NORTHWESTERN UNIVERSITY

Minimal Mathematical Models of Human and Animal Dynamical Systems

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Engineering Sciences and Applied Mathematics

By

Sara M. Clifton

EVANSTON, ILLINOIS

September 2017

ABSTRACT

Minimal Mathematical Models of Human and Animal Dynamical Systems

Sara M. Clifton

Minimal mathematical models are used to understand complex phenomena in the physical, biological, and social sciences. This modeling philosophy never claims, nor even attempts, to fully capture the mechanisms underlying the phenomena, and instead offers insights and predictions not otherwise possible. Here, we build and explore minimal dynamical systems models to understand three complex animal and human systems. First, we incorporate the assumptions of Zahavi's handicap principle into a mathematical model of ornament evolution and show that this existing hypothesis is sufficient to explain the previously puzzling observation of bimodally distributed ornament sizes in a variety of species. Second, we propose a 'return-to-setpoint' model of chronic pain dynamics in sickle cell disease patients with the goal of offering personalized, data-driven recommendations for treating chronic pain. Third, we present a conceptual model of restaurant competition that predicts the existence of a critical gratuity rate threshold at which restaurant owners will disallow tipping to maximize their profits. Because of their simplicity, these models

of complex human and animal systems offer new connections between existing ideas, give optimized solutions with limited data, and provide qualitative predictions of future events.

Acknowledgements

This dissertation comprises three distinct projects that were only possible through the supervision and encouragement of my advisor Danny Abrams, whose enthusiasm and creativity inspired my love of applied mathematics research. When I began my graduate degree, I was not sure if a career in the mathematical sciences would be rewarding or even realistic for me. After four years of unwavering support and mentorship, I discovered my own interests in the field, and I learned to ask and answer new questions. I cannot thank him enough for leading me through my development from a student to a scholar.

I also want to extend my gratitude to my other primary collaborators, Rosemary Braun (Northwestern University) and Chaeryon Kang (University of Pittsburgh), whose hard work and creativity sets a high standard for my future research, and whose patience and kindness sets a high standard for my future advising. I feel lucky to have worked with such inspiring role models.

Thanks are also due to the National Science Foundation. My NSF Graduate Research Fellowship (DGE-1324585) and travel grants allowed me to pursue research without distraction, share my work with scholars across the country, and take advantage of numerous professional development opportunities.

In addition to my applied mathematics research, several faculty supported me through pedagogical research and professional development. I thank Danny again for helping me mentor my own students on a research project of my invention. I cannot express enough

Preface

The research presented in this dissertation was performed collaboratively with my advisor, Daniel M. Abrams (DMA), and several other scientists and mathematicians. My contributions to the research and its presentation are noted below.

Chapter 2 is based on "Handicap principle implies emergence of dimorphic ornaments" published in the *Proceedings of the Royal Society B* [1]. The original model was proposed by DMA. I performed the vast majority of analysis, data collection, and manuscript writing under the guidance of DMA. Rosemary Braun suggested the least-unimodal unimodal statistical method and wrote the statistical analysis section. The animal and plant data was extracted from over a dozen publications, referenced in the main text. Data and code are available on the Dryad repository: `http://dx.doi.org/10.5061/dryad.vb1pp` [2].

Chapter 3 is based on "Hybrid statistical and mechanistic mathematical model guides mobile health intervention for chronic pain" published in the *Journal of Computational Biology* [3]. The original model was proposed by DMA, Chaeryon Kang (CK), Qi Long, and Jessica Li. Nirmish Shah ran the clinical trial that provided data. Jude Jonassaint and Charles Jonassaint created and maintained the mobile health application used by patients. I performed the majority of the dynamical systems analysis, and CK performed the majority of the statistical analysis. DMA, CK, and I wrote the manuscript on which this chapter is based. Data and code are available on the Northwestern University ARCH

repository: `https://doi.org/10.21985/N24S3T` [4]. Real patient data is not included to protect patient privacy, but all methods may be tested on synthetic data.

Chapter 4 is based on "The tipping point: mathematical model predicts restaurants will soon abandon tipping en masse" submitted to the *Northwestern Undergraduate Research Journal* [5]. I proposed the original model and supervised two undergraduates, Eileen Herbers and Jack Chen, as they explored the model for a summer research project. DMA consulted with me and the undergraduate researchers throughout the summer. While based on the initial exploration, this chapter is entirely my own work. I performed the analysis and wrote the manuscript.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The essence of mathematics is not to make simple things complicated,

but to make complicated things simple.

Stan Gudder, "A Mathematical Journey"

1.1. Investigating complex biological and social systems with simple models

Minimal mathematical models bring clarity to complex natural and social phenomena, from disease spread [**6, 7, 8**] to fishery dynamics [**9, 10, 11**] to bird flocking [**12, 13, 14**]. Though these models do not capture the full complexity of the systems, their analytical tractability and simplicity offer insights into first order principles not obvious in more detailed models. Moreover, these minimal models provide useful predictions and guidance: compartmental disease models inform vaccination protocols, population dynamics models influence resource management, and flocking models inspire the field of artificial swarm intelligence. The aforementioned models are only a small sample of existing minimal models that give new perspectives on complex systems.

1.2. Mathematical modeling process

Mathematical modeling is the translation of a real world system into a quantitative language. There are as many ways to build a mathematical model as there are math modelers, but Figure 1.1 illustrates one strategy for constructing testable math models.

Modelers typically cycle between the real world and the math world. The real world provides an interesting phenomenon, paradox, or question that evades easy explanation. Modelers take the often vaguely stated real world problem and propose a concrete qualitative hypothesis to be tested. Then the qualitative theory is translated from words into mathematical formulas; these governing equations may be exactly solved or approximated using numerical methods and/or asymptotic analysis. Once model behavior is well understood, the model can be validated or undermined by real world data. In either case, the model offers insight into the forces underlying the real world phenomenon. If the model does not satisfactorily answer the initial question, or if new questions arise after an iteration of the modeling process, the whole cycle may begin again.

TYPICAL MATH MODELING PROCESS
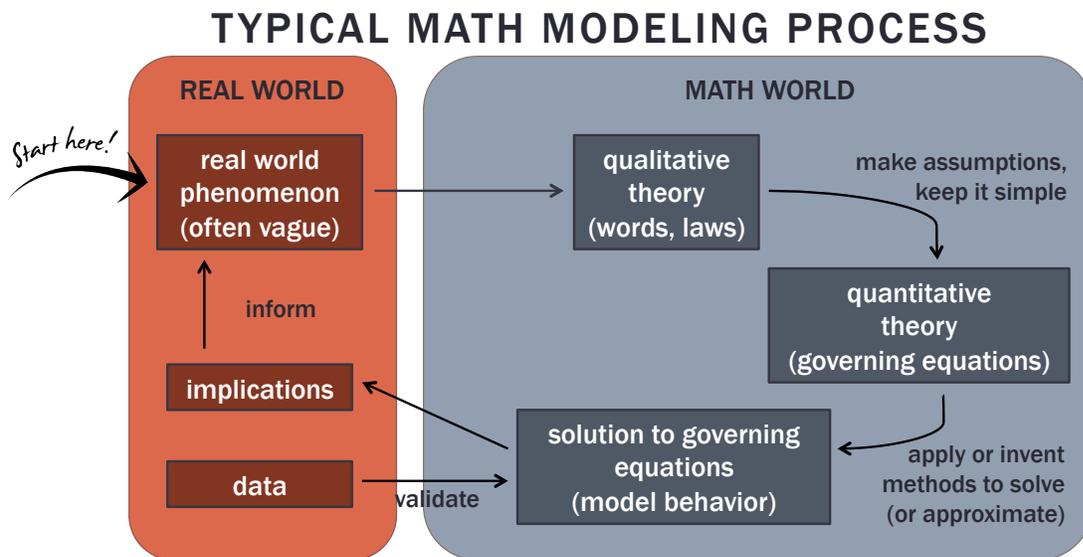
Figure 1.1. Typical mathematical modeling process. Modified with permission from flowchart by Vicky Chuqiao Yang (`http://www.vcyang.com/`).

1.2.1. Example of the math modeling process

As an illustration of the math modeling process, we will briefly walk through the creation of a mathematical model for the evolution of extravagant mating displays, such as peacock feathers and deer antlers. The real world provides an interesting phenomenon: species spanning the animal kingdom grow large or flashy ornamental features that not only require extra resources to grow and maintain, but also hinder the ability to gather resources and avoid predators.

Charles Darwin was the first to attribute this phenomenon to a balance between natural and sexual selection [15], turning a real world observation into a qualitative theory. Evolutionary biologist Amotz Zahavi later refined Darwin's theory with his handicap principle [16]. This hypothesis argues that, because costly ornaments hinder survival, only the highest quality individuals can afford significant investment in them. Thus the cost of an ornament truthfully advertises the quality of an individual, making mate selection more efficient.

Zahavi's handicap principle, a qualitative theory for the evolution of ornaments, serves as the basis for a quantitative theory that can be tested. Because Darwin's theory suggests that survival potential decreases and mating potential increases as ornament size (or flashiness) increases, we create a simple net reproductive benefit function to capture this behavior. Zahavi adds an additional constraint to our formulation that higher quality mates will experience relatively lower costs for a particular ornament size. For the full model derivation, see Chapter 2.

Due to the model's simplicity, we can find exact steady state solutions and their corresponding stability. We can numerically integrate the system to see change in ornament size

over time and investigate relationships among model parameters and final distributions of ornament sizes. Most significantly, we can compare model behavior with real world data. Our model predicts ornament sizes will evolve into a bimodal distribution, which is seen in many ornamented species. Though the model is too simple to perfectly reproduce the ornament size distributions seen in the real world, we find that model predictions are qualitatively consistent with real world data.

This modeling process offers new implications for the evolution of ornaments: Zahavi's handicap principle may explain the dimorphic ornaments seen in many species. This new connection can now be tested quantitatively in the field or in the lab.

1.3. Dissertation overview

In this dissertation, three new models offer fresh ways of approaching problems in three different fields: evolutionary biology, medicine, and behavioral economics. Chapter 2 provides a new connection between an existing biological hypothesis, the handicap principle, and the puzzlingly common existence of dimorphic ornaments. Chapter 3 exploits limited mobile health data to recommend optimal clinical interventions for chronic pain in sickle cell disease patients. Chapter 4 uses non-traditional methods to offer a new qualitative prediction that restaurants will abandon tipping to maximize profits.

CHAPTER 2

Handicap principle implies emergence of dimorphic ornaments

2.1. Background

Darwin was the first to suggest that both natural and sexual selection play a role in the evolution of mating displays [15]. Natural selection is the shift in population traits based on an individual's ability to survive and gather resources, while sexual selection is the shift in population traits based on an individual's ability to mate with more or better partners. Natural selection alone cannot explain ornaments because they hinder survival and provide little to no benefit to the individual [17, 18, 19]. Darwin hypothesized that female preference for exaggerated mating displays drives the evolution of male ornamentation, but he was unable to explain why females prefer features which clearly handicap the males.

Zahavi's handicap principle attempts to resolve the paradox proposed by Darwin [16]. It argues that, because costly ornaments hinder survival, only the highest quality individuals can afford significant investment in them. Thus the cost (often correlated with size) of an ornament truthfully advertises the quality of an individual, which makes mate selection easier. There is a large body of evidence that ornaments are indeed costly to the bearer (e.g. [20, 21, 22]), that ornaments are honest signals of quality (e.g., [23, 24]), and that females prefer mates with larger ornaments (e.g. [25, 26, 27]).

A variety of theoretical approaches have been used to model the handicap principle [19, 28, 29, 30, 31]. Broad categories include game theoretical approaches (e.g., [32, 33]), quantitative genetics (e.g., [34, 35]), and phenotypic dynamics (e.g., [36, 37]). Borrowing and expanding upon ideas from all three methods, we propose a new dynamical systems approach to understanding the evolution of ornaments within a population. Our model differs from some that search for a single evolutionarily stable strategy (ESS) (e.g., [33]) in that we do not require a unique phenotype for a particular male quality; our method allows for the possibility that an optimal *distribution* of strategies may emerge for a population—even a population of equal quality males.

Curiously, it has been observed that ornament sizes frequently have bimodal distributions, resulting in distinct small- and large- "morphs" in many ornamented species (e.g., [38, 39, 40]). Figure 2.1 illustrates a classic example of ornament dimorphism, the horned dung beetle [38]. While in some cases researchers have identified genetic and environmental factors associated with ornament size variation (e.g., [41, 42]), the splitting into two *distinct* large- and small-ornamented subpopulations (morphs) remains a contentious area of study.

Some evolutionary theories suggest that variety within the sexes may be due to varied mating strategies such as mimicry, sneaking, or fighting [43, 44]. However, our model suggests that the handicap principle alone may be sufficient to explain the origin of the observed ornament bimodality.
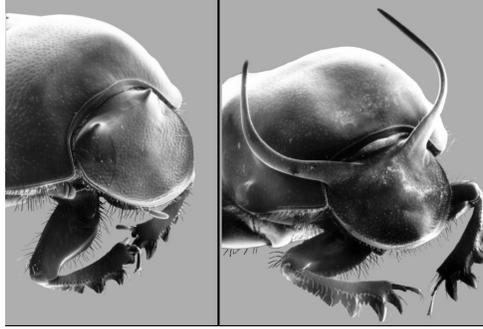
Figure 2.1. Example of a dimorphic ornament: dung beetles with differing horn lengths (*Onthophagus taurus*, Coleoptera: Scarabaeidae), reprinted from [**38**] with permission.

2.2. Model

With the goal of examining the quantitative implications of the handicap principle, we construct a minimal dynamical systems model for the evolution of extravagant and costly ornaments on animals. This proposed model incorporates two components of ornament evolution: an intrinsic cost of ornamentation to an individual (natural selection), and a social benefit of relatively large ornaments within a population (sexual selection). We show that on an evolutionary time scale, identically healthy animals can be forced to split into two morphs, one with large ornaments and one with small.

To express our model, we introduce the idea of a "reproductive potential" φ. This can be thought of as similar to fitness, though our definition differs from the fitness function commonly used in the replicator equation [**36, 45**]. See Section 2.4.3.3 for the precise relationship. Over long time scales the effect of evolution is to select for individuals with higher reproductive potential.

Consider an *individual* reproductive potential $\varphi^{(\text{ind})}$ of a solitary male with ornament size a (e.g., a deer with ornamental antlers). Some ornaments have practical as well

as ornamental value (e.g., anti-predation [**46, 47**]), but have a deleterious effect beyond a certain size. We therefore expect that there exists an optimal ornament size (possibly zero), for which individual potential is maximum, and thus take this to be a singly-peaked function of ornament size. For simplicity we assume the quadratic form[1]

$$(2.1) \qquad \varphi^{(\mathrm{ind})} = a(2a_{\mathrm{opt}} - a).$$

Following the handicap principle, we expect the optimal ornament size $a_{\mathrm{opt}} = a_{\mathrm{opt}}(h)$ to be an increasing function of "intrinsic health" h—i.e., healthier individuals can afford larger ornaments. See Figure 2.2 (a) for the general shape of the individual reproductive potential function.

Next, we consider a *social* reproductive potential $\varphi^{(\mathrm{soc})}$ that captures the effects of competition for partners (i.e., sexual selection). We assume social potential is an increasing function of ornament size[2] because sexual selection often favors larger or more elaborate ornaments [**26**]. For simplicity, and motivated by the ubiquity of power laws in nature [**48, 49**], we choose social potential to be a power of the difference between a male's ornament size and the average herd ornament size. To ensure monotonicity, we force the social reproductive potential to be antisymmetric about the average ornament size. The social potential is then

$$(2.2) \qquad \varphi^{(\mathrm{soc})} = \mathrm{sgn}\,(a - \bar{a})|a - \bar{a}|^{\gamma},$$

---

[1]This is a generic form for an arbitrary smooth peaked function approximated close to its peak.
[2]This assumption applies most naturally to inter-sexual selection, ignoring alternative reproductive strategies associated with intra-sexual selection (e.g., cryptic males).

where the positive parameter γ quantifies the rate at which deviations from the mean influence reproductive potential, sgn is the sign function, and $\bar{a}$ represents the average ornament size in the population. Loosely speaking, the parameter γ tunes female choice; we take this "female choice" parameter to be effectively constant because female choice may evolve on a slower time scale than male ornamentation [35]. Refer to Figure 2.2 (b) for an example of the social reproductive potential function.

Because both natural and sexual selection play a role in the evolution of ornaments [35], we take total reproductive potential to be the weighted average

$$
(2.3) \qquad\qquad \varphi = s\varphi^{(\mathrm{soc})} + (1-s)\varphi^{(\mathrm{ind})},
$$

where s tunes the relative importance of competitive social effects (sexual selection) versus individual effects (natural selection). We show in the Appendix that a weighted product [50] produces identical qualitative results, so we focus on this case for simplicity of calculations. See Figure 2.2 (c),(d) for examples of total potential functions.

Assuming that evolutionary forces optimize overall reproductive potential at a rate proportional to the marginal benefit of ornamentation, ornament sizes will follow the dynamics

$$
(2.4) \qquad\qquad \frac{\mathrm{d}a}{\mathrm{d}t} = c\frac{\partial\varphi}{\partial a}
$$

with time-scaling parameter $c > 0$. Note that this model does not presume that individual ornaments explicitly change size: the "phenotype flux" $\mathrm{d}a/\mathrm{d}t$ is simply a way of describing how the distribution of ornament sizes in a large animal population changes over long time scales as a result of selection processes.
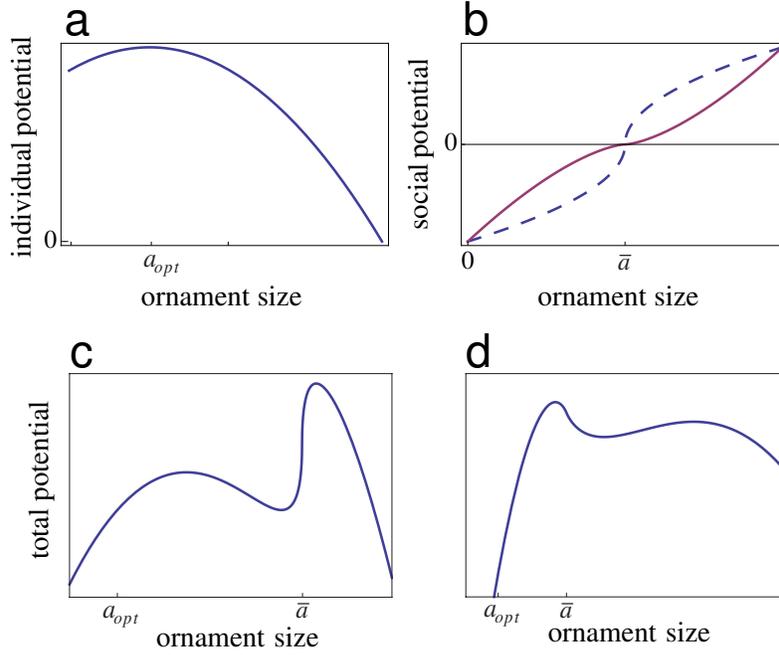
Figure 2.2. Model derivation. **(a)** Example individual potential function, singly peaked at a_{opt}. We use a quadratic function. **(b)** Example social potential function, antisymmetric about the population mean $\bar{a}$. We use an anti-symmetrized power law such that the shape depends on the social sensitivity γ (blue dashed is $\gamma = 0.5$; maroon solid is $\gamma = 1.5$). **(c)** Example total reproductive potential function at equilibrium for $\gamma < 1$. There are two local maxima corresponding to two distinct morphs, with the larger ornament morph having the highest potential (here $\gamma = 0.5$). **(d)** Example total potential function at equilibrium for $1 < \gamma < 2$. There are two local maxima corresponding to two distinct morphs, with the smaller ornament morph having the highest potential (here $\gamma = 1.5$). Note that the fitness landscape is distinct for each population representative, and representatives are not assumed to be identical.

This produces a piecewise-smooth ordinary differential equation for ornament size flux,

$$(2.5) \qquad \frac{\mathrm{d}a}{\mathrm{d}t} = c\left[s\gamma \left(1 - \frac{1}{N}\right) |a - \bar{a}|^{\gamma - 1} + 2(1-s)(a_{\mathrm{opt}} - a) \right],$$

where N is the population size. If we wish to model the evolution of all N population representatives, then we would index a by i, producing a system of N ordinary differential

equations. We omit the indices for notational clarity. Plugging (2.5) into the continuity equation yields a replicator equation for the evolution of the ornament size distribution (see Section 2.4.3.3).

2.3. Numerical exploration

For biologically relevant values of the social sensitivity parameter γ (details to follow), our model predicts stratification into distinct phenotypes for a population of identically healthy individuals (i.e., individuals of identical quality). See Figure 2.3 for the time evolution of ornament size for two representative values of γ.
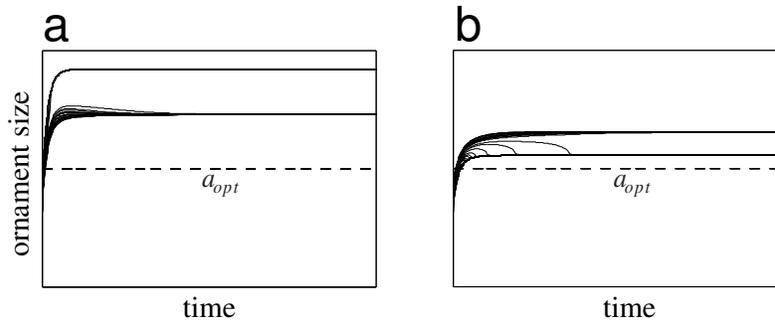
Figure 2.3. Evolution of ornament size. **(a)** Evolution of $N = 100$ population representatives over time for $\gamma = 0.5$ and **(b)** $\gamma = 1.5$. The initial conditions were sampled randomly from a normal distribution with mean 0.75 and standard deviation 0.25. The optimal ornament size $a_{\mathrm{opt}} = 1.0$, maximum simulation time $t_{max} = 50$, time scaling constant $c = 1.0$, and $s = 0.5$.

For $0 < \gamma < 1$, the ornament sizes stratify into large-ornament and small-ornament groups, with the majority possessing a large-ornament "morph." For $1 < \gamma < 2$, the population stratifies into large- and small-ornament morphs, but the majority have small ornaments. The case $\gamma \geq 2$ is not a reasonable option because we have selected a quadratic form for the local approximation of the individual potential function; any power γ

exceeding 2 implies sexual selection is the dominant evolutionary force even for extremely large ornaments, an unreasonable assumption.

These qualitative results are consistent for all a_{opt} and $0 \leq s < 1$. While for clarity we have presented predictions of a specific minimal model, the qualitative results hold for a wide range of models. See Section 2.6.2 for the generality of model predictions.

2.4. Analytical results

Because numerical integration shows that the uniform and two-morph steady states are of interest, we concentrate our analysis on these equilibria. However, we can show graphically that uniform and two-morph steady states are the only possible solutions for a wide range of potential functions, including our potential function (2.3) (see Appendix).

2.4.1. Uniform steady state

To investigate the uniform equilibrium with an identically healthy population, we set $a = \bar{a}$ producing the single ordinary differential equation[3],

$$(2.6) \qquad \frac{\mathrm{d}a}{\mathrm{d}t} = 2c\,(1-s)(a_{\text{opt}} - a).$$

The steady state (i.e., $\mathrm{d}a/\mathrm{d}t = 0$) is clearly $a = a_{\text{opt}}$. Linear stability analysis within this identical ornament manifold shows the fixed point $a = a_{\text{opt}}$ is stable for all γ and $0 \leq s < 1$, but numerical simulation suggests that the uniform fixed point is only stable for $\gamma \geq 2$. To resolve this apparent discrepancy, we investigate the uniform fixed point of (2.5) in the continuum limit, and evaluate stability without restriction to the uniform

---

[3]For $\gamma \leq 1$, we set $\varphi^{(\text{soc})} = 0$ before setting $a = \bar{a}$ to avoid an undefined right-hand side of (2.5).

manifold. We are then able to find γ-dependence that agrees with simulations. See Section 2.4.3 for more details.

2.4.2. Two-morph steady state

To investigate the two-morph equilibrium, we assume all males have one of two ornament sizes a_1 and a_2. Taking x to be the fraction of males with ornament size a_1, and $N \to \infty$, the dynamical system becomes

(2.7)
$$\frac{\mathrm{d}a_1}{\mathrm{d}t} = c\left[s\gamma\Big((1-x)|a_1 - a_2|\Big)^{\gamma-1} + 2(1-s)(a_{\mathrm{opt}} - a_1)\right]$$
$$\frac{\mathrm{d}a_2}{\mathrm{d}t} = c\left[s\gamma\Big(x|a_1 - a_2|\Big)^{\gamma-1} + 2(1-s)(a_{\mathrm{opt}} - a_2)\right].$$

There exists one two-morph steady state (i.e., solution to $\mathrm{d}a_1/\mathrm{d}t = \mathrm{d}a_2/\mathrm{d}t = 0$):

(2.8)
$$a_1 = a_{\mathrm{opt}} + \left(\frac{s\gamma}{2(1-s)}\right)^{\frac{1}{2-\gamma}} \left((1-x)\left|\frac{(1-x)^\gamma x - x^\gamma + x^{1-\gamma}}{(1-x)x}\right|^{\frac{1}{2-\gamma}}\right)^{\gamma-1}$$
$$a_2 = a_{\mathrm{opt}} + \left(\frac{s\gamma}{2(1-s)}\right)^{\frac{1}{2-\gamma}} \left(x\left|\frac{(1-x)^\gamma x - x^\gamma + x^{1-\gamma}}{(1-x)x}\right|^{\frac{1}{2-\gamma}}\right)^{\gamma-1}.$$

Figure 2.4 (a),(b) shows how two-morph equilibria vary with the morph fractionation x. Within the shaded region, the fixed point is stable. To be clear, the model predicts that a bimodal population will emerge, with the fraction x of the individuals within the population possessing ornaments of size a_1. We are *not* claiming that a proportion x of populations will evolve to ornament size a_1.

The eigenvalues for the linearized system constrained to this two-morph manifold are $\lambda_1 = -2(1-s)/s$ and $\lambda_2 = 2(\gamma-2)(1-s)/s$. Clearly, the two-morph equilibrium is stable (within the two-morph manifold) for $0 < \gamma < 2$ and unstable for $\gamma > 2$, when $\lambda_2 > 0$.

Curiously, the stability of the two-morph equilibrium does not depend on x, the morph fractionation. This presents an apparent problem because numerical simulation suggests that only certain ranges of x are stable: see Figure 2.4 (c). Similarly to the uniform fixed point analysis, we investigate the fixed points of the model in the continuum limit, and evaluate stability without restriction to any manifold. We are then able to find x-dependence that agrees well with simulations: see Figure 2.4 (d). Refer to Section 2.4.3 for more details.
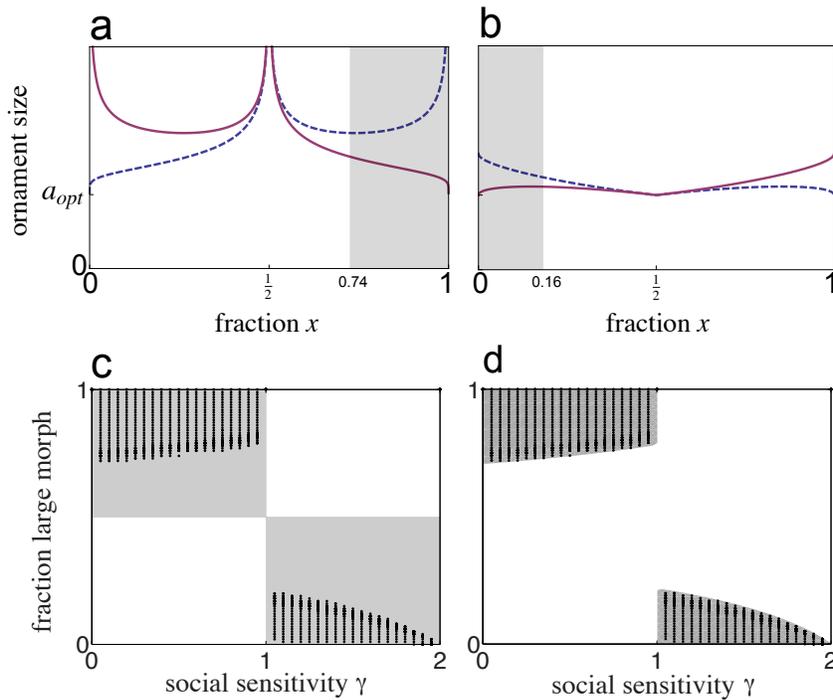
Figure 2.4. Stability regions for two-morph steady states ($N = 100$, $s = 1/2$). The ornament size for morph 1 is blue (dashed line), and the ornament size for morph 2 is maroon (solid line). The shaded regions are stable. **(a)** Two morph steady state for various morph fractionation x and $\gamma = 0.5$ **(b)** Two morph steady state for various morph fractionation x and $\gamma = 1.5$. **(c)** Analytical stability region (grey shading) for finite N model within two-morph manifold with numerical stability region (dots) superimposed. **(d)** Analytical stability region (grey shading) from continuum model with numerical stability region (dots) superimposed.

2.4.3. Continuum limit

Up until this point, we have modeled a system of N population representatives whose dynamics follow equation (2.5). The fixed points of this system are a discrete set of ornament sizes. Now we take $N \to \infty$, which turns the N ordinary differential equations into a partial integro-differential equation for a continuous distribution of ornament sizes $p(a, t)$. The equation we derive is the replicator function for continuous phenotypes [45].

We use conservation of probability to find the governing equation for the probability density function $p(a, t)$. The probability of a male having an ornament size in $(a, a + \mathrm{d}a)$ for small $\mathrm{d}a$ is approximately $p(a, t) \, \mathrm{d}a$. Given our assumption that individuals are neither created nor destroyed in $(a, a + \mathrm{d}a)$, we have

$$\frac{\partial p}{\partial t} \, \mathrm{d}a = p \left. \frac{\mathrm{d}a}{\mathrm{d}t} \right|_a - p \left. \frac{\mathrm{d}a}{\mathrm{d}t} \right|_{a+\mathrm{d}a}.$$

In other words, the change in individuals in the sliver $(a, a + \mathrm{d}a)$ is equal to the number that enter the sliver minus the number that leave. In the limit $\mathrm{d}a \to 0$, we get the continuity equation

$$\text{(2.9)} \qquad \frac{\partial p}{\partial t} = -\frac{\partial}{\partial a} \left(p \frac{\mathrm{d}a}{\mathrm{d}t} \right).$$

The dynamics of a follow (2.5) in the limit $N \to \infty$

$$\text{(2.10)} \qquad \frac{\mathrm{d}a}{\mathrm{d}t} = c \left[s \, \gamma |a - \bar{a}|^{\gamma - 1} + 2(1 - s) \left(a_{\mathrm{opt}} - a \right) \right],$$

where the mean ornament size is

$$\bar{a} = \int_{-\infty}^{\infty} a(t)\, p(a,t)\, \mathrm{d}a.$$

We substitute (2.10) into (2.9) to get a partial integro-differential equation for the probability density function $p(a,t)$ for ornament size

(2.11) $$\frac{\partial p}{\partial t} = -c\frac{\partial}{\partial a}\left(p\left[s\,\gamma|a-\bar{a}|^{\gamma-1} + 2(1-s)\,(a_{\text{opt}} - a)\right]\right).$$

2.4.3.1. Continuum limit uniform steady state. Now that we have established the continuum limit of the discrete model, we wish to investigate the uniform fixed point we found in Section 2.4.1. Within this continuum framework, the uniform fixed point $a = a_{\text{opt}}$ is the delta distribution

(2.12) $$p(a,t) = \delta(a - a_{\text{opt}}).$$

Previously, we investigated the stability of the uniform steady state by perturbing every member of the population by the same arbitrary, small amount. If we wished to repeat this investigation for the continuum model, we would shift the peak of the delta function by an arbitrary small amount from a_{opt} to some a_0. To make stability analysis more general, we also consider widening the delta function into a narrow Gaussian with an arbitrary small standard deviation σ. Figure 2.5 (a),(b) illustrates this idea.

We now wish to confirm that this continuum representation (2.10) of the model is consistent with our discrete model (2.5), at least near the simplest fixed point (the uniform state). Based on our previous stability analysis, we expect that a_0 will shift back to a_{opt}

and the width of the peak will shrink to 0 for $\gamma \geq 2$. However, we do not know how quickly these shifts occur relative to each other.

We will first investigate the dynamics of a_0 assuming that σ is effectively constant on the time scale of interest. Then the "perturbed" distribution is the narrow Gaussian

$$(2.13) \qquad p(a, t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(a-a_0(t))^2/2\sigma^2}$$

with constant $\sigma \ll 1$ and $a_0(t)$ near the fixed point a_{opt}.

Plugging (2.13) into the continuity equation (2.9), and solving for the highest order (fastest) dynamics of a_0, we see

$$(2.14) \qquad \frac{\mathrm{d}a_0}{\mathrm{d}t} = s\gamma|a - a_0|^{\gamma-1} + 2(1 - s)(a_{\mathrm{opt}} - a).$$

Note that (2.14) is only true if $\sigma \to 0^+$ faster than $a \to a_0$. If we instead assume $\sigma \to 0^+$ slower than $a \to a_0$, the right-hand side of (2.14) is unbounded, and therefore inconsistent with the discrete model. Taking $a \to a_0$ in (2.14), we see as expected

$$\frac{\mathrm{d}a_0}{\mathrm{d}t} = 2(1 - s)(a_{\mathrm{opt}} - a_0).$$

As we see that σ shrinks to 0 faster than $a \to a_0$, we investigate the dynamics of $\sigma(t) \ll 1$ for $a_0 = a_{\mathrm{opt}}$. Again, we take $p(a, t)$ to be a narrow Gaussian distribution

$$(2.15) \qquad p(a, t) = \frac{1}{\sigma(t)\sqrt{2\pi}} e^{-(a-a_0)^2/2\sigma(t)^2}.$$

Substituting (2.15) into (2.9) and Taylor expanding about $\sigma = 0$ gives

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \left[\frac{\gamma |a - a_0|^{\gamma-1}}{a - a_0} + 2\frac{1 - s}{s}\frac{a_{\mathrm{opt}} - a}{a - a_0} \right]\sigma + \mathcal{O}(\sigma^3).$$

We see that as $a \to a_0 = a_{\mathrm{opt}}$ for $\gamma < 2$, the uniform fixed point is unstable (coefficient of σ is ∞). For $\gamma > 2$, the fixed point is stable (coefficient of σ is $-2\frac{1 - s}{s}$). The fixed point for $\gamma = 2$ is conditionally stable (coefficient of σ is $\pm 2 - 2\frac{1 - s}{s}$). These results agree with the finite N model.

2.4.3.2. Continuum limit two-morph steady state. Next, we investigate the stability of the two-morph steady state. Similar to our investigation of the uniform steady state, we "perturb" the two-morph steady state to the weighted sum of two narrow Gaussian distributions

(2.16) $$p(a, t) = \frac{x}{\sigma_1(t)\sqrt{2\pi}}e^{-(a-a_1)^2/2\sigma_1(t)^2} + \frac{1 - x}{\sigma_2(t)\sqrt{2\pi}}e^{-(a-a_2)^2/2\sigma_2(t)^2},$$

where a_1 and a_2 are given by the two-morph fixed point (2.8). Figure 2.5 (c),(d) illustrates this idea.

Plugging (2.16) into the continuity equation (2.9) and using $\bar{a} = xa_1 + (1 - x)a_2$, we get a system of two ordinary differential equations for the evolution of σ_1 and σ_2:

(2.17)
$$\frac{\mathrm{d}\sigma_1}{\mathrm{d}t} = \lambda_1\sigma_1 + \mathcal{O}(\sigma_1^3)$$
$$\frac{\mathrm{d}\sigma_2}{\mathrm{d}t} = \lambda_2\sigma_2 + \mathcal{O}(\sigma_2^3),$$

where λ_1 and λ_2 depend on $a_{\mathrm{opt}}, s, x,$ and γ (expressions omitted due to length). Setting $a_{\mathrm{opt}} = 1$ and $s = 1/2$ for instance, we plot the stability region (i.e., where $\lambda_1, \lambda_2 < 0$) for
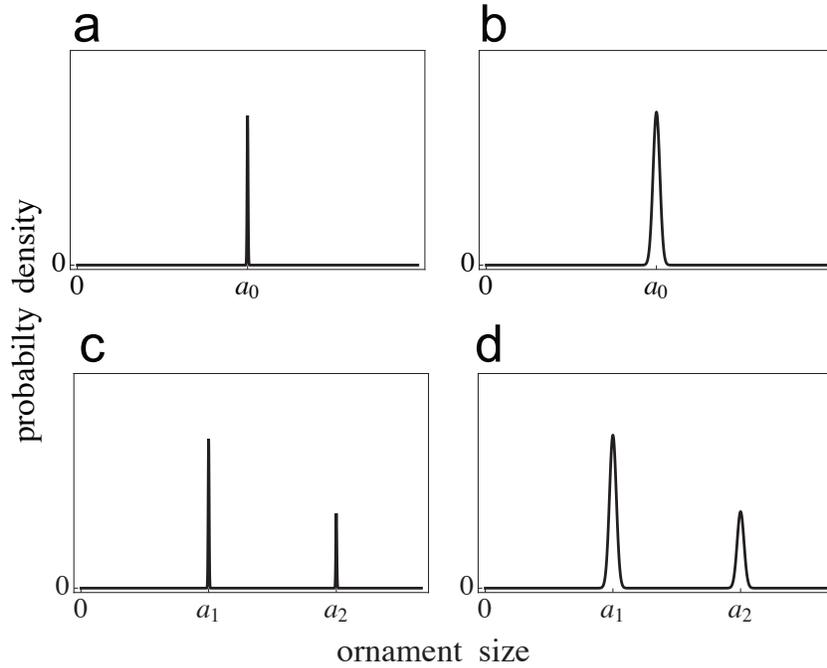
Figure 2.5. We consider perturbations to the uniform fixed point $a = a_{\text{opt}}$ and the two-morph fixed point in equation (2.8) such that the peaks of the distribution are centered at the fixed point solution, and the widths of the peaks are nearly 0. **(a)** Shift peak of the delta uniform solution to a_0. **(b)** Perturb peak width of the delta uniform solution. **(c)** Two-morph steady state. **(d)** Perturb peak widths of the delta two-morph solution.

the two-morph steady state in terms of social sensitivity γ and the proportion of males in the large-ornamented group. See Figure 2.4 (d). This is the same stability region we found numerically, which resolves the apparent discrepancy we saw when perturbing the locations of the peaks (but not the widths of the peaks) of the two-morph steady state distribution.

2.4.3.3. The connection between potential and fitness. Expressing our model in the continuum limit also allows us to connect our phenotype flux to the more traditional replicator equation. Many evolutionary dynamics problems begin with the replicator

equation [**36**], which in the continuum limit is as follows:

$$(2.18) \qquad \frac{\partial p}{\partial t} = p(a,t) \left[f(a,p) - \bar{f}(p) \right],$$

where p is the probability distribution of a continuous phenotype a at time t, f is the fitness of a phenotype (say, ornament size or brightness) given a population state, and $\bar{f} = \int_{-\infty}^{\infty} f(a,p)p(a,t) \, \mathrm{d}a$ is the average population fitness [**45**].

Given that probability must be conserved, the distribution of phenotypes must also follow the continuity equation

$$(2.19) \qquad \frac{\partial p}{\partial t} = -\frac{\partial}{\partial a} \left(p \frac{\mathrm{d}a}{\mathrm{d}t} \right).$$

This formulation differs from the replicator equation (2.18) in that it requires specification of the phenotype flux $\mathrm{d}a/\mathrm{d}t$ rather than fitness f. Our approach treats this flux as derivable from some potential function, which we refer to as φ, the net "reproductive potential" (see equation (2.3)).

Intuitively, the relationship between our phenotype flux $\mathrm{d}a/\mathrm{d}t$ and the more commonly used replicator equation approach (the upward distribution flux) can be seen in Figure 2.6. These reflect interchangeable ways of viewing the evolutionary process of optimizing the probability distribution $p(a,t)$.

We can express the relationship between the two approaches mathematically simply by equating the right-hand-sides of equations (2.18) and (2.19), yielding

$$(2.20) \qquad f - \bar{f} = -\frac{1}{p} \frac{\partial}{\partial a} \left(p \frac{\mathrm{d}a}{\mathrm{d}t} \right) = -c \left(\frac{1}{p} \frac{\partial p}{\partial a} \frac{\partial \varphi}{\partial a} + \frac{\partial^2 \varphi}{\partial a^2} \right)$$
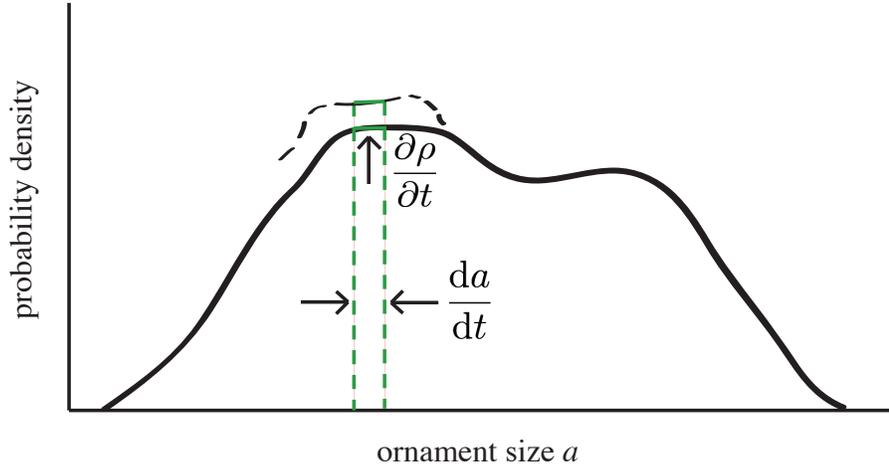
Figure 2.6. Consider an infinitesimal sliver (dashed green) of the probability density function at a particular time (solid black). After an infinitesimal time increment, the probability density function changes a small amount (dashed black). Because probability is conserved, the flux $\mathrm{d}a/\mathrm{d}t$ of population ornament sizes into (or out of) the sliver increases (or decreases) the height of the probability density function.

where the last equality makes use of equation (2.4). Integrating equation (2.20) once with respect to a and using equation (2.4) yields an integro-differential equation for φ in terms of f:

$$(2.21) \qquad \frac{\partial \varphi}{\partial a} = -\frac{1}{cp} \int_{-\infty}^{a} p \left(f - \bar{f} \right) \mathrm{d}a,$$

assuming $p \, \mathrm{d}a/\mathrm{d}t \to 0$ as $a \to -\infty$.

2.4.4. Eigenvalues of system as $N \to \infty$

When investigating the stability of the two morph steady state, we chose to take the continuum limit of the model and then investigate the dynamics of the standard deviation of a Gaussian perturbation to the two morph equilibrium. Now we look at the eigenvalues

of the finite N system in the limit $N \to \infty$. Scaling time such that $c = 1$, the Jacobian for the system (2.5) has diagonal elements

$$J_{ii} = s\gamma(\gamma - 1)\left(1 - \frac{1}{N}\right)^2 \mathrm{sgn}\,(a_i - \bar{a})|a_i - \bar{a}|^{\gamma - 2} - 2(1 - s),$$

and off-diagonal elements

$$J_{ij} = s\gamma(\gamma - 1)\left(-\frac{1}{N}\right)\left(1 - \frac{1}{N}\right)\mathrm{sgn}\,(a_i - \bar{a})|a_i - \bar{a}|^{\gamma - 2}.$$

As $N \to \infty$,

$$J_{ii} \to s\gamma(\gamma - 1)\,\mathrm{sgn}\,(a_i - \bar{a})|a_i - \bar{a}|^{\gamma - 2} - 2(1 - s)$$

$$J_{ij} \to 0,$$

indicating that for large N, the Jacobian matrix is approximately diagonal. Therefore, the diagonal elements are approximately the eigenvalues. Plugging in the two morph fixed point (2.8), we get two eigenvalues λ_1 and λ_2 with multiplicity xN and $(1 - x)N$ respectively. If we plot the stability region (i.e. where $\lambda_1, \lambda_2 < 0$), we see that it's the same as that of the continuum model seen in Figure 2.4 (d).

2.4.5. Fixed point basins of attraction

Numerically integrating with initial conditions very close to the fixed points verifies the analytical stability region. However, not all stable equilibria have equally large basins of attraction. Numerically, we found that most initial conditions lead to steady states close to the "frontline" of stability onset. This can be seen in a simple numerical experiment; we tried a range of increasingly perturbed initial conditions centered at the fixed points and looked at the eventual outcomes. See Figure 2.7.
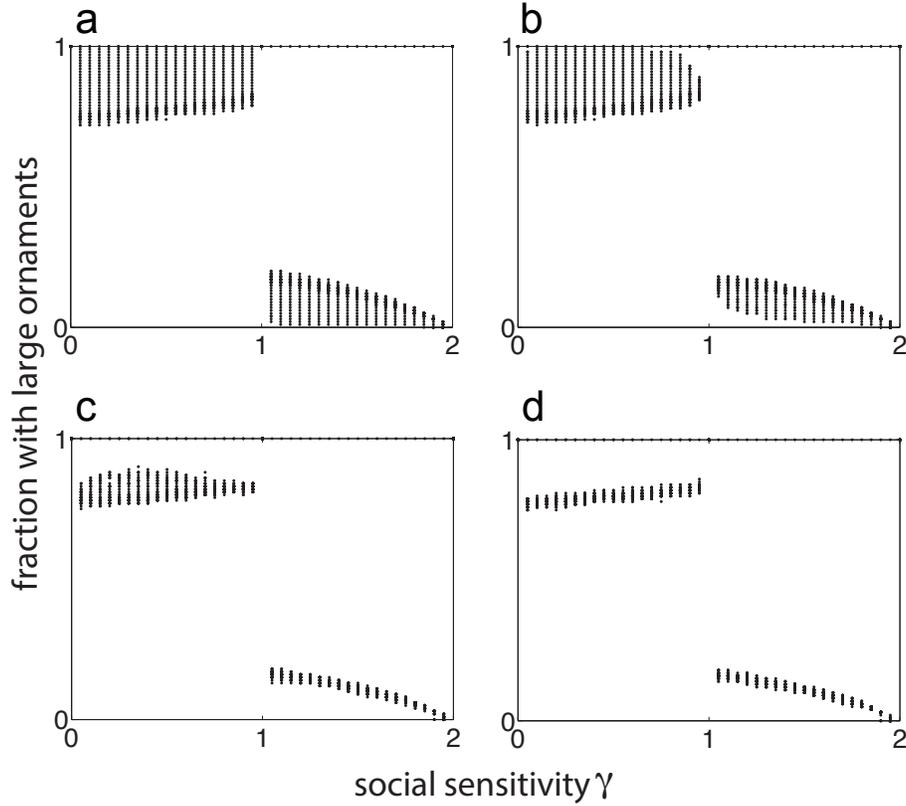
Figure 2.7. Numerical stability region for initial conditions near the analytical fixed points with increasing variances ($N = 100, s = 1/2$) **A.** $\sigma = 10^{-3}$ **B.** $\sigma = 10^{-2}$ **C.** $\sigma = 10^{-1}$ **D.** $\sigma = 10^0$

2.5. Model validation

We now revisit our simplifying assumption that all males are equally healthy. More realistically, we allow the intrinsic health h to be taken from some distribution (perhaps set by genetic, developmental, or environmental factors). Suppose this distribution is such that the individual optimal ornament size $a_{\text{opt}}(h)$ is normally distributed. Then the stable two-morph steady state changes from a weighted sum of perfectly narrow Dirac delta functions to a distribution roughly resembling the sum of two Gaussians—usually

a *bimodal* distribution. Marginal histograms in Figure 2.11 (a),(b) shows examples of steady states with varied intrinsic health.

These examples resemble data from many species that grow ornaments. Figure 2.11 (c),(d) shows two examples of real-world ornament distributions that exhibit bimodality. Note that we do not expect the exact shape of the real-world distributions to match our simulations because the measured quantities will not necessarily be linear in cost. However, bimodality will be preserved regardless of the measured quantity.

In a literature search [**51, 52, 53, 54, 55, 56, 26, 57, 58, 59, 25, 60, 61, 62, 40, 63**], we found a number of published data sets showing size distributions of suspected ornaments; 23 were of sufficient quality for testing agreement with this model. In 13 of those data sets we found some evidence for rejecting the hypothesis of unimodality: the data were more consistent with a mixture of two or more Gaussian distributions than with a single Gaussian. In six data sets, we found stronger evidence: non-parametric tests rejected the hypothesis of unimodality. Note that other data sets were not inconsistent with bimodality, but small sample sizes often limited the power of statistical testing. All histograms and statistical test results of data sets are in the Appendix.

2.5.1. Statistical analysis of ornamentation data

Our model for the evolution of costly mating displays predicts the emergence of two distinct morphs of ornament sizes. We tested whether the two-morph state was detectable in a variety of ornament datasets (Figures A.2, A.3). Three approaches were used: a parametric mixture model fit; the nonparametric but highly conservative Hartigans' Dip

Test for bimodality [**64**]; and a simulation–based nonparametric test which improves upon the Hartigan test sensitivity.

2.5.1.1. Parametric two-morph test. All count and size measurement data were log-transformed prior to analysis (as is typical for physical measurements) to account for the bounded support of size distributions. Here, we make the assumption that ornament sizes within a morph will be log-normally distributed, and that a multi-morph state will exhibit a mixture of distributions. We thus fit Gaussian mixture models with 1–5 components of unequal variance to the log-transformed data and find the number of components that yields the best Bayes Information Criterion (BIC) [**65**]. In the absence of a social fitness pressure, we expect the best fit to be a single Gaussian (corresponding to the one morph state), while the two morph state predicted from our model will have the best fit with two or more components.

2.5.1.2. Hartigans' dip test. An essential drawback of using the above mixture model fit to assess the number of ornament–size morphs in the data is that it is extremely sensitive to deviations from the parametric assumption that a one morph state will be well described by a single Gaussian. False positives are likely when those assumptions are violated; if a single morph state has a skewed (or otherwise non-normal) distribution, a mixture of two or more Gaussians will generally give a higher BIC than a single component distribution.

A more conservative approach is to look for evidence of strict multimodality (with dips in the distribution), rather than a mixture which may not exhibit a "dip". Hartigan and Hartigan define the dip statistic D as the maximum difference between the empirical

cumulative distribution function and the CDF of the unimodal distribution that minimizes that maximum difference. The reference distribution is customarily taken to be the uniform distribution, the least singly peaked of all unimodal distributions. The p-value for the dip is calculated by comparing D to those obtained from repeated samples of the same size drawn from a uniform distribution. The dip test thus measures whether the empirical distribution of the data exhibits greater departure from unimodality than would be expected from a sample of the same size if the underlying distribution were uniform.

2.5.1.3. Bootstrap dip test. While the mixture test may be overly sensitive in detecting deviations from a single morph, Hartigans' dip test is likely to be excessively conservative and insensitive at small sample sizes. A finite sample drawn from a uniform distribution will, with high probability, have a larger dip by chance than a finite sample drawn from a two morph distribution such as those shown in Figure 2.11 (a),(b).

To address this problem, we propose a bootstrap dip test which takes as its reference distribution the "least unimodal" unimodal density estimate of the sample. Given a finite sample, we construct a kernel density estimate (KDE) using a Gaussian kernel at various bandwidths. At very large bandwidths, the KDE will be unimodal; as the bandwidth is reduced, the KDE will approach a multimodal distribution with as many modes as there are unique values in the dataset. We define the least–unimodal unimodal (LUU) distribution to be that obtained from the smallest bandwidth for which the KDE is still strictly unimodal. Figure 2.8 demonstrates this step of the method visually.

From this LUU density estimate, we generate random samples of the same size as the original data and compute their dip statistics. These bootstrapped samples serve as the reference distribution against which the dip statistic of the data is compared. This test

thus measures whether the empirical distribution of the data exhibits greater departure from unimodality than would be expected from a sample of the same size if the underlying distribution were *the unimodal distribution best fit to the sample.* Figure 2.9 demonstrates this step of the method visually. Figure 2.10 illustrates that this bootstrap dip test is more sensitive to bimodality than Hartigans' Dip Test.
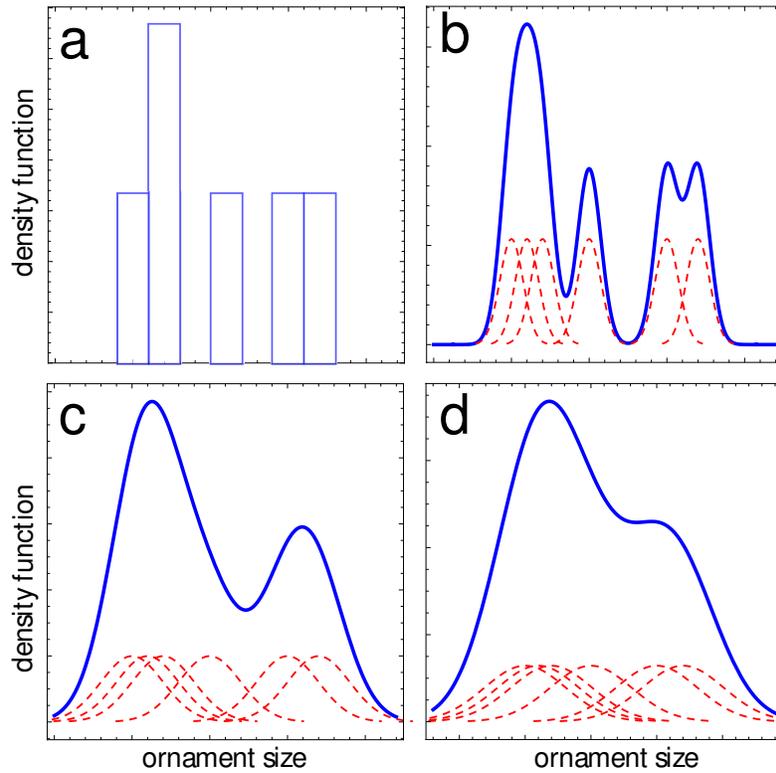
Figure 2.8. Illustration of least-unimodal unimodal (LUU) bootstrap method. **(a)** A hypothetical data set histogram with small sample size. **(b)-(c)** Normal kernel density estimates (solid blue) of the hypothetical data set with increasing standard deviations. As the width of the kernel increases, the kernel estimate is multimodel (b), bimodal (c), and least-unimodal unimodal (d).
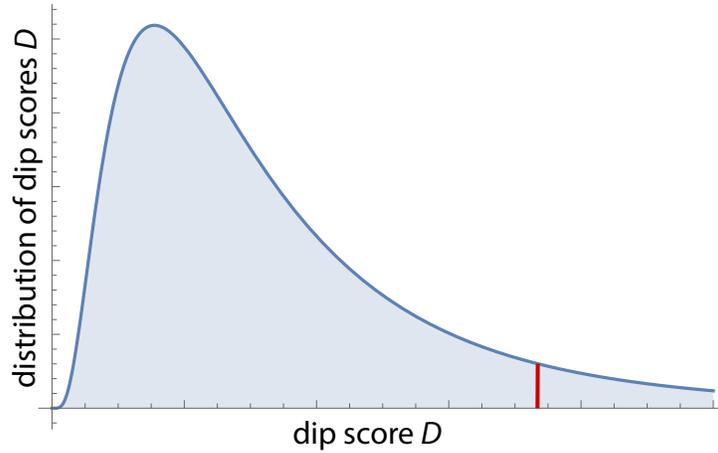
Figure 2.9. Illustration of least-unimodal unimodal (LUU) bootstrap method. Given a real data set of size N and a large number of bootstrapped samples of size N drawn from the LUU distribution, we plot the dip score of our data against the distribution of dip scores from all LUU samples. In this illustration, our real data dip score (solid red) has more 'dip' than the vast majority of bootstrapped data. Therefore, it is likely that our data is multimodal.

We present test results for Hartigans' Dip Test and the LUU (Least Unimodal Unimodal) test in Table A.1. Test results for the Gaussian mixture model fit are in Table A.2.

2.6. Discussion

2.6.1. Implications for honest signaling

Assuming this model adequately represents the handicap principle, we may ask if ornament size really does honestly advertise quality. In other words, if a female can choose among all the males, is she able to detect the healthiest (or weakest) males simply by
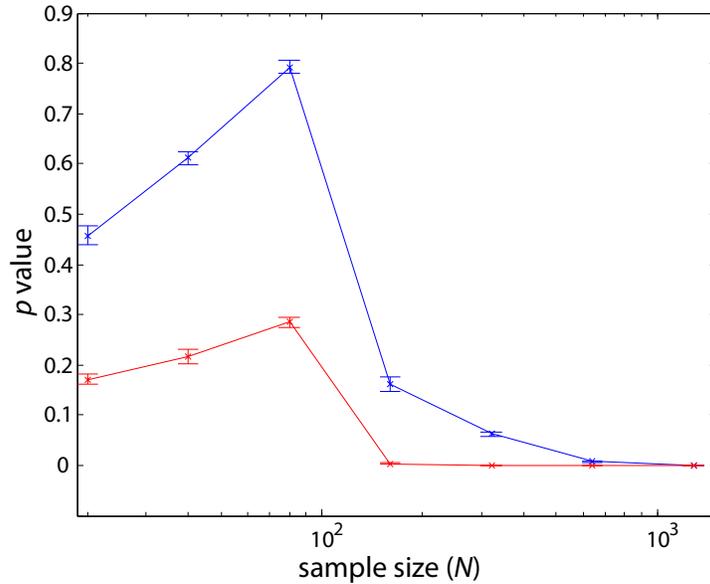
Figure 2.10. For small sample sizes of synthetic bimodal data, like we have for many of our animal data sets, the p-values for bimodality using Hartigans' Dip Test (blue) are larger than our bootstrap dip test (red). As the sample size increases, we gain significance using our test first and Hartigans' Dip Test eventually, showing our test is less conservative. The data used here are equilibrium states of our model (2.5) for $\gamma = 1.5, s = 0.5$, and a_{opt} drawn from a normal distribution with mean 1 and standard deviation 0.25. We know these samples are bimodal. Error bars are standard deviations from 10 trials.

looking at ornament size? Again taking the optimal ornament size a_{opt} to be normally distributed, we examine the Kendall rank correlation between intrinsic health (as indicated by our proxy a_{opt}) and equilibrium ornament size.

We find that the advertising is mostly honest, at least for large enough variance in health. Both observational and experimental work supports this finding [23]. Figure 2.11 (a),(b) show examples of ornament size versus intrinsic health based on our model.
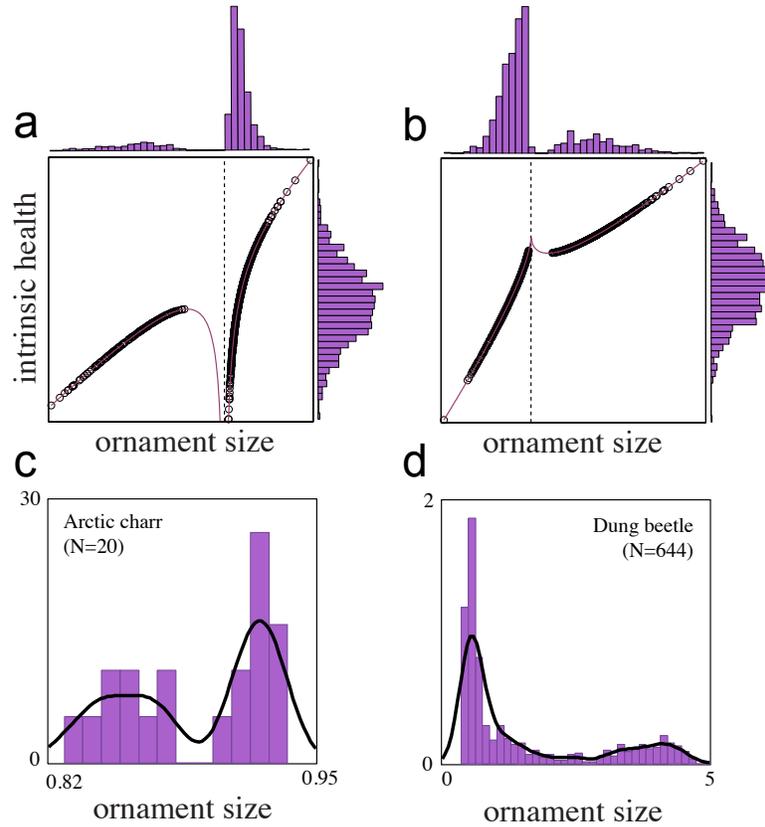
Figure 2.11. Ornament size distributions in model and real-world data. Due to smaller sample sizes in real-world data, we superimpose a kernel density estimate (KDE) over the histograms as a visual aid (solid black line). **(a)** Simulation of model with $N = 1000$ individuals, $\gamma = 0.5$, $s = 1/2$ (Kendall's rank correlation $\tau = 0.9149$). **(b)** Simulation of model with $N = 1000$ individuals, $\gamma = 1.5$, $s = 1/2$ (Kendall's rank correlation $\tau = 0.9998$). In both (a) and (b), black dashed line ($a = \bar{a}$) shows division between morphs, solid maroon curve shows analytical solution. Marginal histograms illustrate that normal distribution of a_{opt} (proxy for intrinsic health) leads to bimodal distribution of a. **(c)** Normalized histogram for arctic charr brightness [**63**] (N=20, KDE bandwidth=0.01). **(d)** Normalized histogram for dung beetle horn length [**60**] (N=644, KDE bandwidth=0.2).

2.6.2. Generality

It is natural to wonder about the generality of the results we have presented here. For a reasonable set of potential functions (described below), the only possible stable equilibria are multimodal distributions of ornament size. Again we consider a potential function

$$\varphi = s\,\varphi^{(\mathrm{soc})} + (1 - s)\,\varphi^{(\mathrm{ind})}, \quad s \in [0, 1]$$

where $\varphi^{(\mathrm{soc})}$ is a continuous and differentiable increasing function of ornament size, and $\varphi^{(\mathrm{ind})}$ is a continuous, singly-peaked function of ornament size. Assuming that the dynamics are such that ornaments grow on an evolutionary time scale at a rate proportional to marginal potential gain,

$$\frac{\mathrm{d}a}{\mathrm{d}t} \propto \frac{\partial}{\partial a}\varphi,$$

then we have $\dfrac{\mathrm{d}a}{\mathrm{d}t} = 0$ only for $a \geq a_{\mathrm{opt}}$. In other words, equilibrium ornament sizes will not be smaller than the optimal size for survival.

We further assume that the following two criteria are satisfied:

(1) Individual effects dominate reproductive potential for large ornament sizes. Specifically,

$$(2.22) \qquad (1 - s)\left|\frac{\partial}{\partial a}\varphi^{(\mathrm{ind})}\right| > s\left|\frac{\partial}{\partial a}\varphi^{(\mathrm{soc})}\right| \quad \text{as } a \to \infty.$$

This prevents ornament size from growing without bound, as can occur in equation (2.5) for $\gamma \geq 2$.

(2) Social effects dominate reproductive potential for at least some range of ornament sizes greater than the population mean. In other words,

$$(2.23) \qquad (1-s)\left|\frac{\partial}{\partial a}\varphi^{(\text{ind})}\right| < s\left|\frac{\partial}{\partial a}\varphi^{(\text{soc})}\right|$$

for at least some range of $a > \bar{a}$. Failure to meet this criterion could be considered "false" ornamentation, as can occur in equation (2.5) for $\gamma = 1$.

Assuming that the two-sided limits exist everywhere for both potential functions (a less strict requirement than continuity), these criteria guarantee that two or more morphs will emerge. See Figure 2.12 for graphical proof.
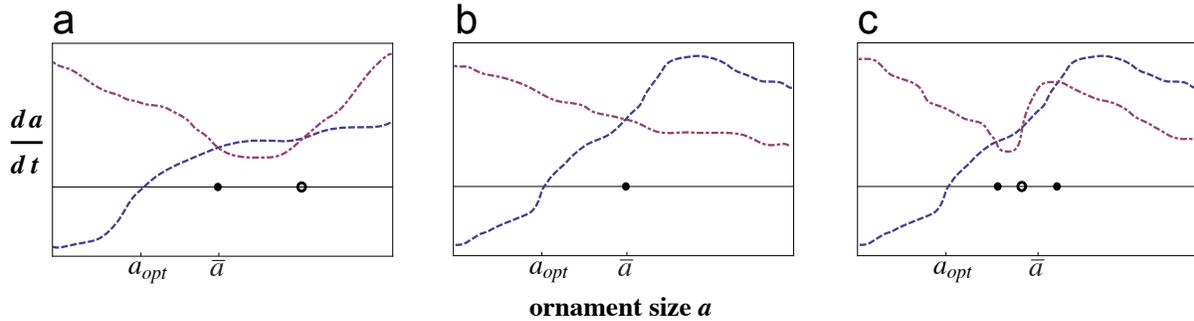
Figure 2.12. Schematic of derivatives of negated individual potential (dashed blue) and social potential (dot dashed maroon) for a single male in a population near equilibrium. The derivative of total potential is proportional to da/dt, so intersections of individual and social potentials are the fixed points. Stable fixed points are marked with a filled black dot, and unstable fixed points are marked with an unfilled black dot. **(a)** An example of potential functions that satisfy restriction (2.23), but not restriction (2.22). In this case, both a stable uniform state and unbounded growth are possible. **(b)** An example of potential functions that satisfy restriction (2.22), but not restriction (2.23). In this case, the population will evolve to a uniform state. **(c)** An example of potential functions satisfying both restrictions (2.22) and (2.23). These conditions guarantee that the population will evolve into at least two morphs.

One benefit of our modeling approach is that it is agnostic to the genetic mechanism by which the two-morph distribution is maintained. That is, we make no assumptions about the genetics other than presuming that maintenance of such a distribution is possible (by some mechanism). Rather, we demonstrate that the bimodal distribution—and thus a mechanism to maintain it—is an emergent, evolutionarily favorable consequence of Zahavi's handicap principle.

There are several molecular mechanisms by which a population with multiple optima can be maintained, including over-dominance, complex polygenic or epistatic relationships, or epigenetic modifications. Any of these (amongst others) could be involved in the maintenance of the predicted dimorphic trait. Because our model makes no assumptions (or predictions) about which of these mechanisms maintain the two-morph state, it is general to any organismal trait where Zahavi's handicap principle applies, and is insensitive to assumptions about the genetic architecture.

Of course, we do not wish to imply that our model is the only possible explanation (or even necessarily the dominant effect) where polymorphism is observed. The importance of this effect probably varies from species to species and ornament to ornament. Our model applies most naturally to inter-sexual selection (female choice as the dominating force), and in the interest of simplicity we ignore alternative reproductive strategies (e.g., female mimicry by males). We believe that it may be possible to generalize our model to include effects like negative-frequency dependent selection (e.g., as another type of social effect that would impact the shape of the social potential function in our model), but we leave that for future work.

2.6.3. Extension to speciation models

We speculate that the mechanism we describe here may also have implications for speciation. Models of speciation presented in Lande [35] and Stewart [66] are similar to our ornamentation model in both form and outcome. Stewart claims that for an all-to-all system of behaviorally identical individuals (like ours), the population will split into two species for most environmental conditions. Like our social sensitivity γ, Stewart's environmental factor λ varies on a slow time scale relative to the dynamical system. Also like our model, Stewart's model exhibits similar fractionation (simulating 100 individuals, the population splits into "clumps" of 84 and 16).

Lande uses quantitative genetics techniques to show that sexual selection may lead to speciation. Our model is quite similar to Lande's model interpreted on a logarithmic scale. Like our model, Lande's sexual selection alone would lead to runaway ornament sizes, but natural selection stabilizes growth. Unlike our model, Lande states that "natural selection on mating preferences also creates the possibility of evolutionary oscillations." Because we ignore the long time scale effects of female choice, our model precludes the possibility of oscillations.

2.7. Conclusions

The independent evolution of costly ornamentation across species has puzzled scientists for over a century. Several general evolutionary principles have been proposed to explain this phenomenon. Among the prominent hypotheses is the handicap principle, which posits that only the healthiest individuals can afford to grow and carry large ornaments, thereby serving as honest advertising to potential mates. We base a minimal model

on this idea and find that, surprisingly, it predicts two-morph stratification of ornament size, which appears to be common in nature.

Importantly, the two morphs both have ornament sizes larger than the optimum for lone individuals. This means that the population survival potential, as indicated by the population average of individual potential $\overline{\varphi}^{(ind)}$, is reduced. Due to the presence of ornaments, we conclude that the evolutionary benefits of honest advertising must outweigh the net costs of ornamentation when the displays exist in nature.

CHAPTER 3

Hybrid statistical and mechanistic model guides mobile health intervention for chronic pain

3.1. Introduction

In the fields of physics, chemistry, and engineering, models are often derived from mechanistic fundamental laws expressed in the form of differential equations. Resulting dynamical systems models can be used both to gain intuition into the expected behavior of the system, and to make specific predictions about results of experiments (e.g., see [67]). In fields such as social sciences, bioinformatics, and medicine, models are often constructed from data via statistical inference, without direct derivation from fundamental principles (e.g., see [68]). The mechanistic and statistical approaches to mathematical modeling have different advantages. The former allows prior knowledge to be introduced and validated or rejected based on the success of the model. The latter requires almost no a-priori information about how the system is expected to behave.

Here we present a hybrid approach to mathematical modeling that incorporates both mechanistic and statistical elements, with the goal of gaining a deeper understanding of the human experience of subjective pain. Specifically, we hope to predict how patient-reported pain levels vary over time based on medication dosage information and other patient characteristics.

3.1.1. Application to pain

Sickle cell disease (SCD) is a chronic illness associated with frequent medical complications and hospitalizations. Approximately 90% of acute care visits are for pain events, and 30-day hospital re-utilization rates are alarmingly high [**69**]. While factors influencing these high re-utilization rates are poorly understood, close follow-up and continued use of pain medications has been shown to decrease re-hospitalization rates. Mobile technology has become an integral part of health care management, and our recently self-developed mobile application (Sickle cell Mobile Application to Record symptoms via Technology, or SMART app—see Figure 3.1) for SCD assists with documentation and intervention of pain.

Pain in particular is difficult to quantify and has never before been monitored at the temporal scale we report here across so many patients. It is known that subjective pain, though indeed subjective, is correlated with objective measurable stimuli qualities in experiments (see, e.g., [**70, 71, 72**]). Thus there is reason to believe that subjective pain may follow understandable dynamics in time, especially when mitigated by opioid or non-opioid drugs. Our approach to the problem is motivated by the hope that a reasonable model for pain dynamics will yield some level of predictive power, despite the clear expectation that there will also be significant noise within and across patients. We can attribute the stochastic variation to sources like patient mood, temporal changes in patient state, weather, etc. In contrast, we hope that patient attributes like age, gender, SCD disease type, etc. will remain roughly constant on the time scale of the experiment and allow us to explore possible correlation of these attributes with model parameters.

3.1.2. Data source: mobile health application

We seek to understand the temporal dynamics of chronic pain as experienced by SCD patients. To that end, we have developed a mobile phone application that allows patients to record medication usage and subjective pain levels (measured on a 0-10 scale) in real time [**73, 74**].

Figure 3.1 shows several images of the application interface, while Figure 3.2 shows a typical data set resulting from a single patient's use of the app over the course of several weeks.
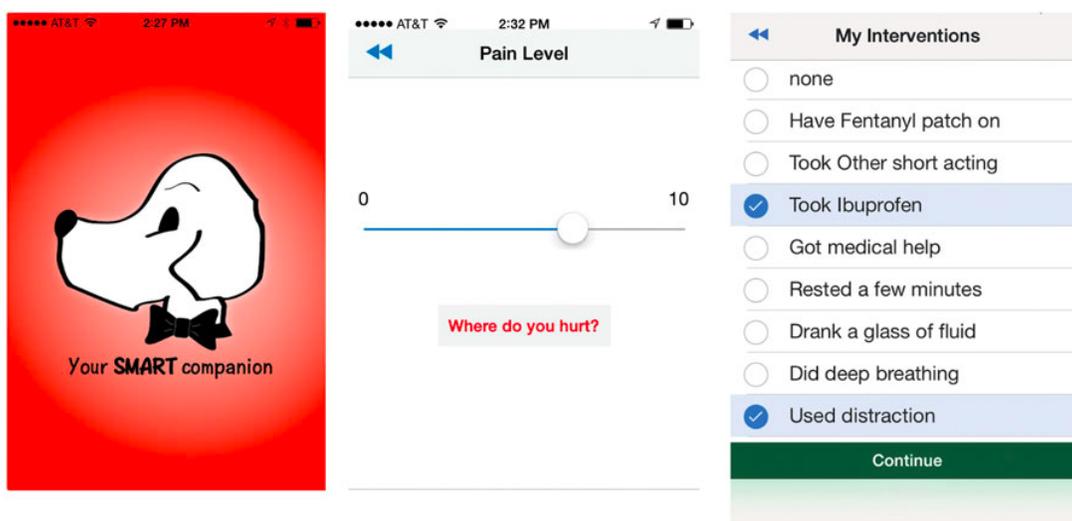
Figure 3.1. **Smartphone app.** Sample images of SMART application for iPhone/Android smartphone devices.

3.2. Materials and methods

3.2.1. Data

As of October 2016, data were available from 47 patients using the SMART app. Data sets from 8 of those patients were excluded because of excessive sparsity based on the
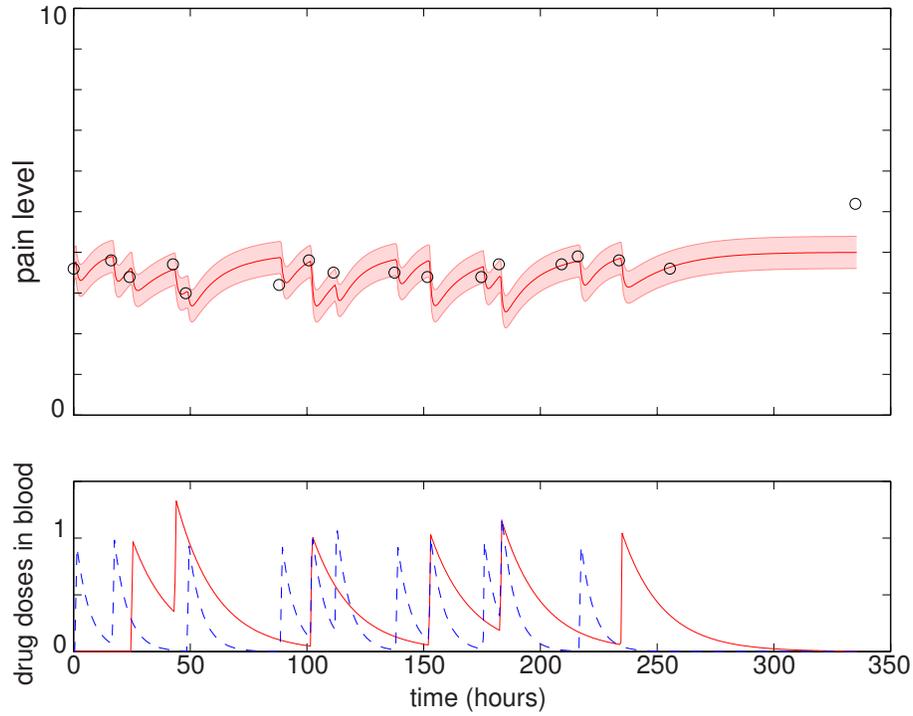
Figure 3.2. **Sample pain and medication data from a single patient.** Upper panel: patient reported pain (black circles) and model fit (red solid line); red shading indicates model fit plus/minus one standard deviation. Lower panel: long-acting methadone (red solid line) and short-acting oxycodone (blue dashed line) medication concentrations in patient bloodstream as inferred from medication usage reported via the SMART application.

following criteria: (1) total number of reports ≤ 5; or (2) pain reports never exceeded zero during the period under consideration. See Table 3.1 for demographic details of included patients. We denote the sample size $n = 39$.

Demographic characteristics

		N	(%)	
Institution				
	A	14	(35.9)	
	B	17	(43.6)	
	C	8	(20.5)	
Gender				
	Male	16	(41.0)	
	Female	23	(59.0)	
Age at baseline (years)				
	18-34	24	(61.5)	
	> 34	15	(38.5)	
SCD disease type				
	Hemoglobin SC	8	(20.5)	
	Hemoglobin SS	22	(56.4)	
	Hemoglobin SB+ (Beta) Thalassemia	5	(12.8)	
	Beta-Zero Thalassemia	3	(7.7)	
	SO−Ara	1	(2.6)	
Hydroxyurea user		27	(69.2)	
Folic acid vitamin user		26	(66.7)	
Long-acting opioid user		29	(74.4)	
Short-acting opioid user		35	(89.7)	
Non-opioid user		29	(74.4)	
		Mean	SD	(Min, Max)
Number of pain reports		67.2	60.4	(9.0, 257.0)
Days of pain reports		164.6	109.6	(10.3, 435.1)
Within-patient average VAS score		4.7	2.1	(0.3, 9.4)
		Mean	SD	(Min, Max)
Number of pain reports (first 2 weeks)		13.2	9.6	(2.0, 45.0)
Number of long-acting opioid doses (first 2 weeks)		6.0	8.4	(0.0, 35.0)
Number of short-acting opioid doses (first 2 weeks)		7.2	7.5	(0.0, 35.0)
Number of non-opioid doses (first 2 weeks)		2.1	3.1	(0.0, 12.0)

Table 3.1. Patient demographic information and the number of pain reports supplied by patients across entire study.

3.2.2. Predictive model

In order to develop a hybrid model that incorporates both a mechanistic a-priori knowledge-driven component and a statistical data-driven component, we divide tasks into two disjoint sets that fit these two categories; see Section 3.5.1 for more context.

We begin with a dynamical systems model for subjective pain motivated by the hypothesis that human sensory systems function on a roughly "return to setpoint" basis [**75, 76, 77, 78**]. Any model of human pain response, however, will inevitably require specification of a variety of parameters determining the time scale(s) and degree of severity of the response. The statistical modeling tasks employ patient data to infer parameters (1) from patient characteristics and population distributions and (2) from patient-specific pain and medication response history.

To make this more concrete, in Figure 3.3 we present a flow chart summarizing our approach to the hybrid modeling problem. Steps I_2 and A comprise the statistical modeling component; steps B and C comprise the mechanistic modeling component. A further optimization step D builds on the predictions of the hybrid model to allow for a balance between competing demands of pain reduction and medication usage minimization. This work details steps I_1, I_2, and A – E. We leave the remaining steps for future work.

3.2.2.1. Mechanistic component. We propose and evaluate two related mechanistic models based on a set of coupled ordinary differential equations (ODEs), either (a) deterministic or (b) stochastic. The stochastic differential equation (SDE) model comprises a Langevin equation, which can be converted into a Fokker-Planck partial differential equation (PDE) for the evolution of the probability distribution for pain $\rho(P, t)$ [**79**]. This allows for prediction of both the expected pain level for a patient at any point in the future *and* an assessment of the confidence in (and a confidence interval for) that prediction.
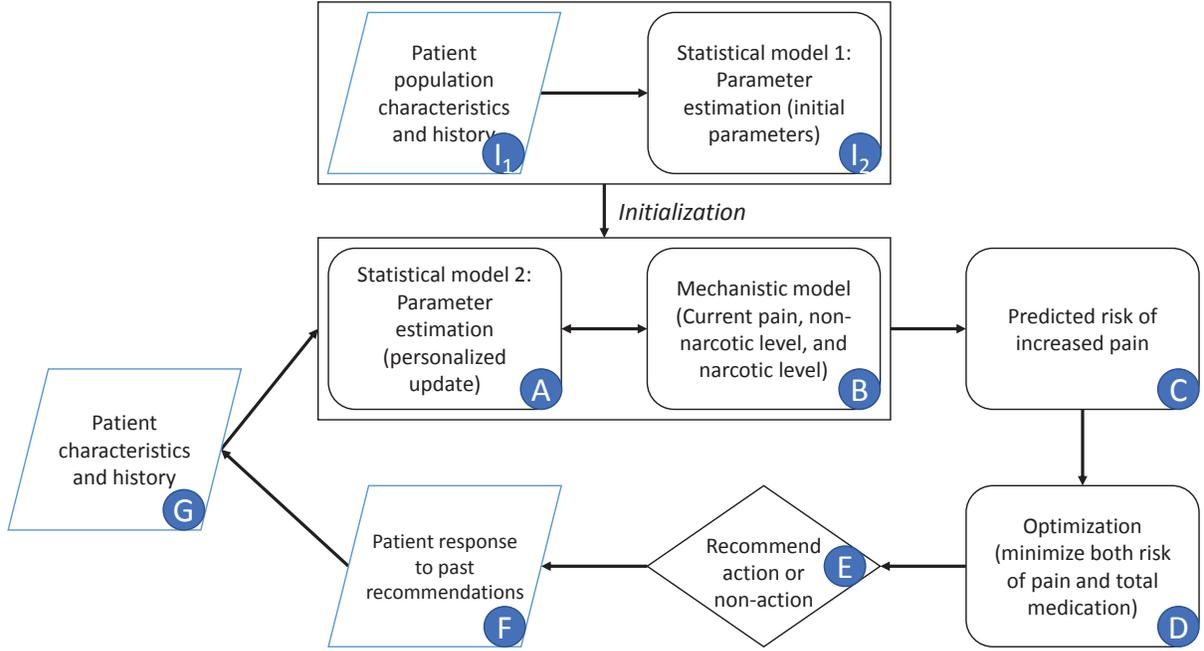
Figure 3.3. **Schematic flowchart showing model framework.**
Rounded rectangles represent modeling or computation steps, rhombuses
represent data inputs or outputs, and diamond represents decision step.
Items I_1 and I_2 are only necessary for initialization of the model. Items A
through E are the focus of this work.

Mathematically, the deterministic mechanistic model we propose is the following, for
a single patient:

<div style="text-align:left">(3.1)</div>

$$\frac{\mathrm{d}P}{\mathrm{d}t} = -(k_0 + k_1 D_1 + k_2 D_2 + k_3 D_3)P + k_0 u$$

$$\frac{\mathrm{d}D_1}{\mathrm{d}t} = -k_{D_1} D_1 + \sum_{j=1}^{N_1} \delta(t - \tau_{1,j})$$

$$\frac{\mathrm{d}D_2}{\mathrm{d}t} = -k_{D_2} D_2 + \sum_{j=1}^{N_2} \delta(t - \tau_{2,j})$$

$$\frac{\mathrm{d}D_3}{\mathrm{d}t} = -k_{D_3} D_3 + \sum_{j=1}^{N_3} \delta(t - \tau_{3,j}),$$

where P is the patient pain level (on a scale of 1–10), k_0 is the pain relaxation rate without drugs, k_i is the marginal effect on the pain relaxation rate due to drug i ($i = 1, 2, 3$), u is the unmitigated pain level (i.e. without drug intervention), D_i is the amount of standard drug i doses within the patient, k_{D_i} is the elimination rate of drug i within the patient, $\{\tau_{i,j}\}_{j=1}^{N_i}$ are the drug i dosage times, and N_i is the number of doses of drug i taken. δ represents the Dirac delta function. Note that the parameters and variables will in general need to be indexed with distinct values for each patient in a population, though we omit those indices here for clarity. Tables 3.2 and 3.3 summarize the meanings of model variables and parameters, respectively.

Variable	Meaning	Units
$P(t)$	Instantaneous pain level on 0–10 scale	[pain]
$D_1(t)$	Concentration of drug 1 (long-acting opioid) in the body	[standard doses]
$D_2(t)$	Concentration of drug 2 (short-acting opioid) in the body	[standard doses]
$D_3(t)$	Concentration of drug 3 (non-opioid) in the body	[standard doses]
$\rho(P, t)$	Instantaneous probability distribution of pain level P	[probability]

Table 3.2. Variables in mechanistic models.

Parameter	Meaning	Units
u	unmitigated pain level	[pain]
k_0	rate of decrease of pain in the absence of drugs or acute sources of pain	$[T^{-1}]$
k_1	effect of drug 1 (long-acting opioid) on pain relaxation rate	$[T^{-1}]$
k_2	effect of drug 2 (short-acting opiod) on pain relaxation rate	$[T^{-1}]$
k_3	effect of drug 3 (non-opioid) on pain relaxation rate	$[T^{-1}]$
k_{D_1}	rate of decay of drug 1 (long-acting opioid) in body due to metabolism	$[T^{-1}]$
k_{D_2}	rate of decay of drug 2 (short-acting opioid) in body due to metabolism	$[T^{-1}]$
k_{D_3}	rate of decay of drug 3 (non-opioid) in body due to metabolism	$[T^{-1}]$
ε	amplitude of intrinsic variability in human subjective pain reports	$[T^{1/2}]$
N_i	number of standard drug i doses taken	[count]
$\{\tau_{i,j}\}$	drug i dose times (indexed by j)	$[T]$

Table 3.3. Parameters in mechanistic models.

In this simple model for pain dynamics (3.1), pain is expected to relax at rate k_0 to unmitigated level u set by aggravating factors (like sickle cell disease) in the absence of intervention through opioids (drugs 1 and 2) or non-opioids (drug 3). When drugs are present in the patient's body, pain drops at a faster rate and the short-term equilibrium pain level (not the unmitigated pain level u) is reduced. Note that we treat all parameters as constant over the time period of interest, which we take to be two weeks (based on clinical heuristic experience).

In the model for drug concentrations, medication in the body is assumed to be metabolized at a constant rate. Rates can be determined from existing substantiated pharmacokinetic models (e.g., [**80, 81**]); Dirac delta function onset of medication serum concentration is a good approximation to the fast rise typical of the medications under consideration. See Figure 3.2 for a sample deterministic model output.

Note that we deliberately chose to employ an extremely simple conceptual model for pain dynamics. More sophisticated versions might be developed to incorporate higher order dynamics for P, or to include nonlinear or nonautonomous effects (e.g., allowing for explicit parameter variation with time of day or year), but currently available data are insufficient to constrain a model of greater complexity.

The stochastic differential (Langevin) equation version of our mechanistic model is as follows:

$$dP = -(k_0 + k_1 D_1 + k_2 D_1)P dt + k_0 u(dt + \varepsilon dW)$$

(3.2)
$$dD_1 = \left(-k_{D1}D_1 + \sum_{j=1}^{N_1} \delta(t - \tau_{1,j})\right) dt,$$

$$dD_2 = \left(-k_{D2}D_2 + \sum_{j=1}^{N_2} \delta(t - \tau_{2,j})\right) dt$$

$$dD_3 = \left(-k_{D3}D_3 + \sum_{j=1}^{N_3} \delta(t - \tau_{3,j})\right) dt,$$

where a hypothesis of uncorrelated additive white noise has been made. From this we derive the Fokker-Planck equation for the probability distribution of pain over time $\rho(P, t)$:

$$(3.3) \quad \frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial P}\left[\left(-(k_0 + k_1 D_1 + k_2 D_2 + k_3 D_3)P + k_0 u\right)\rho(P, t)\right] + \frac{\partial^2}{\partial P^2}\left[\frac{1}{2}(\varepsilon k_0 u)^2 \rho(P, t)\right].$$

Absent any pain medication, this Fokker-Planck equation implies the steady-state pain distribution

$$(3.4) \qquad \rho^*(P) = \sqrt{\frac{1}{\pi k_0 u^2 \varepsilon^2}} \exp\left[-\frac{(P - u)^2}{k_0 u^2 \varepsilon^2}\right],$$

a Gaussian distribution with mean u and standard deviation $u\varepsilon\sqrt{k_0/2}$. See Figure 3.4 for a sample stochastic model output.

3.2.2.2. Statistical component. In order to account for the variation among patients and improve prediction of the fitting parameters, we associate patient characteristics and history with each fitting parameter. The total number of fitting parameters varies
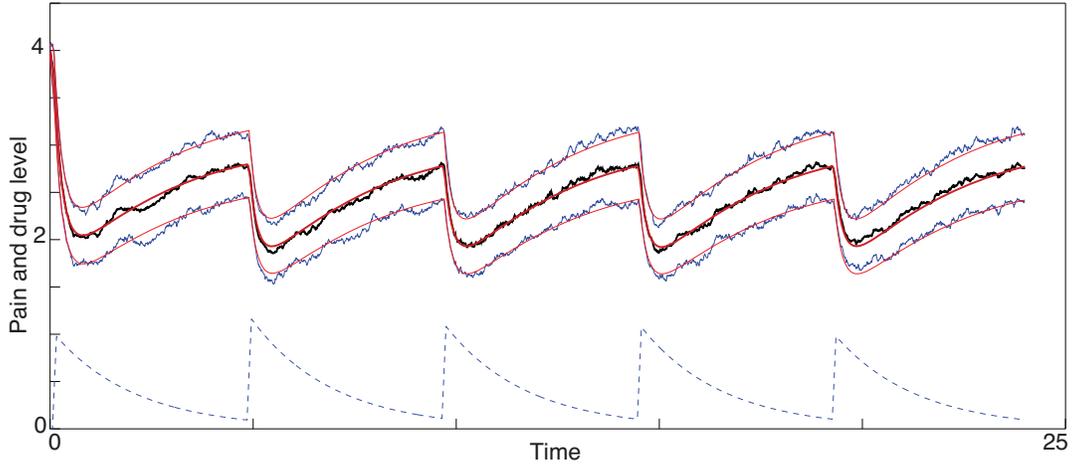
Figure 3.4. **Sample output from stochastic differential equation model** (3.2)**.** Red thick line: theoretical mean pain; red thin lines: $\pm$ one theoretical standard deviation; black thick line: mean of pain distribution in ensemble of 100 stochastic simulations; blue thin lines: $\pm$ one standard deviation in ensemble of 100 stochastic simulations; blue dashed line: drug 1 dose in bloodstream. Spikes occur when patient takes recommended dosage.

among patients: those who have taken no drugs will only have one fitting parameter, the unmitigated pain level u, and those who have taken more classes of drugs will have more drug fitting parameters.

To illustrate the statistical component, we focus on the unmitigated pain level u because, unlike drug parameters, all patients will have an estimate for u. In this case, we associate patient characteristics with u, an n-dimensional vector with u_j corresponding to the jth patient's unmitigated pain level, using a linear model.

Let X be an $n \times p$ design matrix containing the covariates of patients (i.e., patient characteristics). We write $X = (X_1, \ldots, X_n)^T$, with X_j corresponding to the p-dimensional covariates of patient j. Then we formulate the relationship between between u and the p

predictors as:

$$(3.5) \qquad\qquad u = X\beta + \epsilon,$$

where β is a p-dimensional coefficient vector, and ϵ is an n-dimensional vector of zero-mean random errors. When p is small, the estimate for β is obtained using the ordinary least squares procedure: $\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \|u - X\beta\|_2^2$, where $\|\cdot\|_q$ denotes the ℓ_q norm. Then the unmitigated pain level u_j is updated by $u_j^{new} = X_j^T \hat{\beta}$, $j = 1, \ldots, n$.

Because the unmitigated pain levels are not observable from patient pain reports, the initial u_j's are independently sampled from a uniform distribution between 0 and 10, i.e. $u_j^0 \sim U(0, 10)$. After using $\{u_j^0\}$ as initial values to fit the mechanistic model, the resulting estimated $\{u_j\}$ will be updated by the linear model (3.5) as $\{u_j^{new}\}$, which will then be used as initial values in the next round of fitting of the mechanistic model. See Section 3.2.3 for more detail on the hybridization of the statistical component with the mechanistic component.

Given a high-dimensional set of patient characteristics, we need to select a subset of patient characteristics that are significantly associated with u by minimizing the penalized loss function. In this study, we select patient characteristics using the LASSO (Least Absolute Shrinkage and Selection Operator) [82], by minimizing the penalized loss function $\Gamma(\beta) = \|u - X\beta\|_2^2 + \lambda\|\beta\|_1$ with respect to β. The penalty parameter λ is determined using 5-fold cross-validation. The selected p features are then used to fit the linear model (3.5) by ordinary least squares.

If time-varying unmitigated pain levels and time-varying covariates are present, the regression model (3.5) can be extended to the linear mixed model [83, 84]: $u = X\beta +$

$Z\delta + \epsilon$, where $Z = (Z_1, \ldots, Z_n)^T$ is an $n \times r$ design matrix for r random effect factors and $\delta = (\delta_1, \ldots, \delta_r)^T$ is a vector of random effects. Patient characteristics can be selected by maximizing the penalized log-likelihood: $\ell_{\mathrm{pen}}(\beta, \delta) = \ell(\beta, \delta) - \lambda \|\beta\|_1$ [85]. Such an extension of the proposed hybrid model to allow for time-varying unmitigated pain levels and covariates will be considered in a future study with more data available.

3.2.3. Model fitting

We fit our model to real patient data by minimizing the residual sum of squares between model predictions and patient reports provided within the first two weeks of reporting. We expect that the assumption of constant model parameters breaks down after approximately two weeks (clinical heuristics). Minimization over parameters u, k_1, k_2, and k_3 was done via the Nelder-Mead simplex algorithm [86]. Parameter k_0 was fixed at $2\ln(2) \approx 1.4$ corresponding to a pain equilibration half-life time scale of 30 minutes in the absence of medication. If a patient did not take all three classes of drugs, the model and fitting only included the consumed drugs.

We initialize the parameter optimization in n mechanistic models (one per patient) with random values during a first iteration, then we feed the optimization output into the statistical model (for all patients). Once the statistical model is run, it results in a new set of parameter estimates that can then be employed as initial parameter seeds for a second round of optimization in n mechanistic models (to minimize the residual sum of squares). Proceeding iteratively in this fashion (see Figure 3.3), we find convergence to a consistent set of parameters for each patient (details below).

3.2.4. Method verification

Before applying our hybrid model to real-world patient data, we verify the soundness of the approach with synthetic data constructed to resemble real-world data, but generated by the model itself with high sampling frequency. The synthetic data used for verification of the method are generated directly from the mechanistic model with an assumed parameter set generated in the following way: unmitigated pain $u =$ (patient age)$/10$, initial pain level $P(t = 0) = u - 2$, and drug parameters $k_1, k_2, k_3 \sim N(0.75, 0.25)$. Each patient reports pain every $1/2$ hour for 336 hours (two weeks). At each report time, the probability of the patient taking a particular drug (among three drug classes) is $1/16$; in other words, the patients take each drug on average every 8 hours. White noise of magnitude 1 is added to each pain report.

As an illustration using *real* patient drug times (specifically those of Patient A3), we create synthetic data generated using $u = 5, k_1 = 3$, and $k_2 = 2$: see Figure 3.5. When the initial parameter search is seeded with random parameter values, the mechanistic model fit can lead to convergence to either the true optimum $(5, 3, 2)$ or to other spurious optima with incorrect values of u, k_1, and k_2.

In this illustrative example, the method converges to $u = 5.01, k_1 = 3.19$, and $k_2 = 1.84$. The relevant root-mean-square (RMS) error is 1.01; this is close to the lowest possible expected error given the unit magnitude noise added to the synthetic data. This numerical experiment shows that the mechanistic model fitting method can converge even in the presence of significant amounts of noise. However, with *only* the mechanistic model

it can be quite difficult to find a good set of initial parameter seeds[1]: that is one motivation for introducing the statistical model.
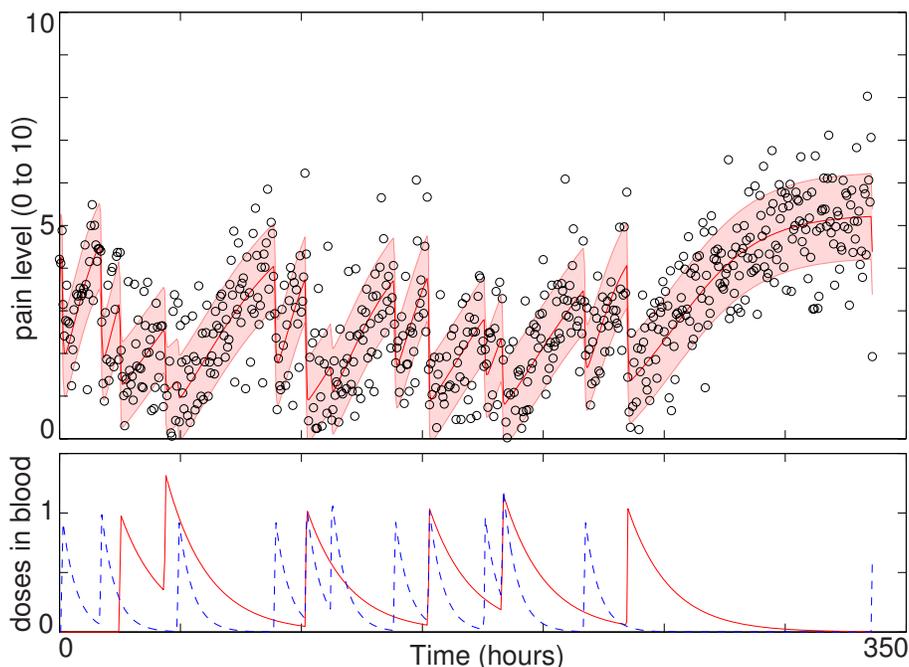
Figure 3.5. **Model fitting demonstration for densely reported noisy synthetic data.** Upper panel: hypothetical densely-reported patient pain (black circles) and model fit (red solid line); red shading indicates model fit plus/minus one standard deviation. Lower panel: long-acting opioid (red solid line) and short-acting opioid (blue dashed line) medication concentration in patient bloodstream.

To test our hybrid method using both the mechanistic model for fitting and the statistical model for parameter estimation, we create a synthetic patient database of 39 patients as described above. We then iterate rounds of fitting between mechanistic and statistical models, starting with uniform random guesses for all patient parameters (u, k_1, k_2, k_3). Figure 3.6 demonstrates how the parameter u converges to a value with small error after

[1]The seeding problem becomes exponentially harder as the dimension of the parameter space increases.

just a few iterations steps, even in the presence of significant noise. In order to evaluate the performance of the model on new data, we use the hold-out validation method by splitting the dataset into a training set (first week) and a test set (second week). Model fit error and hold-out validation error, as well as other parameters values, converge similarly: see Figure 3.7.
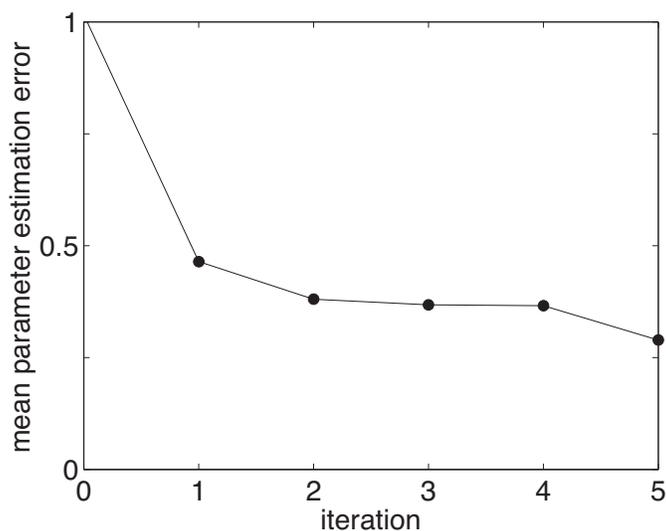
Figure 3.6. **Hybrid model fitting demonstration for ensemble of densely reported noisy synthetic data.** For an ensemble of 39 synthetic patient data sets, the average absolute error in u gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical + mechanistic) fits.

3.3. Results

3.3.1. General results

One key result is that our model for chronic pain does indeed have some predictive value (see Figure 3.8). This is an improvement over the current state of the art, since no other predictive model exists of which we are aware. Furthermore, fitted parameter values
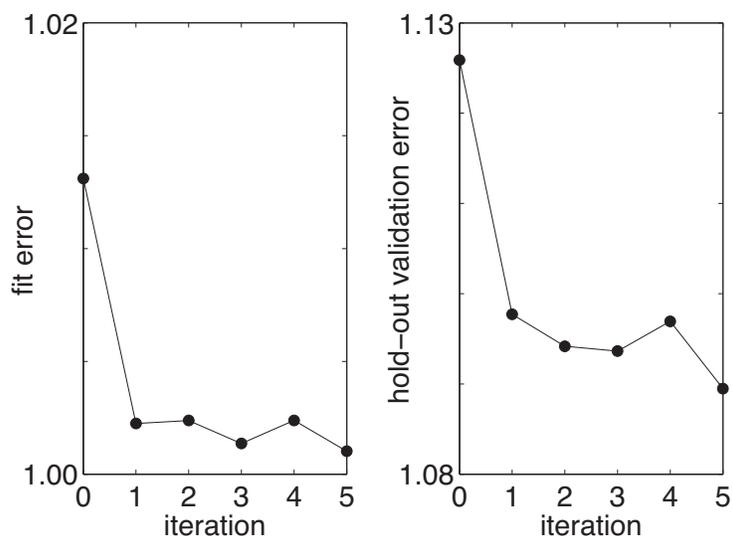
Figure 3.7. **Hybrid model fitting demonstration for ensemble of densely reported noisy synthetic data.** For an ensemble of 39 synthetic patient data sets, the average root-mean-squared (RMS) error in patient pain levels gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical+mechanistic) fits. Training error (or fit error) is on the left; test error (or validation error) is on the right. Due to the additive white noise of magnitude 1, the smallest testing or training error we could expect is 1.

correlate significantly with patient characteristics, suggesting that meaningful information is captured by this minimal plausible model. It may be possible to motivate new clinical insight on the basis of the observed correlations, perhaps leading to differential treatment of SCD sufferers with differing characteristics.

3.3.2. Statistical results

We use the following baseline patient characteristics to predict the unmitigated pain levels in the statistical modeling step: age, gender, SCD disease type, hydroxyurea use, folic acid vitamin use, long-acting opioid use, short-acting opioid use, and non-opioid use. We
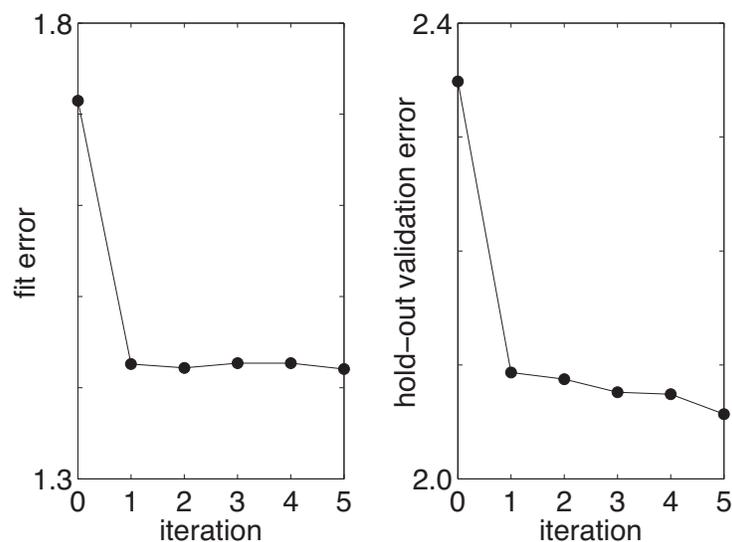
Figure 3.8. **Hybrid model fitting on real patient data.** For an ensemble of 39 real patient data sets, the average root-mean-squared (RMS) error in patient pain levels gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical+mechanistic) fits. Training error (or fit error) is on the left; test error (or validation error) is on the right.

explore the marginal effects of these characteristics and their possible pairwise two-way interactions using the LASSO. The model (3.5) can be extended to include time-varying covariates such as temperature, weather, patient's walking/social activities, and patient's mood at time t, once these data become available in a future study.

The statistical model that resulted from the LASSO variable selection is given by

$$u = \beta_0 + \beta_1 \mathbf{1}\{\text{SCD disease type} = \text{HgbSC}\}$$

$$+ \beta_2 \mathbf{1}\{\text{SCD disease type} = \text{SB+Thal or SO-Ara}\}$$

(3.6)
$$+ \beta_3(\text{age} - 18) + \beta_4 \mathbf{1}\{\text{Hydroxyurea user}\}$$

$$+ \beta_5 \mathbf{1}\{\text{Non-opioid user }\}$$

$$+ \beta_6 \mathbf{1}\{\text{SCD disease type} = \text{SB+Thal or So-Ara}\} \times (\text{age} - 18) + \epsilon,$$

where $\epsilon_j \sim N(0, \sigma^2)$, $j = 1, \ldots, n$, and $\mathbf{1}\{\cdot\}$ is the indicator function.

Table 3.4 summarizes the results from one round of fitting of the regression model (3.6). Adjusting for the effect of other terms in the regression model, SCD disease type of SB+Thal or So-Ara (with coefficient β_2), non-opioid use (with coefficient β_5), and the interaction term between SCD disease type of SB+Thal or So-Ara and age (with coefficient β_6) are important predictors of the unmitigated pain levels at the significant level of 0.05. Using non-opioid medication is associated with decreased unmitigated pain levels. Unmitigated pain levels increase with patients' age for SB+Thal or So-Ara patients.

Variable	Estimate	Std Err	T-value	P-value	
Intercept	7.646	1.228	6.228	0.000	***
HgbSC	-1.566	0.890	-1.761	0.088	
SB+Thal or So-Ara	-5.479	2.332	-2.349	0.025	*
Age at baseline -18	0.001	0.034	0.290	0.773	
Hydroxyurea user	-1.205	0.839	-1.437	0.160	
Non-opioid user	-2.523	0.842	-2.995	0.005	**
(SB+Thal or So-Ara) $\times$ (Age at baseline -18)	0.241	0.010	2.419	0.021	*

Table 3.4. Result of the prediction model of the unmitigated pain using the linear regression model. Significance levels: $^*(p < 0.05)$,$^{**}(p < 0.01)$,$^{***}(p < 0.001)$

3.3.3. Mechanistic model validation

With such sparse data and up to four fitting parameters, one may worry that the model (3.1) is being overfitted. To test this concern, we propose 6 related alternative models with fewer fitting parameters, and we compare cross-validation error and Akaike information criterion (AIC) among the models. See Table 3.5 for model descriptions. Neither measure selected a best-fit model across all patients, but none of these simple models is overfitting the data. See Figure 3.9 for AIC results and Figure 3.10 for cross validation results.

Model name	Description	Fitting parameters
Full model	Include all drugs taken (model (3.1))	up to 4
No drugs	Include no drug dosing information	1
Merge drugs	Combine all drugs into one drug class with same response	up to 2
LA only	Include only long-acting opioid doses	up to 2
SA only	Include only short-acting opioid doses	up to 2
NO only	Include only non-opioid doses	up to 2
Threshold	Include drug class only if drug is taken at least n times*	up to 4

Table 3.5. Mechanistic model variations. Fitting parameters include unmitigated pain level u and drug response parameters k_i for all drugs consumed. Therefore some patients have fewer fitting parameters than listed if they consumed fewer than three types of drugs. *In our tests, $n = 5$ was the drug dose threshold.

3.3.4. Biased pain reporting

Perhaps the most significant limitation of our model lies in a potential bias in our data set. Patients typically report pain levels when taking medication, but many of them only take medication when pain levels rise. Thus we suspect a selection bias of unknown significance,
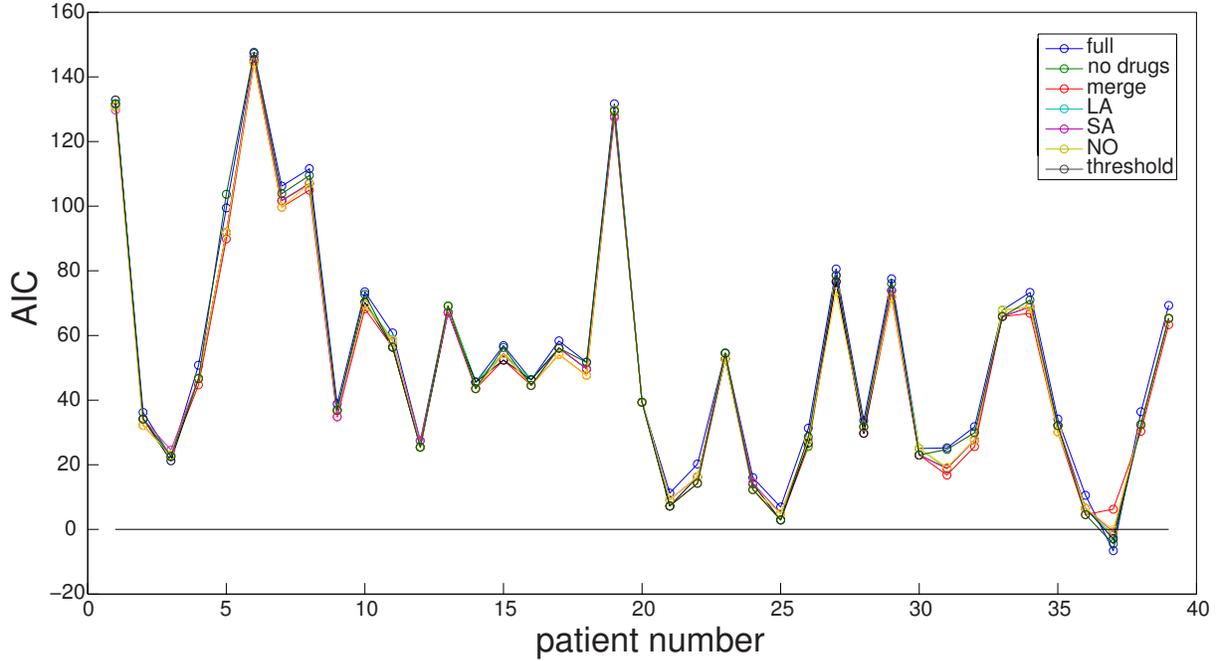
Figure 3.9. Akaike information criterion (AIC) for alternative models listed in Table 3.5. Most models perform equally well; among patients with differing model performance, there exists no clear 'best' model for all patients.

causing higher pain levels to be reported at a disproportionately high rate. To test this concern, we compare the unbiased model (3.1) with a similar model incorporating biased pain reporting.

Suppose the probability density function of pain at a particular time is a normal distribution with mean μ and variance σ^2:

$$(3.7) \qquad \rho(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Integrating model (3.1) gives the expected pain value μ at any point in time.

If higher pain is disproportionately reported through the mobile health application, then we will be much more likely to see higher pain levels from this normal distribution.
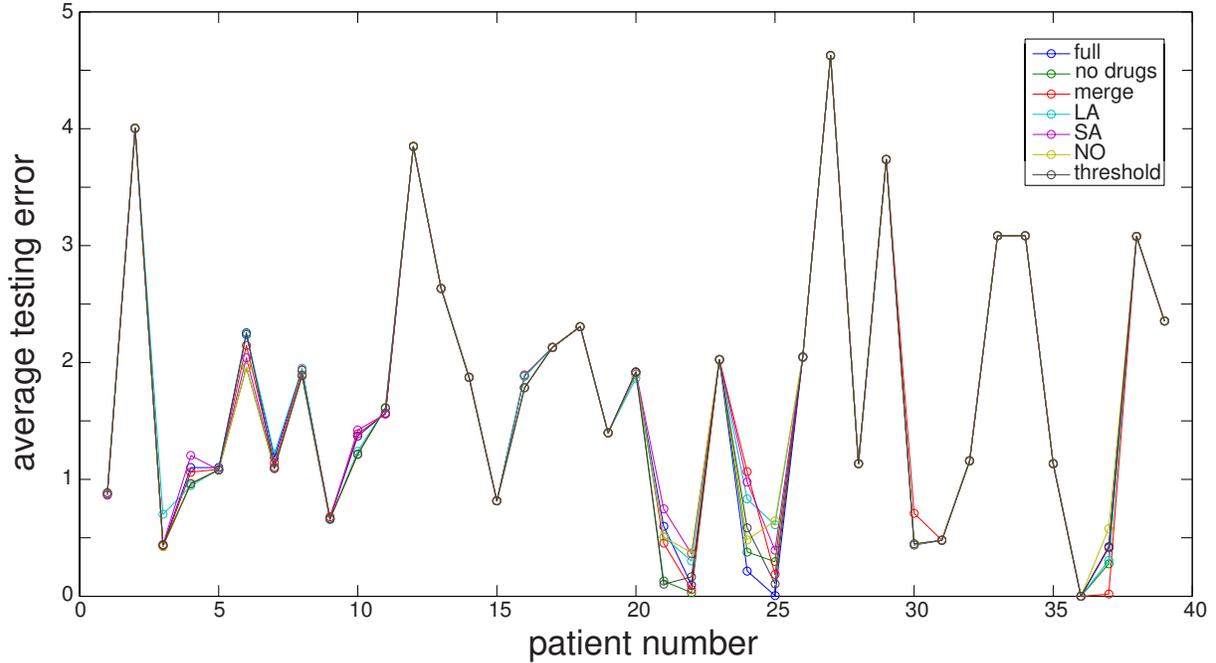
Figure 3.10. Two-fold cross validation testing error for alternative models listed in Table 3.5. For every patient, each model was independently fitted to the first and second half of the time series pain report data (training). Then the fitted models were used to test the other half of the data. This figure shows the average root-mean-square testing error for the two tests. Most models perform equally well; among patients with differing model performance, there exists no clear 'best' model for all patients. Note that patient #36 did not have enough data to fit any models, so zero error is misleading.

As a first approximation, we assume the reporting bias is linear:

$$(3.8) \qquad \rho_r(x \,|\, \mu, \sigma) = \alpha(ax + b) \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] H(x),$$

where α normalizes the distribution, a, b tune the probabilities of reporting a pain value x, and the Heaviside function $H(x)$ prevents negative pain values. Figure 3.11 shows both real and reported pain distributions at a particular time.
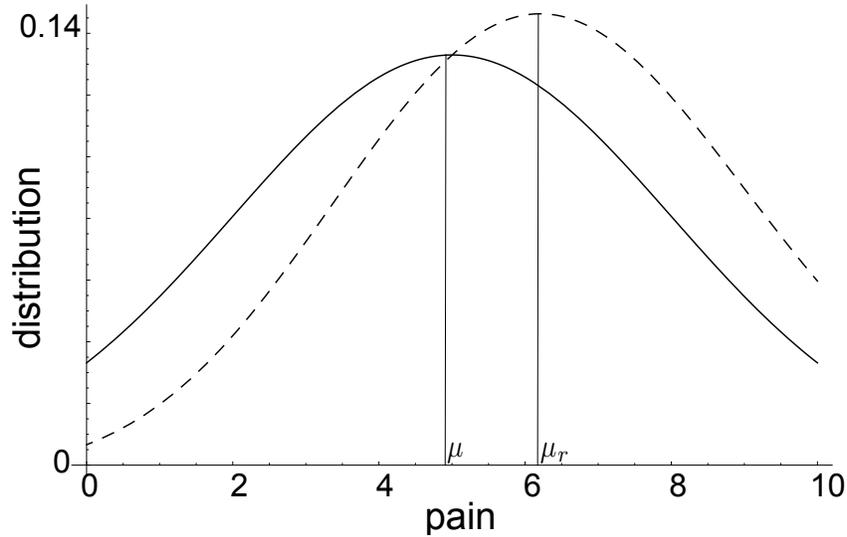
Figure 3.11. Probability density distributions for unbiased (solid) and biased (dashed) pain reporting at a particular time t_i. The standard deviation (here, $\sigma = 3$) has been exaggerated for illustrative purposes. In real data, the typical standard deviation is around $\sigma = 1.8$.

We need a way to connect these distributions because we want to control real pain described by (3.7), but the patient only provides data from the reported distribution (3.8). In other words, real pain is important but invisible, and reported pain is unimportant but visible. One way to connect the distributions is through their means and variances[2]:

$$(3.9) \qquad \mu_r(\mu, \sigma) = \int_{-\infty}^{\infty} x \, \rho_r(x \mid \mu, \sigma) \, \mathrm{d}x$$

$$(3.10) \qquad \sigma_r^2(\mu, \sigma) = \int_{-\infty}^{\infty} x^2 \, \rho_r(x \mid \mu, \sigma) \, \mathrm{d}x.$$

---

[2]Note that the means and variances also change in time. We omit time dependence for notational clarity.

Assuming σ_r is approximately constant in time[3], we can also estimate σ_r^2 using the definition of variance:

$$(3.11) \qquad \sigma_r^2 = \frac{1}{M} \sum_{i=1}^{M} \left(P_i - \mu_r(t_i) \right)^2,$$

where M is the number of pain reports, P_i is the ith reported pain value, and $\mu_r(t_i)$ is the expected reported pain given distribution (3.8) at time t_i.

Given a proposed model for real pain, we can solve this system of three equations for the three unknowns: $\mu_r(t_i)$, $\sigma(t_i)$, and σ_r. We can then compute the likelihood of the reported pain using

$$(3.12) \qquad \mathcal{L} = \prod_{i=1}^{M} \rho_r(P_i \,|\, \mu(t_i), \sigma(t_i)).$$

Because we can compute the likelihood of the supplied data given any proposed model for real pain, we can tune the model parameters to maximize likelihood (technically, we minimize the negative logarithm of likelihood). This results in a best-fit model under the assumption of biased pain reporting. We compare the best-fits of the model under both biased and unbiased reporting assumptions, and find that neither model is a better fit for most patients. See Figure 3.12.

3.4. Pain and medication optimization

A key goal of the modeling of human pain dynamics is to develop predictions that allow optimized treatment: both pain and medication use should be minimized. Excess

---

[3]It is not possible to verify this with data because patients only report one pain value at a particular time. However, a Kolmogorov-Smirnov normality test on residuals over the first two weeks of data rejects normality ($p < 0.05$) for only 2 of the 39 patients.
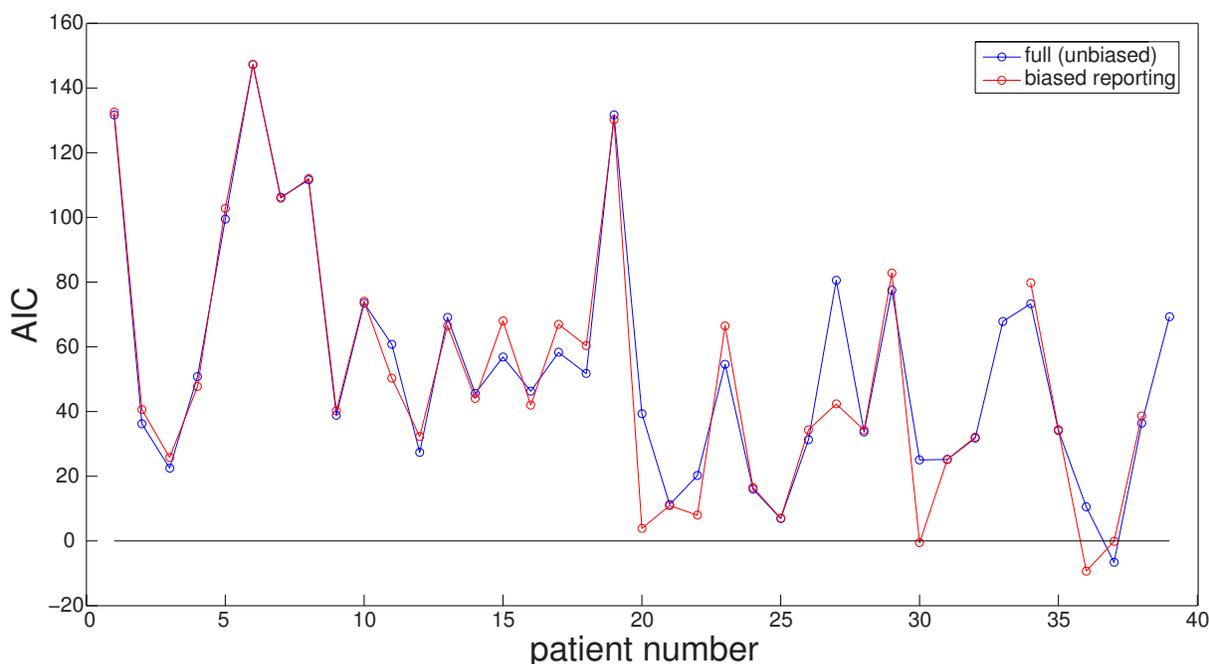
Figure 3.12. Akaike information criterion for unbiased and biased pain reporting. Because both models have an equal number of fitting parameters, AIC is a proxy for model likelihood (lower AIC implies higher likelihood). Again, it is not clear that one model performs universally better than the other. Note that missing biased reporting model fits indicate that the fitting algorithm did not converge ($a = 0.08, b = 0.1$).

medication carries particular long-term risks for chronic pain sufferers [**87, 88, 89, 90**], but pain mitigation is also a primary goal of SCD treatment. How can these contradictory objectives be balanced?

Our model allows us to forecast the probability distribution of pain for a patient at a point in the near future, given past data and future drug dosage protocol. This information may be useful to a physician, allowing him or her to make an optimized, data-driven decision balancing medication and pain for the patient in real time.

We propose several tools that may be useful to a physician. First, we find the optimal drug timing given that a certain amount of each drug will be taken over a certain time

period (say, within 24 hours). For instance, if the patient will take two short-acting opioids and one long-acting opioid within 24 hours, then the algorithm will offer the best times to take those three drug doses in order to minimize the expected average pain. We provide the physician with the expected optimal average pain for all drug combinations up to a certain maximum number of safe drug doses. See Figure 3.13.
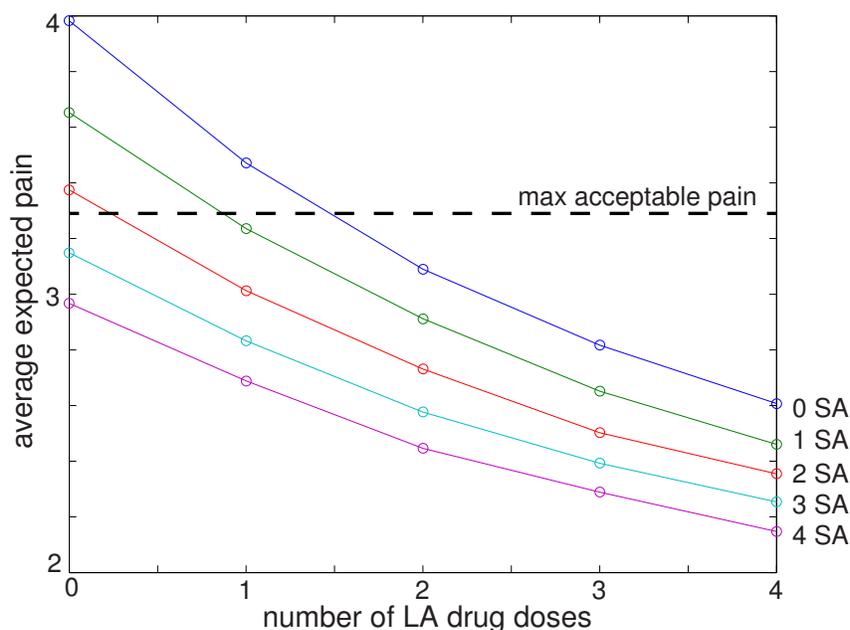
Figure 3.13. Example expected pain given optimal drug dosage protocol (in this case, for Patient A3). For each set of drug dosage protocols, from no drugs to 4 doses of each drug, a physician can see the expected average pain over a certain time period. Given a patient's maximum acceptable pain level, the physician can select the best compromise between drug doses and expected pain. In this case, the physician may tell the patient to take no long-acting (LA) drugs but take 3-4 short-acting (SA) drugs. Alternatively, the physician might tell the patient to take one LA and 1-2 doses of SA medication. The timing of the LA and SA drugs is provided by the optimization algorithm.

Second, we select the best drug dosing protocol (both number of drugs and dose timings) given an objective function balancing pain and medication. There are an unlimited

number of possible objective functions that balance pain and medication, but we propose the following:

(3.13)

$$m(\bar{P}, d1, d2, d3) = w_P\left((P_{\max} - \bar{P})^{-\alpha} - P_{\max}^{-\alpha}\right) + w_{d1}\left((d1_{\max} - d1)^{-\beta} - d1_{\max}^{-\beta}\right)$$
$$+ w_{d2}\left((d2_{\max} - d2)^{-\gamma} - d2_{\max}^{-\gamma}\right) + w_{d3}\left((d3_{\max} - d3)^{-\eta} - d3_{\max}^{-\eta}\right),$$

where $\bar{P}$ is the average expected pain; $\{d1, d2, d3\}$ are the number of drug doses of each type; $\{w_P, w_{d1}, w_{d2}, w_{d3}\}$ are the weights of pain and drugs; $\{P_{\max}, d1_{\max}, d2_{\max}, d3_{\max}\}$ are the maximum safe levels of pain and drugs; and $\{\alpha, \beta, \gamma, \eta\}$ tune the steepness of the objective function near those dangerous levels.

Figure 3.14 shows the contributions of the pain and one drug component to the objective function m. The contribution to the objective function is zero if no pain exists or if no drugs are taken. As pain or drug doses approach dangerous levels, the contribution to the objective function blows up. After a physician has made sufficient recommendations to a patient, a machine learning algorithm could select the weight parameters for each physician/patient pair. At that point, the algorithm could propose the optimal dosing protocol without much effort on the physician's part. See Figure 3.15.

3.5. Discussion

3.5.1. Reflection on hybrid modeling

Statistical models and mechanistic models have both been successfully applied to various aspects of human behavior. The inference of "black box" statistical models from empirical data has the advantage that it obviates the need for a-priori knowledge of system
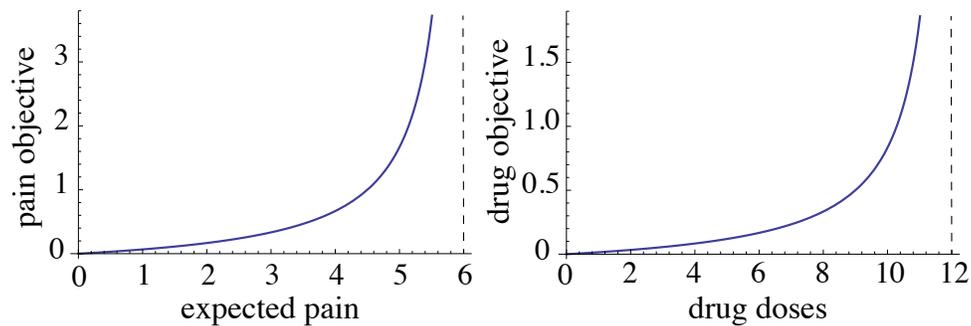
Figure 3.14. Contribution of the pain and one drug component to the objective function (3.13). The contribution to the objective function is 0 if no pain exists or if no drugs are taken. As pain or drug doses approach dangerous levels, the contribution to the objective function blows up.
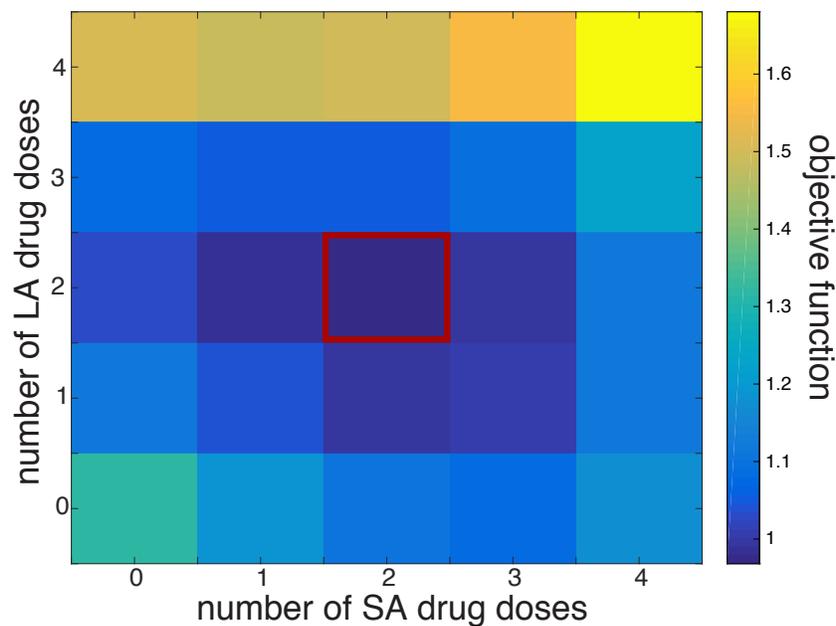
Figure 3.15. Example patient intervention recommendation (in this case, for Patient A3). For a set of personalized optimization parameters (selected by a physician or machine learning algorithm), the optimal drug timing minimizes the objective function (3.13) for each number of drug doses per 24-hour period. In this case, the patient is advised to take two standard doses of the long-acting (LA) opioid and two standard doses of the short-acting (SA) opioid, indicated by the red box.

dynamics. However, mechanistic models (sometimes referred to as "white box" or "clear box") can easily incorporate such knowledge when available.

Perhaps because of the often distinct educational backgrounds of practitioners or distinct typical applications, statistical and mechanistic approaches are not frequently combined in addressing a single problem. Compared with our work, the most similar hybrid modeling idea was developed by Sheiner and colleagues in the field of pharmacokinetics, where they proposed models to estimate population characteristics of pharmacokinetic parameters [91, 92, 93]. In their work, the pharmacokinetic models (i.e., mechanistic models) are well established, and the novelty and focus was the introduction of statistical models for pharmacokinetic parameter estimation. On the contrary, in our study the mechanistic model is not known before but developed by us based on clinical knowledge and reasonable assumptions, and our focus is the prediction of pain levels rather than parameter estimation.

Other attempts based on the hybrid modeling idea in the scientific literature have appeared in the context of neural networks (e.g., [94, 95, 96]) and chemical engineering (e.g., [95, 97, 98]), where they largely played a computational rather than analytical role. Some attempts have also been made with medical applications: Rosenberg et al. ([99]) and Adams et al. ([100]) developed a model by combining a dynamical systems approach with a statistical model to predict a patient's CD4 cell counts and HIV viral load over time in an HIV study. Timms et al. ([101]) proposed a dynamical systems approach using ODEs to improve self-regulation in a smoking cessation study. Reinforcement learning techniques such as Q-learning (e.g., [102]) also share some commonalities with the hybrid approach.

In this work we make our own attempt at a novel incorporation of statistical inference together with mechanistic dynamical systems modeling to produce a hybrid mathematical model for predicting and explaining human behavior. We apply the new approach specifically to the problem of predicting the dynamics of subjective pain in a population of individuals suffering from sickle cell disease. The rationale behind our method development is that we have prior knowledge of pain trajectories with medication, making the problem suitable for mechanistic modeling; meanwhile, we do not know the relationship between patient health characteristics and pain levels, so we would like to investigate this using a statistical model.

3.5.2. Limitations and future work

This hybrid dynamical systems/statistical approach appears to have great potential for improving patient care. The utility of this approach is currently limited by the low frequency of pain reporting, but additional high-frequency pain correlates like blood pressures, heart rate, activity level, etc., via wearable medical devices (e.g. the "Fitbit") may alleviate this limitation. Furthermore, application of similar methods to more data-rich forecasting problems (e.g. insulin levels) may also expand the utility of our work.

Another important limitation to our current model lies in the mechanistic component. Here we presented what we considered to be the simplest plausible model: pain fluctuates about an "unmitigated" equilibrium u, and medication reduces pain below this equilibrium; pain returns as medication is metabolized and removed from the bloodstream. This simple model does not capture long-term changes in the unmitigated pain level, and hence its forecast validity is likely limited to short time scales (days to weeks).

In the future, clinicians will collect an overwhelming amount of medical patient data. It is imperative that we learn to take advantage of this information to improve patient treatments beyond the traditional standard of care. The approach we report here not only addresses the specific challenge of chronic pain mitigation in SCD patients, but also provides a testbed for new ways of dealing with big, ever-growing data sets in real time.

With denser temporal data, the mechanistic model of pain dynamics within a single patient could be expanded to include diurnal rhythms (such as sleep, work, meals, etc), nonlinear drug interactions, fluctuating mood, weather, and more. With more long term data, the mechanistic model could incorporate gradual resistance to pain management medications or slowly varying demographic data (such as age, weight, blood pressure, or chronic medication). With data from larger populations, the statistical model could incorporate more demographic data to predict parameters in the mechanistic model.

A feedback loop between physicians and modelers through the mobile health application would allow for future machine learning algorithms to automatically optimize and recommend treatment to patients (see Figure 3.3). With enough data, each patient could receive a personalized "pain forecast" for the next few days, including recommendations for drug protocols and quantified risks for pain crisis and hospitalization.

3.6. Conclusions

We have successfully demonstrated the hybrid application of statistical and mechanistic mathematical modeling with application to understanding the dynamics of subjective human pain. Our model explains real-world data on human pain and can generate predictions of future pain dynamics.

We expect that similar methods could be used to incorporate disease-specific knowledge and modeling with statistical inference in a variety of medical applications. Given the coming deluge of data from wearables (including clinical trial NCT02895841 already underway) and mobile health applications, there is a clear need for new mathematical methods to take advantage of the opportunity for personalizable, data-driven medical treatments.

CHAPTER 4

The tipping point: a mathematical model for the profit-driven abandonment of restaurant tipping

Background

Tipping for restaurant service has gone in and out of fashion in America since its introduction from Europe in the 19th century [103]. Tipping has always been a controversial social convention, for both scholars and the public. The practice has been consistently tied to the worst of human nature: racism, sexism, and classism [104, 105, 106, 107]. At many points in time, tipping has been considered downright anti-democratic [108]. Yet the practice persists because the vast majority of Americans prefer to choose how much gratuity they leave after a meal [109, 110].

Economists have traditionally struggled to explain the practice of tipping in terms of rational costs and benefits because a rational economic agent would not incur a monetary cost that provides no present or future benefit [111]. Sociologists and psychologists have appealed to negative feelings of guilt, embarrassment, or anxiety to explain why people conform to the social convention of tipping [112, 113, 114], while others have appealed to altruistic feelings of generosity and empathy [115, 116, 117].

While much scholarly attention has been paid to the consumers who tip and employees who receive tips (e.g. [118, 119, 120, 121]), relatively little has been paid to restaurant owners who employ tipped staff. Theoretical models of tipping often see restaurant owners

as inefficient judges of service quality [122]. The natural conclusion to this rationale is that restaurant owners should allow tips in their restaurants in order to efficiently evaluate the quality of their wait staff [123]. This conclusion breaks down when we consider the real-world factors that influence restaurant owners' decisions, such as remaining profitable while also adhering to minimum wage restrictions, retaining talented staff, and meeting customer expectations for food and service quality [124]. By accounting for these factors and appealing only to rational profit-maximization motivations, we show that restaurants owners should play a more active role in determining tip rates in their restaurants.

As the conventional tip rate gradually increases in the US (see Figure 4.1), waiters' take home pay steadily increases, while back-of-house employees' pay remains stagnant [125]. Despite the low federal minimum wage for tipped workers ($2.13 as of 2017 [126]), waiters consistently earn more than cooks [127]. As the wage disparity increases, talented cooks may defect to restaurants where profits are shared more equitably among staff, and talented waiters may defect to restaurants with higher tips. A rational restaurant owner interested only in maximizing profit might take control of the tip rate in his/her restaurant in order to retain the most talented front-of-house (tipped) and back-of-house (untipped) staff. We show in a conceptual model of two competing restaurants that a critical tipping rate exists at which a rational restaurant owner will abandon tipping to maximize profit.

Methods

4.2.1. Model derivation

As a simple conceptual model, consider two restaurants competing for diners, waiters, and cooks. For notational ease, we will focus on one restaurant (*our restaurant*, Restaurant
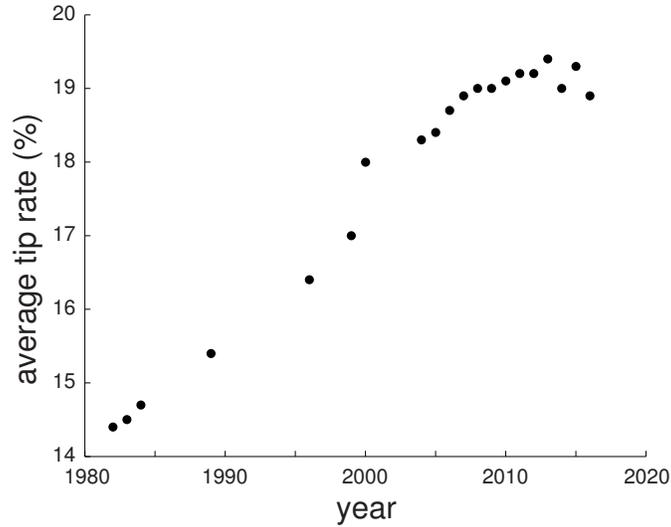
Figure 4.1. Average reported tip rate in American restaurants over time, according to NPD Group (1982-84) and Zagat annual surveys (1989-present) [128]. Note that both surveys are aimed at diners who patronize full-service midscale and upscale restaurants. Full source list in Supplemental Materials.

1). We assume all diners, waiters, and cooks intend to eat or work, respectively, at either our restaurant or the competing restaurant. Following a dynamical systems approach by Abrams et. al for modeling generic social group competition [129], people will transition between restaurants at a rate proportional to the relative utility of being at each restaurant (either as a customer or an employee). This general model assumes that the popularity of a social group also influences the transition rates, but for our purpose of modeling purely rational behavior, we focus only on utility as a driving force. Our simplification reduces the model by Abrams et. al to

(4.1)
$$\frac{\mathrm{d}x}{\mathrm{d}t} = y\, P_{yx}(u_x) - x\, P_{xy}(u_x),$$

where x is the fraction of people in group X, P_{yx} is the probability of transitioning from group Y to group X, u_x is the utility of group X, and transition rates are symmetric under exchange of x and y. In our case, X is our restaurant and Y is the competing restaurant. Alternatively, the "competing restaurant" could be viewed as a reservoir of all other dining options (including home), but that extension is left for future work.

4.2.1.1. Dynamics of cooks. Suppose our system has a number N_C of cooks who must choose between our restaurant and the competing restaurant. Because cooks do not receive tips, cooks choose where to work based only on the relative wage at the two restaurants. Assuming for simplicity that the transition probabilities are linear in relative base cook pay, the change in the number of cooks $\tilde{C}$ at our restaurant is

$$
(4.2) \qquad \frac{\mathrm{d}\tilde{C}}{\mathrm{d}t} = \underbrace{(N_C - \tilde{C})\frac{b_{C1}}{b_{C1} + b_{C2}}}_{\substack{\text{switch from} \\ \text{competitor to us}}} - \underbrace{\tilde{C}\frac{b_{C2}}{b_{C1} + b_{C2}}}_{\substack{\text{switch from} \\ \text{us to competitor}}},
$$

where b_{C1} and b_{C2} are the hourly base cook pay at our restaurant and the other restaurant, respectively, and time has arbitrary units. For example, if both restaurants offer the same base pay ($b_{C1} = b_{C2}$), then eventually half the cooks will be at our restaurant, and the other half will go to the competitor. Tilde notation will be removed later when the model is normalized.

4.2.1.2. Dynamics of waiters. Because waiters receive both hourly wages and gratuity, the transition rate between restaurants depends on the relative hourly take home (total) pay at the two restaurants. The hourly gratuity at our restaurant is

$$
(4.3) \qquad g_1 = \frac{m_1 \tilde{D} T_1}{\tilde{W}},
$$

where m_1 is the hourly menu price, $\tilde{D}$ is the number of diners, T_1 is the tip rate, and $\tilde{W}$ is the number of waiters that must split the total tips. The gratuity g_2 at the competing restaurant is similarly defined.

The hourly take home pay at our restaurant is then $b_{W1} + g_1$, where b_{W1} is the hourly base waiter pay at our restaurant. The change in the number of waiters $\tilde{W}$ at our restaurant is then

$$(4.4) \qquad \frac{\mathrm{d}\tilde{W}}{\mathrm{d}t} = \underbrace{(N_W - \tilde{W}) \frac{b_{W1} + g_1}{b_{W1} + g_1 + b_{W2} + g_2}}_{\substack{\text{switch from} \\ \text{competitor to us}}} - \underbrace{\tilde{W} \frac{b_{W2} + g_2}{b_{W1} + g_1 + b_{W2} + g_2}}_{\substack{\text{switch from} \\ \text{us to competitor}}},$$

where N_W is the number of waiters in the system, and time has arbitrary units.

4.2.1.3. Dynamics of diners. Assuming all diners intend to eat at a restaurant, they must chose between our restaurant and our competitor. Many factors influence a person's decision to eat at a particular restaurant, but we will focus on food and service quality versus menu cost. There are also many ways to measure food and service quality [**130, 131, 132**], but we will use the number of cooks and waiters who choose to work at our restaurant as a basic proxy. For instance, if our restaurant attracts more waiters, then diners will receive more personal attention and perceived service quality will increase. Suppose for simplicity that the quality q_1 of the meal and service at our restaurant is a linear combination of the number of cooks $\tilde{C}$ and waiters $\tilde{W}$ working at our restaurant:

$$(4.5) \qquad\qquad\qquad q_1 = \alpha_W \tilde{W} + \alpha_C \tilde{C},$$

where α_W and α_C are the weights placed on service and food, respectively, when evaluating our restaurant. The quality q_2 of the competing restaurant is defined similarly.

We define the value v_1 of our restaurant as the quality q_1 over the menu cost (including tips):

$$(4.6) \qquad v_1 = \frac{\alpha_W \tilde{W} + \alpha_C \tilde{C}}{m_1(1 + T_1)},$$

where m_1 is the hourly menu cost and T_1 is the tip rate at our restaurant. The value v_2 of the other restaurant is defined similarly.

A rational diner chooses a restaurant based on the perceived relative value of each restaurant. The change in the number of diners $\tilde{D}$ at our restaurant is then

$$(4.7) \qquad \frac{\mathrm{d}\tilde{D}}{\mathrm{d}t} = \underbrace{(N_D - \tilde{D})\frac{v_1}{v_1 + v_2}}_{\substack{\text{switch from} \\ \text{competitor to us}}} - \underbrace{\tilde{D}\frac{v_2}{v_1 + v_2}}_{\substack{\text{switch from} \\ \text{us to competitor}}},$$

where N_D is the number of diners in the system, and time has arbitrary units. Naturally, the overall transition rates may vary for diners, waiters, and cooks; customers may switch dining locations more rapidly than employees switch jobs. However, we are only interested in equilibrium states, so we ignore this detail.

4.2.1.4. Profitability. Given the flow of employees and customers to and from our restaurant, a rational restaurant owner will maximize hourly profit

$$(4.8) \qquad \tilde{P} = \underbrace{m_1 \tilde{D}}_{\text{revenue}} - \underbrace{b_{W1} \tilde{W}}_{\text{waiter pay}} - \underbrace{b_{C1} \tilde{C}}_{\text{cook pay}}.$$

We ignore fixed costs because we are only concerned with maximizing profitability, not absolute profits.

4.2.2. Normalized model

We now normalize and nondimensionalize the system (4.2)-(4.7) to reduce the number of parameters. We make the following substitutions

$$(4.9) \qquad D = \frac{\tilde{D}}{N_D}, \ \ W = \frac{\tilde{W}}{N_W}, \ \ C = \frac{\tilde{C}}{N_C}$$

$$(4.10) \qquad r = \frac{\alpha_C}{\alpha_W}, \ \ r_{DW} = \frac{N_D}{N_W}, \ \ r_{CW} = \frac{N_C}{N_W},$$

so that D, W, and C are the fraction diners, waiters and cooks at our restaurant, r is the ratio of food to service importance for customers, and r_{DW} and r_{CW} are the ratios of diners and cooks to waiters, respectively. Then the fraction of diners at our restaurant follows the dynamics

$$(4.11) \qquad \frac{\mathrm{d}D}{\mathrm{d}t} = (1 - D)\frac{v_1}{v_1 + v_2} - D\frac{v_2}{v_1 + v_2}$$

$$(4.12) \qquad v_1 = \frac{W + r\, r_{CW}\, C}{m_1(1 + T_1)}, \ \ v_2 = \frac{(1 - W) + r\, r_{CW}\,(1 - C)}{m_2(1 + T_2)}.$$

The fraction of waiters at our restaurant follows the dynamics

$$(4.13) \qquad \frac{\mathrm{d}W}{\mathrm{d}t} = (1 - W)\frac{b_{W1} + g_1}{b_{W1} + g_1 + b_{W2} + g_2} - W\frac{b_{W2} + g_2}{b_{W1} + g_1 + b_{W2} + g_2}$$

$$(4.14) \qquad g_1 = \frac{m_1 r_{DW} DT_1}{W}, \ \ g_2 = \frac{m_2 r_{DW}(1 - D)T_2}{W}.$$

Finally, the fraction of cooks at our restaurant follows the dynamics

$$(4.15) \qquad \frac{\mathrm{d}C}{\mathrm{d}t} = (1 - C)\frac{b_{C1}}{b_{C1} + b_{C2}} - C\frac{b_{C2}}{b_{C1} + b_{C2}}.$$

All variables and parameters are described in Table 4.1.

Variable	Meaning	Units	Range	Baseline
D	fraction of diners at our restaurant	–	[0, 1]	–
W	fraction of waiters at our restaurant	–	[0, 1]	–
C	fraction of cooks at our restaurant	–	[0, 1]	–
r	relative importance of food quality versus service quality, typically a value exceeding one	–	[1,20]	12◇
r_{CW}	ratio of total cooks to waiters in the system	–	[0.25, 4]	1
r_{DW}	ratio of total diners to waiters in the system	–	[8, 32]	12
m_1	average menu cost per hour at our restaurant	$/hr	[5, 20]	10
b_{W1}	waiters' base pay per hour at our restaurant	$/hr	[2.13*, 25]	5.00†
b_{C1}	cooks' base pay per hour at our restaurant	$/hr	[7.25*, 1]	10.40†
T_1	average tip rate at our restaurant, determined by either social convention or mandated by restaurant owner	–	[0.1, 0.25]	0.19‡
v_1	meal value perceived by customers at our restaurant	1/$	–	–
g_1	gratuity per hour at our restaurant	$/hr	–	–

Table 4.1. Description of model variables and parameters for our restaurant, Restaurant 1. The competing restaurant (Restaurant 2) has similarly defined parameter values subscripted with 2. We present a range of plausible values for each parameter and a baseline value for midscale and upscale restaurants like those reviewed by Zagat. (◇ crude estimate based on customer surveys [130]; *federal minimum wage as of 2017 [126]; †average waiter and cook pay as of 2015 [133]; ‡average self-reported tip rate as of 2016 [128]; other baseline values are guesses based on author experience).

With this change of variables, hourly profit $\tilde{P}$ becomes the hourly profit per waiter in the system

$$(4.16) \qquad P = m_1\, r_{DW}\, D - b_{W1}\, W - b_{C1}\, r_{CW}\, C.$$

Results

4.3.3. Numerical exploration

Numerical integration suggests that one stable steady state solution exists for each set of parameters regardless of the initial condition, so long as the initial condition is physically meaningful. Because cooks only switch restaurants in response to base pay (constant parameter), the distribution of cooks equilibrates first. Diners and waiters respond to everyone else in the system, so the distribution of diners and waiters equilibrates later. See Figure 4.2 for several examples of model behavior.

4.3.4. Equilibrium stability analysis

Fixed point analysis shows that four steady states exist. Only one fixed point is meaningful (i.e. $D^*, W^*, C^* \in [0,1]$). The steady state for cooks is $C^* = b_{C1}/(b_{C1} + b_{C2})$. The steady states for waiters and diners have closed forms but are too long to include. For all reasonable parameter values (listed in Table 4.1), the eigenvalues of the Jacobian evaluated at the fixed point are real and negative. This implies that the equilibrium is a stable sink. See Figure 4.3.

4.3.5. Equilibrium sensitivity analysis

Global sensitivity and uncertainty analysis using Latin Hypercube Sampling (LHS) of parameter space and Partial Rank Correlation Coefficients (PRCC) [134] reveal that equilibrium distributions of diners and waiters depend significantly ($p < 0.001$) on tip rates and cook pay. Equilibrium distributions of waiters also depend significantly on waiter pay. Note that the parameters that significantly influence these distributions describe the
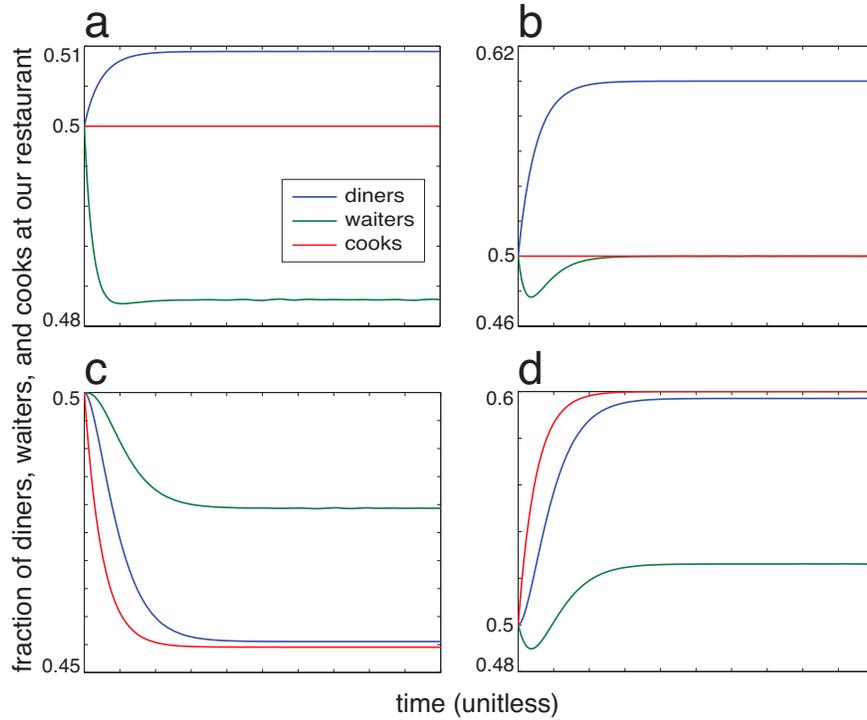
Figure 4.2. Numerical simulation of system (4.11)-(4.15). **(a)** For otherwise identical restaurants, if the tip rate at our restaurant is lower than the competitor, then waiters will leave our restaurant because they get paid less, but diners will prefer our restaurant because they pay less ($T_2 = 0.25$). **(b)** If the menu price at our restaurant is lower than the competitor, then diners will flock to our restaurant (they paid less), and waiters will temporarily leave our restaurant because lower menu prices lead to lower tips. However, after our restaurant has a large share of diners, waiters return because the density of diners balances the lower menu prices ($m_2 = 15$). **(c)** If we pay our cooks less than our competitor, then cooks will leave our restaurant because they get paid less; as food quality decreases, diners will leave our restaurant, and then waiters will leave our restaurant as their hourly tips decrease ($b_{c2} = 12$). **(d)** If we pay our cooks more but pay our waiters less to compensate, cooks will flock to our restaurant followed by diners; waiters will temporarily leave because they are paid lower wages, but eventually they will come back as diners flood our restaurant ($b_{w2} = 10, b_{c1} = 15$). Unless otherwise noted, $m_1 = m_2 = 10, T_1 = T_2 = 0.2, b_{w1} = b_{w2} = 5, b_{c1} = b_{c2} = 10, r = 12$.
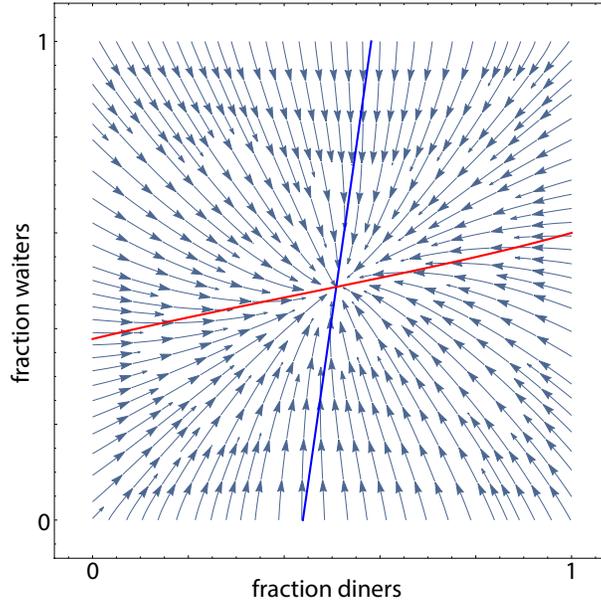
Figure 4.3. Phase portrait of two identical restaurants with differing tip rates. Our restaurant (shown) enforces an automatic gratuity of $T_1 = 0.15$, and the competing restaurant allows the conventional tip rate of $T_2 = 0.2$. The steady state is $(D^*, W^*, C^*) = (0.51, 0.49, 0.50)$. Nullclines $dD/dt = 0$ (blue) and $dW/dt = 0$ (red) are superimposed. For this example, $m_1 = m_2 = 10, r = 12, b_{W1} = b_{W2} = 5, b_{C1} = b_{C2} = 10, r_{DW} = 1, r_{CW} = 1$.

differences between restaurants and do not describe the system as a whole. See Figure 4.3.5.

Discussion

4.4.6. Tip abandonment threshold

Suppose our restaurant is attempting to maximize hourly profit (4.16) at equilibrium. We assume our restaurant is competing with a typical American restaurant that is not making dynamic changes to staff pay, menu prices, or tipping policies. Given the choices the other restaurant has made, our restaurant can choose base pay for cooks and waiters (within legal limits) and a gratuity policy. Both restaurants maintain identical menu

Figure 4.4. Global sensitivity and uncertainty analysis for equilibrium state. **(a)** Partial Rank Correlation Coefficient (PRCC) between model parameters and diner equilibrium. **(b)** PRCC between model parameters and waiter equilibrium. Asterisks indicate that the correlation is significant ($***p < 0.001, N = 100$ samples). Note that we use PRCC because numerical tests suggest that the relationships between parameters and equilibria are monotonic.

prices to ensure the restaurants are true competitors; fine dining establishments do not typically compete with casual restaurants.

If the competing restaurant allows the conventional tipping rate, then there exists a critical tip rate threshold T_c at which a rational restaurant owner would forbid tipping to maximize profit. Figure 4.5 shows the conventional tip rate at which a hypothetical restaurant should switch from allowing the conventional tip to abandoning tipping in their establishment. Assuming the typical tip rate continues to increase in the US, we predict that restaurants will eventually forbid tipping when it become more profitable to do so.

Global sensitivity and uncertainty analysis shows that this critical tipping threshold depends significantly on the menu price shared by both restaurants, the ratio of customers
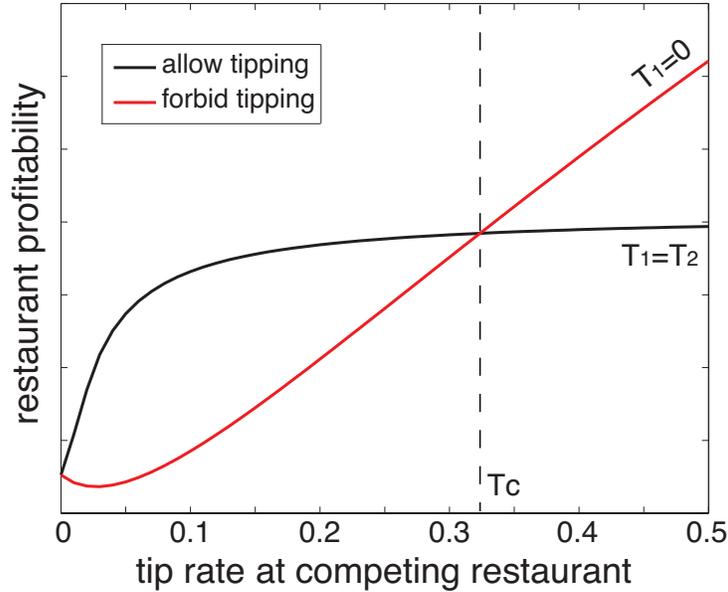
Figure 4.5. Example of critical tip rate threshold. For conventional tip rates below some critical threshold T_c, a rational restaurant owner would allow diners to leave gratuity to maximize profitability (black curve). Beyond that critical threshold, a rational restaurant owner would disallow tipping in their restaurant (red curve). Both curves assume that the restaurant owner selects staff pay (within legal limits) to maximize profit. For this example, $m_1 = m_2 = 10, r = 12, b_{W2} = 5, b_{C2} = 10, r_{DW} = 12, r_{CW} = 0.5$, the minimum wage for tipped workers is 2.13, and the minimum wage for untipped workers is 7.25.

to waiters and cooks to waiters, and the ratio of food quality to service quality in the eyes of the customer. See Figure 4.6. Note that the parameters that significantly influence the critical tip rate T_c describe the type or "class" of restaurant system we are considering. For instance, fine dining restaurants maintain a low diner to waiter ratio r_{DW} and high menu prices m. It is also likely that diners at fine dining establishments place more value on service than at casual restaurant, decreasing r.

Local sensitivity analysis about 'typical' American restaurant parameters suggests that increased menu price, increased service importance, increased diner-to-waiter ratio,
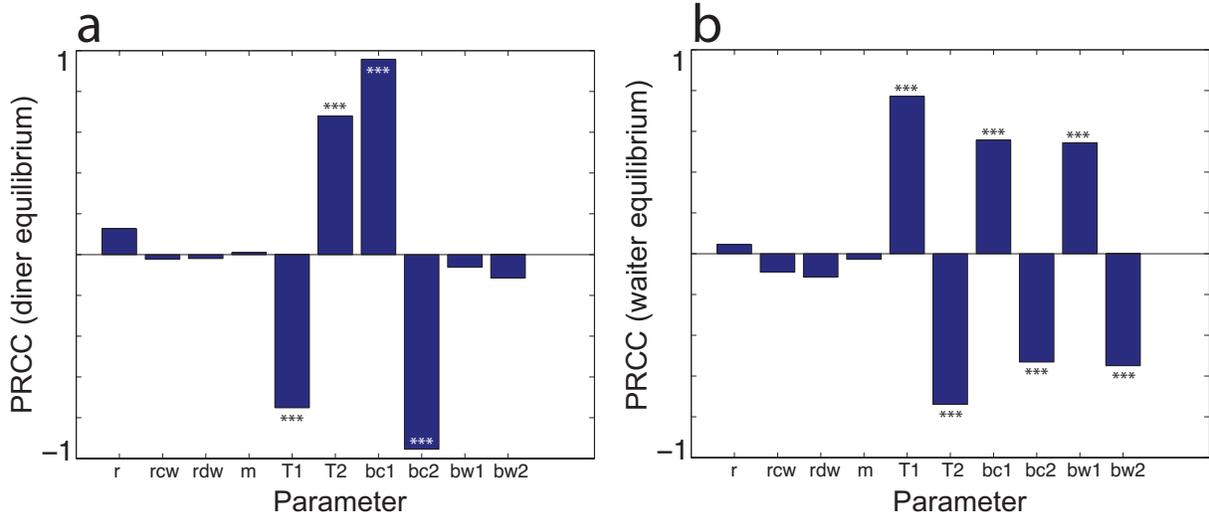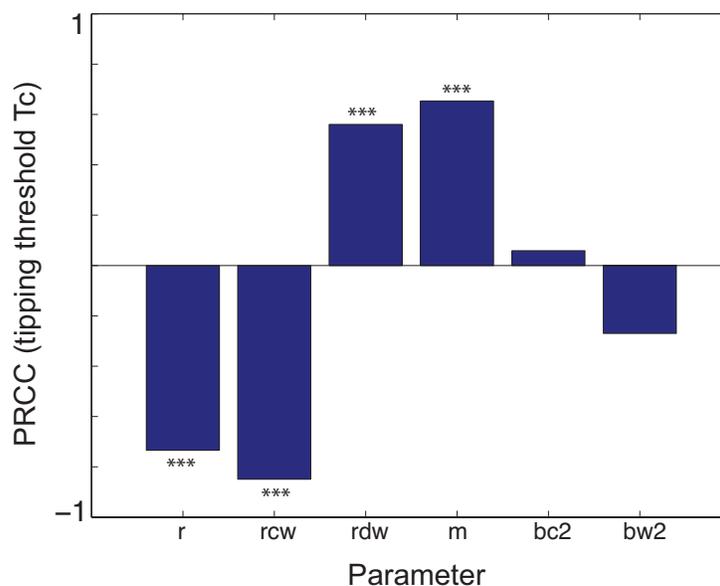
Figure 4.6. Global sensitivity and uncertainty analysis for tipping threshold T_c. The Partial Rank Correlation Coefficient (PRCC) between model parameters and the tipping threshold T_c are show. Asterisks indicate that the correlation is significant (***$p < 0.001, N = 100$ samples). Note that we use PRCC because numerical tests suggest that the relationship between parameters and T_c is monotonic.

and increased waiter-to-cook ratio all increase the critical tipping rate. See Figure 4.7. Because no type or class of restaurant increases all these parameters, we cannot say with certainty that a certain type of restaurant should abandon tipping before another. However, the three strongest correlated parameters (r, r_{CW}, m) support the prediction that fine dining establishments should be the last restaurants to abandon tipping.

This prediction is surprising because the most vocal advocates for eliminating tipping in America have been owners of upscale restaurants. However, fine dining restaurant owners cite social justice as the primary motive for eliminating tipping in their establishments [135]. This claim is consistent with our prediction because many restauranteurs

have been forced to reinstate tipping in their restaurants in order to remain profitable [**136**].
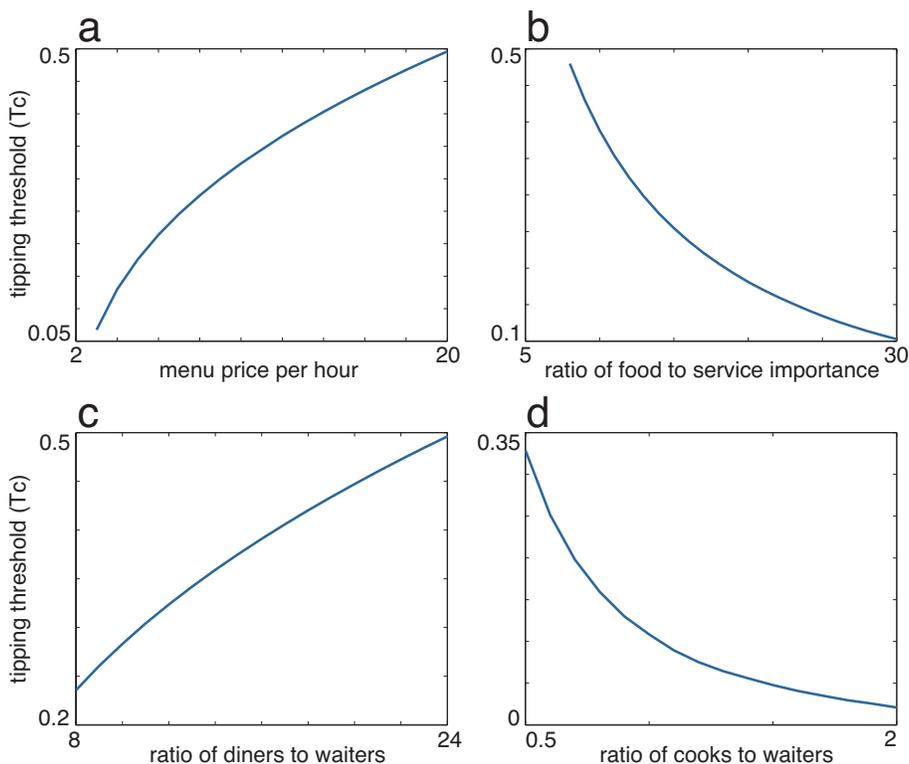
Figure 4.7. Local sensitivity analysis for tipping threshold T_c. **(a)** Holding all else constant, higher menu price implies a larger tip threshold. This indicates that fine dining restaurants should be the last the abandon tipping if all other parameters are the same. **(b)** As diners place more relative importance on food than service, the critical tip rate decreases.Because customers at fine dining establishments likely place more value on service, we again expect that fine dining restaurants will be the last to abandon tipping. **(c)** In contrast, as the ratio of diners to waiters increases, the critical tipping rate increases. All else held constant, this would imply that casual dining establishments would drop tipping last. **(d)** As the ratio of cooks to waiters increases, the critical tip rate decreases. Though it is difficult to know this ratio for full service restaurants, it makes sense that the large cook-to-waiter ratio seen at counter service restaurants implies little to no tipping. For this example, $m_1 = m_2 = 10, r = 12, b_{W2} = 5, b_{C2} = 10, r_{DW} = 12, r_{CW} = 0.5$, unless otherwise noted on the independent axis.

4.4.7. Limitations

As a conceptual model, system (4.11)-(4.15) cannot offer quantitative predictions with confidence. One limitation of this model is the lack of competition among many restaurants or eating at home, though this could be addressed by considering the "competing restaurant" as a pool of competition. In contrast to the real world, this model also does not allow waiters and cooks to change roles. In fact, many restaurants struggle to retain talented cooks because they often become waiters to make more money [137]. Additionally, the model assumes that the benefit of more employees does not have diminishing returns. More realistically, restaurant food or service will only benefit from more employees up to a certain point; after the restaurant is fully staffed, more employees will be a waste of money and may even impede service.

Our model also ignores both the federal law that requires restaurant owners to supplement tipped worker wages if their hourly tips do not exceed the federal minimum wage [138] and many state laws that impose larger minimum wages for tipped employees [126]. We also do not provide a mechanism by which the conventional tip rate increases and merely assume that the increasing trend will continue; however, the increasing trend is supported by theoretical economic models [111]. Finally and most importantly, this model assumes that humans behave rationally when spending or earning money, a false assumption common among economic models [139]. Restaurant owners may choose to abandon or maintain tipping regardless of profit, citing economically irrational reasons or responding to irrational customer feelings.

In spite of these limitations, the qualitative prediction that a critical tipping threshold exists at which restaurant owners may abandon tipping is supported by previous trends

both in America and internationally. Tipping has gone in and out of fashion around the world, and though customers normally drive the introduction (or reintroduction) of the trend, restaurant owners or governments typically end the practice [**140**].

Conclusion

The conceptual model presented here takes a new direction towards understanding the complex service industry. The oscillating popularity of tipping has previously been attributed to social contagion and irrational responses to classism. Using a new approach to modeling the social convention of tipping, we show that rational decisions to maximize profit may drive the cycle of the tipping trend. We predict that there exists a critical tip rate threshold at which restaurant owners would be wise to eliminate tipping in their establishments. Furthermore, we expect that casual restaurants should be the first to abandon tipping, and fine dining restaurants should be the last.

The simplicity of the model does not allow for quantitative predictions, such as when tipping will go out of fashion in the US or what the threshold tip rate will be. However, the model serves as a base for more sophisticated models and could direct economic data collection to better answer quantitative questions. This effort would be important not only to restaurant owners, but also to economists, sociologists, policy makers, and all people who play a role in or interact with the service industry.

References

[1] S. M. Clifton, R. I. Braun, and D. M. Abrams, "Handicap principle implies emergence of dimorphic ornaments," *Proceedings of the Royal Society B*, vol. 283, no. 1843, 2016.

[2] S. M. Clifton, R. I. Braun, and D. M. Abrams, "Data from: Handicap principle implies emergence of dimorphic ornaments."

[3] S. M. Clifton, C. Kang, J. J. Li, Q. Long, N. Shah, and D. M. Abrams, "Hybrid statistical and mechanistic mathematical model guides mobile health intervention for chronic pain," *Journal of Computational Biology*, vol. 24, no. 7, 2017.

[4] S. M. Clifton, C. Kang, J. J. Li, Q. Long, N. Shah, and D. M. Abrams, "Data from: Hybrid statistical and mechanistic mathematical model guides mobile health intervention for chronic pain."

[5] J. Chen, E. Herbers, D. M. Abrams, and S. M. Clifton, "The tipping point: mathematical model predicts restaurants will soon abandon tipping en masse," *Northwestern Undergraduate Research Journal (submitted)*, 2016.

[6] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London A*, vol. 115, no. 772, pp. 700–721, 1927.

[7] O. Diekmann, J. A. P. Heesterbeek, and J. A. Metz, "On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations," *Journal of mathematical biology*, vol. 28, no. 4, pp. 365–382, 1990.

[8] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.

[9] T. R. Malthus, *An essay on the principle of population: or, A view of its past and present effects on human happiness*. Reeves & Turner, 1888.

[10] F. Richards, "A flexible growth function for empirical use," *Journal of experimental Botany*, vol. 10, no. 2, pp. 290–301, 1959.

[11] M. B. Schaefer, "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries," *Inter-American Tropical Tuna Commission Bulletin*, vol. 1, no. 2, pp. 23–56, 1954.

[12] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," *ACM SIGGRAPH computer graphics*, vol. 21, no. 4, pp. 25–34, 1987.

[13] L. Spector, J. Klein, C. Perry, and M. Feinstein, "Emergence of collective behavior in evolving populations of flying agents," in *Genetic and Evolutionary ComputationGECCO 2003*, pp. 200–200, Springer, 2003.

[14] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Physical review letters*, vol. 75, no. 6, p. 1226, 1995.

[15] C. Darwin, *The descent of man and selection in relation to sex*. John Murray, 1871.

[16] A. Zahavi, "Mate selection—a selection for a handicap," *Journal of Theoretical Biology*, vol. 53, no. 1, pp. 205–214, 1975.

[17] M. Andersson and L. W. Simmons, "Sexual selection and mate choice," *Trends in Ecology & Evolution*, vol. 21, no. 6, pp. 296–302, 2006.

[18] T. Clutton-Brock, "Sexual Selection in Males and Females," *Science*, vol. 318, no. 5858, pp. 1882–1885, 2007.

[19] A. G. Jones and N. L. Ratterman, "Mate choice and sexual selection: what have we learned since Darwin?," *Proceedings of the National Academy of Sciences*, vol. 106, no. Supplement 1, pp. 10001–10008, 2009.

[20] B. J. Allen and J. S. Levinton, "Costs of bearing a sexually selected ornamental weapon in a fiddler crab," *Functional Ecology*, vol. 21, no. 1, 2007.

[21] M. R. Evans and A. L. Thomas, "The aerodynamic and mechanical effects of elongated tails in the scarlet-tufted malachite sunbird: measuring the cost of a handicap," *Animal Behaviour*, vol. 43, no. 2, pp. 337–347, 1992.

[22] J. Goyens, S. Van Wassenbergh, J. Dirckx, and P. Aerts, "Cost of flight and the evolution of stag beetle weaponry," *Journal of The Royal Society Interface*, vol. 12, no. 106, pp. 20150222–20150222, 2015.

[23] R. A. Johnstone, "Sexual selection, honest advertisement and the handicap principle: reviewing the evidence," *Biological Reviews*, vol. 70, no. 1, pp. 1–65, 1995.

[24] J. D. Blount, N. B. Metcalfe, T. R. Birkhead, and P. F. Surai, "Carotenoid modulation of immune function and sexual attractiveness in zebra finches," *Science*, vol. 300, no. 5616, pp. 125–127, 2003.

[25] P. M. West, "Sexual Selection, Temperature, and the Lion's Mane," *Science*, vol. 297, no. 5585, pp. 1339–1343, 2002.

[26] M. Petrie and T. Halliday, "Experimental and natural changes in the peacock's (Pavo cristatus) train can affect mating success," *Behavioral Ecology and Sociobiology*, vol. 35, no. 3, pp. 213–217, 1994.

[27] M. Andersson, "Female choice selects for extreme tail length in a widowbird," *Nature*, vol. 299, no. 5886, pp. 818–820, 1982.

[28] B. Kuijper, I. Pen, and F. J. Weissing, "A guide to sexual selection theory," *Annual Review of Ecology, Evolution, and Systematics*, vol. 43, pp. 287–311, 2012.

[29] S. Collins, "Is there only one type of male handicap?," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 252, no. 1335, pp. 193–197, 1993.

[30] H. Kokko, M. D. Jennions, and R. Brooks, "Unifying and testing models of sexual selection," *Annual Review of Ecology, Evolution, and Systematics*, pp. 43–66, 2006.

[31] G. E. Hill and K. Yasukawa, "The evolution of ornaments and armaments," *Animal Behavior: How and Why Animals Do the Things They Do*, vol. 2, pp. 145–172, 2014.

[32] H. Gintis, E. A. Smith, and S. Bowles, "Costly signaling and cooperation," *Journal of theoretical biology*, vol. 213, no. 1, pp. 103–119, 2001.

[33] A. Grafen, "Biological signals as handicaps," *Journal of theoretical biology*, vol. 144, no. 4, pp. 517–546, 1990.

[34] Y. Iwasa, A. Pomiankowski, and S. Nee, "The evolution of costly mate preferences ii. the'handicap'principle," *Evolution*, pp. 1431–1442, 1991.

[35] R. Lande, "Models of speciation by sexual selection on polygenic traits," *Proceedings of the National Academy of Sciences*, vol. 78, no. 6, part 2, pp. 3721–3725, 1981.

[36] M. A. Nowak, *Evolutionary Dynamics*. Harvard University Press, 2006.

[37] U. Dieckmann and R. Law, "The dynamical theory of coevolution: a derivation from stochastic ecological processes," *Journal of mathematical biology*, vol. 34, no. 5-6, pp. 579–612, 1996.

[38] D. J. Emlen and H. F. Nijhout, "Hormonal control of male horn length dimorphism in the dung beetle Onthophagus taurus(Coleoptera: Scarabaeidae)," *Journal of Insect Physiology*, vol. 45, no. 1, pp. 45–53, 1999.

[39] A. Aisenberg and F. G. Costa, "Reproductive isolation and sex-role reversal in two sympatric sand-dwelling wolf spiders of the genus Allocosa," *Canadian Journal of Zoology*, vol. 86, no. 7, pp. 648–658, 2008.

[40] J. L. Tomkins, J. S. Kotiaho, and N. R. LeBas, "Matters of Scale: Positive Allometry and the Evolution of Male Dimorphisms," *The American Naturalist*, vol. 165, no. 3, pp. 389–402, 2005.

[41] K. A. Glover, O. T. Skilbrei, and Ø. Skaala, "Stock-specific growth and length frequency bimodality in brown trout," *Transactions of the American Fisheries Society*, vol. 132, no. 2, pp. 307–315, 2003.

[42] K. A. Glover, C. Skår, K. E. Christie, J. Glette, H. Rudra, and Ø. Skaala, "Size-dependent susceptibility to infectious salmon anemia virus (ISAV) in Atlantic salmon (Salmo salar L.) of farm, hybrid and wild parentage," *Aquaculture*, vol. 254, no. 1-4, pp. 82–91, 2006.

[43] M. J. West-Eberhard, "Sexual selection and social behavior," *Man and Beast Revisited*, pp. 159–172, 1991.

[44] M. R. Gross, "Alternative reproductive strategies and tactics: diversity within sexes," *Trends in Ecology & Evolution*, vol. 11, no. 2, pp. 92–98, 1996.

[45] G. P. Karev, "On mathematical theory of selection: continuous time population dynamics," *Journal of mathematical biology*, vol. 60, no. 1, pp. 107–129, 2010.

[46] P. Galeotti and D. Rubolini, "Head ornaments in owls: what are their functions?," *Journal of Avian Biology*, vol. 38, no. 6, pp. 731–736, 2007.

[47] V. van den Brink, V. Dolivo, X. Falourd, A. N. Dreiss, and A. Roulin, "Melanic color-dependent antipredator behavior strategies in barn owl nestlings," *Behavioral Ecology*, vol. 23, no. 3, pp. 473–480, 2012.

[48] M. E. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[49] W. J. Reed and B. D. Hughes, "From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature," *Physical Review E*, vol. 66, no. 6, p. 067103, 2002.

[50] T. Getty, "Sexually selected signals are not similar to sports handicaps," *Trends in Ecology & Evolution*, vol. 21, no. 2, pp. 83–88, 2006.

[51] W. A. Searcy, "Species recognition of song by female red-winged blackbirds," *Animal Behaviour*, vol. 40, no. 6, pp. 1119–1127, 1990.

[52] K. J. Norris, "Female choice and the quality of parental care in the great tit Parus major," *Behavioral Ecology and Sociobiology*, vol. 27, no. 4, pp. 275–281, 1990.

[53] G. R. Bortolotti, J. Blas, J. J. Negro, and J. L. Tella, "A complex plumage pattern as an honest social signal," *Animal Behaviour*, vol. 72, no. 2, pp. 423–430, 2006.

[54] A. Badyaev, "Evolution of sexual dichromatism: contribution of carotenoid- versus melanin-based coloration," *Biological journal of the Linnean Society*, vol. 69, no. 2, pp. 153–172, 2000.

[55] M. Niecke, S. Rothlaender, and A. Roulin, "Why do melanin ornaments signal individual quality? Insights from metal element analysis of barn owl feathers," *Oecologia*, vol. 137, no. 1, pp. 153–158, 2003.

[56] S. Andersson, S. R. Pryke, J. Örnborg, M. J. Lawes, and M. Andersson, "Multiple Receivers, Multiple Ornaments, and a Trade-off between Agonistic and Epigamic Signaling in a Widowbird," *The American Naturalist*, vol. 160, no. 5, pp. 683–691, 2002.

[57] A. Loyau, M. Saint Jalme, C. Cagniant, and G. Sorci, "Multiple sexual advertisements honestly reflect health status in peacocks (Pavo cristatus)," *Behavioral Ecology and Sociobiology*, vol. 58, no. 6, pp. 552–557, 2005.

[58] H. L. Mays, Jr, K. J. McGraw, G. Ritchison, S. Cooper, V. Rush, and R. S. Parker, "Sexual dichromatism in the yellow-breasted chat Icteria virens: spectrophotometric analysis and biochemical basis," *Journal of Avian Biology*, vol. 35, no. 2, pp. 125–134, 2004.

[59] C. Pélabon and L. van Breukelen, "Asymmetry in antler size in roe deer (Capreolus capreolus): an index of individual and population conditions," *Oecologia*, vol. 116, no. 1-2, pp. 1–8, 1998.

[60] A. P. Moczek and H. F. Nijhout, "Developmental mechanisms of threshold evolution in a polyphenic beetle," *Evolution & Development*, vol. 4, no. 4, pp. 252–264, 2002.

[61] I. Barber, D. Nairn, and F. A. Huntingford, "Nests as ornaments: revealing construction by male sticklebacks," *Behavioral Ecology*, vol. 12, no. 4, pp. 390–396, 2001.

[62] G. W. Hyatt and M. Salmon, "Combat in the fiddler crabs Uca pugilator and U. pugnax: a quantitative analysis," *Behaviour*, vol. 65, no. 1, pp. 182–211, 1978.

[63] F. Skarstein and I. Folstad, "Sexual dichromatism and the immunocompetence handicap: an observational approach using arctic charr," *Oikos*, vol. 76, pp. 359–367, 1996.

[64] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *The Annals of Statistics*, vol. 13, no. 1, pp. 70–84, 1985.

[65] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[66] I. Stewart, "Speciation: a case study in symmetric bifurcation theory," *Universitatis Iagellonicae Acta Mathematica*, vol. 41, pp. 67–88, 2003.

[67] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* Westview press, 2014.

[68] D. A. Freedman, *Statistical models: theory and practice.* cambridge university press, 2009.

[69] O. S. Platt, B. D. Thorington, D. J. Brambilla, P. F. Milner, W. F. Rosse, E. Vichinsky, and T. R. Kinney, "Pain in sickle cell disease: rates and risk factors," *New England Journal of Medicine*, vol. 325, no. 1, pp. 11–16, 1991.

[70] Y. Granovsky, M. Granot, R.-R. Nir, and D. Yarnitsky, "Objective correlate of subjective pain perception by contact heat-evoked potentials," *The Journal of Pain*, vol. 9, no. 1, pp. 53–63, 2008.

[71] A. Hughes, A. Macleod, J. Growcott, and I. Thomas, "Assessment of the reproducibility of intradermal administration of capsaicin as a model for inducing human pain," *Pain*, vol. 99, no. 1, pp. 323–331, 2002.

[72] S. Stevens, A. Carton, and G. Shickman, "A scale of apparent intensity of electric shock.," *Journal of Experimental Psychology*, vol. 56, no. 4, p. 328, 1958.

[73] C. R. Jonassaint, N. Shah, J. Jonassaint, and L. De Castro, "Usability and feasibility of an mhealth intervention for monitoring and managing pain symptoms in sickle cell disease: The sickle cell disease mobile application to record symptoms via technology (smart)," *Hemoglobin*, no. ahead-of-print, pp. 1–7, 2015.

[74] N. Shah, J. Jonassaint, and L. De Castro, "Patients welcome the sickle cell disease mobile application to record symptoms via technology (smart)," *Hemoglobin*, vol. 38, no. 2, pp. 99–103, 2014.

[75] D. T. McRuer and E. S. Krendel, "Mathematical models of human pilot behavior," tech. rep., DTIC Document, 1974.

[76] U. G. Fors, L. G. Edwall, and G. A. Haegerstam, "The ability of a mathematical model to evaluate the effects of two pain modulating procedures on pulpal pain in man," *Pain*, vol. 33, no. 2, pp. 253–264, 1988.

[77] N. Britton, S. Skevington, and M. Chaplain, "Mathematical modelling of acute pain," *Journal of Biological Systems*, vol. 3, no. 04, pp. 1119–1124, 1995.

[78] G. Stepan, "Delay effects in the human sensory system during balancing," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1891, pp. 1195–1212, 2009.

[79] C. W. Gardiner *et al.*, *Handbook of stochastic methods*, vol. 3. Springer Berlin, 1985.

[80] F. Yang, X. Tong, D. G. McCarver, R. N. Hines, and D. A. Beard, "Population-based analysis of methadone distribution and metabolism using an age-dependent physiologically based pharmacokinetic model," *Journal of pharmacokinetics and pharmacodynamics*, vol. 33, no. 4, pp. 485–518, 2006.

[81] P. Poulin and F.-P. Theil, "Prediction of pharmacokinetics prior to in vivo studies. ii. generic physiologically based pharmacokinetic models of drug disposition," *Journal of pharmaceutical sciences*, vol. 91, no. 5, pp. 1358–1370, 2002.

[82] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[83] P. Diggle, *Analysis of longitudinal data.* Oxford University Press, 2002.

[84] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, *Applied longitudinal analysis*, vol. 998. John Wiley & Sons, 2012.

[85] A. Groll and G. Tutz, "Variable selection for generalized linear mixed models by l 1-penalized estimation," *Statistics and Computing*, vol. 24, no. 2, pp. 137–154, 2014.

[86] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[87] B. Bannwarth, "Risk-benefit assessment of opioids in chronic noncancer pain," *Drug safety*, vol. 21, no. 4, pp. 283–296, 1999.

[88] R. J. Gatchel, "A biopsychosocial overview of pretreatment screening of patients with pain," *The Clinical journal of pain*, vol. 17, no. 3, pp. 192–199, 2001.

[89] S. Savage, "Opioid therapy of chronic pain: assessment of consequences," *Acta Anaesthesiologica Scandinavica*, vol. 43, no. 9, pp. 909–917, 1999.

[90] D. Brookoff, "Chronic pain: 2. the case for opioids," *Hospital Practice*, vol. 35, no. 9, pp. 69–84, 2000.

[91] L. B. Sheiner, B. Rosenberg, and V. V. Marathe, "Estimation of population characteristics of pharmacokinetic parameters from routine clinical data," *Journal of pharmacokinetics and biopharmaceutics*, vol. 5, no. 5, pp. 445–479, 1977.

[92] L. B. Sheiner and S. L. Beal, "Evaluation of methods for estimating population pharmacokinetic parameters. i. michaelis-menten model: routine clinical pharmacokinetic data," *Journal of pharmacokinetics and biopharmaceutics*, vol. 8, no. 6, pp. 553–571, 1980.

[93] J. W. Mandema, D. Verotta, and L. B. Sheiner, "Building population pharmacokineticpharmacodynamic models. i. models for covariate effects," *Journal of pharmacokinetics and biopharmaceutics*, vol. 20, no. 5, pp. 511–528, 1992.

[94] D. C. Psichogios and L. H. Ungar, "A hybrid neural network-first principles approach to process modeling," *AIChE Journal*, vol. 38, no. 10, pp. 1499–1511, 1992.

[95] M. L. Thompson and M. A. Kramer, "Modeling chemical processes using prior knowledge and neural networks," *AIChE Journal*, vol. 40, no. 8, pp. 1328–1340, 1994.

[96] H.-T. Su, N. Bhat, P. Minderman, and T. McAvoy, "Integrating neural networks with first principles models for dynamic modeling," in *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes (DYCORD+'92):*

Selected Papers from the 3rd IFAC Symposium, Maryland, USA, 26-29 April 1992, p. 327, Elsevier, 2014.

[97] J. Schubert, R. Simutis, M. Dors, I. Havlík, and A. Lübbert, "Hybrid modelling of yeast production processes–combination of a priori knowledge on different levels of sophistication," *Chemical engineering & technology*, vol. 17, no. 1, pp. 10–20, 1994.

[98] B. P. Duarte and P. M. Saraiva, "Hybrid models combining mechanistic models with adaptive regression splines and local stepwise regression," *Industrial & engineering chemistry research*, vol. 42, no. 1, pp. 99–107, 2003.

[99] E. S. Rosenberg, M. Davidian, and H. T. Banks, "Using mathematical modeling and control to develop structured treatment interruption strategies for hiv infection," *Drug and alcohol dependence*, vol. 88, pp. S41–S51, 2007.

[100] B. Adams, H. Banks, M. Davidian, and E. Rosenberg, "Estimation and prediction with hiv-treatment interruption data," *Bulletin of mathematical biology*, vol. 69, no. 2, pp. 563–584, 2007.

[101] K. P. Timms, D. E. Rivera, L. M. Collins, and M. E. Piper, "A dynamical systems approach to understanding self-regulation in smoking cessation behavior change," *nicotine & tobacco research*, vol. 16, no. Suppl 2, pp. S159–S168, 2014.

[102] L. G. Jaimes, M. Llofriu, and A. Raij, "A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation," in *Proceedings of*

the 9th International Conference on Body Area Networks, pp. 197–203, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[103] O. H. Azar, "The history of tipping–from sixteenth-century england to united states in the 1910s," *The Journal of Socio-Economics*, vol. 33, no. 6, pp. 745–764, 2004.

[104] M. Lynn, M. Sturman, C. Ganley, E. Adams, M. Douglas, and J. McNeil, "Consumer racial discrimination in tipping: A replication and extension," *Journal of Applied Social Psychology*, vol. 38, no. 4, pp. 1045–1060, 2008.

[105] B. Rind and P. Bordia, "Effect on restaurant tipping of male and female servers drawing a happy, smiling face on the backs of customers' checks," *Journal of Applied Social Psychology*, vol. 26, no. 3, pp. 218–225, 1996.

[106] M. Lynn and T. Simons, "Predictors of male and female servers' average tip earnings," *Journal of Applied Social Psychology*, vol. 30, no. 2, pp. 241–252, 2000.

[107] Y. Margalioth, "The case against tipping," *U. Pa. J. Lab. & Emp. L.*, vol. 9, p. 117, 2006.

[108] K. Segrave, *Tipping: An American social history of gratuities.* McFarland, 1998.

[109] M. Lynn and R. J. Kwortnik, "The effects of tipping policies on customer satisfaction: A test from the cruise industry," *International Journal of Hospitality Management*, vol. 51, pp. 15–18, 2015.

[110] O. H. Azar, "Tipping motivations and behavior in the us and israel," *Journal of Applied Social Psychology*, vol. 40, no. 2, pp. 421–457, 2010.

[111] O. H. Azar, "What sustains social norms and how they evolve?: The case of tipping," *Journal of Economic Behavior & Organization*, vol. 54, no. 1, pp. 49–64, 2004.

[112] J. Elster, "Social norms and economic theory," *The Journal of Economic Perspectives*, vol. 3, no. 4, pp. 99–117, 1989.

[113] M. Lynn, "Individual differences in self-attributed motives for tipping: Antecedents, consequences, and implications," *International Journal of Hospitality Management*, vol. 28, no. 3, pp. 432–438, 2009.

[114] M. Lynn, G. M. Zinkhan, and J. Harris, "Consumer tipping: A cross-country study," *Journal of Consumer Research*, vol. 20, no. 3, pp. 478–488, 1993.

[115] A. E. Greenberg, "On the complementarity of prosocial norms: The case of restaurant tipping during the holidays," *Journal of Economic Behavior & Organization*, vol. 97, pp. 103–112, 2014.

[116] D. B. Strohmetz, B. Rind, R. Fisher, and M. Lynn, "Sweetening the till: The use of candy to increase restaurant tipping," *Journal of Applied Social Psychology*, vol. 32, no. 2, pp. 300–309, 2002.

[117] S. F. Fong, *The socio-economic motives underlying tipping behaviour*. PhD thesis, University of Saskatchewan, 2005.

[118] M. Lynn and A. Grassman, "Restaurant tipping: an examination of three 'rational' explanations," *Journal of Economic Psychology*, vol. 11, no. 2, pp. 169–181, 1990.

[119] W. A. Gibson *et al.*, "Economics and restaurant gratuities: determining tip rates," *American Journal of Economics and Sociology*, vol. 56, no. 2, pp. 187–203, 1997.

[120] B. Rind and P. Bordia, "Effect of server's "thank you" and personalization on restaurant tipping," *Journal of Applied Social Psychology*, vol. 25, no. 9, pp. 745–751, 1995.

[121] M. Lynn and K. Mynier, "Effect of server posture on restaurant tipping," *Journal of Applied Social Psychology*, vol. 23, no. 8, pp. 678–685, 1993.

[122] M. Conlin, M. Lynn, and T. O'Donoghue, "The norm of restaurant tipping," *Journal of Economic Behavior & Organization*, vol. 52, no. 3, pp. 297–321, 2003.

[123] M. Lynn, "Restaurant tipping and service quality: A tenuous relationship," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 42, no. 1, pp. 14–20, 2001.

[124] C. A. Enz, "Issues of concern for restaurant owners and managers," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 45, no. 4, pp. 315–332, 2004.

[125] P. Wachter, "Why tip?," October 2008. [Online; posted 09-October-2008].

[126] USDoL, "Wage and hour division (whd): Minimum wages for tipped employees," January 2017. [Online; posted 1-January-2017].

[127] M. Lynn, "Should us restaurants abandon tipping? a review of the issues and evidence.," *Psychosociological Issues in Human Resource Management*, vol. 5, no. 1, 2017.

[128] Y. Margalioth, "The Social Norm of Tipping, Its Correlation with Inequality, and Differences in Tax Treatment Across Countries," *Theoretical Inquiries in Law*, 2010.

[129] D. M. Abrams, H. A. Yaple, and R. J. Wiener, "Dynamics of social group competition: modeling the decline of religious affiliation," *Physical Review Letters*, vol. 107, no. 8, p. 088701, 2011.

[130] J. M. Sulek and R. L. Hensley, "The relative importance of food, atmosphere, and fairness of wait: The case of a full-service restaurant," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 45, no. 3, pp. 235–247, 2004.

[131] Y. Namkung and S. Jang, "Are highly satisfied restaurant customers really different? a quality perception perspective," *International Journal of Contemporary Hospitality Management*, vol. 20, no. 2, pp. 142–155, 2008.

[132] S. Gagić, D. Tešanović, and A. Jovičić, "The vital components of restaurant quality that affect guest satisfaction," *Turizam*, vol. 17, no. 4, pp. 166–176, 2013.

[133] PayScale, "Restaurant report key stats," January 2015. [Online; posted 1-January-2015].

[134] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *Journal of theoretical biology*, vol. 254, no. 1, pp. 178–196, 2008.

[135] S. Estreicher and J. R. Nash, "The case for tipping and unrestricted tip-pooling," *Social Science Research Network*, 2016.

[136] A. Chandler, "Why some restaurants are walking back their no-tipping policies," May 2016. [Online; posted 10-May-2016].

[137] R. A. Ferdman, "The problem that's tearing restaurants apart," August 2015. [Online; posted 20-August-2015].

[138] USDoL, "Tips," June 2017. [Online; posted 2-June-2017].

[139] D. L. McFadden, "The new science of pleasure," tech. rep., National Bureau of Economic Research, 2013.

[140] W. M. Lynn, "Tipping in restaurants and around the globe: An interdisciplinary review," *Handbook Of Contemporary Behavioral Economics: Foundations And Developments*, 2006.

[141] J. E. Thorpe, "Bimodal distribution of length of juvenile Atlantic salmon (Salmo salar L.) under artificial rearing conditions," *Journal of Fish Biology*, vol. 11, no. 2, pp. 175–184, 1977.

[142] M. P. Eichhorn, "Spatial organisation of a bimodal forest stand," *Journal of Forest Research*, vol. 15, no. 6, pp. 391–397, 2010.

[143] M. Yokozawa and T. Hara, "Foliage profile, size structure and stem diameter-plant height relationship in crowded plant populations," *Annals of Botany*, vol. 76, no. 3, pp. 271–285, 1995.

[144] D. J. Emlen, "Artificial selection on horn length-body size allometry in the horned beetle Onthophagus acuminatus (Coleoptera: Scarabaeidae)," *Evolution*, vol. 50, no. 3, pp. 1219–1230, 1996.

[145] J. S. Huxley, *Problems of Relative Growth.* New York: L. MacVeagh, The Dial Press, 1932.

APPENDIX A

Appendix

A.1. General class of reproductive potential functions with no more than two stable states

Consider a general total potential function

(A.1)
$$\varphi = s\,\varphi^{(\text{soc})} + (1-s)\,\varphi^{(\text{ind})}, \quad s \in [0,1]$$

where $\varphi^{(\text{soc})}$ is a continuous and differentiable increasing function of ornament size, and $\varphi^{(\text{ind})}$ is a continuous, singly-peaked function of ornament size. Similar to our previous general class of potential functions,

$$\frac{\mathrm{d}a}{\mathrm{d}t} \propto \frac{\partial}{\partial a}\varphi,$$

we conclude that $\dfrac{\mathrm{d}a}{\mathrm{d}t} = 0$ only for $a \geq a_{\text{opt}}$. This implies that equilibrium ornament sizes (if an equilibrium exists) will all be at least as large as the optimal. Because this is a first order ordinary differential equation model, we also know that oscillations are not possible.

We further assume that

(A.2)
$$\frac{\partial^3}{\partial a^3}\varphi^{(\text{ind})} \equiv 0$$
$$\frac{\partial^3}{\partial a^3}\varphi^{(\text{soc})} > 0 \text{ or } \frac{\partial^3}{\partial a^3}\varphi^{(\text{soc})} < 0, \quad a \neq \bar{a}.$$

In other words, individual potential is quadratic, and the derivative of social potential is either concave up or concave down, except possibly at the mean. With these additional restrictions on the potential function, only uniform and two-morph stable fixed points are possible. See Figure A.1 for graphical proof. Our model (2.5) satisfies all restrictions, so we conclude that our exploration of the one- and two-morph steady states is a thorough investigation of all possible fixed points.
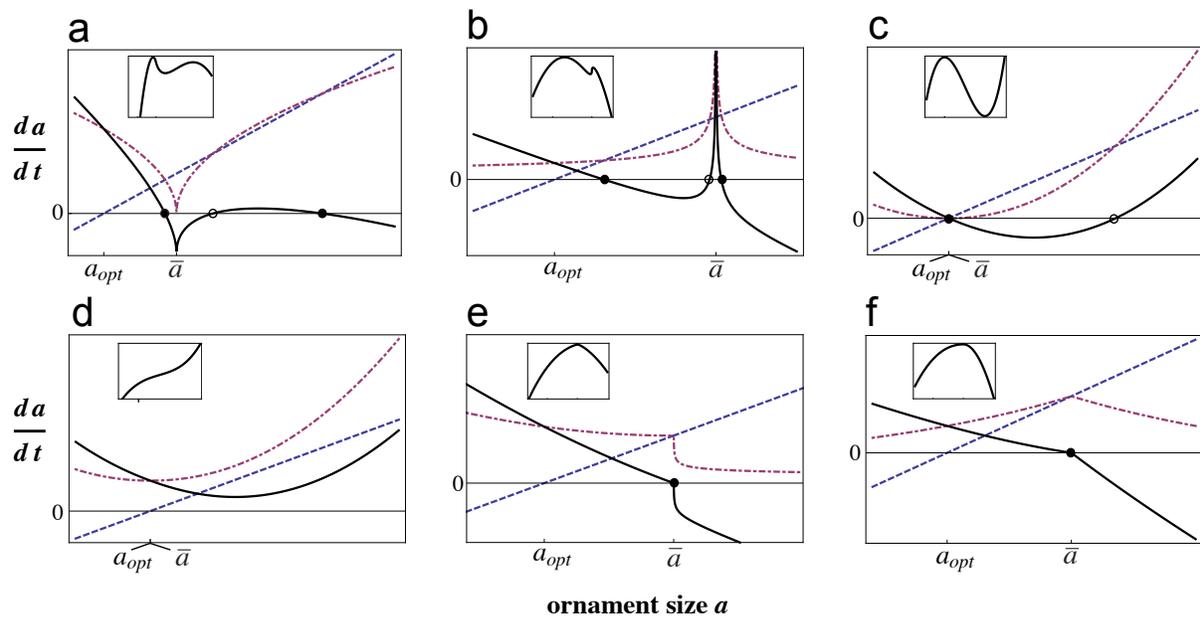
Figure A.1. Examples of derivatives of negated individual potential (dashed blue) and social potential (dot dashed maroon) for a single male in a population near equilibrium. The derivative of total potential (solid black) is proportional to da/dt, so intersections of individual and social potentials are the fixed points. Stable fixed points are marked with a filled black dot, and unstable fixed points are marked with an unfilled black dot. The total potential is inset. With restrictions (A.2), the only possible stable steady states (filled black dots) are one- or two-morphs. Note that the system may or may not have an unstable node (unfilled black dots), or it may have no fixed points.

A.2. Alternative multiplicative form for reproductive potential

Rather than using a weighted sum to construct a total potential as in equation (2.3) of the main text, we could have considered a weighted product as advocated in [50] and [35], where the authors argue that viability and mating success contribute multiplicatively. We chose to present the mathematical analysis in the context of an additive reproductive potential because of the greater simplicity, but a multiplicative potential produces the same qualitative results and can even be considered in the same way by reinterpreting quantities in question on a logarithmic scale (in which case multiplicative terms become additive).

Still, it may be of interest to see how an explicitly multiplicative model plays out. To create such a model, we first need to scale the reproductive potential functions such that $\varphi^{(\text{soc})}, \varphi^{(\text{ind})} \in [0, 1]$ and $a \in [0, 1]$.

To retain a quadratic individual potential function with maximum $\varphi^{(\text{ind})} = 1$ and roots at $a = \{0, 1\}$, we can choose

$$(A.3) \qquad \varphi^{(\text{ind})} = 4a(1 - a), \quad a \in [0, 1].$$

Note that this choice implies[1] $a_{\text{opt}} = 1/2$. Next, to retain the monotonically increasing social potential function tuned with the social sensitivity γ, we set

$$(A.4) \qquad \varphi^{(\text{soc})} = \frac{\operatorname{sgn}(a - \bar{a})|a - \bar{a}|^{\gamma} + \bar{a}^{\gamma}}{(1 - \bar{a})^{\gamma} + \bar{a}^{\gamma}}, \quad a \in [0, 1].$$

---

[1] A more general form allowing for arbitrary a_{opt} would map $a \to a^{\alpha}$, where $\alpha = -\ln(2)/\ln(a_{\text{opt}})$.

Assuming a weighted product of the individual and social potential terms, the total reproductive potential becomes

$$(A.5) \qquad \varphi = \left(\varphi^{(\text{soc})}\right)^{s}\left(\varphi^{(\text{ind})}\right)^{1-s},$$

where $\varphi^{(\text{soc})}$ and $\varphi^{(\text{ind})}$ are (A.4) and (A.3), respectively, and s tunes the relative importance of each term. Plugging this into equation (2.4) of the main text, we get

$$(A.6) \qquad \frac{\mathrm{d}a}{\mathrm{d}t} = c\left[s\left(\frac{\varphi^{(\text{ind})}}{\varphi^{(\text{soc})}}\right)^{1-s}\frac{\partial}{\partial a}\varphi^{(\text{soc})} + (1-s)\left(\frac{\varphi^{(\text{ind})}}{\varphi^{(\text{soc})}}\right)^{-s}\frac{\partial}{\partial a}\varphi^{(\text{ind})}\right]$$

or

$$(A.7) \qquad \frac{\mathrm{d}a}{\mathrm{d}t} = c\left(\frac{\varphi^{(\text{ind})}}{\varphi^{(\text{soc})}}\right)^{-s}\left[s\left(\frac{\varphi^{(\text{ind})}}{\varphi^{(\text{soc})}}\right)\frac{\partial}{\partial a}\varphi^{(\text{soc})} + (1-s)\frac{\partial}{\partial a}\varphi^{(\text{ind})}\right].$$

We observe that (A.7) retains the form of a weighted sum of two terms, though the split of natural and sexually selective forces is now more complicated. The weights are different from (2.5), but remain positive, and the general arguments for existence of multimodal equilibrium distributions may be extended to this system in a straightforward way.

A.3. Additional data and analysis for ornamentation model

We have additional data sets of ornament distribution from various species in Figures A.2 and A.3. The kernel density curves are superimposed for reference. If body size is a form of advertising, then we may also use data of salmon [141], trout [41], wolf spiders [39], and other bimodally distributed species. See Figure A.4.

While the work presented in Chapter 2 is based on mating displays in the animal kingdom, we hypothesize that similar forces operate on plants that compete within their

Data set	N	p-value (Dip test)	p-value (LUU test)	p-value (Dip test - log data)	p-value (LUU test - log data)	Tests reject unimodality?
Dung beetle horn length (Emlen [**144**])	223	0.0011**	0.0001***	0.0035**	0.0000***	yes
Yellow-breasted chat plumage coloration (Mays [**58**])	62	0.1932	0.0530	0.5479	0.2652	no
Peacock eye spots (Loyau [**57**])	24	0.6390	0.3793	0.5965	0.3187	no
Peacock eye spots (Petrie [**26**])	24	0.9183	0.7682	0.8809	0.6963	no
Peacock eye spots (Loyau/Petrie merged)	48	0.9016	0.6699	0.9006	0.6587	no
Arctic charr skin brightness (Skarstein [**63**])	20	0.2633	0.1558	0.2802	0.1658	no
Salmon body size (Glover [**42**])	72	0.6206	0.1467	0.7432	0.2497	no
Widowbird tail length (Anderson [**56**])	107	0.9992	0.9700	0.9972	0.9594	no
Widowbird red collar patch size (Anderson [**56**])	107	0.0046**	0.0002***	0.0317*	0.0030**	yes
Barn owl spottiness (Nieche [**55**])	20	0.6476	0.3858	0.7196	0.5157	no
Finch carotenoid coloration (Badyaev [**54**])	68	0.5295	0.1927	NA	NA	no
Stickleback nest compactness (Barber [**61**])	38	0.6085	0.2221	NA	NA	no
Partridge black ventral area (Bortolotti [**53**])	29	0.9032	0.6652	0.8704	0.5812	no
Roe deer antler length (Pelabon [**59**])	242	0.0341*	0.0012**	0.0232*	0.0001***	yes
Lion >2.2 yrs mane length (West [**25**])	441	0.8687	0.4134	0.9873	0.9521	no
Lion >2.2 yrs mane darkness (West [**25**])	442	0.9078	0.6698	0.9602	0.9033	no
Lion >5 yrs mane length (West [**25**])	257	0.8085	0.4779	0.8557	0.5356	no
Lion >5 yrs mane darkness (West [**25**])	257	0.8285	0.4129	0.8567	0.5173	no
Dung beetle horn length - WA (Moczek [**60**])	644	0.0000***	0.0000***	0.0000***	0.0000***	yes
Dung beetle horn length - NC (Moczek [**60**])	1016	0.0000***	0.0000***	0.0000***	0.0000***	yes
Earwig forceps length (Tomkins [**40**])	134	0.0000***	0.0000***	0.0000***	0.0000***	yes
Great tit stripe length (Norris [**52**])	63	0.2034	0.0781	NA	NA	no
Fiddler crab fight duration (Hyatt [**62**])	80	0.7059	0.2601	0.6362	0.3312	no
Fiddler crab fight acts (Hyatt [**62**])	80	0.8966	0.5273	0.9006	0.5714	no

Table A.1. Unimodality test results for animal ornamentation data sets. Hartigans' Dip Test (Dip test) is more conservative than our bootstrap dip test (LUU test); therefore our LUU test is more likely to reject unimodality. We performed both tests on log-transformed data because tissue measurements are often log-normally distributed [**145**]. We note in the rightmost column if the unimodality tests reject the null hypothesis that the distributions of ornament size are unimodal. Note that we exclude p-values for log-transformed data (NA) if the original data is not a straight-forward measurement of tissue investment.

own species for resources. For instance, a tree's height could be analogous to ornament size in our model, in that growing taller incurs costs to the individual, but being relatively taller in a forest has competitive benefits. In fact, certain tree species exhibit bimodal height distributions [**142, 143**]. See Figure A.5.

Data set	N	fractionation	morph means	morph variances	fractionation (log data)	morph means (log data)	morph variances (log data)
Dung beetle horn length (Emlen [144])	223	0.2372	0.2631	0.0055	0.0448	-2.8934	0.0005
		0.2677	1.0576	0.0112	0.2103	-1.3094	0.0509
		0.2414	0.7280	0.0142	0.3299	-0.3286	0.0553
		0.2156	0.1204	0.0018	0.1950	-2.0101	0.0412
		0.0380	0.5126	0.0000	0.2200	0.0629	0.0082
Yellow-breasted chat plumage coloration (Mays [58])	62	0.7247	40.2987	58.3743	0.2924	3.1794	0.0084
		0.2753	23.7743	4.9154	0.7076	3.6963	0.0302
Peacock eye spots (Loyau [57])	24	1.0000	152.0645	46.7236	1.0000	5.0233	0.0021
Peacock eye spots (Petrie [26])	24	1.0000	145.9515	95.9004	1.0000	4.981	0.0046
Peacock eye spots (Loyau/Petrie merged)	48	1.0000	149.0080	80.6543	1.0000	5.0021	0.0038
Arctic charr skin brightness (Skarstein [63])	20	0.4505	2.3538	0.0015	0.4507	0.8559	0.0003
		0.5495	2.5160	0.0004	0.5493	0.9226	0.0001
Salmon body size (Glover [42])	72	0.1383	9.3169	0.5107	0.1388	2.2296	0.0056
		0.8617	14.6375	1.0055	0.8612	2.6814	0.0046
Widowbird tail length (Anderson [56])	107	1.0000	221.5356	796.5005	1.0000	5.3920	0.0179
Widowbird red collar patch size (Anderson [56])	107	1.0000	222.1704	2419.6	1.0000	5.3779	0.0526
Barn owl spottiness (Nieche [55])	20	1.0000	1.2436	0.4555	1.0000	0.0695	0.3068
Finch carotenoid coloration (Badyaev [54])	68	1.0000	1.7732	3.1678	NA	NA	NA
Stickleback nest compactness (Barber [61])	38	0.8947	37.7314	99.9319	NA	NA	NA
		0.1053	90.2335	0.0131			
Partridge black ventral area (Bortolotti [53])	29	1.0000	21.1812	56.7020	1.0000	2.9779	0.1728
Roe deer antler length (Pelabon [59])	242	0.0903	12.1801	7.7178	0.1235	2.5521	0.0693
		0.9097	18.1135	5.1123	0.8765	2.8933	0.0144
Lion > 2.2 yrs mane length (West [25])	442	0.1936	0.6800	0.0192	0.7171	0.2489	0.0166
		0.8064	1.2663	0.0338	0.2829	-0.2681	0.0827
Lion > 2.2 yrs mane darkness (West [25])	442	1.0000	1.1008	0.0562	0.6464	0.1695	0.0217
					0.3536	-0.1118	0.0673
Lion> 5 yrs mane length (West [25])	257	1.0000	1.2977	0.0319	0.0383	-0.1331	0.0814
					0.9617	0.2652	0.0145
Lion > 5 yrs mane darkness (West [25])	257	1.0000	1.2021	0.0363	0.3205	0.0484	0.0351
					0.6795	0.2283	0.0142
Dung beetle horn length - WA (Moczek [60])	644	0.3546	0.5105	0.0033	0.4784	-0.6237	0.0224
		0.0837	2.0758	0.4042	0.2111	1.3512	0.0152
		0.1616	1.1310	0.0782	0.3105	0.1371	0.2152
		0.1910	0.6517	0.0110			
		0.2091	3.9032	0.2139			
Dung beetle horn length - NC (Moczek [60])	1016	0.2301	2.6811	0.6706	0.2423	0.1279	0.2082
		0.1633	0.9594	0.0686	0.1907	1.1523	0.0295
		0.2268	0.5430	0.0097	0.2292	-0.6075	0.0418
		0.3799	4.0161	0.1330	0.3378	1.4015	0.0064
Earwig forceps length (Tomkins [40])	134	0.3165	5.9727	0.7099	0.2964	1.8033	0.0144
		0.2501	7.3120	0.1154	0.2460	1.9901	0.0020
		0.4333	3.5705	0.0982	0.4576	1.2796	0.0098
Great tit stripe length (Norris [52])	63	0.5789	-14.1532	77.5214	NA	NA	NA
		0.4211	17.1432	60.9468			
Fiddler crab fight duration (Hyatt [62])	80	0.0500	482.1489	5555.8303	1.0000	3.8413	1.1077
		0.4433	19.5720	65.1629			
		0.1848	51.5698	33.6134			
		0.3219	125.9917	2189.0050			
Fiddler crab fight acts (Hyatt [62])	80	0.1103	53.5474	555.7647	1.0000	2.7213	0.5112
		0.2370	26.7320	14.2841			
		0.6526	11.3968	22.7231			

Table A.2. We fit Gaussian mixture models with 1–5 components of unequal variance to the animal ornamentation data sets and find the number of components that yields the best BIC [65]. We performed this fit on log-transformed data because tissue measurements are often log-normally distributed [145]. Note that we exclude Gaussian mixture models for log-transformed data (NA) if the original data is not a straight-forward measurement of tissue investment.

A.4. Data sources for tipping model

The tip rate data presented in Figure 4.1 of Chapter 4 is compiled from Zagat and NPD Research Group surveys. Primary sources proved difficult to obtain, so the majority of data points are reported in secondary sources. Table A.3 shows all sources.

Year	Average tip	Source
2016	18.9	Zagat State of American Dining report
		`https://goo.gl/sGBQQy`
2015	19.3	Zagat State of American Dining report
		`https://goo.gl/YhbDRW`
2014	19.0	Zagat Dining Trends survey
		`https://goo.gl/xEXE7R`
2013	19.4	Zagat State of American Dining report
		`https://goo.gl/iHilWI`
2012	19.2	Zagat State of American Dining report
		`https://goo.gl/tzvv0j`
2011	19.2	Zagat State of American Dining report
		`https://goo.gl/cmrHC5`
2010	19.1	Zagat State of American Dining report
		`https://goo.gl/vxxb19`
2009	19.0	Zagat State of American Dining report
		`http://a-r-n.net/Jin_e/ContentDetail.aspx?id=307754295`
2008	19.0	Zagat State of American Dining report
		`https://goo.gl/NIWO8R`
2007	18.9	Zagat State of American Dining report
		`https://goo.gl/qQIs8d`
2006	18.7	Zagat State of American Dining report
		`https://goo.gl/xC2u1P`
2005	18.4	Zagat State of American Dining report
		`http://www.hotel-online.com/News/PR2004_4th/Nov04_Zagat.html`
2004	18.3	Zagat State of American Dining report
		`https://goo.gl/lB4UiD`
2000	18	Zagat State of American Dining report
		`https://goo.gl/e4j57k`
1999	17	Zagat State of American Dining report
		`http://a-r-n.net/Jin_e/ContentDetail.aspx?id=307754295`
1996	16.4	Zagat State of American Dining report
		`https://goo.gl/pcFH8m`
1995	15*	Zagat State of American Dining report
		`https://goo.gl/yrok0h`
1989	15.4	Zagat State of American Dining report
		`https://goo.gl/38tLlc`
1984	14.7	NPD research
		`https://goo.gl/yrok0h`
1983	14.5	NPD research
		`https://goo.gl/yrok0h`
1982	14.4	NPD research
		`https://goo.gl/yrok0h`

Table A.3. Average tip rate sources, including original source and website (typically secondary source). *Rounding error suspected.
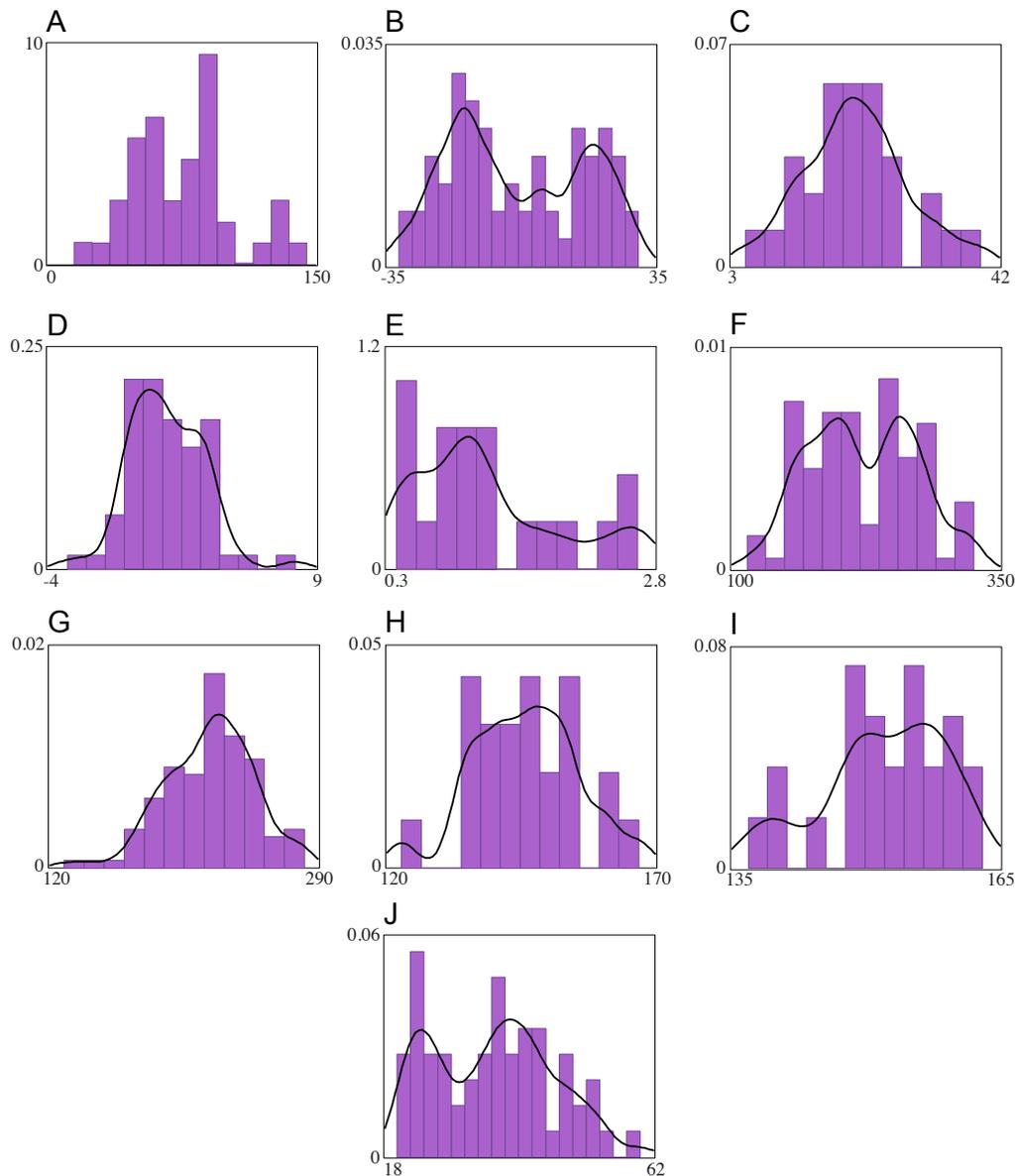
Figure A.2. Additional ornament data sets (birds) **A.** Blackbird song pulse repetition rate [**51**] (data extracted from histogram, so sample size uncertain) **B.** Great tit stripe size [**52**] (N=63) **C.** Partridge black ventral area [**53**] (N=29) **D.** Finch carotenoid coloration [**54**] (N=68) **E.** Barn owl spottiness [**55**] (N=20) **F.** Widowbird collar patch size [**56**] (N=107) **G.** Widowbird tail length [**56**] (N=107) **H.** Peacock eye spots [**26**] (N=24) **I.** Peacock eye spots [**57**] (N=24) **J.** Yellow-breasted chat plumage color [**58**] (N=62)
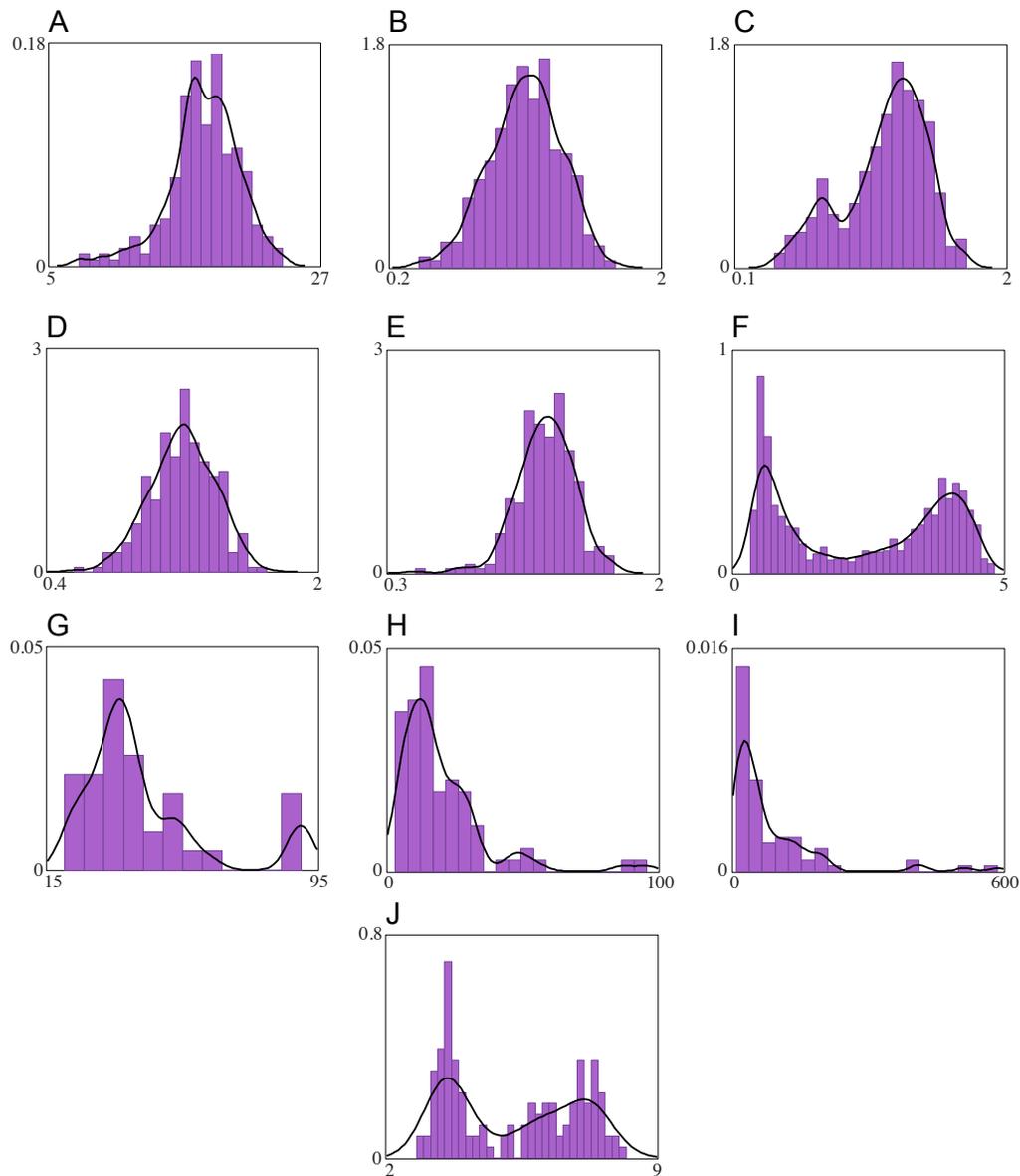
Figure A.3. Additional ornament data sets **A.** Roe deer antler length [**59**] (N=242) **B.** Mature (> 2.2 yr) lion mane darkness [**25**] (N=442) **C.** Mature (> 2.2 yr) lion mane length [**25**] (N=442) **D.** Older (> 5 yr) lion mane darkness [**25**] (N=257) **E.** Older (> 5 yr) lion mane length [**25**] (N=257) **F.** Dung beetle horn length (North Carolina) [**60**] (N=1016) **G.** Stickleback nest compactness [**61**] (N=38) **H.** Fiddler crab fight acts [**62**] **I.** Fiddler crab fight duration [**62**] **J.** Earwig forceps length [**40**] (N=134)
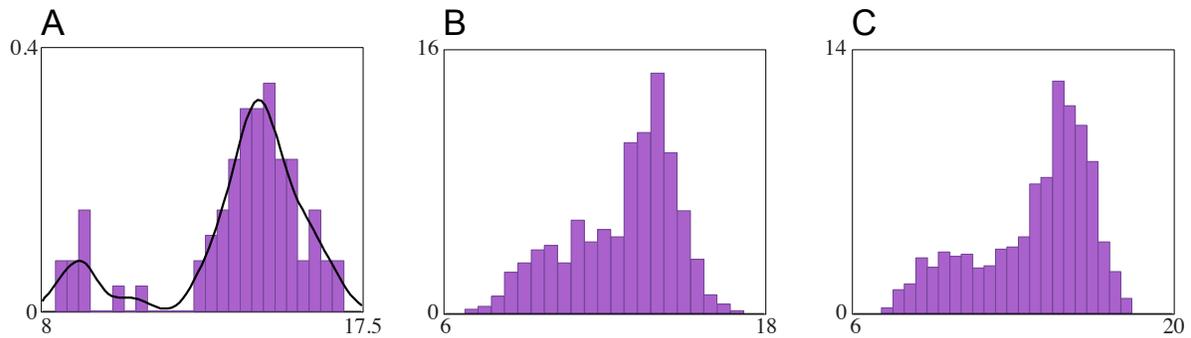
Figure A.4. Bimodal body size data sets **A.** Salmon body size [**42**] (N=72) **B.** Trout body size (early season) [**41**] (data extracted from histogram, so sample size uncertain) **C.** Trout body size (late season) [**41**] (data extracted from histogram, so sample size uncertain)
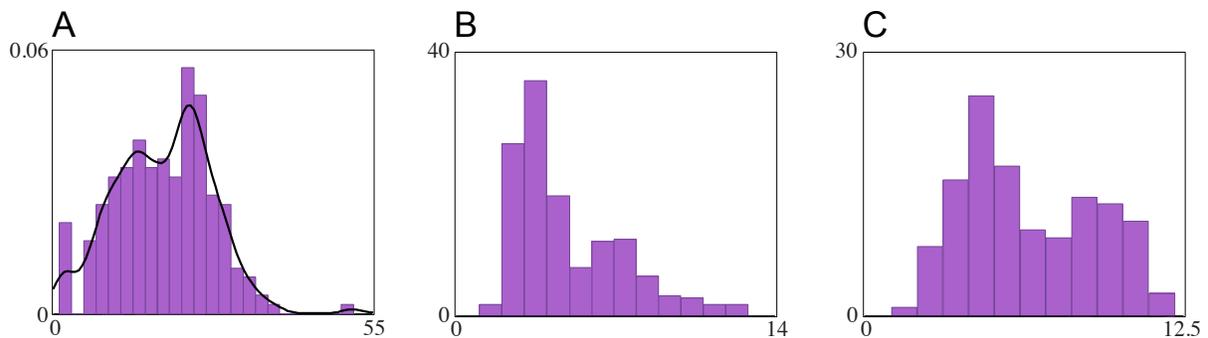

Figure A.5. Bimodal forest data sets **A.** Diameter at breast height for B. platyphylla trees [**142**] (N=217) **B.** Diameter at breast height for B. ermanii (11-16 yrs old) [**143**] (data extracted from histogram, so sample size uncertain) **C.** Height of B. ermanii (11-16 yrs old) [**143**] (data extracted from histogram, so sample size uncertain)