*).

is 15.0 pp lower (SE = 7.4). The full posterior estimates for these differences, relative to the uninterested-robot treatment, are displayed in figure 4*B*, 4*C*.

In contrast, we found that males' within-bout interval lengths were basically indistinguishable for the two robot treatments (difference = 0.1 ± 0.4 s; table B4). Compared with the robot treatments, males' interstrut intervals were slightly shorter in the presence of a real female and slightly longer when no female was present (table B4).

The value of $p_{\text{continuing}}$ also varies with male-to-female distance. For example, the average male in close courtship (5 m from a female stimulus) will continue a bout 94.4% of the time, while a more distant male (25 m away) tends to continue displaying 89.8% of the time. A 20-m change in male-to-female distance can thus nearly double the probability of ending a bout (e.g., from 5.6% to 10.2%). As with the type of female stimulus, however, the effect of distance on interval length within a given interval type (short vs. long) was minimal (changing by less than 0.5% for a 20-m change in male-to-female distance). Time-of-day effects showed a pattern different from the effects discussed above: the model was very uncertain about the sign and magnitude of an effect on bout length (i.e., $p_{\text{continuing}}$; table B1) but found that the time between struts for a given interval type increased by 9% (SE = 4%) for each hour after nautical dawn.

Mating Success

Male mating success was highly skewed in 2012: 35% of identified males were never observed mating, while the most successful individual mated 77 times ($N = 51$ identified males, 291 successful copulations; see fig. A1). In our mating analysis, we modeled males' mating success as a count variable (color scale in fig. 5*A*, 5*C*). However, to facilitate the interpretation of the results, we also show expected values for all successfully mated versus unmated males (fig. 5*B*, 5*D*). We observed several noteworthy relationships between males' display behavior and their mating success. In the absence of real or robotic female stimuli (i.e., during pretrial periods), we found a negative relationship between $p_{\text{continuing}}$ and male

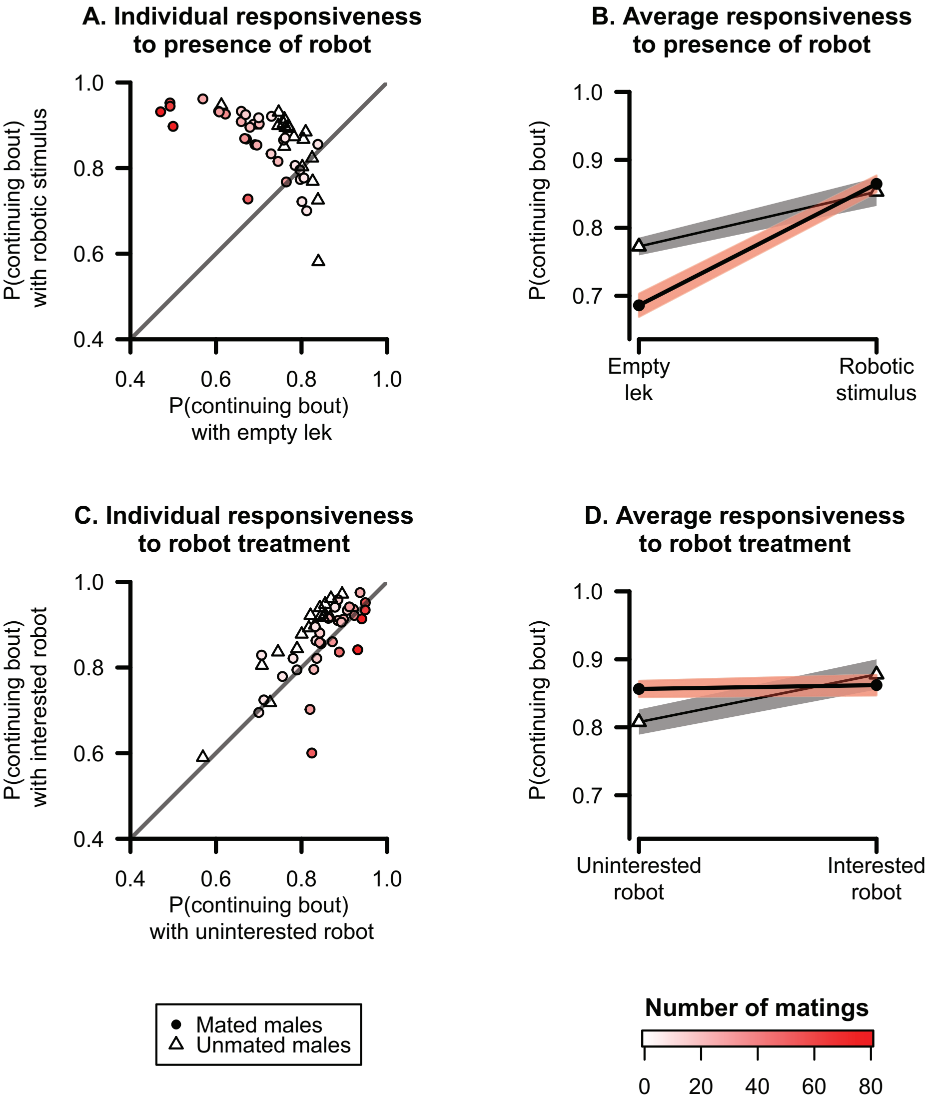


Figure 5: Males' responsiveness to the social context of courtship, including the presence/absence of the robotic stimulus (*A*, *B*) and the two robot treatments (*C*, *D*). The scatterplots (*A*, *C*) show individual males' responsiveness (individual-level posterior means) and mating success (color indicates number of observed matings; unmated males are marked by triangles). Line graphs (*B*, *D*) compare these mean posterior estimates for successfully mated and unmated males. The red and gray bands in *B* and *D* indicate the standard errors of the posterior means for successfully mated and unmated males, respectively. *A*, *B*, Successful males increase their display effort more for a robotic female (vs. empty lek) than unmated males do. Mated males have a greater average height above the 1 : 1 line (*A*) and a steeper slope (*B*). In contrast, unmated males make greater adjustments in response to robot behavior (*C*, *D*), whereas mated males tend to be closer to the 1 : 1 line (*C*).

mating success (fig. 5*A*, 5*B*). When comparing males' responses to the robot treatments, we found that successful males generally produced long bouts under both behavioral treatments, whereas males with lower mating success were more responsive to robot behavior, increasing their probability of continuing a bout when the robot was "interested" (fig. 5*C*, 5*D*). A visual inspection of figure 5*C* may suggest that some highly successful males have posterior means below the 1 : 1 line (i.e., display effort biased toward the uninterested treatment); however, all of these males except one had 95% credible intervals spanning the line. Overall, our results indicate that unmated males tend to be more responsive to female behaviors than successfully mating males, which is consistent with the predictions of the differential-constraints hypothesis.

Evaluating the HMM's Treatment Effect

The WAIC weights and rankings supported the inclusion of the experimental treatment as a factor in our candidate HMMs (ΔWAIC against model without treatment > 5.6; table B5). However, the standard error of the WAIC difference was relatively large at the level of a 180-s observation window (SE = 11.72; table B5). This reflects the fact that most of the variation in the data arises not from treatment but rather from individual male variation and other factors (table B1). Thus, we conclude both that treatment affected the display effort of many males and that there are other factors with larger effects on display effort. We also found support for the inclusion of a treatment effect in the model predicting male mating success (ΔWAIC against mating model without male responsiveness to treatment effect > 4; SE of difference = 3.23; table B6). In both cases, the DIC metric produced the same qualitative results (tables B5, B6).

HMM versus Alternative Analyses

Tables B7 and B8 summarize the output for the bout-agnostic and BCI-based models, respectively, including mean posterior values, standard errors, and effective sample sizes for all parameters. We found that these simpler models made untenable assumptions that prevented them from accurately describing the distribution of interdisplay intervals in our data set, resulting in biased estimates of males' display effort (fig. 6). Moreover, the bout-agnostic model could not estimate the fixed effects associated with males' social or environmental context with the same precision as the HMM ("Likelihood for Linear Mixed Model Predicting Males' Harmonic Mean Interval Lengths" in app. A). The BCI-based model did return fairly precise estimates ("Likelihood for Generalized Linear Mixed Model Using a 'Bout Criterion Interval'" in app. A); however, Perry et al. (2017*a*) showed that this precision is often misleading because this type of model converges on the wrong values for coefficients.

Discussion

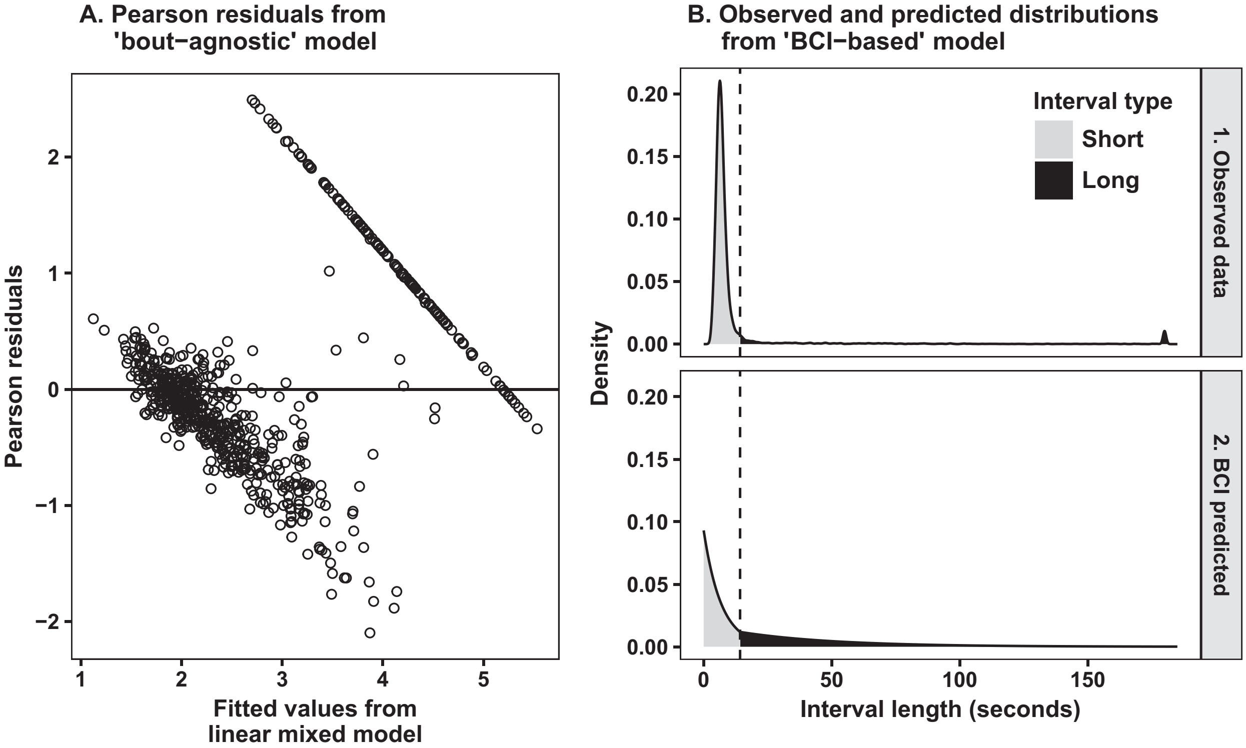
The HMM and Sage Grouse Display Effort

We used a state-based statistical analysis to assess the display effort of male sage grouse, whose displays are grouped in bouts. The HMM divided males' display effort into two separate components (fig. 3), which equipped us to make finer-scale distinctions about their display behavior than previous analyses of sage grouse. For example, we found that variation in males' overall display rates was mainly driven by changes in the number of consecutive displays (i.e., bout length), rather than within-bout display rates (i.e., interval length). In retrospect, this makes sense: within-bout display rates are relatively constrained (e.g., 95% of intervals are between 5 and 9 s; fig. 2*B*), but bout length varies widely (from 2 struts to more than 20). In general, prior sage grouse studies (as well as the bout-agnostic model that we fit) have combined these two components of display effort into an overall display rate (Wiley 1973*b*; Gibson and Bradbury 1985; Patricelli and Krakauer 2010). However, "display rates" may not be the most appropriate metric for this system, as the number of display events per bout appears to be a better predictor of male mating success in this species.

The HMM also enabled us to make detailed estimates of males' behavioral plasticity. Because we found that variation in sage grouse males' display effort is largely determined by differences in the number of struts per bout, we focus our discussion on males' display persistence (rather than the within-bout rates that differed by only fractions of a second). On average, we found that males produced longer bouts (i.e., had higher $p_{\text{continuing}}$ values) when a real or robotic stimulus was present on the lek and when the robotic female maintained an upright (or "interested") posture (fig. 4*A*); accounting for this male plasticity improved the predictive accuracy of our HMM (table B5). Moreover, we used the varying-slopes (i.e., random-slopes) estimates for each identified male in our HMM to show that the degree of plasticity varied among males and was related to their mating success (fig. 5).

Male Display Effort in the Absence of Female Stimuli

Most males performed shorter bouts in the absence of female stimuli (during pretrial periods). However, we found that unmated males decreased their bout length less than successfully mated males did (fig. 5*A*, 5*B*). In other words, unmated males had a higher "baseline" display activity level when females were absent. This finding is also consistent with trends from another sage grouse experiment using a robotic female (Patricelli and Krakauer 2010) and field observations in other lekking Galliformes, including black grouse (*Tetrao tetrix*; Höglund et al. 1997) and greater prairie chickens (*Tympanuchus cupido*; Nooker and Sandercock 2008).



A. Pearson residuals from 'bout−agnostic' model

B. Observed and predicted distributions from 'BCI−based' model

Figure 6: Untenable assumptions made by alternative display analyses. *A*, Pearson residuals from the log-harmonic mean model. There would be no visible patterns in these residuals if this model's assumptions were appropriate. The diagonal line in the upper right corresponds to 180-s-long intervals (which are censored observations from sampling periods where a given male did not display). *B*, Observed and predicted distributions from "BCI-based" model. In both plots, the dashed vertical line at 14.21 s shows the length of our bout criterion interval (BCI), as determined by maximum likelihood estimation (see "Likelihood for Generalized Linear Mixed Model Using a 'Bout Criterion Interval'" in app. A). Short, within-bout intervals are shown in gray, whereas longer, between-bout intervals are shown in black. Plot *B1* shows our real interval data classified using that BCI value, while plot *B2* shows the distribution of both interval types as predicted by the BCI procedure (see "Likelihood for Generalized Linear Mixed Model Using a 'Bout Criterion Interval'" in app. A). This procedure underestimated the average length of within-bout intervals and overestimated the frequency of longer, between-bout intervals (*B2*). This discrepancy is due to the fact that exponential distributions (an underlying assumption of the most common BCI procedures) are always peaked at 0, while the interdisplay intervals of sage grouse are clustered around an intermediate duration (*B1*).

Why would unsuccessful males continue to strut when there are no females to court? In lekking species (e.g., sage grouse), courtship displays often play a role in long-distance attraction and lek advertisement, even when no females are nearby (Gibson 1989, 1996; Höglund and Alatalo 1995). Like many species (Berglund et al. 1996; Galeotti 1998), male sage grouse also use the same display for both courtship and male-male competition (Wiley 1973*a*, 1973*b*), although the role of struts in the context of territoriality requires further study. It is therefore possible that unsuccessful males (who may be younger or lower in quality) need to allocate more display effort toward defending their territory and thus have less energy to expend during courtship (Patricelli and Krakauer 2010). Another possibility is that successfully mated males may be more skillful at tactically adjusting their display effort to invest more when females are present and in close proximity (Patricelli and Krakauer 2010).

Male Responsiveness to Robot Treatment

Males' responsiveness to robot behavior was also correlated with their mating success. When female proximity and other factors were controlled for, females mated more often with males that produced long bouts regardless of the robotic female's apparent interest. In other words, females were less likely to mate with males that showed high persistence only when the robot indicated some interest in mating (fig. 5*C*, 5*D*). These findings are more consistent with the differential-constraints hypothesis than with the differential-social-skills hypothesis: sage grouse females seem to prefer males that show greater display persistence over males that are more responsive to females' outward behavior.

Tactical adjustments in male display behavior—for example, producing shorter display bouts in certain social conditions—may be especially important for individuals with greater energetic constraints or lower endurance. In species like sage grouse, repeated displays can become extremely costly over time (Vehrencamp et al. 1989; Seymour and Sozou 2009; Clark 2012). Given these costs, display persistence may reflect differences in male vigor or stamina and, ultimately, underlying genetic or developmental differences (Byers et al. 2010). We hypothesize that unmated males tend to have fewer resources to allocate to courtship, perhaps because they display more in the absence of females. With limited resources, these males would need to reduce their display effort for foraging females, which are unlikely to solicit in the immediate future (Perry et al. 2017*b*). From this perspective, our results indicate that only the most successful sage grouse males would have sufficient resources/endurance to court foraging females until those females transition to interested behaviors (either during the current interaction or during that female's future lek visits). This general pattern is consistent with what has been found in other species. For example, in the scorpionfly (*Panorpa cognata*), high-condition males are indiscriminate in their choice of females, whereas males in poorer condition preferentially invest in females of higher reproductive value (Engqvist and Sauer 2001).

In the absence of direct information about males' energetic expenditure or body condition, we cannot directly test the possibility that males' display tactics and mating success reflect variation in their underlying state. However, the sharp decline in bout length over the breeding season (fig. 4) is consistent with the premise that males could not sustain extended display bouts after weeks of effort. We also find that males' within-bout intervals lengthened slightly over the course of a day, which is also consistent with fatigue. Ongoing efforts to tie male condition and off-lek foraging behaviors to on-lek courtship effort will address this possibility directly (Forbey et al. 2017).

Modeling Variation among Display Events

In this article, we assumed that all display events were equivalent, and we focused on the intervals between them. However, this would not be a reasonable assumption in other taxa, where displays vary in length (or other properties, such as signal amplitude). It would be straightforward to extend our two-state HMM to include an additional fully observed variable describing variation among display events. Unlike the two hidden states in the HMM that we presented, which could be distinguished only probabilistically, properties such as signal duration or amplitude can be measured directly. As a result, a "partially hidden" Markov model could be used, and the likelihood would be more tractable to calculate than with a fully hidden HMM with additional states (Scheffer and Wrobel 2001; Stan Development Team 2016, sec. 9.6). This type of model would be particularly advantageous for behavioral ecologists interested in the covariance between the properties of animals' individual display events and the interdisplay intervals that separate them.

Final Modeling Considerations

We conclude with a few considerations for researchers modeling complex display sequences. First, a modeling framework that can approximate the data-generating process for bout-structured data will yield better estimates (Perry et al. 2017*a*). For example, our HMM could provide better estimates than either of the simpler display models that we fit (i.e., the bout-agnostic and BCI-based models), because those models made untenable assumptions about the distribution of our interval data (fig. 6). Second, a modeling framework that can accommodate censored observations (like our HMM) will provide less biased estimates when evaluating animals' display effort from time series data. For our analyses, incorporating censored values was especially important for modeling sampling periods during which males performed few, or no, displays (e.g., when females were absent from the lek). The inclusion of these samples, which may be underrepresented or excluded in traditional analyses, was particularly illuminating for inferring different patterns of display among males. Finally, this article highlights the biological value of a statistical model capable of simultaneously estimating changes in multiple components of display effort. Although we found that most of the between-individual variation in sage grouse males' display behavior can be attributed to changes in the number of consecutive displays produced per bout, other species that display in bouts may show different patterns (e.g., orthopterans and anurans; Gerhardt and Huber 2002). Some animals, for example, may respond to different social situations by simultaneously adjusting both their bout length and their within-bout display rates. In this scenario, an animal that incurs higher costs from longer display bouts could potentially compensate by relaxing their within-bout display rate or extending their breaks between bouts. In such cases, using a modeling framework capable of estimating the covariance structure between model parameters describing both bout duration and display rate, like our HMM, could prove especially enlightening for studying trade-offs in multiple, dynamic components of display effort (Perry et al. 2017*a*).

Acknowledgments

This material is based on work supported by the University of California, Davis (UC Davis; USDA Hatch Funds Project

CA-D-EVE-2264-H), and the National Science Foundation (IOS-0925038 to G.L.P. and A.H.K.; IOS-1258217 to G.L.P., A.H.K., and Jennifer S. Forbey; and a Graduate Research Fellowship, DGE-1148897, awarded to A.C.P.). D.J.H. was supported by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative (through grant GBMF4563 to Ethan White). For robot design and fabrication, we thank Jonathan Tran, Jose Mojica, and Scott McKibben. For female sage grouse specimens to cover the exterior of the robots, we thank Tom Christiansen, Stan Harter, Pat Hnilicka, Andrew Sutphin, Tom Maechtle, Tom Easterly, and Tom Rinkes. For advice and field assistance, we thank the Lander offices of the Bureau of Land Management and the Wyoming Game and Fish Department, Tom Christiansen, Stan Harter, Sue Oberlie, and our field technicians: Melissa Breyer, Julia Gillis, Rebecca Koch, and Elizabeth Payne. For video data collection, we thank our many dedicated UC Davis undergraduates, especially Rebecca Ehrlich, Marty Martinez, and Carolyn Tran. We thank members of the Patricelli Lab, the UC Davis Animal Behavior Graduate Group, Andrew Sih, and our reviewers (E. McKenna Kelly and Alfredo Sánchez-Tójar) for helpful comments on the manuscript.

Literature Cited

Berglund, A., A. Bisazza, and A. Pilastro. 1996. Armaments and ornaments: an evolutionary explanation of traits of dual utility. Biological Journal of the Linnean Society 58:385–399.

Bird, K. L., C. L. Aldridge, J. E. Carpenter, C. A. Paszkowski, M. S. Boyce, and D. W. Coltman. 2013. The secret sex lives of sage-grouse: multiple paternity and intraspecific nest parasitism revealed through genetic analysis. Behavioral Ecology 24:29–38.

Boyce, M. S. 1990. The Red Queen visits sage grouse leks. American Zoologist 30:263–270.

Byers, J., E. Hebets, and J. Podos. 2010. Female mate choice based upon male motor performance. Animal Behaviour 79:771–778.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: a probabilistic programming language. Journal of Statistical Software 76(1):1–32.

Clark, C. J. 2012. The role of power versus energy in courtship: what is the "energetic cost" of a courtship display? Animal Behaviour 84:269–277.

Delaney, K. J., J. A. Roberts, and G. W. Uetz. 2007. Male signaling behavior and sexual selection in a wolf spider (Araneae: Lycosidae): a test for dual functions. Behavioral Ecology and Sociobiology 62:67–75.

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

Engqvist, L., and K. P. Sauer. 2001. Strategic male mating effort and cryptic male choice in a scorpionfly. Proceedings of the Royal Society B 268:729–735.

Fiske, P., P. Rintamaki, and E. Karvonen. 1998. Mating success in lekking males: a meta-analysis. Behavioral Ecology 9:328–338.

Forbey, J. S., G. L. Patricelli, D. M. Delparte, A. H. Krakauer, P. J. Olsoy, M. R. Fremgen, J. D. Nobler, et al. 2017. Emerging technology to measure habitat quality and behavior of grouse: examples from studies of greater sage-grouse. Wildlife Biology 2017:wlb00238. doi:10.2981/wlb.00238.

Galeotti, P. 1998. Correlates of hoot rate and structure in male tawny owls *Strix aluco*: implications for male rivalry and female mate choice. Journal of Avian Biology 29:25–32.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models: analytical methods for social research. Cambridge University Press, Cambridge.

Gelman, A., J. Hwang, and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Gerhardt, H. C., and F. Huber. 2002. Acoustic communication in insects and anurans: common problems and diverse solutions. University of Chicago Press, Chicago.

Gibson, R. M. 1989. Field playback of male display attracts females in lek breeding sage grouse. Behavioral Ecology and Sociobiology 24:439–443.

———. 1996. Female choice in sage grouse: the roles of attraction and active comparison. Behavioral Ecology and Sociobiology 39:55–59.

Gibson, R. M., and J. W. Bradbury. 1985. Sexual selection in lekking sage grouse: phenotypic correlates of male mating success. Behavioral Ecology and Sociobiology 18:117–123.

Gibson, R. M., J. W. Bradbury, and S. L. Vehrencamp. 1991. Mate choice in lekking sage grouse revisited: the roles of vocal display, female site fidelity, and copying. Behavioral Ecology 2:165–180.

Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. Kruuk. 2010. The misuse of BLUP in ecology and evolution. American Naturalist 175:116–125.

Hartzler, J. E. 1972. An analysis of sage grouse lek behavior. PhD dissertation. University of Montana, Missoula.

Höglund, J., and R. V. Alatalo. 1995. Leks. Monographs in Behavior and Ecology. Princeton University Press, Princeton, NJ.

Höglund, J., T. Johansson, and C. Pelabon. 1997. Behaviourally mediated sexual selection: characteristics of successful male black grouse. Animal Behaviour 54:255–264.

Klein, J., and M. Moeschberger. 2003. Survival analysis: statistical methods for censored and truncated data. Springer, New York.

Krakauer, A. H., M. Tyrrell, K. Lehmann, N. Losin, F. Goller, and G. L. Patricelli. 2009. Vocal and anatomical evidence for two-voiced sound production in the greater sage-grouse *Centrocercus urophasianus*. Journal of Experimental Biology 212:3719–3727.

Lagakos, S. 1979. General right censoring and its impact on the analysis of survival data. Biometrics 35:139–156.

MacDonald, I. L., and D. Raubenheimer. 1995. Hidden Markov models and animal behaviour. Biometrical Journal 37:701–712.

Martin, P., and P. Bateson. 2007. Measuring behavior: an introductory guide. Cambridge University Press, New York.

McElreath, R. 2016. Statistical rethinking: a Bayesian course with examples in R and Stan. CRC, Boca Raton, FL.

Milner, R. N. C., M. D. Jennions, and P. R. Y. Backwell. 2011. Keeping up appearances: male fiddler crabs wave faster in a crowd. Biology Letters 8:176–178.

Monnahan, C. C., J. T. Thorson, and T. A. Branch. 2016. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution 8:339–348.

Mowles, S. L., and T. J. Ord. 2012. Repetitive signals and mate choice: insights from contest theory. Animal Behaviour 84:295–304.

Murai, M., and P. R. Y. Backwell. 2006. A conspicuous courtship signal in the fiddler crab *Uca perplexa*: female choice based on display structure. Behavioral Ecology and Sociobiology 60:736–741.

Nooker, J., and B. Sandercock. 2008. Phenotypic correlates and survival consequences of male mating success in lek-mating greater prairie-chickens (*Tympanuchus cupido*). Behavioral Ecology and Sociobiology 62:1377–1388.

Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. Journal of Animal Ecology 78:1113–1123. doi:10.1111/j.1365-2656.2009.01583.x.

Patricelli, G. L., and A. H. Krakauer. 2010. Tactical allocation of effort among multiple signals in sage grouse: an experiment with a robotic female. Behavioral Ecology 21:97–106.

Patricelli, G. L., A. H. Krakauer, and C. C. Taff. 2016. Variable signals in a complex world: shifting views of within-individual variability in sexual display traits. Advances in the Study of Behavior 48:319–386.

Patricelli, G. L., J. A. C. Uy, and G. Borgia. 2004. Female signals enhance the efficiency of mate assessment in satin bowerbirds (*Ptilonorhynchus violaceus*). Behavioral Ecology 15:297–304.

Patricelli, G. L., J. A. C. Uy, G. Walsh, and G. Borgia. 2002. Sexual selection: male displays adjusted to female's response. Nature 415:279–280.

Perry, A. C., D. J. Harris, R. McElreath, and G. L. Patricelli. 2017*a*. Best practices for modeling display effort from bouts of display activity in animal communication research. Pages 53–103 *in* Behavioral tactics on the lek: manipulating and modeling the social environment of courtship in greater sage-grouse (*Centrocercus urophasianus*). PhD dissertation. University of California, Davis.

Perry, A. C., A. H. Krakauer, R. McElreath, D. J. Harris, and G. L. Patricelli. 2019. Data from: Hidden Markov models reveal tactical adjustment of temporally clustered courtship displays in response to the behaviors of a robotic female. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.sn0c503.

Perry, A. C., K. M. Long, D. J. Harris, A. H. Krakauer, and G. L. Patricelli. 2017*b*. Inferring internal state from behavioral sequences: an analysis of female courtship behavior in the greater sage-grouse. Pages 104–152 *in* Behavioral tactics on the lek: manipulating and modeling the social environment of courtship in greater sage-grouse (*Centrocercus urophasianus*). PhD dissertation. University of California, Davis.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286.

R Development Core Team. 2015. R: a language and environment for statistical computing. Version 3.2.0. R Foundation for Statistical Computing, Vienna.

Robbins, M. B. 1998. Display behavior of male Sprague's pipits. Wilson Bulletin 110:435–438.

Rodríguez, R. L., C. Haen, R. B. Cocroft, and K. D. Fowler-Finn. 2012. Males adjust signaling effort based on female mate-preference cues. Behavioral Ecology 23:1218–1225.

Scheffer, T., and S. Wrobel. 2001. Active learning of partially hidden Markov models. Pages 1–15 *in* Proceedings of the 12th ECML/PKDD Workshop on Instance Selection (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Freiburg, 2001).

Semple, K., R. K. Wayne, and R. M. Gibson. 2001. Microsatellite analysis of female mating behaviour in lek-breeding sage grouse. Molecular Ecology 10:2043–2048.

Seymore, K., A. McCallum, and R. Rosenfeld. 1999. Learning hidden Markov model structure for information extraction. Pages 37–42 *in* Proceedings of the AAAI-99 Workshop on Machine Learning for Information Extraction. Technical Report WS-99-11. Association for the Advancement of Artificial Intelligence, Orlando, FL.

Seymour, R. M., and P. D. Sozou. 2009. Duration of courtship effort as a costly signal. Journal of Theoretical Biology 256:1–13.

Sibly, R., H. Nott, and D. Fletcher. 1990. Splitting behaviour into bouts. Animal Behaviour 39:63–69.

Spurrier, M. F., M. S. Boyce, and B. F. Manly. 1991. Effects of parasites on mate choice by captive sage grouse. Pages 389–398 *in* J. E. Loye and M. Zuk, eds. Bird-parasite interactions: ecology, behavior and evolution. Oxford University Press, Oxford.

———. 1994. Lek behaviour in captive sage grouse *Centrocercus urophasianus*. Animal Behaviour 47:303–310.

Stan Development Team. 2015*a*. RStan: the R interface to Stan. R package version 2.6.0.

———. 2015*b*. Stan modeling language: user's guide and reference manual. Version 2.6.0.

———. 2016. Stan modeling language: user's guide and reference manual. Version 2.14.0. http://mc-stan.org.

Sullivan-Beckers, L., and E. A. Hebets. 2011. Modality-specific experience with female feedback increases the efficacy of courtship signalling in male wolf spiders. Animal Behaviour 82:1051–1057.

———. 2014. Tactical adjustment of signalling leads to increased mating success and survival. Animal Behaviour 93:111–117.

Vehrencamp, S. L., J. W. Bradbury, and R. M. Gibson. 1989. The energetic cost of display in male sage grouse. Animal Behaviour 38:885–896.

Vehtari, A., and A. Gelman. 2014. WAIC and cross-validation in Stan. http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11:3571–3594.

Wiley, R. H. 1973*a*. The strut display of male sage grouse: a fixed action pattern. Behaviour 47:129–152.

———. 1973*b*. Territoriality and non-random mating in sage grouse (*Centrocercus urophasianus*). Animal Behaviour Monographs 6:85–169.

Zucchini, W., D. Raubenheimer, and I. L. MacDonald. 2008. Modeling time series of animal behavior by means of a latent-state model with feedback. Biometrics 64:807–815.

References Cited Only in the Online Enhancements

Bates, D. M. 2010. lme4: mixed-effects modeling with R. http://lme4.r-forge.r-project.org/book/front.pdf.

Bivand, R., and N. Lewin-Koh. 2016. maptools: tools for reading and handling spatial objects. R package version 0.8-39. https://CRAN.R-project.org/package=maptools.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, NJ.

Grolemund, G., and H. Wickham. 2011. Dates and times made easy with lubridate. Journal of Statistical Software 40(3):1–25.

Labra, A., P. Carazo, E. Desfilis, and E. Font. 2007. Agonistic interactions in a *Liolaemus* lizard: structure of head bob displays. Herpetologica 63:11–18.

Luque, S. P., and C. Guinet. 2007. A maximum likelihood approach for identifying dive bouts improves accuracy, precision and objectivity. Behaviour 144:1315–1332.

Mowles, S. L. 2014. The physiological cost of courtship: field cricket song results in anaerobic metabolism. Animal Behaviour 89:39–43.

Myung, I. J. 2003. Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology 47:90–100.

Stan Development Team. 2018*a*. RStan: the R interface to Stan. R package version 2.17.3.

———. 2018*b*. rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.2. http://mc-stan.org/.

Associate Editor: Alex Jordan
Editor: Russell Bonduriansky

A male greater sage grouse (*Centrocercus urophasianus*; *left*) performs a “strut display” to court a female (*right*). We studied how sage grouse males tactically adjust the timing of these temporally clustered displays during interactions with a robotic female grouse. Photo credit: Anna C. Perry.