## NORTHWESTERN UNIVERSITY

# Multimodality in Social and Biological 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

Joseph D. Johnson

## EVANSTON, ILLINOIS

September 2020
(c) Copyright by Joseph D. Johnson 2020

All Rights Reserved


#### Abstract

Multimodality in Social and Biological Systems

Joseph D. Johnson

Perhaps because of the influence of the central limit theorem, it is common for scientists to assume distributions in the real world are singly peaked and unimodal. However, many quantities in nature are actually better represented by multimodal distributions. One must provide an explanation for this disconnect between the central limit theorem and reality. In this thesis, I investigate how multimodality may arise in three distinct systems by developing mathematical models.


First, I investigate how multimodality arises in an idealized system of coupled oscillators. I demonstrate that multimodality naturally emerges when oscillators repel each other for a broad class of interaction functions.

Second, I examine the impact of advertising on commodity product prices in the free market. I show that this system can support a hierarchical structure where the market segments into "name-brand" companies, advertising a significant amount and selling costly goods, and "generic-brand" companies, advertising minimally and selling cheap goods.

Third, I offer a theory that explains how anisogamy, size differences between sex cells, could have arisen. A common theory is that anisogamy factored into the development of sexual dimorphism. Using a dynamical systems model, I show that competition among zygotes in a population can lead to the development of anisogamy.

Finally, data can be skewed due to poor measurements and/or direct manipulation. Psychological pricing, where prices are set so that they are either aesthetically pleasing to the customer or set so that prices appear cheaper due to psychological factors, is an example of such distortion. I develop a method to correct for this predisposition to skew the documented value away from the "true" value of the data.

## Acknowledgements

I want to acknowledge my Mom and Dad. I would not have been able to attain many of my achievements without their love and support throughout my life. I want to thank Danny for the support over the last four years. Also, I thank the undergraduates who assisted with this work: Adam Redlich, Nathan White, and Alain Kangabire.

## Table of Contents

ABSTRACT ..... 3
Acknowledgements ..... 5
Table of Contents ..... 6
List of Tables ..... 9
List of Figures ..... 10
Chapter 1. Introduction ..... 22
1.1. Dissertation overview ..... 22
Chapter 2. The development of multimodality via coupled oscillators ..... 23
2.1. Introduction ..... 23
2.2. Model with antisymmetric repulsive coupling ..... 25
2.3. Concrete example ..... 30
2.4. Generalization to asymmetric interaction functions ..... 34
2.5. Generalization to non-identical oscillators ..... 35
2.6. Discussion ..... 37
2.7. Conclusions ..... 39
Chapter 3. Modeling advertising competition ..... 40
3.1. Introduction ..... 40
3.2. Results ..... 41
3.3. Model and analysis ..... 43
3.4. Discussion ..... 56
Chapter 4. Size dimorphism ..... 59
4.1. Introduction ..... 59
4.2. Results and discussion ..... 60
4.3. Model development ..... 61
4.4. Model implications ..... 65
Chapter 5. Debiasing data ..... 70
5.1. Introduction ..... 70
5.2. Method ..... 71
5.3. Numerical experiment ..... 74
5.4. Spatially varying smoothing ..... 75
5.5. Removing bias from real data ..... 77
5.6. Conclusions ..... 78
References ..... 79
Appendix A. The development of multimodality via coupled oscillators: appendix ..... 93
A.1. Trimodal equilibria ..... 93
A.2. Additional coupling functions ..... 97
A.3. Basins of attraction for multimodal states ..... 102
A.4. Critical coupling strength ..... 103
Appendix B. Modeling advertising competition: appendix ..... 107
B.1. Case with sigmoidal advertising payoff ..... 107
B.2. Welfare analysis ..... 112
B.3. Comparing model predictions and data ..... 114
B.4. Seeding the minimization algorithm ..... 115
B.5. Data adjustment ..... 118
B.6. Data Availability ..... 120
Appendix C. Size dimorphism: appendix ..... 122
C.1. Sex ratios ..... 122
C.2. Numerical simulations ..... 123
C.3. Stability tests ..... 124
C.4. Nonidentical individuals ..... 126
C.5. Nonzero size for small gamete group ..... 127
C.6. Absolute gamete fitness ..... 128
Appendix D. Debiasing data: appendix ..... 130
D.1. Derivation of the partial differential equation ..... 130

## List of Tables

3.1 Parameter definitions. Table of parameters used in the model with descriptions. 47

## List of Figures


from Eq. (2.12). (a) When initial fractionation is in the stable range (here $0.4<x_{\text {initial }}=0.55<0.6$ ) perturbations shrink and the solution returns to its initial state. (b) When initial fractionation is outside stable band (here $x_{\text {initial }}=0.65>0.6$ ) perturbations grow for some oscillators until system evolves to a different fractionation state. (c) When $x_{\text {initial }}=1 / 2$ and $a<a_{\text {crit }}$, perturbations shrink and the solution returns to its initial state. (d) When $x_{\text {initial }}=1 / 2$ and $a>a_{\text {crit }}$, perturbations grow and the system moves away from the unstable bimodal state until it reaches a new trimodal equilibrium.
Sample asymmetric interaction function. This function (solid blue curve)

| 2.5 |
| :---: |
| does not satisfy $f(\psi)=-f(-\psi)$. Existence of bimodal equilibria requires |
| that it intersect its mirror reflection (dotted blue curve) or a scaled version |
| of it (see Eq. (2.15)). The fixed points of the system for $x=1 / 2$ are marked |
| by black dots. |


| $\square$ |
| :--- |


| 2.6 | Numerical experiments with heterogeneous oscillators. Here, $N=1000$ |  |
| :---: | :---: | :---: |
|  | and oscillators' frequencies are drawn from the distribution $\mathcal{N}(0,100)$ |  |
|  | and the perturbations, $\xi_{i}, i=1 \ldots N$, are drawn from $\mathcal{N}(0,0.01)$. |  |
|  | Using example from Eq. (2.11), panels (a) and (b) show the results for |  |
|  | $x_{\text {initial }}=1 / 2$ and $a=\pi / 2<a_{\text {crit }}$ (compare to Fig. $2.4(\mathrm{a})$ ). Panels (c) and |  |
|  | (d) show the results for $x_{\text {initial }}=1 / 2$ and $a=\pi / \sqrt{2}+0.1>a_{\text {crit }}$ (compare |  |
|  | to Fig. $2.4(\mathrm{~d})$ ). | 37 |


green indicates that a state is stable and red indicates that a state is unstable.
The middle column, where $\mathrm{B}(\bar{a} \mid \vec{a})<k_{\mathrm{a}}<\max _{a} \mathrm{~B}(a \mid \vec{a})$ is where both states are stable. 52
3.5 Simulation of the system. In this figure we give snapshots of the numerical integration of the system from initial condition to equilibrium.

In panel (a), the system starts from the uniform randomly distributed state with the advertising initial condition set as $\mathcal{U}(7.5,12.5)$. In panel (b), the separation into two groups has begun. Companies change their spending until the lower group is far away from the mean, as seen in panel
(c). Finally, in panel (d), a bimodal equilibrium has been reached, with one group representing generic brand companies ( $a \approx 0$ ) and the other representing name-brand companies (advertising at a nonzero value at $\left.a=a_{\text {name }}\right)$. The green areas indicate where companies will increase their advertising and the red areas indicate areas where companies will decrease their advertising. In this simulation we set the number of companies to

$$
N=1000, k_{\mathrm{a}}=k_{\mathrm{P}}=k_{\mathrm{Q}}=\lambda=1, \text { and } Q_{\min }=\Delta Q_{\mathrm{ad}}=10 \text { (see Methods }
$$

section for parameter definitions). Additionally, advertising and production cost functions are linear.
$4.1 \quad$ Example individual reproductive potential function. Here we show the reproductive potential function defined by Eq. (4.4) in arbitrary units (a.u.).

Two maxima are apparent, one at zero and another at a nonzero value $s^{*}$.
Dynamics given by Eq. (4.5) are illustrated by color with red indicating
regions where gamete size decreases and blue indicating where gamete size increases. For this illustration, we set $w=1 / 10, \alpha=1$, and $\bar{s}=2$.

| 4.2 | Geometric argument for anisogamy. We illustrate a case where |
| :---: | :---: |
|  | gamete reproductive potential, $\varphi_{\mathrm{g}}$, and gamete production, $N_{\mathrm{g}}$, satisfy the |
|  | conditions set out in the section "Geometric argument" When gametes |
|  | are small, the relative gains due to the ability to produce more of them |
|  | $\left\|N_{\mathrm{g}}^{\prime} / N_{\mathrm{g}}\right\|$ outweigh the relative drop in reproductive potential $\left\|\varphi_{\mathrm{g}}^{\prime} / \varphi_{\mathrm{g}}\right\|$. In |

some intermediate range, reproductive potential gains dominate, and then
as gametes become very large the production terms again dominate as
reproductive potential gains saturate.
$5.1 \quad$ Ricker wavelet example. We visualize a Ricker wavelet (see Eq. (5.6p) with width $w=1$ and height $h=3$. 73
$5.2 \quad$ Setting smoothing to remove disturbance: theory versus numerics. We generate the integrated squared error between the model equilibrium and true distribution $f(x)$ given the initial condition $\rho_{0}(x)=f(x)+\xi(x)$, where $\xi(x)$ a Ricker wavelet defined in Eq. (5.6) and $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. The color indicates the amount of error with blue error corresponding to large error and yellow to small error. The $k$ value that yields the minimum error (red, dashed) for a given height of the Ricker wavelet is equivalent to the $k$ value derived from the Eq. (5.5) and (5.10) (black, solid). Here, $\sigma=w=5$ and $N=1000$
$5.3 \quad$ Heatmap of the integrated squared error when varying the height and the width of the perturbation. We compute the integrated squared error between the model result and true distribution $f(x)$ when the initial condition the sum $\rho_{0}(x)=f(x)+\xi(x)$, where $\xi(x)$ is the added perturbation. Here, the true distribution $f(x)$ is a Gaussian with variance $\sigma^{2}=25$ and mean $\mu=0$ and the perturbation $\xi(x)$ is a Ricker wavelet (see (5.6p) with width $w$ and height $h$. The error is minimized when $w=\sigma$ and $h=1 /\left(\sqrt{2 \pi} q^{2} \sigma^{3}\right)=0.38$, consistent with Eq. (5.10). The smoothing strength $k$ ranges from 0.01 to 0.99 with stepsize 0.005 and the height of the Ricker wavelet ranged from 0.0008 to 0.1596 with stepsize 0.0008 . 76
5.4 Removing perturbations. Panels (a)-(d) show the evolution from the perturbed distribution $\rho_{0}(x, t)=(1-\epsilon) \mathcal{N}\left(0, \sigma^{2}\right)+\epsilon \mathcal{N}\left(0, \delta^{2}\right)$, where $\epsilon \ll 1, \delta \ll \sigma$ to the equilibrium distribution when the smoothing parameter $k_{i}\left(y_{i, 0}, y_{i}\right)$ varies in space. The equilibrium distribution (red, solid) shows good agreement with the true distribution $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. The bias sensitivity parameter $\lambda$ was set by minimizing the integrated squared error, with the $\lambda \approx 1.8624 \times 10^{3}$. In this simulation, $\sigma=50, \delta=0.5, \epsilon=0.01$, and $N=1000$.
5.5 Removing bias from height data. We apply the our method with a spatial varying smoothing parameter $k_{i}\left(y_{i, 0}, y_{i}\right)$ set by Eq. (5.11). We see that the equilibrium generated by the model (red, circles) removes the second mode at 72 inches, which we assume to be due to overestimation.
A. 1 Numerical experiments testing the threshold for trimodality. Panel
(a): parameter value is $a=a_{\text {tricrit }}+0.1$, and the trimodal state appears tobe unstable (as expected). Panel (b): parameter va;ie is $a=a_{\text {tricrit }}-0.1$,and the trimodal state appears to be stable (as expected). Both panelsuse the example interaction function from Eq. (2.11), and both use equalfractionation $(x=y=z=1 / 3)$ and equal spacing between clusters( $\psi_{1}=2 \pi / 3$ ) in initial conditions.97

| A. 2 | Numerical experiments testing bistability. Panel (a) and (b): we set |  |
| :---: | :---: | :---: |
|  | $a=1.43 \pi \in\left(\sqrt{2} \pi, a_{\text {tricrit }}\right)$ and both the bimodal state and the trimodal state |  |
|  | are stable (as predicted). Panel (c): we set $a=1.43 \pi-0.1<\sqrt{2} \pi<a_{\text {crit }}$ |  |
|  | and we see that the bimodal state is unstable (we have added black dashed |  |
|  | lines so that one can see that the clusters away from the origin are not at |  |
|  | $\pm \pi$ ). Panel (d): we set $a=1.43 \pi+0.1>a_{\text {tricrit }}>\sqrt{2} \pi$ and the trimodal |  |
|  | state is unstable (as predicted). In all panels $N=300$ and the initial |  |
|  | conditions are equally spaced and have equal fractionation with a random |  |
|  | perturbation to all the phases of the oscillators. | 98 |

A. 3 Additional interaction functions. Solid blue curve: triangle wave from Eq. (A.14); solid red curve: antisymmetrized variant of the von Mises distribution from Eq. (A.15) with $\kappa<0$; dashed red curve: antisymmetrized variant of the von Mises distribution from Eq. (A.15) with $\kappa>0$. Panels (a) and (b) of Fig. $\mid$ A. 4 use the triangle wave. Panels (c) and (d) use the antisymmetrized von Mises function, with positive $\kappa$ (dashed red) in panel (c) and negative $\kappa$ (solid red) in panel (d). We note that for $\kappa>0$ the slope
at the $\pm \pi$ is never steeper when compared to the origin and for $\kappa<0$ the slope at the origin is never steeper when compared to the slope at $\pm \pi$.

| A. 4 | Numerical experiments using additional interaction functions. We test |  |
| :---: | :---: | :---: |
|  | the stability of the bimodal equilibria for alternative coupling functions |  |
|  | shown in Fig. $\mid$ A.3 ( a) Triangle wave coupling with initial fractionation in |  |
|  | predicted stable range. (b) Triangle wave coupling with initial fractionation |  |
|  | outside predicted stable range. (c) Von Mises coupling with $\kappa>0$ |  |
|  | (expected to be unstable). (d) Von Mises coupling with $\kappa>0$ (expected |  |
|  | to be stable). In all panels $N=100$ and oscillators' natural frequencies |  |
|  | are drawn from the distribution $\mathcal{N}(0,100)$. Initial phases are bimodally |  |
|  | distributed with modes at 0 and $\pi$, with perturbations $\xi_{i}, i=1, \ldots, N$, are |  |
|  | drawn from $\mathcal{N}(0,0.01)$. | 100 |


| Basins of attraction. We plot the fraction of uniform random initial |
| :--- |
| conditions that end up in bimodal (blue circles), trimodal (orange asterisks), <br>  <br> or higher order multimodal (purple xs) states for the concrete system <br> examined given by Eq. (2.11). Here $N=100, K=-10000$ and <br> oscillators' natural frequencies are drawn from the distribution $\mathcal{N}(0,100)$. <br> We performed 100 unique simulations for each value of $a$. Final states <br> (presumed equilibria) were identified automatically via $k$-means clustering. <br>  <br> Thresholds given by Eqns. (2.12) and (A.13) for stability of bimodality <br> and the antiphase state are given by the solid black line and the dot-dashed <br> green line, respectively. The threshold for the necessary condition for |

stability of the trimodal state, see Eq. (A.12), is given by the vertical dashed magenta line. 101
A. 6 Critical coupling strength. We perform numerical experiments to

| demonstrate the existence of a critical coupling strength for our system |
| :--- |
| and evaluate its dependence on parameter $a$ using the interaction <br> function defined in Eq. (2.11). Here $N=100$, the natural frequency <br> distribution is given by $\mathcal{N}\left(0, \sigma^{2}\right)$, and the initial phase distribution is | $\rho(\theta)=0.5 \delta(\theta)+0.5 \delta\left(\theta-\psi_{0}\right)$, where $\psi_{0}$ is the predicted phase separation given by the stable fixed points of Eq. (11). Here, each curve represents a different value of $a$ (values indicated in legend). As in the standard Kuramoto model, the critical coupling strength is dependant on the size of the standard deviation of the distribution, but unlike the standard Kuramoto model, it appears to also depend on $a$, which sets the shape of the interaction function.

B. 1 Optimized consumer welfare, profit and total welfare given marginal advertising costs. In (a) we display the optimal total consumer welfare generated by the market across all possible differentiated equilibria for a given marginal advertising cost $k_{\mathrm{a}}$ (blue, solid) and the total consumer welfare generated by the undifferentiated state (red, dashed). The black line in all three panels indicates when $k_{\mathrm{a}}=\max \mathrm{B}\left(a_{i} \mid \vec{a}\right)$ and thus, past that point the differentiated state ceases to be stable. In (b) we display the optimal total profit generated by the market across all possible differentiated equilibria for a given marginal advertising cost $k_{\mathrm{a}}$ (blue) and the total
profit generated by the undifferentiated state (red, dashed). In (c) we display the optimal total welfare generated by the market across all possible differentiated equilibria for a given marginal advertising $\operatorname{cost} k_{\mathrm{a}}$ (blue, solid) and the total welfare generated by the undifferentiated state (red, dashed). We set $N=100, Q_{\text {min }}=30, \Delta Q_{\mathrm{ad}}=5$, and $\lambda=\mu=\nu=k_{\mathrm{Q}}=k_{\mathrm{P}}=1$, with $k_{\mathrm{a}}$ ranging over a range 0 to 45 . 113
B. 2 Histograms of Fitting Statisitics. Distributions of $p$ values from tests for unimodality (a-b) and consistency of price data with model predictions (c). The red dashed line indicates the significance level of .05. Top and middle row: LUU and Hartigan's Dip Test (respectively). Rejection (low $p$-value) means the price distribution is not consistent with unimodal null hypothesis. Bottom row: KS test. Rejection (low $p$-value) means the price distribution is not consistent with same-distribution null hypothesis (i.e., model and data not from same distribution).
B. 3 Raw and Smoothed Price Data. Example of the raw price data with spikes situated near certain ending digits (blue) and the smoothed, debiased data that retains the quantitative properties of the underlying distribution while removing the spikes (orange). Red arrows point to example spikes located at $\$ 2.99, \$ 3.49$, and $\$ 3.99$.
C. 1 Possible sex ratios. The solid black curve shows the threshold for existence of the anisogamous state given by Eq. (C.1). The anisogamous equilibrium exists below the threshold (blue shaded region) and ceases to exist above
the threshold (red shaded region). Here the fraction "male" refers to the fraction with small gametes.
C. 2 Simulation of the system. Panels (a)-(d) show the evolution of the system
from an isogamous state to an anisogamous state. Here, the blue curve shows the reproductive potential landscape given by Eq. (4.4), the red circles indicate gamete sizes, and the yellow crosses give positions of gametes along the reproductive potential landscape. Panel (a) captures the isogamous initial condition $\mathcal{U}(1,3)$. Panel (b) shows the individuals moving along the landscape in the direction that increases reproductive potential. Panel (c) shows the beginning of two groups forming. In Panel (d), the simulation has arrived at an anisogamous equilibrium, with gamete sizes converging to zero or $s^{*}$ as given by Eq. (4.8). The final fraction of organisms that produce small gametes is $x=0.1$. For this numerical experiment, we set $\alpha=1, N=100$, and $w=1 / 10$.
C. 3 Numerical test of the stability of anisogamy. We test the stability of the anisogamous state. Panel (a) shows the large gamete group being perturbed and then returning to its equilibrium value, $s^{*}$. Panel (b) shows the small gamete group being perturbed and then returning to its equilibrium value zero. Panel (a) and (b) demonstrate the stability of the anisogamous equilibrium and are consistent with the asymptotic theory from Eqns (4.9) and (4.11). In both panels, we set $\alpha=1 / 3, N=1000$, and $w=2$, with an initial fractionation $x=1 / 2$. 126
C. 4 Simulation with a heterogeneous population. We display the evolution from isogamy to anisogamy population for a heterogeneous population. The large gamete group widens out when adding noise to the width of the sigmoid in Eq. (4.3). The final fraction of organisms that produce small gametes is $x=0.1$. For this simulation, we set $\alpha=1, N=100$, and $w$ was sampled from the distribution $\mathcal{N}\left(1 / 10,1 / 50^{2}\right)$.
C. 5 Simulation with a nonzero minimum gamete size. We display a simulation where the individual reproductive potential is multiplied by $e^{-k / s}$, where $k>0$. The initial isogamous population moves to an anisogamous population $\rho(s)=x \delta\left(s-s_{\text {small }}^{*}\right)+(1-x) \delta\left(s-s^{*}\right)$, $0<s_{\text {small }}^{*}<s^{*}$. Here, $N=100, w=1, \alpha=1, k=1$, and the final fraction of small gametes is $x=0.16$. The initial isogamous population was drawn from $\mathcal{U}(1,3)$. 128
C. 6 Simulation with both absolute and relative reproductive potential. In
this simulation the individual reproductive potential was a weighted sum of two sigmoidal functions, one as in the Eq. (4.3) (i.e., centered at $\bar{s}$ ), and the other identical but centered at $c=1$. Weight was $90 \%$ absolute, $10 \%$ relative. The population converges to an anisogamous state with 40\% small gametes. Here, $N=100, w=0.1+\mathcal{N}\left(0,0.01^{2}\right), \alpha=1$, and the initial population was drawn from $\mathcal{U}(0,1.5)$.

## CHAPTER 1

## Introduction

The act of modeling consists of taking the real world and boiling it down to a simplified version that focuses on aspects that one deems important. Given the way they simplify the world, models are not a one-to-one representation of reality, but, rather, a model's value should be judged by its ability to do two things: explain phenomena and make predictions.

A model allows one to make sense of the world and a simple model gives a clear and concise story as to how the world works. The models laid out in this dissertation give insight into the inner workings of various social and biological phenomena.

### 1.1. Dissertation overview

This thesis is divided into four main chapters encapsulating my PhD research at Northwestern University. Each chapter corresponds to a project resulting in a manuscript that has been or will be submitted for publication.

In Chapter 2, I use coupled oscillators as a testbed case to explore how multimodality emerges. In Chapter 3, I demonstrate that advertising can partition the market into "namebrand" and "generic-brand" companies, where "name-brand" companies advertise a consequential amount, and "generic-brand" companies advertise a negligible amount. In Chapter 4, I develop theory that gives an account as to how sex cell size dimorphism came to be. Finally, in Chapter 5, I put forth a method to correct data skewed due to measurement error and/or intentional distortion.

## CHAPTER 2

## The development of multimodality via coupled oscillators

### 2.1. Introduction

Synchronization is a widespread phenomenon observed in biological [1-3], chemical [4-6], physical [7-10], and social settings [11-14]. A paradigmatic mathematical model that can explain synchronization in many contexts is the Kuramoto model [15-19]. Much work has been done on understanding the complex and surprising dynamics of the Kuramoto model and its variants, but the vast majority of that research focuses on the case of attractive coupling; here we are interested in the case where the coupling is repulsive.

Repulsive (or inhibitory) coupling is of physical interest as it arises frequently in the context of neuronal networks (e.g., see refs. [2, 20]), chemical interactions (e.g., refs. [4, 21, 22]), and many other systems (see refs. [23-28]). Some coupled oscillator models have examined repulsive coupling: Giver et al. developed a local variant of the Kuramoto model with repulsive coupling based on the interaction between water micro-droplets with reactants of the BelousovZhabotinsky reaction [29]. Hong and Strogatz developed two variants of the Kuramoto model that involved mixes of positive and negative coupling [30,31].

The relationship between network structure and repulsive coupling has also been analyzed, with Levnajić [42,43] showing that, given the network coupling structure, many different phase configurations can arise. Recently, it has been shown that synchronization can arise in both repulsive and attractive coupling scenarios subject to common noise [44-47]. Gong et al. [47],


Figure 2.1. Selected examples of bimodality. Histograms (normalized) for (a) size of beetle horns [mm], [32-34] (b) Atlantic salmon body mass [g] [35,36] (c) color of galaxies at redshift 0.1 [37-39] (d) inverse growth rates of bacteria $\left[\min ^{-1}\right]$ 40, 41].
inspired by the work of Gil et al. [46], studied instances where common noise can lead to clustering in the phase distribution of oscillators for repulsive coupling.

Nakamura et al. [48] investigated the effect of time-delayed nearest-neighbor coupling in the Kuramoto model and found that it could lead to the development of clustered states for both attractive and repulsive coupling. Mishra et al. [49] demonstrated that "chimeralike" states could arise with globally coupled Lienard systems incorporating both attractive and repulsive mean-field feedback. Yeldesbay et al. [50] established that chimeralike states can arise in the Kuramoto-Sakaguchi model. They also considered a model with oscillators that could be synchronous (attractive coupling) or asynchronous (repulsive coupling) depending on their natural frequencies. They found that in this case a chimera state arises.

Golomb et al. [51] showed that clustering is possible in a coupled oscillator model with repulsive coupling that is suited for strong interactions between the limit-cycle oscillators. They further provided theory for when a frequency locked stationary phase distribution and when a nonperiodic attractor can arise.

Tsimring et al. [52] showed that heterogeneous globally coupled oscillators obeying the standard Kuramoto model can cluster with all configurations having a zero order parameter, but this clustering breaks down as the number of oscillators increases. They also showed that, with local coupling, clustering can occur for nonidentical oscillators given sufficiently large coupling strength.

Closest to the work we present here, Okuda [53] looked at the effect that an arbitrary coupling function may have on oscillators and developed theory as to when an $n$-cluster state, with all clusters being the same size, can arise. He found that harmonics in the coupling function are necessary for clusters to arise.

The central limit theorem [54] may influence us to expect that distributions in nature should tend to a singly-peaked, unimodal shape akin to the Gaussian normal distribution. Yet, bimodality and multimodality can be observed in biological [55-57], social [58-60], and chemical [61-64] contexts and beyond [65-67] (see Fig. 2.1]for selected examples). In this chapter we demonstrate that multimodality may arise as a result of repulsive or inhibitory coupling dynamics and we give an in-depth explanation of how it can arise for a range of coupling functions.

### 2.2. Model with antisymmetric repulsive coupling

We begin by considering a system of $N$ phase oscillators characterized by natural frequencies $\omega_{i}, i=1 \ldots N$. The oscillators are globally coupled with coupling strength $K$ through an
interaction function $f$ that depends only on the phase difference between each pair of oscillators:

$$
\begin{equation*}
\dot{\theta}_{i}=\omega_{i}+\frac{K}{N} \sum_{i=1}^{N} f\left(\theta_{j}-\theta_{i}\right), \quad i=1, \ldots, N . \tag{2.1}
\end{equation*}
$$

Here $K>0$ represents attractive coupling and $K<0$ represents repulsive coupling.
We consider interaction functions $f(u), u \in(-\pi, \pi]$, that satisfy the following conditions:

$$
\begin{align*}
& f(0)=0  \tag{2.2a}\\
& f^{\prime}(0)>0 \tag{2.2b}
\end{align*}
$$

$$
\begin{align*}
& f(u)=-f(-u)  \tag{2.2c}\\
& f^{\prime}(u) \text { continuous }  \tag{2.2d}\\
& f(\pi)=\lim _{u \rightarrow-\pi^{+}} f(u) . \tag{2.2e}
\end{align*}
$$

These conditions impose: 2.2a no coupling effects between oscillators in sync; 2.2b) locally attractive (repulsive) coupling near sync state for $K>0(K<0)$; 2.2c) odd interaction function; (2.2d) no discontinuities in $f^{\prime}(u)$; (2.2e) $2 \pi$-periodic interaction function on $(-\pi, \pi]$ domain. We point out that conditions (2.2c) and (2.2e) lead to $f(\pi)=\lim _{u \rightarrow-\pi^{+}} f(u)=0$.

### 2.2.1. Identical Oscillators

We assume that oscillators frequencies are drawn from a known frequency distribution $g(\omega)$.
For simplicity we first consider the case of identical oscillators, i.e., we set the distribution to


Figure 2.2. Sample interaction functions. Two cases of coupling functions that we consider. Case 1 (red, dashed) is an odd, $2 \pi$-periodic function with a continuous derivative, no zeros in between 0 and $\pi$, and has a positive slope at 0 . Case 2 (blue, solid) is similar to case 1 but has a zero of order 1 in between 0 and $\pi$.
be $g(\omega)=\delta\left(\omega-\omega_{0}\right)$, so the system becomes

$$
\begin{equation*}
\dot{\theta}_{i}=\omega_{0}+\frac{K}{N} \sum_{j=1}^{N} f\left(\theta_{j}-\theta_{i}\right), \quad i=1, \ldots, N . \tag{2.3}
\end{equation*}
$$

### 2.2.2. Bimodal equilibria

We assume that the number of oscillators is large, $N \gg 1$, and we look for bimodal equilibria by making the ansatz of an oscillator phase distribution $h(\theta)=x \delta\left(\theta-\theta_{1}\right)+(1-x) \delta\left(\theta-\theta_{2}\right)$, where $0<x<1$ describes the fraction in cluster 1 . Note that this constitutes an explicit restriction to a bimodal manifold within the broader space of all possible oscillator phase distributions.

[^0]Then system (2.3) can be reduced to two coupled ordinary differential equations (ODEs):

$$
\begin{align*}
\dot{\theta}_{1} & =\omega_{0}+\frac{K}{N}\left(\sum_{i=1}^{x N} f\left(\theta_{1}-\theta_{1}\right)+\sum_{i=x N+1}^{N} f\left(\theta_{2}-\theta_{1}\right)\right) \\
& =\omega_{0}+K(1-x) f\left(\theta_{2}-\theta_{1}\right)  \tag{2.4}\\
\dot{\theta}_{2} & =\omega_{0}+\frac{K}{N}\left(\sum_{i=1}^{x N} f\left(\theta_{1}-\theta_{2}\right)+\sum_{i=x N+1}^{N} f\left(\theta_{2}-\theta_{2}\right)\right) \\
& =\omega_{0}-K x f\left(\theta_{2}-\theta_{1}\right) . \tag{2.5}
\end{align*}
$$

We define a new phase-difference variable $\psi=\theta_{2}-\theta_{1}$ and write its dynamical system by subtracting Eq. (2.4) from Eq. (2.5):

$$
\begin{equation*}
\dot{\psi}=-K f(\psi) . \tag{2.6}
\end{equation*}
$$

We observe that the fixed points of the system for $\psi$ are fully determined by the zeros of $f(\psi)$. From the assumptions above $f(\psi)$ must have zeros at $\psi=0$ and $\psi=\pi$. Furthermore, if conditions 2.2a 2.2e hold and $f(\psi)$ has no other zeros (as in the case of the red dashed curve from Fig. 2.2, then it is implied that $f^{\prime}(\pi) \leq 0$. Hence, within the bimodal manifold, the fixed point at $\psi=\pi$ should be stable with $\psi=0$ being unstable. $\psi=\pi$ corresponds to a bimodal equilibrium with two clusters of oscillators separated by $180^{\circ}$ of phase.

If additional roots of $f(\psi)$ exist between 0 and $\pi$, these will also correspond to bimodal fixed points with alternating stability (again restricted to the bimodal manifold). We focus on the cases where there are no other fixed points or there is exactly one other fixed point $\psi_{0}$ in $(0, \pi)$; other cases are similarly tractable. Figure 2.2 illustrates the typical general shapes of the interaction functions that we consider.
2.2.2.1. Stability of bimodal equilibrium. To investigate the broader stability of solutions to perturbations outside the bimodal manifold, we consider the perturbation of a single oscillator by a small amount $\epsilon$. Because $N \gg 1$, we approximate the dynamics of the two clusters as unaffected by this perturbation. We examine the evolution of distance between the perturbed oscillator and the group from which it was perturbed, $\epsilon(t)$, to evaluate whether the system returns to its initial state.

For convenience, we move into a rotating frame by redefining $\theta_{i} \rightarrow \theta_{i}+\omega_{0} t$, which is equivalent to setting $\omega_{0}=0$. Without loss of generality we choose oscillator index $N$ from the $\theta_{2}$ cluster for the perturbation and assume $\theta_{1}=0$, and thus $\theta_{2}=\psi_{0} \leq \pi$ (assuming for now that our interaction function has only one or zero fixed points in $(0, \pi)$ ). Then $\theta_{N}=\theta_{2}-\epsilon=\psi_{0}-\epsilon$, and

$$
\begin{aligned}
\dot{\epsilon} & =-\frac{K}{N}\left[\sum_{i=1}^{x N} f\left(\theta_{1}-\psi_{0}+\epsilon\right)+\sum_{i=x N+1}^{N-1} f\left(\theta_{2}-\psi_{0}+\epsilon\right)\right] \\
& =-K x f\left(-\psi_{0}+\epsilon\right)-K(1-x) f(\epsilon) .
\end{aligned}
$$

We expand the functions $f$ in a Taylor series to linear order:

$$
\dot{\epsilon} \approx-\epsilon K\left[x f^{\prime}\left(\psi_{0}\right)+(1-x) f^{\prime}(0)\right]
$$

Assuming that $K<0$ (repulsive coupling, our case of interest in this chapter), this implies stability if and only if

$$
\begin{equation*}
x f^{\prime}\left(\psi_{0}\right)+(1-x) f^{\prime}(0)<0 . \tag{2.7}
\end{equation*}
$$

A nearly identical calculation starting with the perturbation of a single oscillator from the $\theta_{1}$ (zero phase) cluster leads to a similar equation,

$$
\begin{equation*}
(1-x) f^{\prime}\left(\psi_{0}\right)+x f^{\prime}(0)<0 \tag{2.8}
\end{equation*}
$$

Since Eqs. 2.7) and (2.8) must be simultaneously satisfied for stability of the full bimodal distribution, the following inequality must hold:

$$
\begin{equation*}
f^{\prime}(0)<(1-x)\left[f^{\prime}(0)-f^{\prime}\left(\psi_{0}\right)\right]<-f^{\prime}\left(\psi_{0}\right) \tag{2.9}
\end{equation*}
$$

Interestingly, this implies that the slope of the interaction function $f(\psi)$ must be steeper at $\psi=\psi_{0}$ compared to $\psi=0$ if the bimodal state is to be stable. We can also compute explicit bounds on the proportion of the oscillators in each group by isolating $x$ in inequality (2.9) :

$$
\begin{equation*}
\frac{f^{\prime}(0)}{f^{\prime}(0)-f^{\prime}\left(\psi_{0}\right)}<x<\frac{-f^{\prime}\left(\psi_{0}\right)}{f^{\prime}(0)-f^{\prime}\left(\psi_{0}\right)} . \tag{2.10}
\end{equation*}
$$

### 2.3. Concrete example

As a concrete example, we consider a simple class of interaction functions

$$
\begin{equation*}
f(u ; a)=\frac{1}{\pi^{2} a^{2}} u\left(\pi^{2}-u^{2}\right)\left(a^{2}-u^{2}\right) . \tag{2.11}
\end{equation*}
$$

These functions have roots on $(-\pi, \pi]$ at $0, \pi$, and $\pm a$, and satisfy all the conditions set forth earlier in section 2.2. As long as $0<|a|<\pi$ there are three roots in $0 \leq u \leq \pi$, and one can check that $f^{\prime}(0)=1$ for all choices of $a$ (see Fig. 2.3 for example plots). For inequality (2.10)


Figure 2.3. Concrete interaction function. The interaction function defined in Eq. 2.11) plotted for several different values of $a: \sqrt{3} \pi / 4$ (black, dotted), $\sqrt{6} \pi / 4$ (red, dashed), and $3 \pi / 4$ (blue, solid). As the value of $|a|$ approaches $\pi$ the slope at zero stays fixed with slope 1 and the slope at $\pm a$ decreases in magnitude. This relation between $a$ and the slope values at $\pm a$, combined with Eq. (2.10) leads to the threshold for bimodality given by Eq. (2.12).
to be satisfiable, we require

$$
\frac{\pi^{2}}{3 \pi^{2}-2 a^{2}}<\frac{2 \pi^{2}-2 a^{2}}{3 \pi^{2}-2 a^{2}}
$$

which reduces to

$$
\begin{equation*}
|a|<\pi / \sqrt{2} \equiv a_{\mathrm{crit}} . \tag{2.12}
\end{equation*}
$$

We note that symmetry of the roots allows us to consider positive $a$ without loss of generality. Figure 2.4 shows the results of numerical experiments where we test this predicted stability threshold. In each panel, Eq. (2.3) is implemented with the interaction function from Eq. (2.11). We initialize $x N$ oscillators at $\theta_{1}=0$ and $(1-x) N$ at $\theta_{2}=a$, then add a small random perturbation $\xi_{i}$ to each oscillator's initial phase, where $\xi_{i}$ is drawn from the normal distribution $\mathcal{N}\left(0, \delta^{2}\right)$, with $\delta=0.1$ used in Fig. 2.4. We numerically integrate the system using a 4 th $/ 5$ th order Runge-Kutta scheme and consider evidence for stability if it approaches the unperturbed state, i.e. $\psi=\theta_{2}-\theta_{1} \rightarrow a$ with $x_{\text {final }}=x_{\text {initial }}$. We note that in these experiments we set coupling strength $K=-100 q^{2}$.

In panels (a) and (b), we use $N=100$ oscillators, $\omega_{0}=0$, and set $a=\pi / 2$, consistent with the stability threshold from Eq. (2.12), $a<a_{\text {crit }}=\pi / \sqrt{2}$. The stable band of fractionation according to inequality $(2.10)$ is then $2 / 5<x<3 / 5$. In panel (a), we set $x_{\text {initial }}=0.55$, below the band's upper bound; in panel (b), we set $x_{\text {initial }}=0.65$, above the band's upper bound. As expected, the bimodal equilibrium appears stable in panel (a), but unstable in panel (b), where eleven oscillators move between clusters to establish a different equilibrium within the stable fractionation band ( $2 / 5<x_{\text {final }}=0.54<3 / 5$ ).

In panels (c) and (d), we again use $N=100$ oscillators and $\omega_{0}=0$, but here we examine the predicted stability threshold $a_{\text {crit }}=\pi / \sqrt{2}$ from Eq. (2.12). We expect the bimodal state with $\psi^{*}=a$ to be unstable for all positive $a>a_{\text {crit }}$ (but note that this state ceases to exist when $a>\pi)$. We set $x_{\text {initial }}=1 / 2$ since this is within the fractionation stability band from inequality (2.10) for all $a<a_{\text {crit }}$. In panel (c), we set $a=a_{\text {crit }}-0.1$, just below the threshold for stability; in panel (d), we set $a=a_{\text {crit }}+0.1$, just barely in the unstable domain. As expected, the bimodal

[^1]

Figure 2.4. Numerical experiments with identical oscillators. Using example from Eq. 2.11), top two panels show test for stability range of fractionation $x$ from Eq. (2.10; ; bottom two panels show test for critical parameter $a_{\text {crit }}$ from Eq. (2.12). (a) When initial fractionation is in the stable range (here $0.4<$ $\left.x_{\text {initial }}=0.55<0.6\right)$ perturbations shrink and the solution returns to its initial state. (b) When initial fractionation is outside stable band (here $x_{\text {initial }}=0.65>$ 0.6 ) perturbations grow for some oscillators until system evolves to a different fractionation state. (c) When $x_{\text {initial }}=1 / 2$ and $a<a_{\text {crit }}$, perturbations shrink and the solution returns to its initial state. (d) When $x_{\text {initial }}=1 / 2$ and $a>a_{\text {crit }}$, perturbations grow and the system moves away from the unstable bimodal state until it reaches a new trimodal equilibrium.
equilibrium again appears stable in panel (c), but it appears unstable in panel (d). Since no fractionation $x$ will lead to a stable bimodal equilibrium, the system must move to an entirely different state, and it appears to converge to a trimodal distribution of oscillator phases.

We are able to understand why the system converges to a trimodal state by performing a similar analysis for the stability of three-cluster, or trimodal, oscillator distributions. One can show that a necessary condition for stability is:

$$
\begin{align*}
f^{\prime}(0)< & -\left[(x+y) f^{\prime}\left(\psi_{1}\right)+(y+z) f^{\prime}\left(\psi_{2}\right)\right. \\
& \left.+(x+z) f^{\prime}\left(\psi_{1}+\psi_{2}\right)\right] \tag{2.13}
\end{align*}
$$

where $\psi_{i}$ is the angle separating clusters at $\theta_{i}$ and $\theta_{i+1}\left(\theta_{4}\right.$ identified with $\left.\theta_{1}\right)$, and $x, y$, and $z$ are the fractionations of the three clusters at $\theta_{1}, \theta_{2}$, and $\theta_{3}$ respectively. With equal spacing between the clusters $\psi_{1}=\psi_{2}=2 \pi-\psi_{1}-\psi_{2}=2 \pi / 3$, the necessary condition simplifies to

$$
f^{\prime}(0)<-2 f^{\prime}\left(\frac{2 \pi}{3}\right)
$$

For the example function shown in Eq. (2.11) this is

$$
a<\frac{2}{3} \sqrt{\frac{14}{3}} \pi \equiv a_{\text {tricrit }} \approx 1.44 \pi
$$

This implies that a trimodal state remains stable for all $a<\pi$. It stably coexists with the bimodal state for $a<\pi / \sqrt{2}$, and may coexist with other multimodal states for $\pi / \sqrt{2}<a<\pi$.

In general different multimodal states may stably coexist over various parameter ranges. More details of the analysis for trimodality can be found in Section A.1 in the appendix.

### 2.4. Generalization to asymmetric interaction functions

We can relax assumption (2.2c) of an antisymmetric coupling function and still find stability boundaries for multimodal states. In place of Eq. (2.6) (which used oddness of the coupling
function), we find instead

$$
\begin{equation*}
\dot{\psi}=K x f(-\psi)-K(1-x) f(\psi) \tag{2.14}
\end{equation*}
$$

Clearly $\psi^{*}=0$ and $\psi^{*}=\pi$ both remain fixed points. Other fixed points exist if

$$
\begin{equation*}
x f\left(-\psi^{*}\right)=(1-x) f\left(\psi^{*}\right) \tag{2.15}
\end{equation*}
$$

has a solution on $-\pi<\psi^{*} \leq \pi$. Figure 2.5 shows an example of an asymmetric interaction function. Geometrically this condition can be understood as identifying intersections of $f(\psi)$ and its reflection $f(-\psi)$ when $x=1 / 2$ (or scaled versions when $x \neq 1 / 2$ ). Once multimodal fixed points are identified, stability analysis is analogous to that presented earlier.

### 2.5. Generalization to non-identical oscillators

We argue that real-world bimodal or multimodal distributions may result from similar dynamics to those presented in this chapter. Of course, heterogeneity is inevitable in most realworld systems, yet we have focused thus far on the case of identical oscillators. While we leave the more general analysis for future work, we have conducted numerical experiments that appear to show that the predicted behavior occurs even in the presence of oscillator heterogeneity.

Again using the same example interaction function from Eq. 2.11), we now draw frequencies, $\omega_{i}$, from a normal distribution $\mathcal{N}\left(0, \sigma^{2}\right)$ and set the initial phases of the oscillators to $\theta_{i}=\xi_{i}($ fraction $x)$ or $\theta_{i}=a+\xi_{i}($ fraction $1-x)$, where $\xi_{i}$ is a small perturbation drawn from the distribution $\mathcal{N}\left(0, \delta^{2}\right)$. Figure 2.6 shows the results of perturbation experiments analogous to those presented in Fig. 2.4, with analogous results except that the final phase distributions have


Figure 2.5. Sample asymmetric interaction function. This function (solid blue curve) does not satisfy $f(\psi)=-f(-\psi)$. Existence of bimodal equilibria requires that it intersect its mirror reflection (dotted blue curve) or a scaled version of it (see Eq. (2.15)). The fixed points of the system for $x=1 / 2$ are marked by black dots.
phases that cluster about the modes rather than all converging to them precisely (right panels show histograms of final states).

In Fig. 2.6 panels (a) and (b), we use $N=1000$ oscillators and set $a=\pi / 2<a_{\text {crit }}$ and $x_{\text {initial }}=1 / 2$. Even with perturbed initial phases and heterogeneous natural frequencies, the oscillators still remain in the bimodal state as predicted for $a<a_{\text {crit }}$. Specifically, panel (b) shows that the steady state distribution of oscillators has finite-width clustering about the


Figure 2.6. Numerical experiments with heterogeneous oscillators. Here, $N=1000$ and oscillators' frequencies are drawn from the distribution $\mathcal{N}(0,100)$ and the perturbations, $\xi_{i}, i=1 \ldots N$, are drawn from $\mathcal{N}(0,0.01)$. Using example from Eq. (2.11), panels (a) and (b) show the results for $x_{\text {initial }}=1 / 2$ and $a=\pi / 2<a_{\text {crit }}$ (compare to Fig. 2.4(a)). Panels (c) and (d) show the results for $x_{\text {initial }}=1 / 2$ and $a=\pi / \sqrt{2}+0.1>a_{\text {crit }}($ compare to Fig. 2.4(d)).
fixed point positions predicted from the identical-oscillator case. In panels (c) and (d), since $a=\pi / \sqrt{2}+0.1>a_{\text {crit }}$, the bimodal state breaks down (consistent with the prediction of the identical-oscillator theory) and the system appears to converge to a trimodal equilibrium with three finite-width clusters.

### 2.6. Discussion

Coupled oscillators are an excellent testbed for models of synchronization or clustering. Even though real-world variables (e.g., sediment grain size [68], salmon body size [35, 69],
human communication frequency [70], dopamine receptor density [71], neutron star mass [72], galaxy color [39], gamma ray burst duration [73], tree height [74], animal ornament size [33]) may not be oscillatory or confined to a periodic domain, bimodality may emerge for qualitatively similar reasons. In our model, the coupling of one unit's dynamical behavior to that of others is key to the phenomenon.

For clarity of presentation we have focused on a single example of interaction function (Eq. (2.11)), but evaluation of two other classes of interaction functions (triangle waves and antisymmetrized von-Mises kernels) also supports our analytical results-see Section A. 2 in appendix for details. In Sections A. 3 and A.4, we also present further results regarding dependence of bimodal equilibria on coupling strength $K$, as well as some numerical evidence regarding sizes of basins of attraction; each of these topics merits further in-depth study. The analysis we present here focuses exclusively on the case of all-to-all coupling; we leave further investigation of the impact of network structure for future work.

For real-world scenarios where bimodality or multimodality is of interest, the interaction function may not be known exactly. Nevertheless, we expect that it will often be possible to assess whether the conditions expressed in Eqns. (2.2) and (2.9) hold in a particular case. It also seems plausible that functions describing real-world interactions between coupled systems will have no more than a handful of roots, making bimodality and trimodality likely outcomes when repulsive or inhibitory coupling is imposed.

One particularly important case occurs when the interaction function has only roots at zero and $\pi$, with the root at zero having larger or equal magnitude slope. That is the case in the standard Kuramoto model with sinusoidal coupling. In such a case we expect that the incoherent splay state will be stable. In general, the splay state should be stable when the tendency to
cluster (due to large-phase interactions) cannot overcome the oscillators' repulsive small-phase interactions.

### 2.7. Conclusions

We have shown that, when coupling is repulsive, multi-modality of the oscillator distribution can be a stable configuration for a wide range of interaction functions. We showed that bimodality can be expected under repulsive coupling when the slope of the interaction function at the origin is shallower than at the other root(s). We performed numerical experiments for both identical and nonidentical oscillators and observed results consistent with theory.

This demonstration that repulsive coupling can produce clustering under reasonable assumptions about the interaction dynamics is important as repulsive coupling is present in many natural systems. Hence, the theory we present in this chapter provides an argument as to why one might expect multi-modality instead of unimodality or incoherence in systems known to have repulsive coupling.

## CHAPTER 3

## Modeling advertising competition

### 3.1. Introduction

Advertising is an important component of a free market system; it has been estimated that advertising expenditures in the United States exceeded $\$ 200$ billion dollars in 2018 alone [75]. Although the monetary investment is large, it remains unclear exactly how advertising affects demand and what the implications are for market competition. Perhaps advertising leads to increased market efficiency, greater aggregate profit for sellers, or better outcomes for buyers. The opposite could also be argued.

There are three prevailing theories as to how advertising influences the consumer [76]. Advertising can be viewed as persuasive, whereby it changes the tastes of consumers and increases demand (and price) [77-79]; informative, whereby it increases competition and decreases price $[\mathbf{8 0}]$ 82] ; or complementary, whereby it appeals to consumers with specific preferences that complement the consumption of the advertised products $[\mathbf{8 3} \mathbf{8 5}]$. These views have drastically different implications.

In this chapter we focus on persuasive advertising and, as in [86], assume that it increases demand. We look to work by Abernethy and Butler [87] to justify this assumption, where they report that an average TV ad contains just one mention of descriptive information about the displayed product (e.g., price, quality, performance, etc.), and that 37.5 percent contain no descriptive information at all. We take this to mean that a significant portion of TV ads are not
informative, implying that they are persuasive or complementary. Additionally, we make the simplifying assumption that persuasive advertising is always complementary, as Lindstädt and Budzinski argue that the viewer relates with the images and messages for both complimentary and persuasive advertising [88].

A large amount of research has been devoted to using game theory to choose the optimal advertising outlay to maximize profit [ $\mathbf{8 9}-95]$. Often this work focuses on settings where there is a monopoly (only one supplier of a good or service) or an oligopoly (only a small number of suppliers of a good or service) [90, 92 97]. Less research has focused on monopolistic competition, where there are many suppliers of a product or service, but the products or services are differentiated only by brand and/or quality. In this chapter we develop a model for this setting, looking at the expected advertising expenditure distribution for an arbitrary number of firms competing in a single commodity-product sector. Our goal is to develop a qualitative understanding of the expected shape of the distribution in a monopolistic competitive setting.

### 3.2. Results

Using a simple model in which firms work to differentiate themselves through advertising, we find that, when advertising is relatively cheap compared to the benefit of advertising, two groups arise: a "generic brand" group that advertises a minimal amount, and a "name brand" group that advertises at a significantly higher level. We find that this segmentation is stable and only ceases to exist when the marginal cost of advertising becomes too high relative to the marginal benefit of advertising. Although our model is intended chiefly to provide a conceptual "toy" description, fits to real-world price data show good qualitative agreement (see Fig. 3.1).


Figure 3.1. Price distributions. Red histograms show distribution of prices paid for four common household products; blue histograms show best-fit model predictions. Data are from a Nielsen database [98] of over 64 million transactions (purchasing history for 60,000 households).

### 3.2.1. Summary of model

We make the following simplifying approximations:
(1) Companies sell an indistinguishable product (except for brand label).
(2) Companies have linear demand curves.
(3) Demand for a company's product increases when its advertising is above the mean advertising level and decreases when its advertising is below the mean.
(4) Each company sets the price at a level that maximizes its profit.
(5) Companies continuously adjust their advertising so as to maximize profit.

These assumptions lead to a system of ordinary differential equations describing the dynamics of advertising investments for $N$ firms.

We show that when production and advertising costs are linear and the maximum marginal profit of advertising is positive (i.e., when advertising pays off at least in some situations), the
equilibrium where companies split into name brands and generics is stable-a state that we call differentiated. Furthermore, when the maximum marginal profit is negative, all companies choose to abstain from advertising ${ }^{1 /}$ and thus, they all become generic companies—a state that we call undifferentiated. The undifferentiated state is stable when increasing advertising decreases profit for the average company. We find that regions of stability for the undifferentiated state and the differentiated state overlap leading to a region of bistability. Details about stability are given in Section 3.3.2.2.

### 3.3. Model and analysis

### 3.3.1. Model Derivation

Consider $N$ companies (or firms) in a market all selling the same indistinguishable $\int^{2}$ product. The $i^{\text {th }}$ firm purchases a quantity of advertising $a_{]^{3}}$ For simplicity we assume that the firms have linear demand curves of the form:

$$
\begin{equation*}
Q_{i}=Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{P}} P_{i}, \quad i=1,2, \ldots, N \tag{3.1}
\end{equation*}
$$

where $Q_{i}$ is the quantity demanded of firm $i$ 's product, $P_{i}$ is the unit price for firm $i$ 's product, $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is the quantity demanded when the unit price is zero, which may depend on the full distribution of advertising in the market $\vec{a}=\left(a_{1}, a_{2}, \ldots, a_{N}\right)$, and $k_{\mathrm{P}}$ is a constant that sets the market's sensitivity to price.

[^2]One measure of a firm's health is the profit generated, with profit defined here as revenue minus production and advertising costs. We take revenue $R_{i}$ for the $i$ th firm to be solely due to sales of this single product at market price:

$$
\begin{equation*}
R_{i}=Q_{i} P_{i}=Q_{\text {free }}\left(a_{i} \mid \vec{a}\right) P_{i}-k_{\mathrm{P}} P_{i}^{2} \quad i=1,2, \ldots, N . \tag{3.2}
\end{equation*}
$$

In this model we only consider two types of operating costs: the cost of production $C_{\mathrm{Q}}\left(Q_{i}\right)$ and the cost of advertising $C_{\mathrm{a}}\left(a_{i}\right)$, and we assume an additive relationship

$$
\begin{equation*}
C_{i}\left(Q_{i}, a_{i}\right)=C_{\mathrm{Q}}\left(Q_{i}\right)+C_{\mathrm{a}}\left(a_{i}\right), \tag{3.3}
\end{equation*}
$$

where $C_{i}\left(Q_{i}, a_{i}\right)$ is the net operating cost for the $i^{t h}$ firm. We assume that both $C_{\mathrm{Q}}$ and $C_{\mathrm{a}}$ are increasing functions of their arguments, and for simplicity ${ }^{4}$ assume a power law form for each:

$$
\begin{align*}
C_{\mathrm{Q}}\left(Q_{i}\right) & =k_{\mathrm{Q}} Q_{i}^{\mu}  \tag{3.4a}\\
C_{\mathrm{a}}\left(a_{i}\right) & =k_{\mathrm{a}} a_{i}^{\nu} \tag{3.4b}
\end{align*}
$$

where $\mu, \nu>0$ and $k_{\mathrm{Q}}, k_{\mathrm{a}}$ are scale factors and can be interpreted as the marginal costs of production and advertising respectively when $\mu=\nu=1$. Thus, the profit function for the $i^{\text {th }}$ firm is

$$
\begin{equation*}
\pi_{i}=R_{i}-C_{i}=Q_{\text {free }}\left(a_{i} \mid \vec{a}\right) P_{i}-k_{\mathrm{P}} P_{i}^{2}-k_{\mathrm{Q}} Q_{i}^{\mu}-k_{\mathrm{a}} a_{i}^{\nu} \tag{3.5}
\end{equation*}
$$

[^3]

Figure 3.2. Effect of advertising on a firm's demand curve. (a) Demand shifts due to advertising above (red dashed) or below (blue dotted) the mean level (black solid). Vertical-axis intercepts are $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$. (b) A simple piecewise linear form for $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$, the quantity demanded at zero price, which we take to be a non-decreasing function of $a_{i}-\bar{a}$ that saturates at both left and right limits. Here the minimum demand (with advertising far below the mean) is $Q_{\min }$, the maximum demand increase due to advertising is $\Delta Q_{\mathrm{ad}}$, and the advertising needed beyond the mean for saturation is $\lambda$.

Critically, we tie a firm's level of advertising, $a_{i}$, to its ability to capture market power. We do this by assuming $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ to be a non-decreasing function of $a_{i}$ referenced to the mean advertising level $\bar{a}=N^{-1} \sum_{i=1}^{N} a_{i}$, i.e., a non-decreasing function of $a_{i}-\bar{a}$ (in the most general case, however, it might be an arbitrary function of the full advertising distribution $\vec{a}=$ $\left.\left(a_{1}, \ldots, a_{N}\right)\right)$. We assume firms that advertise more than the average firm have their demand curves shift out, and firms that advertise less than average have their demand curves shift in.

We also assume there is a saturation to the amount advertising can influence a firm's ability to capture market share. A plausible smooth, nondecreasing function that saturates is the sigmoid. We present results for that case in Section B.1. For greater algebraic simplicity, we
define $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ here as the following saturating piecewise linear function:

$$
Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)= \begin{cases}Q_{\mathrm{min}}, & a_{i}-\bar{a} \leq-\lambda  \tag{3.6}\\ Q_{\min }+\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda}\left(a_{i}-\bar{a}\right)+\frac{\Delta Q_{\mathrm{ad}}}{2}, & -\lambda<a_{i}-\bar{a} \leq \lambda \\ Q_{\min }+\Delta Q_{\mathrm{ad}}, & a_{i}-\bar{a}>\lambda\end{cases}
$$

where $\Delta Q_{\mathrm{ad}}$ is the maximum demand increase due to advertising, $Q_{\text {min }}$ is the zero-advertising (minimum) quantity demanded at zero price, which we deem "intrinsic demand," and $\lambda$ is the width of $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ (roughly the amount of excess advertising-above or below the meanneeded for benefits to saturate). See Fig. 3.2 for an illustration. Note, however, that for the purpose of simulations, we assume $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ takes on a smooth sigmoidal form:

$$
\begin{equation*}
Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)=\frac{\Delta Q_{\mathrm{ad}}}{2}\left\{\tanh \left[\frac{a_{i}-\bar{a}}{\lambda}\right]+1\right\}+Q_{\mathrm{min}} \tag{3.7}
\end{equation*}
$$

We assume that each firm always chooses the price $P_{i}^{*}$ that maximizes its profit, with corresponding quantity demanded $Q_{i}^{*}$. We introduce dynamics to the model by assuming that firms change their advertising levels at a rate proportional to the amount of profit to be gained, i.e,

$$
\begin{equation*}
\tau \frac{d a_{i}}{d t}=\frac{\partial \pi_{i}}{\partial a_{i}}=\frac{\partial}{\partial a_{i}}\left\{Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right) P_{i}^{*}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{P}}\left[P_{i}^{*}\left(a_{i} \mid \vec{a}\right)\right]^{2}-k_{\mathrm{Q}} Q_{i}^{*}\left(a_{i} \mid \vec{a}\right)^{\mu}-k_{\mathrm{a}} a_{i}^{\nu}\right\}, \tag{3.8}
\end{equation*}
$$

where the constant $\tau$ is sets the time scale for equilibration; we will henceforth set $\tau=1$ (equivalent to rescaling the time axis) without loss of generality. The list of model parameters with definitions is given in Table 3.1.


Figure 3.3. Advertising dynamics given $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$. We plot the the horizontal line $y=k_{\mathrm{a}}$ together with $y=\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ (see Eq. (3.12)) and add color to indicate the direction of change in advertising according to Eq. (3.8). Green regions show where a firm advertising an amount $a_{i}$ would choose to increase its advertising and red regions show where a firm advertising an amount $a_{i}$ would choose to decrease its advertising. Panel (a) shows the case where $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is a piecewise linear function that levels off; (b) shows the case where $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is sigmoidal.

| Parameter | Description |
| :---: | :---: |
| $N$ | Number of companies |
| $Q_{\min }$ | The quantity demanded with minimal advertising at zero price |
| $k_{\mathrm{P}}$ | Decrease in quantity demanded per dollar in unit price increase |
| $k_{\mathrm{Q}}$ | Scale factor for production cost; |
|  | (cost of producing an additional unit when costs are linear) |
| $k_{\mathrm{a}}$ | Scale factor for advertising cost; |
|  | (cost of producing an additional advertisement when costs are linear) |
| $\lambda$ | Amount of excess advertising above/below |
|  | the mean to achieve maximum/minimum advertising benefits |
| $\mu$ | Scaling exponent in the production cost function |
| $\nu$ | Scaling exponent in the advertising cost function |

Table 3.1. Parameter definitions. Table of parameters used in the model with descriptions.

### 3.3.2. A concrete example

As an analytically tractable example, we first consider the case where production and advertising costs grow at a linear rate, i.e, $\mu=\nu=1$. Substituting Eq. (3.1) into Eq. (3.5), setting
$\left[\partial \pi_{i} / \partial P_{i}\right]_{P_{i}=P_{i}^{*}}=0$ and solving for the profit-maximizing price $P_{i}^{*}$ gives

$$
\begin{equation*}
P_{i}^{*}\left(a_{i} \mid \vec{a}\right)=\frac{1}{2}\left[Q_{\text {free }}\left(a_{i} \mid \vec{a}\right) / k_{P}+k_{Q}\right] \tag{3.9}
\end{equation*}
$$

The corresponding profit-maximizing quantity is

$$
\begin{equation*}
Q_{i}^{*}\left(a_{i} \mid \vec{a}\right)=\frac{1}{2}\left[Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)-k_{Q} k_{P}\right] . \tag{3.10}
\end{equation*}
$$

Substituting this into Eq. (3.8) yields the dynamical system

$$
\begin{equation*}
\frac{d a_{i}}{d t}=\mathrm{B}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{a}} \tag{3.11}
\end{equation*}
$$

where $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ is defined as

$$
\mathrm{B}\left(a_{i} \mid \vec{a}\right)= \begin{cases}\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{4 \lambda k_{\mathrm{P}}}\left[\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda}\left(a_{i}-\bar{a}\right)+\frac{\Delta Q_{\mathrm{ad}}}{2}+Q_{\mathrm{min}}-k_{Q} k_{P}\right], & \left|a_{i}-\bar{a}\right|<\lambda  \tag{3.12}\\ 0, & \left|a_{i}-\bar{a}\right|>\lambda\end{cases}
$$

$\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ represents the marginal benefit of advertising and $k_{\mathrm{a}}$ the marginal cost of advertising. For any firm with advertising close enough to the mean $\left(\left|a_{i}-\bar{a}\right|<\lambda\right)$, the function B is simply a line of positive slope $(N-1) \Delta Q_{\mathrm{ad}}^{2} /\left(8 N \lambda^{2} k_{\mathrm{P}}\right) \xrightarrow[N \rightarrow \infty]{\longrightarrow} \Delta Q_{\mathrm{ad}}^{2} / 8 \lambda^{2} k_{\mathrm{P}}$. Firms with $\mathrm{B}>k_{\mathrm{a}}$ have $d a_{i} / d t>0$ and increase their advertising budgets, while firms with $\mathrm{B}<k_{\mathrm{a}}$ decrease their advertising budgets. For all firms far from the mean $\left(\left|a_{i}-\bar{a}\right|>\lambda\right), \mathrm{B}=0$ and thus $d a_{i} / d t=-k_{\mathrm{a}}<0$. This flow is illustrated in the left panel of Fig. 3.3. The corresponding flow in the case of a smooth sigmoidal $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is shown in the right panel of the same figure. The intuition drawn from the piecewise case outlined here applies similarly to the sigmoid.
3.3.2.1. Existence of equilibria. For a given $\bar{a}$, there can be at most three fixed points. In Fig. 3.3, three fixed points are located at $a_{i}^{*}=\bar{a}+\lambda$, the intersection where $\mathrm{B}\left(a_{i}^{*} \mid \bar{a}\right)=k_{\mathrm{a}}$ for $a_{i}^{*}<\bar{a}$, and at $a_{i}^{*}=0$ (since advertising cannot be negative). Because stability must alternate for one-dimensional flows, $a_{i}^{*}=0$ and $a_{i}^{*}=\bar{a}+\lambda$ are the only stable fixed points. Thus, any stable equilibrium distribution $\vec{a}$ with mean $\bar{a}$ must have $a_{i}=0$ or $a_{i}=\bar{a}+\lambda$ for all $i$.

We refer to the case when advertising is bimodal as the differentiated state. We note that such a state may only exist when two stable fixed points exist, which requires $\max _{a} \mathrm{~B}(a \mid \vec{a})>$ $k_{\text {a }}$. Letting $N \rightarrow \infty$, one can write this condition explicitly as

$$
\begin{equation*}
\max _{a_{i}} \frac{\partial \pi_{i}}{\partial a_{i}}=\frac{\Delta Q_{\mathrm{ad}}\left(\Delta Q_{\mathrm{ad}}+Q_{\min }-k_{\mathrm{Q}} k_{\mathrm{P}}\right)}{4 \lambda k_{\mathrm{P}}}-k_{\mathrm{a}}>0 . \tag{3.13}
\end{equation*}
$$

Put simply, if advertising does not increase profit anywhere, bimodality cannot arise. Another case that is possible is the state where all firms set their advertising to zero, which we refer to as the undifferentiated state. Clearly from Eq. (3.11), $\max _{a} \mathrm{~B}(a \mid \vec{a})<k_{\mathrm{a}}$ implies that $d a_{i} / d t<0$ for all $a_{i}$. In this case, $a_{i}^{*}=0$ for all $i$ is the only equilibrium.
3.3.2.2. Stability of equilibria. We now consider the stability of the differentiated and undifferentiated states. First, we focus on the stability of the differentiated state. We assume there exists an equilibrium with $N x$ "generic" firms choosing to invest nothing in advertising, and $N(1-x)$ "name-brand" firms choosing to advertise at level $\bar{a}+\lambda$, with $0<x<1$ representing the proportion of "generic" firms. Assuming that $N \gg 11^{5}$ and hence that a small perturbation of a single firm has a negligible impact on the mean $\bar{a}$, we consider perturbation of one "name-brand" firm $a_{i}$ by an amount $\delta$ and track how $\delta(t)$ changes in time. That is, we set

[^4]$a_{i}=\bar{a}+\lambda+\delta(t)$, which yields the system
\[

\frac{d \delta}{d t}= $$
\begin{cases}\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{4 \lambda k_{\mathrm{P}}}\left[\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda} \delta+Q_{\mathrm{min}}+\Delta Q_{\mathrm{ad}}-k_{Q} k_{P}\right]-k_{\mathrm{a}}, & |\delta+\lambda|<\lambda  \tag{3.14}\\ -k_{\mathrm{a}}, & |\delta+\lambda|>\lambda\end{cases}
$$
\]

If the condition for existence of the differentiated state given in Eq. (3.13) holds, sufficiently small $|\delta|$ implies that $d \delta / d t>0$ when $\delta<0$. Additionally, it is clear that $d \delta / d t<0$ when $\delta>0$. Thus, under this type of perturbation the differentiated state is stable. If we similarly perturb one firm from the generic group, i.e., setting $a_{i}=\delta>0$, we find
(3.15) $\frac{d \delta}{d t}=\left\{\begin{array}{ll}\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{4 \lambda k_{\mathrm{P}}}\left[\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda}(\delta-\bar{a})+Q_{\mathrm{min}}+\frac{\Delta Q_{\mathrm{ad}}}{2}-k_{Q} k_{P}\right]-k_{\mathrm{a}}, & |\delta-\bar{a}|<\lambda \\ -k_{\mathrm{a}}, & |\delta-\bar{a}|>\lambda\end{array}\right.$.

If $\bar{a}>\lambda$ then there exists $\delta>0$ small enough such that $d \delta / d t<0$ since $\delta<\bar{a}+\lambda$ implies that $d \delta / d t=-k_{\mathrm{a}}<0$. If $\bar{a}<\lambda$ then $d \delta / d t$ is given by the linear equation in Eq. 3.15) for small $\delta$. Thus, the differentiated state is stable under such a perturbation when

$$
\begin{equation*}
\left.\frac{d \delta}{d t}\right|_{\delta \rightarrow 0^{+}}=\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{4 \lambda k_{\mathrm{P}}}\left[-\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda} \bar{a}+Q_{\min }+\frac{\Delta Q_{\mathrm{ad}}}{2}-k_{Q} k_{P}\right]-k_{\mathrm{a}}=\left.\frac{\partial \pi_{i}}{\partial a_{i}}\right|_{a_{i}=0}<0 . \tag{3.16}
\end{equation*}
$$

This means it must be unprofitable for companies with no advertising to increase their advertising for the differentiated state to be stable.

Now we consider the stability of the undifferentiated state, $a_{i}=0$ for all $i$. As stated in Section 3.3.2.1, if $\max _{a} \mathrm{~B}(a \mid \vec{a})<k_{\mathrm{a}}$, then $d a_{i} / d t<0$ for all values of $a_{i}$. Thus, it is clear that the undifferentiated state exists and is stable in that case. We now focus on the case where $\max _{a} \mathrm{~B}(a \mid \vec{a})>k_{\mathrm{a}}$. If $a_{i}=0$ for all $i$ then $\bar{a}=0$. We consider a perturbation of one firm from
this state. Letting $a_{i}=\delta$ again we get

$$
\frac{d \delta}{d t}=\left\{\begin{array}{ll}
\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{4 \lambda k_{\mathrm{P}}}\left[\frac{\Delta Q_{\mathrm{ad}}}{2 \lambda}(\delta)+Q_{\mathrm{min}}+\frac{\Delta Q_{\mathrm{ad}}}{2}-k_{Q} k_{P}\right]-k_{\mathrm{a}}, & 0 \leq \delta<\lambda  \tag{3.17}\\
-k_{\mathrm{a}}, & \delta>\lambda
\end{array} .\right.
$$

If $d \delta / d t<0$ when $\delta=0$ by continuity of the $d \delta / d t$ in the range of $0<\delta<\lambda$ there must exist some $\delta>0$ sufficiently small such that $d \delta / d t<0$. Therefore, the system is stable under this kind of perturbation when

$$
\begin{equation*}
\left.\frac{d \delta}{d t}\right|_{\delta \rightarrow 0}=\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}\left[Q_{\min }+\Delta Q_{\mathrm{ad}} / 2-k_{\mathrm{Q}} k_{\mathrm{P}}\right]}{4 \lambda k_{\mathrm{P}}}-k_{\mathrm{a}}=\left.\frac{\partial \pi_{i}}{\partial a_{i}}\right|_{a_{i}=\bar{a}}<0 . \tag{3.18}
\end{equation*}
$$

We surmise from this that the undifferentiated state is stable only if increasing advertising is not profitable for the average firm.

The above stability arguments can be generalized to arbitrary infinitesimal perturbations of the advertising distribution in the limit $N \rightarrow \infty$. See work by Clifton, Braun, and Abrams for a description of such an approach in a different context [33].

Figure 3.4 maps the regions of stability for the differentiated and undifferentiated states given by Eqns. (3.13), (3.18), and (3.19). Both the differentiated and undifferentiated states can be simultaneously stable. If Eqns. (3.13) and (3.18) both hold then both states are stable. Thus, we write the condition for bistability as

$$
\begin{equation*}
\mathrm{B}(\bar{a} \mid \vec{a})=\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}\left[Q_{\min }+\Delta Q_{\mathrm{ad}} / 2-k_{\mathrm{Q}} k_{\mathrm{P}}\right]}{4 \lambda k_{\mathrm{P}}}<k_{\mathrm{a}}<\max _{a} \mathrm{~B}(a \mid \vec{a}) . \tag{3.19}
\end{equation*}
$$



Figure 3.4. Regions of stability. We illustrate the regions of stability for the differentiated and undifferentiated states, indicated in the figure by D and U respectively. These are given by Eqns. (3.13), (3.18), and (3.19). Here green indicates that a state is stable and red indicates that a state is unstable. The middle column, where $\mathrm{B}(\bar{a} \mid \vec{a})<k_{\mathrm{a}}<\max _{a} \mathrm{~B}(a \mid \vec{a})$ is where both states are stable.

In more intuitive terms,

$$
\begin{equation*}
\left.\frac{\partial \pi_{i}}{\partial a_{i}}\right|_{a_{i}=\bar{a}}<0<\max _{a} \frac{\partial \pi_{i}}{\partial a_{i}} . \tag{3.20}
\end{equation*}
$$

Thus, bistability of the differentiated and the undifferentiated states occurs when the maximum marginal profit is positive, but it is profitable for the average firm to decrease its advertising. The regions of stability of the undifferentiated and differentiated states are defined similarly when $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is sigmoidal (and hence, $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ altered appropriately—see Section B.1 in the appendix).

### 3.3.3. Numerical experiments

In order to test model predictions we perform simple numerical experiments. Figure 3.5 shows an example of a simulation where the benefit of advertising saturates (we assume a sigmoidal


Figure 3.5. Simulation of the system. In this figure we give snapshots of the numerical integration of the system from initial condition to equilibrium. In panel (a), the system starts from the uniform randomly distributed state with the advertising initial condition set as $\mathcal{U}(7.5,12.5)$. In panel (b), the separation into two groups has begun. Companies change their spending until the lower group is far away from the mean, as seen in panel (c). Finally, in panel (d), a bimodal equilibrium has been reached, with one group representing generic brand companies ( $a \approx 0$ ) and the other representing name-brand companies (advertising at a nonzero value at $a=a_{\text {name }}$ ). The green areas indicate where companies will increase their advertising and the red areas indicate areas where companies will decrease their advertising. In this simulation we set the number of companies to $N=1000, k_{\mathrm{a}}=k_{\mathrm{P}}=k_{\mathrm{Q}}=\lambda=1$, and $Q_{\text {min }}=\Delta Q_{\mathrm{ad}}=10$ (see Methods section for parameter definitions). Additionally, advertising and production cost functions are linear.
functional form). Starting from a uniformly distributed initial condition, the firms arrange themselves so that there is a "generic" group at advertising level $a=0$ and a "name-brand" group at $a=a_{\text {name }}>0$. Colors have been added to indicate ranges where firms decrease (red) or increase (green) their advertising.
3.3.3.1. Numerics. In all simulations reported here, we use an explicit 4th order Runge-Kutta scheme to integrate the differential equations. We set $N=1000, k_{\mathrm{P}}=k_{\mathrm{Q}}=\nu=\mu=\lambda=1$,
$Q_{\text {min }}=10, \Delta Q_{\mathrm{ad}}=10, k_{\mathrm{a}}=1$ and draw the initial conditions for advertising randomly from $\mathcal{U}(7.5,12.5)$.

### 3.3.4. Data

We use price data from the Nielsen Corporation. Nielsen's consumer panel data contains annual shopping information from thousands of American households, starting from 2004 with yearly updates. Individuals involved in the study used in-home scanners to record all of their purchases that were designated for personal use. Scanners recorded each product's Universal Product Code (a string of digits that uniquely identify the product) and the product's price. We analyze data from 2014 containing over 64 million transactions from 60,000 households [98].

### 3.3.5. Fitting procedure

To fit our model predictions to data, we first define an objective function $H[f(p), g(p)]$ to quantify the difference between distributions predicted by the model $(f(p))$ and inferred from the data $(g(p)$ ). Specifically, we set our objective function $H[f(p), g(p)]$ to be the square integrated difference between the distributions

$$
\begin{equation*}
H(f, g)=\int_{-\infty}^{\infty}[f(p)-g(p)]^{2} d p \tag{3.21}
\end{equation*}
$$

We use the Nelder-Mead algorithm [110] to minimize this objective function over a subset of parameters that most directly affect the demand curve given in Eq. (3.1): the maximum benefit from advertising $\Delta Q_{\mathrm{ad}}$, the minimum quantity demanded $Q_{\min }$, and consumers' price sensitivity, $k_{\mathrm{P}}$. If the data indicate bimodality, we also optimize over the generic fraction $x$.

We model heterogeneity among firms by adding random variables $\zeta_{i}$ and $\xi_{i}$ to the parameters $\Delta Q_{\text {ad }}$ and $Q_{\text {min }}$ respectively. These random variables are drawn from a normal distribution with mean zero and respective standard deviations $\epsilon_{1}$ and $\epsilon_{2}$. We interpret $\epsilon_{1}$ as the variation in the quality of advertising messaging and $\epsilon_{2}$ as the variation in natural demand for the firms' products, and we also optimize their values.

We must choose starting "seeds" for the Nelder-Mead algorithm since it is a local optimization method. We set the initial parameter values such that the profit maximizing price for companies receiving no demand increase from advertising and companies receiving the maximum demand increase match up with the lower and upper modes, respectively, of the price data. For a more detailed description of the initialization of the algorithm see Section B.4 in appendix.

Figure 3.1 provides a few examples of fits for products that had more than 10,000 transactions. These examples also demonstrate the variety of products within the dataset. We see there is qualitative agreement between the model's predicted price distributions and the empirical price distributions.

### 3.3.6. Statistics

We attempt to validate our model by fitting theoretical price distributions to empirical data provided by Nielsen Corporation [98]. We use two tests, the Kolmogorov-Smirnov (KS) test and Hartigan's Dip Test, to assess the quality of our fits. See Figure 3.1 for a sample of model fits to data.

The KS test generates the probability that two samples come from the same underlying distribution by calculating the maximum absolute difference between their cumulative distribution
functions (CDFs). Here, a large difference implies a low probability that the two datasets come from the same distribution. For a majority ( $58 \%$ ) of our model fits to the top 500 products, we fail to reject the null hypothesis (samples from same underlying distribution) at a significance level of 0.05 : the data and the model prediction may come from the same distribution.

Hartigan's Dip Test assesses whether a distribution is unimodal by comparing the CDF of the distribution to a unimodal test distribution [111]. A large difference between the distribution in question and the test distribution indicates a low probability of the distribution being unimodal. We apply Hartigan's Dip Test to the 500 products with the most entries in the database, and find that $46 \%$ have price distributions inconsistent with unimodality at a significance level 0.05. If price distributions are linked to advertising expenditures, as our model indicates, then almost half the products have a multimodal (bimodal or higher number of modes $\sqrt[6]{ }$ advertising distribution. For other products, unimodality could not be rejected, but data may not be inconsistent with bimodality. See Section B.3 for the full distribution of $p$-values.

### 3.4. Discussion

The theory we present provides a possible explanation for the segmentation of commodityproduct sectors into "name brand" and "generic" products. We speculate that similar explanations might exist for other contexts where hierarchy emerges as a result of competition, or where interactions between individual agents can lead to clustering [99]. For example, competition for a mate $[\mathbf{3 3}, \mathbf{1 0 0}, \mathbf{1 0 1}]$ and competition for resources $[\mathbf{1 0 2}-\mathbf{1 0 4}]$ can both result in hierarchies observed in the natural world. Our model might be adapted to yield insight into such phenomena.

[^5]
### 3.4.1. Limitations

In creating a highly simplified model, we have inevitably made some assumptions that limit its generality. These include:

- We assumed that advertising was persuasive and hence, that quantity demanded increased uniformly across all price levels as advertising increased. In cases where advertising is informative, however, one would expect the slope of the demand curve to increase, instead of simply shifting vertically.
- We chose to leave the development of brand loyalty out of our model. This could presumably be captured through a demand curve that becomes more inelastic as loyalty increases.
- We excluded spillover effects from "generic advertising" whereby advertising leads to increases in demand for all companies selling a similar product [105, 106]. We expect that this would increase profit for all companies but not affect the bimodal segmentation our current model predicts.
- We assumed advertising has stable and lasting impact. Our model treats the benefits of advertising as arising instantaneously, an approximation that is only merited when the time scale of interest is much longer than the advertising's "half-life" in the consumer environment.
- We assumed the existence of many producers selling similar products. Some of our arguments would not be valid in the case of an oligopoly, where there are only a handful of producers.
- We approximated demand curves as linear, but of course these could (and likely do) take on more complex forms for real products.

In addition to the limitations of our modeling approach, the data set we examine also contains some biases that should be pointed out. Most saliently, the price distributions we examine are the result of different vendors selling identical products for different prices: this means that branding is really present at the vendor level, slightly different from most direct and natural interpretation of the model. Also, a large fraction of entries in the database are food and other consumable products, since these are purchased more frequently than durable goods. Consumables might have a different market structure than products in non-food markets (e.g., electronics, health care, housing, etc.).

### 3.4.2. Conclusions

We have presented a simple mathematical model for competition among firms on the basis of advertising. Despite the model's simplicity, a surprisingly robust prediction emerges: products split into "name-brand" and "generic" groups. This prediction appears to be largely consistent with data both in a qualitative sense (many products have non-unimodal price distributions) and a quantitative sense (theoretical price distributions from the model are consistent with empirical price distributions), even without a more detailed and accurate model.

Advertising has a large macroeconomic impact on corporate profits, market efficiency, and consumer welfare. The segmentation we report contrasts starkly with (often implicit) assumptions of smooth, singly-peaked functions for economic metrics. We hope that our work helps refine intuition and inspires further inquiries into this intriguing aspect of free market dynamics.

## CHAPTER 4

## Size dimorphism

### 4.1. Introduction

Anisogamy, the phenomenon where fusion occurs between gametes (sex cells) of different sizes, has long been a topic of study ( $[\mathbf{1 2 2}-\mathbf{1 3 1}])$. Anisogamy is common in complex organisms such as plants, animals, fungi, and certain algae ( $[\mathbf{1 2 5}, \mathbf{1 3 2}-\mathbf{1 3 4}])$. There is a consensus in the literature that anisogamy evolved from isogamy, where sexual reproduction occurs between sex cells that are the same size ( $[\mathbf{1 2 5}, \mathbf{1 2 6}, \mathbf{1 3 5}, \mathbf{1 3 6}])$.

Anisogamy has been theorized to be a factor in the development of differences between sexes. Bateman credits to anisogamy the fact that male Drosophila melanogaster are far more eager than females to mate ( [137]). Lehtonen et al. add theory to this intuition, demonstrating that, as the size ratio between large and small gametes increases, organisms with small gametes will choose to allocate more resources to searching for mates and warding off others with small gametes from potential mates as those with small gametes substantially increase their fitness when the number of opportunities to mate increases. ( [ [131] $)$.

A related question that remains of scientific interest is "Why do most complex organisms have only two sexes?" Almost all animals have two sexes, but, e.g., fungi may have scores or even thousands of "mating types" (the term "sex" is typically not used in this case) ( [ $\mathbf{1 3 3}$, 138, 139]). We do not directly address this here, but a better understanding of the origin of anisogamy might also inform our understanding of this question.

Our work builds off the assumption that the size of gametes affects the fitness of the individual, but we account for competition through "mean-field" coupling, meaning that payoff in terms of offspring is largely due to the presence of a relative, rather than absolute, advantage. We make mild assumptions regarding asymptotic properties of relevant functions and demonstrate that these are sufficient to explain the emergence of anisogamy from an isogamous population.

### 4.2. Results and discussion

We put forth a model that provides a plausible explanation for the development and stability of anisogamy, even without the existence of mating types. This model is based upon the assumption that an individual's overall reproductive potential can be broken down into a "gamete production" term quantifying the number of gametes produced, and a "gamete potential" term quantifying those gametes' likelihood of eventually forming zygotes that reach adulthood. Both of these are assumed to depend upon gamete size, with gamete reproductive potential having a positive relationship with gamete size, and number of gametes having a negative relationship with gamete size. A critical assumption is that size-dependence for gamete reproductive potential is determined relative to the mean of the population, encapsulating the intra-species competition for resources.

Although other models have been proposed to explain anisogamy, ours requires minimal assumptions and accounts for its emergence from an initially isogamous state. We require no assumptions about the existence of mating types. For simplicity and clarity, we have treated individuals in this work as identical without variation, and we have allowed small gametes to
approach zero size. More realistic assumptions do not appear to change the broad results shown here (see Sections C.2 C. 6 for various numerical experiments).

### 4.3. Model development

### 4.3.1. Individual reproductive potential

Consider a population of $N$ organisms with gametes that have sizes $s_{j}, j=1, \ldots, N$. Following the approach used in [33], we denote the "reproductive potential" of the $j$ th individual by $\varphi_{\mathrm{i}}$, defined as some increasing function of the fitness (the expected number of adult offspring it will produce). We assume that this potential can be expressed as a product of $N_{\mathrm{g}}$, the expected number of gametes produced, and $\varphi_{\mathrm{g}}$, the average reproductive potential of its gametes (where gamete reproductive potential is, similarly, an increasing function of gamete fitness-the expected number of adults resulting from that gamete, with upper bound 1 , ignoring monozygotic twinning):

$$
\begin{equation*}
\varphi_{\mathrm{i}}=N_{\mathrm{g}} \varphi_{\mathrm{g}} \tag{4.1}
\end{equation*}
$$

Because we are concerned with anisogamy and hence gamete size distributions, we ignore all factors influencing reproductive potential besides gamete size. Other factors are clearly extremely important, but we model only the effects of gamete sizes on reproductive potential here, and thus write that $N_{\mathrm{g}}=N_{\mathrm{g}}\left(s_{j}\right), \varphi_{\mathrm{g}}=\varphi_{\mathrm{g}}\left(s_{j}\right)$.

### 4.3.2. Gamete production function

We assume that $N_{\mathrm{g}}$ is a decreasing function of gamete size due to the fact that each organism has limited resources (physical, temporal and energetic) to dedicate to gamete production.

Some observational evidence supports this: smaller male sex cells are far more numerous than significantly larger female sex cells ( $[\mathbf{1 4 0}-\mathbf{1 4 3}]$ ); additionally, research has found a negative relationship between clutch size and egg size in the black-backed gull Larus fuscus ( [ [144]), and across various species of snakes ( $[\mathbf{1 4 5}]$ ).

To present a concrete analytical argument, and motivated by their ubiquity in nature ( [ $\mathbf{1 0 8}$, 146, 147]), we choose $N_{\mathrm{g}}$ to be a power law, i.e.,

$$
\begin{equation*}
N_{\mathbf{g}}\left(s_{j}\right)=c_{1} s_{j}^{-\alpha} \tag{4.2}
\end{equation*}
$$

where $c_{1}$ is a constant of proportionality and the constant $\alpha$ is assumed to be positive. In the section 'Geometric argument'' below, we generalize our argument to arbitrary decreasing functions.

### 4.3.3. Gamete reproductive potential

We assume that $\varphi_{\mathrm{g}}$ is an increasing function of gamete size. This is motivated by the idea that increased size indicates increased provisions to promote survival of the gamete and the zygote potentially formed after fusion with another gamete. Some evidence supports this link: associations between between egg size (measured by volume or mass) and positive offspring outcomes have been reported in various avian species ( $[\mathbf{1 4 8}-\mathbf{1 5 1}]$ ).

Critically for our model, we assume that the fitness "payoff" accruing to larger gametes is relative rather than absolute in nature. That is, we assume that a gamete of size $s_{i}$ will have greater reproductive potential in a population where it is among the largest than in a population where it is among the smallest. This is motivated by the hypothesis of zygote competition, and ultimately by the same idea underlying natural selection: if environmental conditions preclude
all viable zygotes from reaching adulthood, those with the greater provisions afforded by larger parental gamete sizes will be more likely to survive. A similar argument can be made if direct competition between gametes plays a role in determining fitness.

Thus, we link the reproductive potential of the $j$ th gamete to the full distribution of gamete sizes in the population. We can express such a link in simple terms by assuming $\varphi_{\mathrm{g}}\left(s_{j}\right)$ is an increasing function of $s_{j}-\bar{s}$, where $\bar{s}=N^{-1} \sum_{j=1}^{N} s_{j}$ is the mean gamete size in the population.

We expect reproductive potential to saturate for both extremely large and extremely small gametes, so we choose a sigmoidal form for our analytical expression of $\varphi_{\mathbf{g}}\left(s_{j}\right)$ :

$$
\begin{equation*}
\varphi_{\mathrm{g}}\left(s_{j} \mid \bar{s}\right)=c_{2}\left(1+\frac{s_{j}-\bar{s}}{w+\left|s_{j}-\bar{s}\right|}\right), \tag{4.3}
\end{equation*}
$$

where $c_{2}$ is a constant of proportionality and $w$ sets the width of the sigmoid. In the section 'Geometric argument'' below, we generalize our argument to arbitrary increasing functions $\varphi_{\mathrm{g}}\left(s_{j}-\bar{s}\right)$.

Substituting Eq. (4.2) and Eq. (4.3) into Eq. (4.1), we obtain the following individual reproductive potential function:

$$
\begin{equation*}
\varphi_{\mathrm{i}}\left(s_{j} \mid \bar{s}\right)=N_{\mathrm{g}}\left(s_{j}\right) \varphi_{\mathrm{g}}\left(s_{j} \mid \bar{s}\right)=c_{3} s_{j}^{-\alpha}\left(1+\frac{s_{j}-\bar{s}}{w+\left|s_{j}-\bar{s}\right|}\right) \tag{4.4}
\end{equation*}
$$

where we have combined the multiplicative constants of proportionality into a single constant $c_{3}=c_{1} c_{2}$.

### 4.3.4. Gamete size evolution

We assume that natural selection acts on the population in such a way that gamete sizes change at a rate proportional to the reproductive potential to be gained. That is, there is a "phenotype
flux"

$$
\begin{equation*}
\frac{d s_{j}}{d t}=\frac{1}{\tau} \frac{\partial \varphi_{\mathrm{i}}}{\partial s_{j}} \tag{4.5}
\end{equation*}
$$

or, in the continuum limit $N \rightarrow \infty$,

$$
\begin{equation*}
\frac{\partial \rho}{\partial t}=\nabla \cdot\left(\frac{d s}{d t} \rho\right) . \tag{4.6}
\end{equation*}
$$

where $d s / d t$ is given by

$$
\begin{equation*}
\frac{d s}{d t}=\frac{1}{\tau} \frac{\partial \varphi_{\mathrm{i}}}{\partial s} . \tag{4.7}
\end{equation*}
$$

Here $\tau$ sets the time scale for the evolution of gamete size. Since this is unknown (and not the focus of this work), we rescale time such that $\tau=1$ without loss of generality.

To be clear, we are not assuming that individual organisms explicitly change their gamete sizes in this model, rather, the "phenotype flux" $d s / d t$ captures how the gamete size distribution $\rho(s, t)$ changes over long time scales. Probability density functions such as $\rho(s, t)$ must follow the continuity equation, seen in Eq. 4.6. In [33], the authors demonstrated how this approach (substitution of the phenotype flux from Eq. (4.7) into the continuity equation) can be considered equivalent to a "replicator equation" approach ( $[\mathbf{1 5 2}, \mathbf{1 5 3}]$ ) for appropriate choices of fitness functions.


Figure 4.1. Example individual reproductive potential function. Here we show the reproductive potential function defined by Eq. (4.4) in arbitrary units (a.u.). Two maxima are apparent, one at zero and another at a nonzero value $s^{*}$. Dynamics given by Eq. (4.5) are illustrated by color with red indicating regions where gamete size decreases and blue indicating where gamete size increases. For this illustration, we set $w=1 / 10, \alpha=1$, and $\bar{s}=2$.

### 4.4. Model implications

### 4.4.1. Existence of the anisogamous equilibrium

It is commonly assumed that natural selection leads to traits that maximize reproductive success (subject to ongoing variation) ( [ [154]). This implies that anisogamy is possible only when the reproductive potential has at least two distinct local maxima.

For $\varphi_{\mathrm{i}}$ as defined in Eq. (4.4), at most two local maxima can exist: one at $s_{j}=0$ and another at a nonzero value $s^{*}$. This is illustrated in Fig. 4.1, where two local maxima can be seen.

We can derive a sufficient condition for the existence of the anisogamous equilibrium by assuming it takes the form $\rho(s)=x \delta(s-0)+(1-x) \delta\left(s-s^{*}\right)^{1]}$ where $\delta$ is the Dirac delta

[^6]function and $0<x<1$ is the proportion of gametes that are small (i.e., the proportion that might be referred to as primitive "male" gametes). This equilibrium must be self-consistent, meaning that the first moment of the distribution is indeed the same as the average gamete size $\bar{s}$. Substituting $\rho(s, t)=x \delta(s-0)+(1-x) \delta\left(s-s^{*}\right), \bar{s}=(1-x) s^{*}$, Eq. (4.4), and Eq. (4.7) into Eq. (4.6) and solving $\partial \rho / \partial t=0$ (or, equivalently, setting $\partial \varphi_{\mathrm{i}} / \partial s=0$ after plugging Eq. (4.4) into Eq. (4.7), with $\left.\bar{s}=(1-x) s^{*}\right)$, we find
\[

$$
\begin{equation*}
s^{*}=w \frac{1-3 \alpha x+\sqrt{\beta}}{4 \alpha x^{2}}, \tag{4.8}
\end{equation*}
$$

\]

where $\beta=\alpha^{2} x^{2}-6 \alpha x+1$. An anisogamous equilibrium thus exists for all positive $\alpha$ and $w$, as long as $\beta>0$.

### 4.4.2. Stability of the anisogamous equilibrium

Consider the perturbation of a single individual from the large gamete group by an amount $\epsilon \ll$ 1 in the limit $N \gg 1$, so this represents an infinitesimal change to the full gamete distribution. We set $\bar{s}=(1-x) s^{*}$ and $s_{\text {pert }}=s^{*}+\epsilon$, where $s^{*}$ is given by Eq. (4.8). Substituting into Eq. (4.7) and Taylor expanding to linear order in $\epsilon$,

$$
\begin{equation*}
\frac{d \epsilon}{d t}=Q(x, \alpha) \epsilon \tag{4.9}
\end{equation*}
$$

where

$$
\begin{equation*}
Q=-\frac{16 x \alpha^{2}\left(s^{*}\right)^{-1-\alpha}(1-x+\sqrt{\beta})}{w(1+x \alpha+\sqrt{\beta})^{2}} \tag{4.10}
\end{equation*}
$$

with $\beta$ defined as in Eq. (4.8). Here $Q<0$ for all allowable parameter values, and thus the anisogamous state is also stable under this kind of perturbation.

A similar perturbation of one individual from the small gamete group is simply $s_{\text {pert }}=\epsilon$, which, when substituted into Eq. (4.7) yields

$$
\begin{equation*}
\frac{d \epsilon}{d t}=-c_{3} \alpha \frac{w}{w+\bar{s}} \epsilon^{-1-\alpha} \tag{4.11}
\end{equation*}
$$

when truncated at leading order. Since $c_{3}, \alpha, w$, and $\bar{s}$ are all positive, the anisogamous state is stable to infinitesimal perturbations of this sort whenever it exists.

A more general examination of stability is difficult, but in Section C.3.1 we show that all eigenvalues of the finite $N$ system are negative for $N \gg 1$, and thus that the anisogamous state is indeed linearly stable.

### 4.4.3. Geometric argument

For clarity and convenience, we earlier assumed specific algebraic forms for $\varphi_{\mathrm{g}}$ and $N_{\mathrm{g}}$. We now show the possible emergence of anisogamy in a system where only the monotonicity and asymptotic properties of those functions are known.

We start by expanding the derivative on the right-hand side of Eq. (4.7):

$$
\begin{equation*}
\frac{d s}{d t}=\frac{1}{\tau} \frac{\partial \varphi_{\mathrm{i}}}{\partial s}=\frac{1}{\tau}\left[\frac{\partial \varphi_{\mathrm{g}}}{\partial s} N_{\mathrm{g}}+\frac{d N_{\mathrm{g}}}{d s} \varphi_{\mathrm{g}}\right] \tag{4.12}
\end{equation*}
$$

At an equilibrium $s=s^{*}$, the net phenotype flux $d s / d t=0$. Assuming $N_{\mathrm{g}}>0$ and $\varphi_{\mathrm{g}}>0$, we find that the following condition must hold at each $s^{*}$ :

$$
\begin{equation*}
\frac{\varphi_{\mathrm{g}}^{\prime}}{\varphi_{\mathrm{g}}}=-\frac{N_{\mathrm{g}}^{\prime}}{N_{\mathrm{g}}} \tag{4.13}
\end{equation*}
$$

where ${ }^{\prime} \equiv \partial / \partial s$.
The left side of Eq. (4.13) is the relative change in gamete reproductive potential and the right side is the magnitude of the relative change in gamete production. Gamete sizes will increase when the reproductive potential gains outweigh the decreased gamete production, and will shrink when the opposite is true.

The existence of anisogamy requires that two distinct intersections must exist between the functions on the left and right-hand sides of Eq. (4.13) (see Fig. 4.2). The following conditions are thus sufficient for anisogamy to exist:
(1) Continuity of $\varphi_{\mathrm{g}}^{\prime}$ and $N_{\mathrm{g}}^{\prime}$.
(2) The gamete production terms dominate as size approaches zero (relative decrease in production larger than relative increase in reproductive potential), i.e.,

$$
\left|\frac{\varphi_{\mathrm{g}}^{\prime}}{\varphi_{\mathrm{g}}}\right|<\left|\frac{N_{\mathrm{g}}^{\prime}}{N_{\mathrm{g}}}\right|, s \rightarrow 0^{+}
$$

This is reasonable if the potential saturates at some minimum (possibly zero) for small gametes.
(3) There exists at least one finite value of $s$ (say $s=a, a>0$ ) at which reproductive potential terms dominate over gamete production terms, i.e.,

$$
\left|\frac{\varphi_{\mathrm{g}}^{\prime}}{\varphi_{\mathrm{g}}}\right|>\left|\frac{N_{\mathrm{g}}^{\prime}}{N_{\mathrm{g}}}\right|, s=a
$$

If this fails, smaller gametes are always better for fitness. As long as there is some "provisioning" advantage to larger gametes at some point, however, this condition should be satisfied.


Figure 4.2. Geometric argument for anisogamy. We illustrate a case where gamete reproductive potential, $\varphi_{\mathrm{g}}$, and gamete production, $N_{\mathrm{g}}$, satisfy the conditions set out in the section "Geometric argument", When gametes are small, the relative gains due to the ability to produce more of them $\left|N_{\mathrm{g}}^{\prime} / N_{\mathrm{g}}\right|$ outweigh the relative drop in reproductive potential $\left|\varphi_{\mathrm{g}}^{\prime} / \varphi_{\mathrm{g}}\right|$. In some intermediate range, reproductive potential gains dominate, and then as gametes become very large the production terms again dominate as reproductive potential gains saturate.
(4) Gamete production terms again dominate as size goes to infinity, i.e.,

$$
\left|\frac{\varphi_{\mathrm{g}}^{\prime}}{\varphi_{\mathrm{g}}}\right|<\left|\frac{N_{\mathrm{g}}^{\prime}}{N_{\mathrm{g}}}\right|, s \rightarrow \infty .
$$

This is reasonable if fitness gains eventually saturate.
In addition to the above, a self-consistency condition must also hold: It must be possible for the function $\varphi_{\mathrm{g}}^{\prime} / \varphi_{\mathrm{g}}$ to satisfy

$$
\left|\frac{\varphi_{\mathrm{g}}^{\prime}}{\varphi_{\mathrm{g}}}\right|=\left|\frac{N_{\mathrm{g}}^{\prime}}{N_{\mathrm{g}}}\right|
$$

given $\bar{s}=(1-x) s^{*}$, for some fractionation $x \in(0,1)$.
Figure 4.2 shows an example of functional shapes for $\varphi_{\mathrm{g}}$ and $N_{\mathrm{g}}$ that satisfy these conditions.

## CHAPTER 5

## Debiasing data

### 5.1. Introduction

Inferring an accurate distribution underlying recorded data has long been an important topic of research. The simplest way to approach this task is to construct a histogram of the data, but that comes with two problems. First, the piecewise nature of a histogram necessarily means it is not a smooth distribution, and, second, the inferred distribution is dependent on many choices of parameters (bin boundaries) that are not easily justified. To deal with both of these issues other techniques are often used, such as kernel density estimation [117] or filtering [118].

A question less frequently posed is how to infer an underlying distribution when the data itself are corrupted in a way that is not explained by noise. Often, the data that we record are affected by some type of bias.

This bias is seen in various settings. Research shows that men and women overestimate their height and underestimate weight [162, 165], leading to a distortion in weight, height, and BMI distributions. Selling prices are also artificially skewed with selling prices that end in zero, five, or nine being over-represented compared to random chance [167]. This is known as "psychological pricing" where companies seek to take advantage of how humans read prices in order to increase sales [161].

In this chapter we develop a method that, given a set of data, infers the underlying distribution while removing any skew due local bias toward particular numbers. We find a system of
differential equations that smooths distributions in such a way that kicak diffusive smoothing is balanced by by "attraction" to the initial data values. We construct a testbed for our method by perturbing a Gaussian distributions with a Ricker wavelet, and find that it can be made to exactly eliminate the perturbation in such a case.

We further extend our model by taking inspiration from adaptive kernel density estimates [155], adding spatial variation according to a "bias" parameter that alters the amount of smoothing given the deviation of the original distribution from the smoothed data. We test this extension on both synthetic and real-world data sets. In both cases the perturbation due to bias appears to be successfully removed from the equilibrium distribution.

### 5.2. Method

One can attain only a finite amount of precision when recording data. A simple way of enforcing this precision is dividing the real number line into bins separated of width $\Delta x$.

We consider a system of $N$ bins where the $i^{\text {th }}$ bin contains $y_{i}$ amount of probability mass (henceforth referred to as simply "mass"). The mass in bin $i$ changes as a result of local diffusion and a restoring "force" that adds (or removes) mass at a rate proportional to the difference of the current mass $y_{i}$ and the initial mass of bin $i, y_{i, 0}$. The rate of change of mass in bin $i$ due to this process is thus given by

$$
\begin{equation*}
\Delta_{i}=-k \frac{\alpha}{\Delta x^{2}}\left(y_{i}-\left[\frac{y_{i+1}+y_{i-1}}{2}\right]\right)-(1-k) \beta\left(y_{i}-y_{i, 0}\right) \tag{5.1}
\end{equation*}
$$

where the first term comes from a discrete diffusion operator and the second term represents the "restoring force" described above, and $k$ is a parameter in $[0,1]$ that sets the relative importance
of each term. $\alpha$ is the diffusion coefficient and $\beta$ sets the restoration speed. $k=1$ corresponds to pure diffusion, and $k=0$ corresponds to a system with only restoring forces and no diffusion.

Mass change in the system is distributed locally where the amount lost (or gained) from bin $i$ is evenly distributed to (or pulled from) bins $i+1$ and $i-1$. This leads to the following system of coupled differential equations:

$$
\begin{equation*}
\frac{d y_{i}}{d t}=\Delta_{i}-\frac{1}{2}\left(\Delta_{i+1}+\Delta_{i-1}\right) \tag{5.2}
\end{equation*}
$$

One can show (see appendix) that the continuum limit is

$$
\begin{equation*}
\frac{\partial \rho}{\partial t}=-\alpha\left(\frac{k}{4}\right) \frac{\partial^{4} \rho}{\partial x^{4}}+\beta\left(\frac{1-k}{2}\right)\left(\frac{\partial^{2} \rho}{\partial x^{2}}-\frac{d^{2} \rho_{0}}{d x^{2}}\right) \tag{5.3}
\end{equation*}
$$

where $\rho_{0}(x)=\rho(x, 0)$. This partial differential equation is similar to the one developed by Bevilaqua et al. [157, 158], the main difference is an added source that is proportional to the second spatial derivative of the initial data.

Setting $\partial \rho / \partial t=0$ gives the following steady state equation

$$
\begin{equation*}
0=\frac{\partial^{4} \rho}{\partial x^{4}}-q^{2}\left(\frac{\partial^{2} \rho}{\partial x^{2}}-\frac{d^{2} \rho_{0}}{d x^{2}}\right) \tag{5.4}
\end{equation*}
$$

where

$$
\begin{equation*}
q^{2}=\frac{2 \beta(1-k)}{\alpha k} \tag{5.5}
\end{equation*}
$$

Since $k$ ranges from zero to one, the parameter $q^{2}$ spans zero to infinity, with $q=0$, being the fully diffusive case and $q^{2} \rightarrow \infty$ corresponds to the case without diffusion.

### 5.2.1. Rectifying perturbations

We are particularly interested in problems where the perturbations are local in the sense that some probability mass has been shifted from its original location to a nearby more "appealing" number (where the appealing numbers might represent, e.g., prices ending in the digit nine, or heights rounded to an even value). As a prototypical example of such a locally perturbing function, we will consider the Ricker wavelet [166] (also known as the Marr wavelet [164]), defined as

$$
\begin{equation*}
\psi(x ; w, h)=\frac{h\left(w^{2}-x^{2}\right) e^{-\frac{x^{2}}{2 w^{2}}}}{w^{2}} \tag{5.6}
\end{equation*}
$$

See Figure 5.1 for a visualization of the Ricker wavelet when $h=3$ and $w=1$.


Figure 5.1. Ricker wavelet example. We visualize a Ricker wavelet (see Eq. (5.6)) with width $w=1$ and height $h=3$.

We are interested in the types of functions $\rho_{0}$ for which the "true" distribution $f(x)$ is the solution to Eq. (5.4). We restrict the analysis to the case when the true distribution is a Gaussian with mean $\mu=0$ and standard deviation $\sigma$.

Substituting $\rho(x, t)=\mathcal{N}\left(0, \sigma^{2}\right)$ and solving for $\rho_{0}$ yields

$$
\begin{equation*}
\rho_{0}(x, 0)=f(x)+\xi(x)=\frac{e^{-\frac{x^{2}}{2 \sigma^{2}}}}{\sigma \sqrt{2 \pi}}+\frac{\left(\sigma^{2}-x^{2}\right) e^{-\frac{x^{2}}{2 \sigma^{2}}}}{\sqrt{2 \pi} q^{2} \sigma^{5}} \tag{5.7}
\end{equation*}
$$

The first term is a Gaussian with mean $\mu=0$ and standard deviation $\sigma$-the true distribution $f(x)$ —and the second term is a Ricker wavelet where

$$
\begin{align*}
& h=1 /\left(\sqrt{2 \pi} q^{2} \sigma^{3}\right)  \tag{5.8}\\
& w=\sigma \tag{5.9}
\end{align*}
$$

One often does not have control in the size of the disturbance in the data. Therefore, we assume that $h$ is constant and attempt to find the amount of smoothing that yields the exact answer. Substituting Eq. (5.5) into Eq. (5.8) and isolating $k$ yields

$$
\begin{equation*}
k=\frac{4 \sqrt{\pi} h w^{3} \beta}{4 \sqrt{\pi} h w^{3} \beta+\sqrt{2} \alpha} . \tag{5.10}
\end{equation*}
$$

For all computations, we set $\alpha=\beta=1$.

### 5.3. Numerical experiment

We numerically test the smoothing required to remove a Ricker wavelet perturbation $\xi(x)$ with width $w=\sigma$ and height $h=1 /\left(\sqrt{2 \pi} q^{2} \sigma^{3}\right)$ from a "true" Gaussian distribution $f(x)$. Figure 5.2 shows the integrated square error between the true distribution $f(x)$ and the numerical equilibrium of Eq. (5.2). The black curve shows the amount of smoothing $k$ set by the theory from Eqns. (5.5) and (5.10). The red curve shows the $k$ value that numerically generates the


Figure 5.2. Setting smoothing to remove disturbance: theory versus numerics. We generate the integrated squared error between the model equilibrium and true distribution $f(x)$ given the initial condition $\rho_{0}(x)=f(x)+\xi(x)$, where $\xi(x)$ a Ricker wavelet defined in Eq. (5.6) and $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. The color indicates the amount of error with blue error corresponding to large error and yellow to small error. The $k$ value that yields the minimum error (red, dashed) for a given height of the Ricker wavelet is equivalent to the $k$ value derived from the Eq. (5.5) and (5.10) (black, solid). Here, $\sigma=w=5$ and $N=1000$
smallest integrated squared error given the height of the Ricker wavelet. The red and black curves appear to be identical.

Figure 5.3 shows the error landscape after a similar set of simulations where we independently varied the height and width of the Ricker wavelet with $k=0.95$ fixed. In this case, we see that, as expected from Eq. (5.9), error is only minimized when $w=\sigma$. The corresponding $h$ value derived from Eqns. (5.8) and (5.5) that minimizes error is $h=1 /\left(\sqrt{2 \pi} q^{2} \sigma^{3}\right) \approx 0.38$.

### 5.4. Spatially varying smoothing

We expand our model to include spatially varying smoothing $k_{i}$. $k_{i}$ will be determined by the amount of "bias" at bin $i$ at time $\mathbf{t}$, where bias is defined as $b_{i}=y_{i, 0}-y_{i}$. If the bias increases then we increase the amount smoothing at $y_{i}$, the amount of smoothing decreases if the opposite is true.


Figure 5.3. Heatmap of the integrated squared error when varying the height and the width of the perturbation. We compute the integrated squared error between the model result and true distribution $f(x)$ when the initial condition the sum $\rho_{0}(x)=f(x)+\xi(x)$, where $\xi(x)$ is the added perturbation. Here, the true distribution $f(x)$ is a Gaussian with variance $\sigma^{2}=25$ and mean $\mu=0$ and the perturbation $\xi(x)$ is a Ricker wavelet (see (5.6) with width $w$ and height $h$. The error is minimized when $w=\sigma$ and $h=1 /\left(\sqrt{2 \pi} q^{2} \sigma^{3}\right)=0.38$, consistent with Eq. (5.10). The smoothing strength $k$ ranges from 0.01 to 0.99 with stepsize 0.005 and the height of the Ricker wavelet ranged from 0.0008 to 0.1596 with stepsize 0.0008 .

In this chapter, we set $k_{i}\left(b_{i}\right)$ as

$$
\begin{equation*}
k_{i}\left(b_{i}\right)=\frac{1}{2}\left[\tanh \left(\lambda b_{i}\right)+1\right] \tag{5.11}
\end{equation*}
$$

where $\lambda$ sets bias sensitivity with $\lambda$ large implying high sensitivity and and $\lambda$ small implying low sensitivity. We have chosen $k_{i}$ to be sigmoidal so that as the bias approaches infinity, the smoothing strength $k_{i} \rightarrow 1$ and the system becomes fully diffusive at bin $i$ and as bias approaches minus infinity, the smoothing strength $k_{i} \rightarrow 0$ and diffusion ceases and the $y_{i}$ is attracted back to its initial value $y_{i, 0}$. Furthermore, we have set $k_{i}$ so that when $b_{i}=0$, the rate of change of $y_{i}$ is the average between local diffusion and the restoring source.

We test the model extension on a composite distribution $\rho_{0}(x, t)=(1-\epsilon) \mathcal{N}\left(0, \sigma^{2}\right)+$ $\epsilon \mathcal{N}\left(0, \delta^{2}\right)$ where $\mathcal{N}\left(\mu, \sigma^{2}\right)$ represents a normal distribution centered at $\mu$ with variance $\sigma^{2}$, and
we take $\epsilon \ll 1, \delta \ll \sigma$. We assume that the true underlying distribution is $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. We fit $\lambda$ so the integrated squared error between the true distribution and the model equilibrium is minimized. We carry out this minimization using the Nelder-Mead algorithm [110].


Figure 5.4. Removing perturbations. Panels (a)-(d) show the evolution from the perturbed distribution $\rho_{0}(x, t)=(1-\epsilon) \mathcal{N}\left(0, \sigma^{2}\right)+\epsilon \mathcal{N}\left(0, \delta^{2}\right)$, where $\epsilon \ll 1$, $\delta \ll \sigma$ to the equilibrium distribution when the smoothing parameter $k_{i}\left(y_{i, 0}, y_{i}\right)$ varies in space. The equilibrium distribution (red, solid) shows good agreement with the true distribution $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. The bias sensitivity parameter $\lambda$ was set by minimizing the integrated squared error, with the $\lambda \approx 1.8624 \times 10^{3}$. In this simulation, $\sigma=50, \delta=0.5, \epsilon=0.01$, and $N=1000$.

Figure 5.4 illustrates this method on the test distribution. Panels (a)-(d) show the evolution of the perturbed distribution (red, solid). The spike in the perturbed distribution diminishes and its mass spreads to its neighbors with the equilibrium showing good agreement with the true distribution $f(x)=\mathcal{N}\left(0, \sigma^{2}\right)$. Here, $\sigma=50, \delta=0.5, \epsilon=0.01, \lambda \approx 1.8624 \times 10^{3}$, and $N=1000$.

### 5.5. Removing bias from real data

People report an overestimation of their height when asked [162]. Such misreporting makes height data an appropriate test dataset. Additionally, there is some evidence suggesting that adult male height is actually normally distributed [156].


Figure 5.5. Removing bias from height data. We apply the our method with a spatial varying smoothing parameter $k_{i}\left(y_{i, 0}, y_{i}\right)$ set by Eq. (5.11). We see that the equilibrium generated by the model (red, circles) removes the second mode at 72 inches, which we assume to be due to overestimation.

We test our method on a dataset containing heights recorded in medical records from hospitals in the Northwestern Medical system. Figure 5.5 illustrates how our method affects the data. The equilibrium distribution (red, solid circles) generated by the method removes the second mode at 72 inches seen in the original data (blue histogram) while keeping the underlying distribution unaffected. In this simulation we set the $\lambda=10$.

### 5.6. Conclusions

We have put forth a dynamical systems approach to remove perturbations from data by balancing diffusion and restoration effects. Our approach works exactly in the case of a specific "Mexican hat" type of perturbation, but we expanded the efficacy of this method by allowing the ratio between the two effect to vary in space. We also applied it to a dataset of adult male heights, removing a spike while keeping the underlying distribution apparently unaffected. This method gives a simple tool to correct data that is biased due to systematic error or artificial distortion. We hope our work invites more dynamical system approaches in data science.

## References

[1] Tetsu Saigusa, Atsushi Tero, Toshiyuki Nakagaki, and Yoshiki Kuramoto. Amoebae anticipate periodic events. Physical Review Letters, 100(1):018101, 2008.
[2] Jihwan Myung, Sungho Hong, Daniel DeWoskin, Erik De Schutter, Daniel B Forger, and Toru Takumi. Gaba-mediated repulsive coupling between circadian clock neurons in the scn encodes seasonal time. Proceedings of the National Academy of Sciences, 112(29):E3920-E3929, 2015.
[3] Annette F Taylor, Mark R Tinsley, Fang Wang, Zhaoyang Huang, and Kenneth Showalter. Dynamical quorum sensing and synchronization in large populations of chemical oscillators. Science, 323(5914):614-617, 2009.
[4] Masahiro Toiya, Hector O González-Ochoa, Vladimir K Vanag, Seth Fraden, and Irving R Epstein. Synchronization of chemical micro-oscillators. The Journal of Physical Chemistry Letters, 1(8):1241-1246, 2010.
[5] Sundarapandian Vaidyanathan. Anti-synchronization of brusselator chemical reaction systems via adaptive control. International Journal of ChemTech Research, 8(6):759768, 2015.
[6] Yan-Ni Li, Lan Chen, Zun-Sheng Cai, and Xue-Zhuang Zhao. Study on chaos synchronization in the belousov-zhabotinsky chemical system. Chaos, Solitons \& Fractals, 17(4):699-707, 2003.
[7] James Pantaleone. Synchronization of metronomes. American Journal of Physics, 70(10):992-1000, 2002.
[8] Ryo Yoshida, Masami Tanaka, Satoko Onodera, Tomohiko Yamaguchi, and Etsuo Kokufuta. In-phase synchronization of chemical and mechanical oscillations in selfoscillating gels. The Journal of Physical Chemistry A, 104(32):7549-7555, 2000.
[9] Nijmeijer Henk and Rodriguez-angeles Alejandro. Synchronization of mechanical systems, volume 46. World Scientific, 2003.
[10] Henning Ulrichs, Andreas Mann, and Ulrich Parlitz. Synchronization and chaotic dynamics of coupled mechanical metronomes. Chaos: An Interdisciplinary Journal of Nonlinear Science, 19(4):043120, 2009.
[11] Bruno H Repp. Sensorimotor synchronization: a review of the tapping literature. Psychonomic bulletin \& review, 12(6):969-992, 2005.
[12] Sebastian Kirschner and Michael Tomasello. Joint drumming: social context facilitates synchronization in preschool children. Journal of Experimental Child Psychology, 102(3):299-314, 2009.
[13] Aureo De Paula. Inference in a synchronization game with social interactions. Journal of econometrics, 148(1):56-71, 2009.
[14] Alessandro Pluchino, Vito Latora, and Andrea Rapisarda. Compromise and synchronization in opinion dynamics. The European Physical Journal B-Condensed Matter and Complex Systems, 50(1-2):169-176, 2006.
[15] Yoshiki Kuramoto. Chemical oscillations, waves, and turbulence. Courier Corporation, 2003.
[16] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In International symposium on mathematical problems in theoretical physics, pages 420422. Springer, 1975.
[17] Steven H Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. Physica D: Nonlinear Phenomena, 143(1-4):1-20, 2000.
[18] Juan A Acebrón, Luis L Bonilla, Conrad J Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. Reviews of modern physics, 77(1):137, 2005.
[19] Yamir Moreno and Amalio F Pacheco. Synchronization of kuramoto oscillators in scalefree networks. EPL (Europhysics Letters), 68(4):603, 2004.
[20] Qingyun Wang, Guanrong Chen, and Matjaž Perc. Synchronous bursts on scale-free neuronal networks with attractive and repulsive coupling. PLoS one, 6(1):e15851, 2011.
[21] Irving R Epstein. Nonlinear oscillations in chemical and biological systems. Physica D: Nonlinear Phenomena, 51(1-3):152-160, 1991.
[22] W Hohmann, M Kraus, and FW Schneider. Learning and recognition in excitable chemical reactor networks. The Journal of Physical Chemistry A, 102(18):3103-3111, 1998.
[23] A Koseska, E Volkov, A Zaikin, and J Kurths. Inherent multistability in arrays of autoinducer coupled genetic oscillators. Physical Review E, 75(3):031916, 2007.
[24] Ekkehard Ullner, Alexei Zaikin, Evgenii I Volkov, and Jordi García-Ojalvo. Multistability and clustering in a population of synthetic genetic oscillators via phase-repulsive cell-to-cell communication. Physical review letters, 99(14):148103, 2007.
[25] Ekkehard Ullner, Aneta Koseska, Jürgen Kurths, Evgenii Volkov, Holger Kantz, and Jordi García-Ojalvo. Multistability of synthetic genetic networks with repressive cell-tocell communication. Physical Review E, 78(3):031904, 2008.
[26] Seth A Marvel, Steven H Strogatz, and Jon M Kleinberg. Energy landscape of social balance. Physical review letters, 103(19):198701, 2009.
[27] Rodrigo Laje and Gabriel B Mindlin. Diversity within a birdsong. Physical review letters, 89(28):288102, 2002.
[28] Gábor Balázsi, Ann Cornell-Bell, Alexander B Neiman, and Frank Moss. Synchronization of hyperexcitable systems with phase-repulsive coupling. Physical Review E, 64(4):041912, 2001.
[29] Michael Giver, Zahera Jabeen, and Bulbul Chakraborty. Phase and frequency entrainment in locally coupled phase oscillators with repulsive interactions. Physical Review E, 83(4):046206, 2011.
[30] Hyunsuk Hong and Steven H Strogatz. Kuramoto model of coupled oscillators with positive and negative coupling parameters: an example of conformist and contrarian oscillators. Physical Review Letters, 106(5):054102, 2011.
[31] Hyunsuk Hong and Steven H Strogatz. Mean-field behavior in coupled oscillators with attractive and repulsive interactions. Physical Review E, 85(5):056210, 2012.
[32] Douglas J Emlen. Artificial selection on horn length-body size allometry in the horned beetle onthophagus acuminatus (coleoptera: Scarabaeidae). Evolution, 50(3):1219-1230, 1996.
[33] Sara M Clifton, Rosemary I Braun, and Daniel M Abrams. Handicap principle implies emergence of dimorphic ornaments. Proceedings of the Royal Society B: Biological Sciences, 283(1843):20161970, 2016.
[34] SM Clifton, RI Braun, and DM Abrams. Data from: Handicap principle implies emergence of dimorphic ornaments. Dryad Digital Repository.(doi: 10.5061/dryad. vblpp), 2016.
[35] Børge Damsgård, Tor H Evensen, Øyvind Øverli, Marnix Gorissen, Lars OE Ebbesson, Sonia Rey, and Erik Höglund. Proactive avoidance behaviour and pace-of-life syndrome in atlantic salmon. Royal Society Open Science, 6(3):181859, 2019.
[36] B Damsgård, TH Evensen, Ø Øverli, M Gorissen, L Ebbesson, S Ray, and E H'aglund. Data from: Proactive avoidance behaviour and pace-of-life syndrome in atlantic salmon, 2019.
[37] Michael R Blanton, David W Hogg, Neta A Bahcall, Ivan K Baldry, J Brinkmann, Istvan Csabai, Daniel Eisenstein, Masataka Fukugita, James E Gunn, Željko Ivezić, et al. The broadband optical properties of galaxies with redshifts $0.02 ; \mathrm{zi}_{\mathrm{i}} 0.22$. The Astrophysical Journal, 594(1):186, 2003.
[38] Kevork Abazajian, Jennifer K Adelman-McCarthy, Marcel A Agüeros, Sahar S Allam, Scott F Anderson, James Annis, Neta A Bahcall, Ivan K Baldry, Steven Bastian, Andreas Berlind, et al. The first data release of the sloan digital sky survey. The Astronomical Journal, 126(4):2081, 2003.
[39] Ivan K Baldry, Karl Glazebrook, Jon Brinkmann, Željko Ivezić, Robert H Lupton, Robert C Nichol, and Alexander S Szalay. Quantifying the bimodal color-magnitude distribution of galaxies. The Astrophysical Journal, 600(2):681, 2004.
[40] Irine Ronin, Naama Katsowich, Ilan Rosenshine, and Nathalie Q Balaban. A long-term epigenetic memory switch controls bacterial virulence bimodality. Elife, 6:e19599, 2017.
[41] Amir Goldberg, Ofer Fridman, Irine Ronin, and Nathalie Q Balaban. Systematic identification and quantification of phase variation in commensal and pathogenic escherichia coli. Genome medicine, 6(11):112, 2014.
[42] Zoran Levnajić. Emergent multistability and frustration in phase-repulsive networks of oscillators. Physical Review E, 84(1):016231, 2011.
[43] Zoran Levnajić. Evolutionary design of non-frustrated networks of phase-repulsive oscillators. Scientific Reports, 2:967, 2012.
[44] Anastasiya V Pimenova, Denis S Goldobin, Michael Rosenblum, and Arkady Pikovsky. Interplay of coupling and common noise at the transition to synchrony in oscillator populations. Scientific Reports, 6:38518, 2016.
[45] Ken H Nagai and Hiroshi Kori. Noise-induced synchronization of a large population of globally coupled nonidentical oscillators. Physical Review E, 81(6):065202, 2010.
[46] Santiago Gil, Yoshiki Kuramoto, and Alexander S Mikhailov. Common noise induces clustering in populations of globally coupled oscillators. EPL (Europhysics Letters), 88(6):60005, 2010.
[47] Chen Chris Gong, Chunming Zheng, Ralf Toenjes, and Arkady Pikovsky. Repulsively coupled kuramoto-sakaguchi phase oscillators ensemble subject to common noise. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(3):033127, 2019.
[48] Y Nakamura, F Tominaga, and T Munakata. Clustering behavior of time-delayed nearestneighbor coupled oscillators. Physical Review E, 49(6):4849, 1994.
[49] Arindam Mishra, Chittaranjan Hens, Mridul Bose, Prodyot K Roy, and Syamal K Dana. Chimeralike states in a network of oscillators under attractive and repulsive global coupling. Physical Review E, 92(6):062920, 2015.
[50] Azamat Yeldesbay, Arkady Pikovsky, and Michael Rosenblum. Chimeralike states in an ensemble of globally coupled oscillators. Physical review letters, 112(14):144103, 2014.
[51] D Golomb, David Hansel, B Shraiman, and Haim Sompolinsky. Clustering in globally coupled phase oscillators. Physical Review A, 45(6):3516, 1992.
[52] LS Tsimring, NF Rulkov, ML Larsen, and Michael Gabbay. Repulsive synchronization in an array of phase oscillators. Physical Review Letters, 95(1):014101, 2005.
[53] Koji Okuda. Variety and generality of clustering in globally coupled oscillators. Physica D: Nonlinear Phenomena, 63(3-4):424-436, 1993.
[54] Patrick Billingsley. Probability and Measure. 1995. John Wiley \& Sons, New York, 1995.
[55] Chhaya Chaudhary, Hanieh Saeedi, and Mark J Costello. Bimodality of latitudinal gradients in marine species richness. Trends in Ecology \& Evolution, 31(9):670-676, 2016.
[56] P Seeman, C Ulpian, C Bergeron, P Riederer, K Jellinger, E Gabriel, GP Reynolds, and WW Tourtellotte. Bimodal distribution of dopamine receptor densities in brains of schizophrenics. Science, 225(4663):728-731, 1984.
[57] Timothy J Lewis and John Rinzel. Dynamics of spiking neurons connected by both inhibitory and electrical coupling. Journal of Computational Neuroscience, 14(3):283-309, 2003.
[58] Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber. Evidence for a bimodal distribution in human communication. Proceedings of the National Academy of Sciences, 107(44):18803-18808, 2010.
[59] Paul DiMaggio, John Evans, and Bethany Bryson. Have american's social attitudes become more polarized? American Journal of Sociology, 102(3):690-755, 1996.
[60] Brian R Flay. Catastrophe theory in social psychology: Some applications to attitudes and social behavior. Behavioral Science, 23(4):335-350, 1978.
[61] P Bonifacio, E Caffau, M Spite, M Limongi, A Chieffi, RS Klessen, P François, P Molaro, H-G Ludwig, S Zaggia, et al. Topos-ii. on the bimodality of carbon abundance in cemp stars implications on the early chemical evolution of galaxies. Astronomy \& Astrophysics, 579:A28, 2015.
[62] Johannes Reiter and Irving R Epstein. Bimodality in the cooperative binding of ligands to molecules with multiple binding sites. Journal of Physical Chemistry, 91(18):48134820, 1987.
[63] Yan Ren-Chang, Bian Jian-Chun, and Fan Qiu-Jun. The impact of the south asia high bimodality on the chemical composition of the upper troposphere and lower stratosphere. Atmospheric and Oceanic Science Letters, 4(4):229-234, 2011.
[64] Shen-Su Sun and Robert W Nesbitt. Chemical heterogeneity of the archaean mantle, composition of the earth and mantle evolution. Earth and Planetary Science Letters, 35(3):429-448, 1977.
[65] Avishai Dekel and Yuval Birnboim. Galaxy bimodality due to cold flows and shock heating. Monthly Notices of the Royal Astronomical Society, 368(1):2-20, 2006.
[66] Chidong Zhang, Brian E Mapes, and Brian J Soden. Bimodality in tropical water vapour. Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography, 129(594):28472866, 2003.
[67] Patricia K Kuhl and Andrew N Meltzoff. The bimodal perception of speech in infancy. Science, 218(4577):1138-1141, 1982.
[68] Donghuai Sun, Jan Bloemendal, David K Rea, Zhisheng An, Jef Vandenberghe, Huayu Lu, Ruixia Su, and Tungsheng Liu. Bimodal grain-size distribution of chinese loess, and its palaeoclimatic implications. Catena, 55(3):325-340, 2004.
[69] JE Thorpe. Bimodal distribution of length of juvenile atlantic salmon (salmo salar 1.) under artificial rearing conditions. Journal of Fish Biology, 11(2):175-184, 1977.
[70] Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber. Evidence for a bimodal distribution in human communication. Proceedings of the National Academy of Sciences, 107(44):18803-18808, 2010.
[71] P Seeman, C Ulpian, C Bergeron, P Riederer, K Jellinger, E Gabriel, GP Reynolds, and WW Tourtellotte. Bimodal distribution of dopamine receptor densities in brains of schizophrenics. Science, 225(4663):728-731, 1984.
[72] Josiah Schwab, Ph Podsiadlowski, and Saul Rappaport. Further evidence for the bimodal distribution of neutron-star masses. The Astrophysical Journal, 719(1):722, 2010.
[73] Shude Mao, Ramesh Narayan, and Tsvi Piran. On the bimodal distribution of gamma-ray bursts. The Astrophysical Journal, 420:171-176, 1994.
[74] MP Eichhorn. Spatial organisation of a bimodal forest stand. Journal of forest research, 15(6):391-397, 2010.
[75] H. Shaban, Digital advertising to surpass print and TV for the first time, report says. The Washington Post (2019). https://www.washingtonpost.com/technology/2019/02/20/digital-advertising-surpass-print-tv-first-time-report-says/.
[76] R. Schmalensee, M. Armstrong, R. D. Willig, R. H. Porter, Handbook of Industrial Organization, vol. 3 (Elsevier, 1989).
[77] F. Bloch, D. Manceau, Persuasive advertising in Hotelling's model of product differentiation. International Journal of Industrial Organization 17, 557-574 (1999).
[78] R. Crisp, Persuasive advertising, autonomy, and the creation of desire. Journal of Business Ethics 6, 413-418 (1987).
[79] A. K. Dixit, V. Norman, Advertising and welfare. Bell Journal of Economics 9, 1-17 (1976).
[80] L. G. Telser, Advertising and competition. Journal of Political Economy 72, 537-562 (1964).
[81] P. Nelson, The economic consequences of advertising. The Journal of Business 48, 213241 (1975).
[82] G. J. Stigler, The economics of information. Journal of Political Economy 69, 213-225 (1961).
[83] G. S. Becker, K. M. Murphy, A simple theory of advertising as a good or bad. The Quarterly Journal of Economics 108, 941-964 (1993).
[84] L. M. Nichols, Advertising and economic welfare. The American Economic Review 75, 213-218 (1985).
[85] G. J. Stigler, G. S. Becker, De gustibus non est disputandum. The American Economic Review 67, 76-90 (1977).
[86] D. Braithwaite, The economic effects of advertisement. The Economic Journal 38, 16-37 (1928).
[87] A. M. Abernethy, D. D. Butler, Advertising information: services versus products. Journal of Retailing 68, 398 (1992).
[88] N. Lindstädt, O. Budzinski, Newspaper vs. online advertising-is there a niche for newspapers in modern advertising markets? Ilmenau Economics Discussion Papers (2011).
[89] B. Crettez, N. Hayek, G. Zaccour, Existence and uniqueness of optimal dynamic pricing and advertising controls without concavity. Operations Research Letters 46, 199-204 (2018).
[90] C. Marinelli, S. Savin, Optimal distributed dynamic advertising. Journal of Optimization Theory and Applications 137, 569-591 (2008).
[91] H. Simon, Adpuls: An advertising model with wearout and pulsation. Journal of Marketing Research 19, 352-363 (1982).
[92] S. P. Sethi, Deterministic and stochastic optimization of a dynamic advertising model. Optimal Control Applications and Methods 4, 179-184 (1983).
[93] J. W. Friedman, Advertising and oligopolistic equilibrium. The Bell Journal of Economics pp. 464-473 (1983).
[94] G. M. Erickson, A differential game model of the marketing-operations interface. European Journal of Operational Research 211, 394-402 (2011).
[95] C. Fershtman, Goodwill and market shares in oligopoly. Economica 51, 271-281 (1984).
[96] L. Esteban, A. Gil, J. M. Hernandez, Informative advertising and optimal targeting in a monopoly. The Journal of Industrial Economics 49, 161-180 (2001).
[97] L. Gori, M. Sodini, L. Fanti, A nonlinear cournot duopoly with advertising. Chaos, Solitons \& Fractals 79, 178-190 (2015).
[98] Kilts Nielsen Center, Kilts Nielsen Consumer Panel Data. The Kilts Center Archive of The Nielsen Company (2014).
[99] S. E. Page, Aggregation in agent-based models of economies. The Knowledge Engineering Review 27, 151-162 (2012).
[100] E. L. McCullough, C. W. Miller, D. J. Emlen, Why sexually selected weapons are not ornaments. Trends in Ecology \& Evolution 31, 742-751 (2016).
[101] M. Kirkpatrick, Sexual selection and the evolution of female choice. Evolution 36, 1-12 (1982).
[102] D. Tilman, Resource competition between plankton algae: an experimental and theoretical approach. Ecology 58, 338-348 (1977).
[103] M. E. Hibbing, C. Fuqua, M. R. Parsek, S. B. Peterson, Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology 8, 15-25 (2010).
[104] P. Auger, R. B. De La Parra, E. Sánchez, Hawk-dove game and competition dynamics. Mathematical and Computer Modelling 27, 89-98 (1998).
[105] G. Norman, L. Pepall, D. Richards, Generic product advertising, spillovers, and market concentration. American Journal of Agricultural Economics 90, 719-732 (2008).
[106] F. M. Bass, A. Krishnamoorthy, A. Prasad, S. P. Sethi, Generic and brand advertising strategies in a dynamic duopoly. Marketing Science 24, 556-568 (2005).
[107] D. Marković, C. Gros, Power laws and self-organized criticality in theory and nature. Physics Reports 536, 41-74 (2014).
[108] A. Clauset, C. R. Shalizi, M. E. Newman, Power-law distributions in empirical data. SIAM Review 51, 661-703 (2009).
[109] M. Tribe, R. Alpine, Scale economies and the " 0.6 rule". Engineering Costs and Production Economics 10, 271-278 (1986).
[110] J. A. Nelder, R. Mead, A simplex method for function minimization. The Computer Journal 7, 308-313 (1965).
[111] J. A. Hartigan, P. Hartigan, The dip test of unimodality. The Annals of Statistics 13, 70-84 (1985).
[112] V. J. Tremblay, C. H. Tremblay, New Perspectives on Industrial Organization (Springer, 2012), pp. 467-484.
[113] H. Jošić, B. Žmuk, The application of Benford's law in psychological pricing detection. Zbornik Radova Ekonomskog Fakulteta Sveučilišta u Mostaru (2018).
[114] T. El Sehity, E. Hoelzl, E. Kirchler, Price developments after a nominal shock: Benford's law and psychological pricing after the euro introduction. International Journal of Research in Marketing 22, 471-480 (2005).
[115] R. M. Schindler, Patterns of price endings used in US and Japanese price advertising. International Marketing Review 26, 17-29 (2009).
[116] M. Stiving, R. S. Winer, An empirical analysis of price endings with scanner data. Journal of Consumer Research 24, 57-67 (1997).
[117] E. Parzen, On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33, 1065-1076 (1962).
[118] H. Mohammad, C. P. Diduch, Y. Biletskiy, L. Chang, Filtering out spikes from sensors in power converters system using discrete wavelet transform. 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) pp. 1-3 (2012).
[119] J. D. Johnson, D. M. Abrams, A dynamical systems approach to debiasing data. Manuscript in preparation.
[120] B. W. Silverman, Density Estimation for Statistics and Data Analysis, vol. 26 (CRC press, 1986).
[121] Charlat, S., Hornett, E. A., Fullard, J. H., Davies, N., Roderick, G. K., Wedell, Hurst, G. D. Extraordinary flux in sex ratio. Science. 2007;317(5835):214-214.
[122] Kalmus H. Über den Erhaltungswert der phänotypischen (morphologischen) Anisogamie und die Entstehung der ersten Geschlechtsunterschiede. Biologisches Zentralblatt. 1932;52:716-736.
[123] Kalmus H, Smith C. Evolutionary origin of sexual differentiation and the sex-ratio. Nature. 1960;186(4730):1004-1006.
[124] Scudo FM. The adaptive value of sexual dimorphism: I, anisogamy. Evolution. 1967; p. 285-291.
[125] Parker GA, Baker RR, Smith V. The origin and evolution of gamete dimorphism and the male-female phenomenon. Journal of Theoretical Biology. 1972;36(3):529-553.
[126] Bell G. The evolution of anisogamy. Journal of Theoretical Biology. 1978;73(2):247270.
[127] Cox PA, Sethian JA. Gamete motion, search, and the evolution of anisogamy, oogamy, and chemotaxis. The American Naturalist. 1985;125(1):74-101.
[128] Hurst LD. Parasite diversity and the evolution of diploidy, multicellularity and anisogamy. Journal of Theoretical Biology. 1990;144(4):429-443.
[129] Bonsall MB. The evolution of anisogamy: The adaptive significance of damage, repair and mortality. Journal of Theoretical Biology. 2006;238(1):198-210.
[130] Blute M. The evolution of anisogamy: more questions than answers. Biological Theory. 2013;7(1):3-9.
[131] Lehtonen J, Parker GA, Schärer L. Why anisogamy drives ancestral sex roles. Evolution. 2016;70(5):1129-1135.
[132] Haig D, Westoby M. A model for the origin of heterospory. Journal of Theoretical Biology. 1988;134(2):257-272.
[133] Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biological Reviews. 2011;86(2):421-442.
[134] Bateman RM, DiMICHELE WA. Heterospory: the most iterative key innovation in the evolutionary history of the plant kingdom. Biological Reviews. 1994;69(3):345-417.
[135] Bulmer M, Parker GA. The evolution of anisogamy: a game-theoretic approach. Proceedings of the Royal Society of London Series B: Biological Sciences. 2002;269(1507):2381-2388.
[136] Hayward A, Gillooly JF. The cost of sex: quantifying energetic investment in gamete production by males and females. PLoS One. 2011;6(1):e16557.
[137] Bateman AJ. Intra-sexual selection in Drosophila. Heredity. 1948;2(3):349-368.
[138] Kues U, Casselton LA. The origin of multiple mating types in mushrooms. Journal of Cell Science. 1993;104(2):227-230.
[139] Kuees U. From two to many: multiple mating types in Basidiomycetes. Fungal Biology Reviews. 2015;29(3-4):126-166.
[140] Wallace WHB, Kelsey TW. Human ovarian reserve from conception to the menopause. PloS One. 2010;5(1).
[141] Johnson L, Petty CS, Neaves WB. Further quantification of human spermatogenesis: germ cell loss during postprophase of meiosis and its relationship to daily sperm production. Biology of Reproduction. 1983;29(1):207-215.
[142] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Eggs. In: Molecular Biology of the Cell, 4th edition. Garland Science; 2002.
[143] Bellastella G, Cooper TG, Battaglia M, Ströse A, Torres I, Hellenkemper B, et al. Dimensions of human ejaculated spermatozoa in Papanicolaou-stained seminal and swim-up smears obtained from the Integrated Semen Analysis System (ISAS®). Asian Journal of Andrology. 2010;12(6):871.
[144] Nager RG, Monaghan P, Houston DC. Within-clutch trade-offs between the number and quality of eggs: experimental manipulations in gulls. Ecology. 2000;81(5):1339-1350.
[145] Hedges SB. At the lower size limit in snakes: two new species of threadsnakes (Squamata: Leptotyphlopidae: Leptotyphlops) from the Lesser Antilles. Zootaxa. 2008;1841(1):1-30.
[146] Brown JH, Gupta VK, Li BL, Milne BT, Restrepo C, West GB. The fractal nature of nature: power laws, ecological complexity and biodiversity. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2002;357(1421):619-626.
[147] Newman ME. Power laws, Pareto distributions and Zipf's law. Contemporary Physics. 2005;46(5):323-351.
[148] Blomqvist D, Johansson OC, Götmark F. Parental quality and egg size affect chick survival in a precocial bird, the lapwing Vanellus vanellus. Oecologia. 1997;110(1):18-24.
[149] Krist M. Egg size and offspring quality: a meta-analysis in birds. Biological Reviews. 2011;86(3):692-716.
[150] Erikstad KE, Tveraa T, Bustnes JO. Significance of intraclutch egg-size variation in common eider: the role of egg size and quality of ducklings. Journal of Avian Biology. 1998; p. 3-9.
[151] Valkama J, Korpimäki E, Wiehn J, Pakkanen T. Inter-clutch egg size variation in kestrels Falco tinnunculus: seasonal decline under fluctuating food conditions. Journal of Avian Biology. 2002;33(4):426-432.
[152] Taylor PD, Jonker LB. Evolutionary stable strategies and game dynamics. Mathematical Biosciences. 1978;40(1-2):145-156.
[153] Schuster P, Sigmund K. Replicator dynamics. Journal of Theoretical Biology. 1983;100(3):533-538.
[154] Roff D. The evolution of life histories: theory and analysis. Springer; 1993.
[155] I. S. Abramson, On bandwidth variation in kernel estimates-a square root law, The annals of Statistics, (1982), pp. 1217-1223.
[156] B. A'hearn, F. Peracchi, and G. Vecchi, Height and the normal distribution: evidence from italian military data, Demography, 46 (2009), pp. 1-25.
[157] L. Bevilacqua, A. C. Galeão, and F. P. Costa, A new analytical formulation of retention effects on particle diffusion processes, Anais da Academia Brasileira de Ciências, 83 (2011), pp. 1443-1464.
[158] L. Bevilacqua, A. C. Galeão, and F. P. Costa, On the significance of higher order differential terms in diffusion processes, Journal of the Brazilian Society of Mechanical Sciences and Engineering, 33 (2011), pp. 166-175.
[159] D. Conroy-Beam and D. M. Buss, Euclidean distances discriminatively predict short-term and long-term attraction to potential mates, Evolution and Human Behavior, 38 (2017), pp. 442-450.
[160] S. Gates and P. F. Donald, Local extinction of british farmland birds and the prediction of further loss, Journal of Applied Ecology, 37 (2000), pp. 806-820.
[161] L. M. Kreul, Magic numbers: psychological aspects of menu pricing, Cornell Hotel and Restaurant Administration Quarterly, 23 (1982), pp. 70-75.
[162] C. Lassale, S. Péneau, M. Touvier, C. Julia, P. Galan, S. Hercberg, and E. Kesse-Guyot, Validity of web-based self-reported weight and height: results of the nutrinet-santé study, Journal of medical Internet research, 15 (2013), p. e152.
[163] J. J. Lee, W. K. Hong, W. N. Hittelman, L. Mao, R. Lotan, D. M. Shin, S. E. Benner, X.-C. Xu, J. S. Lee, V. M. Papadimitrakopoulou, et al., Predicting cancer development in oral leukoplakia: ten years of translational research, Clinical Cancer Research, 6 (2000), pp. 1702-1710.
[164] D. MARR, Vision wh freeman and co, San Francisco, (1982), pp. 54-79.
[165] I. Niedhammer, I. Bugel, S. Bonenfant, M. Goldberg, and A. Leclerc, Validity of self-reported weight and height in the french gazel cohort, International journal of obesity, 24 (2000), p. 1111.
[166] N. Ricker, The form and laws of propagation of seismic wavelets, Geophysics, 18 (1953), pp. 10-40.
[167] R. M. Schindler and P. N. Kirby, Patterns of rightmost digits used in advertised prices: implications for nine-ending effects, Journal of Consumer Research, 24 (1997), pp. 192-201.

## APPENDIX A

## The development of multimodality via coupled oscillators: appendix

## A.1. Trimodal equilibria

We again consider a function $f(u)$ that satisfies conditions 2.2a)- (2.2e). We look for solutions with oscillators distributed according to $h(\theta)=x \delta\left(\theta-\theta_{1}\right)+y \delta\left(\theta-\theta_{2}\right)+(1-x-y) \delta\left(\theta-\theta_{3}\right)$, where $x, y>0, x+y<1$ so that the oscillators will be in three clusters at $\theta_{1}, \theta_{2}$, and $\theta_{3}$ (we again assume that the natural frequencies are identical):

$$
\begin{equation*}
\dot{\theta}_{1}=\omega_{0}+K\left(y f\left(\theta_{2}-\theta_{1}\right)+z f\left(\theta_{3}-\theta_{1}\right)\right) \tag{A.1}
\end{equation*}
$$

$$
\begin{equation*}
\dot{\theta}_{2}=\omega_{0}+K\left(-x f\left(\theta_{2}-\theta_{1}\right)+z f\left(\theta_{3}-\theta_{2}\right)\right) \tag{A.2}
\end{equation*}
$$

$$
\begin{equation*}
\dot{\theta}_{3}=\omega_{0}-K\left(x f\left(\theta_{3}-\theta_{1}\right)+y f\left(\theta_{3}-\theta_{2}\right)\right) . \tag{A.3}
\end{equation*}
$$

Here, $z=1-x-y$. We define two variables $\psi_{1}=\theta_{2}-\theta_{1}$ and $\psi_{2}=\theta_{3}-\theta_{2}$, so that the system reduces to

$$
\begin{equation*}
\dot{\psi}_{1}=-K\left(z\left[f\left(\psi_{2}+\psi_{1}\right)-f\left(\psi_{2}\right)\right]+(x+y) f\left(\psi_{1}\right)\right) \tag{A.4}
\end{equation*}
$$

$$
\begin{equation*}
\dot{\psi}_{2}=-K\left(x\left[f\left(\psi_{2}+\psi_{1}\right)-f\left(\psi_{1}\right)\right]+(y+z) f\left(\psi_{2}\right)\right) . \tag{A.5}
\end{equation*}
$$

We set $\dot{\psi}_{i}=0, i=1,2$ and arrive at the following system of equations:

$$
\begin{align*}
f\left(\psi_{2}\right) & =\frac{x f\left(\psi_{1}\right)}{z}  \tag{A.6}\\
f\left(\psi_{2}+\psi_{1}\right) & =\frac{-y f\left(\psi_{1}\right)}{z} .
\end{align*}
$$

To set bounds on the fractionation of the clusters, we assume that there exists points $\psi_{1}, \psi_{2} \in$ $(-\pi, \pi)$ such that Eqns. A.6) and (A.7) are satisfied. Additionally, we put our system of coupled oscillators into a rotating frame so that $\theta_{i} \rightarrow \theta_{i}+\omega_{0} t$. In the rotating frame, we set $\theta_{1}=0$, $\theta_{2}=\psi_{1}$, and $\theta_{3}=\psi_{1}+\psi_{2}-2 \pi$. As before we perturb an oscillator from one of the three groups. We do this for all three groups and get a system of inequalities

$$
\begin{equation*}
x f^{\prime}(0)+y f^{\prime}\left(\psi_{1}\right)+z f^{\prime}\left(\psi_{1}+\psi_{2}\right)<0 \tag{A.8a}
\end{equation*}
$$

$$
\begin{equation*}
z f^{\prime}(0)+y f^{\prime}\left(\psi_{2}\right)+x f^{\prime}\left(\psi_{1}+\psi_{2}\right)<0 \tag{A.8b}
\end{equation*}
$$

All these must be simultaneously satisfied for stability of a trimodal state. Adding, we find

$$
f^{\prime}(0)<-\left[(x+y) f^{\prime}\left(\psi_{1}\right)+(y+z) f^{\prime}\left(\psi_{2}\right)\right.
$$

This states that the weighted sum of the slopes of the coupling function at $\psi=\psi_{i}$, where the weights are the proportions for the groups separated by $\psi_{i}$, is greater in magnitude than the
slope at the origin. This condition reduces to

$$
\begin{equation*}
f^{\prime}(0)<-2 f^{\prime}\left(\frac{2 \pi}{3}\right) \tag{A.10}
\end{equation*}
$$

if $\psi_{1}=\psi_{2}=2 \pi-\psi_{1}-\psi_{2}=2 \pi / 3$. As an example, we return to the class of interaction functions that we introduced in Section 2.2. We relax the assumption that $|a|<\pi$ and consider the case when $\psi_{1}=\psi_{2}=2 \pi-\psi_{1}-\psi_{2}$. To satisfy inequality (A.10), this means that

$$
\begin{equation*}
1<\frac{56 \pi^{2}+54 a^{2}}{81 a^{2}} \tag{A.11}
\end{equation*}
$$

which reduces to

$$
\begin{equation*}
|a|<\frac{2}{3} \sqrt{\frac{14}{3}} \pi \equiv a_{\text {tricrit }} \approx 1.44 \pi \tag{A.12}
\end{equation*}
$$

Figure A.1 shows the results of a numerical experiment where we test this threshold. In both panels we use $N=99, x=y=z=1 / 3, \psi_{1}=\psi_{2}=2 \pi-\psi_{1}-\psi_{2}=2 \pi / 3$, and set $\omega_{0}=0$. We expect the trimodal state to be unstable for $a>a_{\text {tricrit. }}$. In panel (a) we set $a=a_{\text {tricrit }}+0.1$ and perturb the oscillators by amount $\xi_{i}$, with values drawn from the distribution $\mathcal{N}(0,0.01)$. We can see that this perturbation leads to the system leaving the trimodal state and going to a bimodal state with $180^{\circ}$ phase difference.

One might be interested in why the bimodal state is stable in panel (a). Since there are only zeros at $\psi=0$ and $\psi=\pi$, one may check the stability by evaluating the derivative of $f(\psi)$ at these points. One can show that if
(A.13)

$$
|a|>\sqrt{2} \pi
$$

the $180^{\circ}$ antiphase state is stable. Thus, when $a=a_{\text {tricrit }}+0.1>\sqrt{2} \pi$ the trimodal state becomes unstable and perturbations lead to the stable bimodal state.

For the case, when $\sqrt{2} \pi<a<a_{\text {tricrit }}$, both the trimodal state and the bimodal state are stable configurations. Figure A. 2 shows the result of the numerical experiment where we place the parameter inside the previously stated interval and outside of the interval. In all panels we use $N=300$, and set $\omega_{0}=0$. As before, in all panels we perturb the oscillators by amount $\xi_{i}$ from the predicted fixed points, whose values are drawn from the distribution $\mathcal{N}(0,0.01)$. In panels (a) and (b) we set $a=1.43 \pi \in\left(\sqrt{2} \pi, a_{\text {tricrit }}\right)$. In these cases we expect both the bimodal state and the trimodal state to be stable for this value of $a$. In panels (a) and (b), we set the fractionation to be equal in all groups, and we set the spacing between groups to be equal. As expected, we see that the trimodal state and the bimodal state are stable under perturbation.

In panel (c) we set $a=1.43 \pi-0.1<\sqrt{2} \pi<a_{\text {crit. }}$. As expected, we see that the bimodal state is unstable and the system goes in to trimodal state. Given the proximity of the clusters to $\pm \pi$, we have added black dashed lines that at $\pm \pi$, so that one can see that the difference between the final state and $\pm \pi$. In panel (d), we set $a=1.43 \pi+0.1>a_{\text {tricrit }}>\sqrt{2} \pi$. We also observe an expected result, as trimodality appears to be unstable and the system converges to a bimodal equilibrium, which is stable given that $a>\sqrt{2} \pi$.

In summary, we have a necessary condition for the stability of the trimodal equilibrium. Although, this condition is only necessary for stability, not sufficient, numerical experiments seems to point to it being an accurate threshold in examples we have considered. Also, theory and numerical experiments demonstrate that multistability of different multimodal equilibria is possible over parameter space. The theory for the stability of higher modes we leave for future work.


Figure A.1. Numerical experiments testing the threshold for trimodality. Panel (a): parameter value is $a=a_{\text {tricrit }}+0.1$, and the trimodal state appears to be unstable (as expected). Panel (b): parameter va;ie is $a=a_{\text {tricrit }}-0.1$, and the trimodal state appears to be stable (as expected). Both panels use the example interaction function from Eq. (2.11), and both use equal fractionation ( $x=y=z=1 / 3$ ) and equal spacing between clusters ( $\psi_{1}=2 \pi / 3$ ) in initial conditions.

## A.2. Additional coupling functions

Figure A. 3 illustrates two additional coupling functions that we examined. We used a variant of the triangle wave (blue, solid) given by the equation

$$
f_{\mathrm{tri}}(u ; c)=\left\{\begin{array}{ll}
\frac{2 u}{c} & |u|<c  \tag{A.14}\\
\frac{2 u}{c-\pi}-\operatorname{sign}(u)\left(\frac{2 \pi}{c-\pi}\right) & c \leq|u| \leq \pi
\end{array},\right.
$$



Figure A.2. Numerical experiments testing bistability. Panel (a) and (b): we set $a=1.43 \pi \in\left(\sqrt{2} \pi, a_{\text {tricrit }}\right)$ and both the bimodal state and the trimodal state are stable (as predicted). Panel (c): we set $a=1.43 \pi-0.1<\sqrt{2} \pi<a_{\text {crit }}$ and we see that the bimodal state is unstable (we have added black dashed lines so that one can see that the clusters away from the origin are not at $\pm \pi$ ). Panel (d): we set $a=1.43 \pi+0.1>a_{\text {tricrit }}>\sqrt{2} \pi$ and the trimodal state is unstable (as predicted). In all panels $N=300$ and the initial conditions are equally spaced and have equal fractionation with a random perturbation to all the phases of the oscillators.
assuming that $0<c<\pi$, and an antisymmetrized variant of the von Mises distribution (red curves) given by

$$
\begin{equation*}
f_{\mathrm{vM}}(u ; \mu, \kappa)=\sin (u-\mu) \frac{e^{\kappa \cos (u-\mu)}}{2 \pi I_{0}(\kappa)} . \tag{A.15}
\end{equation*}
$$



Figure A.3. Additional interaction functions. Solid blue curve: triangle wave from Eq. A.14; solid red curve: antisymmetrized variant of the von Mises distribution from Eq. A.15) with $\kappa<0$; dashed red curve: antisymmetrized variant of the von Mises distribution from Eq. (A.15) with $\kappa>0$. Panels (a) and (b) of Fig. A. 4 use the triangle wave. Panels (c) and (d) use the antisymmetrized von Mises function, with positive $\kappa$ (dashed red) in panel (c) and negative $\kappa$ (solid red) in panel (d). We note that for $\kappa>0$ the slope at the $\pm \pi$ is never steeper when compared to the origin and for $\kappa<0$ the slope at the origin is never steeper when compared to the slope at $\pm \pi$.

We numerically probe the stability of the bimodal equilibrium using these interaction functions in Fig. A.4. Here $N=100$, the oscillators' frequencies are drawn from a distribution $\mathcal{N}(0,100)$, the phase perturbation, $\xi_{i}$, is drawn from the distribution $\mathcal{N}(0,0.01)$ and we set $K=-1000$. In panels (a) and (b) we take the triangle wave defined in Eq. A.14) and set $c=3 \pi / 4$; this gives a stable fractionation threshold $1 / 4<x<3 / 4$. We test that threshold


Figure A.4. Numerical experiments using additional interaction functions. We test the stability of the bimodal equilibria for alternative coupling functions shown in Fig. A.3. (a) Triangle wave coupling with initial fractionation in predicted stable range. (b) Triangle wave coupling with initial fractionation outside predicted stable range. (c) Von Mises coupling with $\kappa>0$ (expected to be unstable). (d) Von Mises coupling with $\kappa>0$ (expected to be stable). In all panels $N=100$ and oscillators' natural frequencies are drawn from the distribution $\mathcal{N}(0,100)$. Initial phases are bimodally distributed with modes at 0 and $\pi$, with perturbations $\xi_{i}, i=1, \ldots, N$, are drawn from $\mathcal{N}(0,0.01)$.
numerically by setting $x_{\text {initial }}=7 / 10<3 / 4$ in panel (a) and $x_{\text {initial }}=8 / 10>3 / 4$ in panel (b).
As expected, we see that the fractionation is stable in panel (a) and is unstable in panel (b).
In panels (c) and (d) we use the antisymmetrized von-Mises function from Eq. A.15) with $\mu=0$ and $x_{\text {initial }}=1 / 2$. In panel (c) we set $\kappa=10$, and, as expected, we see that the bimodal equilibrium appears unstable; this is because there does not exist a range of $x$ such that Eq. (10) can be satisfied given that the slope at the origin is far steeper than the slope at the $\pm \pi$. We note that in (c) the system appears to tend to the incoherent state. In panel (d) we set $\kappa=-10$ and


Figure A.5. Basins of attraction. We plot the fraction of uniform random initial conditions that end up in bimodal (blue circles), trimodal (orange asterisks), or higher order multimodal (purple xs) states for the concrete system examined given by Eq. 2.11. Here $N=100, K=-10000$ and oscillators' natural frequencies are drawn from the distribution $\mathcal{N}(0,100)$. We performed 100 unique simulations for each value of $a$. Final states (presumed equilibria) were identified automatically via $k$-means clustering. Thresholds given by Eqns. (2.12) and (A.13) for stability of bimodality and the antiphase state are given by the solid black line and the dot-dashed green line, respectively. The threshold for the necessary condition for stability of the trimodal state, see Eq. A.12), is given by the vertical dashed magenta line.
observe that the bimodal state appears to be stable under perturbation, which is expected given that the slope at the $\pm \pi$ is steeper when compared to the origin.

## A.3. Basins of attraction for multimodal states

We have conducted some preliminary numerical exploration of the sizes of basins of attraction for various equilibria for the example interaction function given in Eq. (2.11). We simulated the system one hundred times with initial phases chosen independently at random from the uniform distribution over the circle, i.e. $\mathcal{U}(-\pi, \pi]$, and evaluated the fraction of the time that the system converged to each distinct equilibrium state. Results are shown in Fig. A.5, with $N=100, K=-10000$, and oscillator natural frequencies drawn from the distribution $\mathcal{N}(0,100)$.

Fig. A.5also shows the stability thresholds described in Eqns. (2.12) (bimodal state), A.13) (antiphase state), and trimodal state (A.12), visualized by the solid black, and dot-dashed green, and magenta vertical lines respectively. In order to classify the observed equilibria, we use a $k$-means algorithm on the unit circle, with the number of clusters, $k$, being decided by the gap statistic. We say that a equilibrium state is bimodal if $k=2$, trimodal if $k=3$, and so on.

We note that the results are consistent with our analysis in that the probability of a configuration is always zero in ranges of $a$ where it is excluded. Although, we have not analyzed equilibria with more than three modes, we observe that such modes are unlikely to be observed for most values of $a$, and thus have apparently small basins of attraction.

Given that this experiment was conducted with heterogeneous oscillators, this lends plausibility to the idea that the system will end up in a multimodal state for sufficiently large coupling. More formal analysis of the basin size of the bimodal and trimodal state are left as possible extensions of the model.


Figure A.6. Critical coupling strength. We perform numerical experiments to demonstrate the existence of a critical coupling strength for our system and evaluate its dependence on parameter $a$ using the interaction function defined in Eq. 2.11). Here $N=100$, the natural frequency distribution is given by $\mathcal{N}\left(0, \sigma^{2}\right)$, and the initial phase distribution is $\rho(\theta)=0.5 \delta(\theta)+0.5 \delta\left(\theta-\psi_{0}\right)$, where $\psi_{0}$ is the predicted phase separation given by the stable fixed points of Eq. (11). Here, each curve represents a different value of $a$ (values indicated in legend). As in the standard Kuramoto model, the critical coupling strength is dependant on the size of the standard deviation of the distribution, but unlike the standard Kuramoto model, it appears to also depend on $a$, which sets the shape of the interaction function.

## A.4. Critical coupling strength

In the standard Kuramoto model with attractive coupling, there exists a critical coupling strength $K_{c}$ at which the system bifurcates from an incoherent state to the ordered state. To look for $K$ dependence in the system detailed in Eq. 2.11), we examine the simplest cases of $N=2$ and $N=3$, and also conduct several numerical experiments with results shown in Fig. A.6, though we leave more thorough exploration for future work.

Figure A. 6 shows how order varies as we increase coupling strength among nonidentical oscillators with the concrete interaction function used in Eq. 2.11. Here, we set $N=100$ and draw the frequencies from the distribution $\mathcal{N}\left(0, \sigma^{2}\right)$. From here, we vary the quantity $K / \sigma$ so that $\log _{10}(K / \sigma)$ runs from -2 to 4 . Each curves shown above represents the result of an experiment for a given value of $a$. Here, the order parameter is defined as follows:

$$
\begin{equation*}
R=\max \left\{\left|\sum_{j} \frac{e^{2 i \theta_{j}}}{N}\right|,\left|\sum_{j} \frac{e^{3 i \theta_{j}}}{N}\right|,\left|\sum_{j} \frac{e^{\frac{2 \pi}{a} i \theta_{j}}}{N}\right|\right\} \tag{A.16}
\end{equation*}
$$

Defining the order parameter in this fashion sets the value of the order parameter to be 1 whenever the final configuration is bimodal or an equally spaced trimodal solution. Just as in the standard Kuramoto model, if the coupling strength $K$ is not sufficiently large in magnitude, the system goes to the incoherent state due to intrinsic oscillator heterogeneity. We observe that the critical coupling strength appears to be proportional to the standard deviation of the frequency distribution, similar to the result in the standard Kuramoto analysis, but we point out that the critical coupling strength $K_{c}$ also appears to have dependence on the value of $a$. We believe that some insight into this dependence can be gained from examining the simple $N=2$ and $N=3$ cases, though more rigorous analysis is left for future work.

For $N=2$, the system reduces to

$$
\begin{equation*}
\dot{\psi}=\Delta \omega-K f(\psi) \tag{A.17}
\end{equation*}
$$

where $\Delta \omega=\omega_{2}-\omega_{1}$. Setting $\dot{\psi}=0$, we find that a fixed point $\psi_{0}$ must satisfy the equation:

$$
\begin{equation*}
\frac{\Delta \omega}{K}=f\left(\psi_{0}\right) \tag{A.18}
\end{equation*}
$$

Note, this fixed point does not always exist, but if the coupling function $f$ has zeros, a fixed point must arise as $|K| \rightarrow \infty$.

Even without explicitly defining $\psi_{0}$, we can observe scaling dependencies for the critical coupling strength $K_{c}$, which is defined such that

$$
\begin{equation*}
f\left(\psi_{\max }\right)=\frac{\Delta \omega}{K_{c}} \tag{A.19}
\end{equation*}
$$

where $\psi_{\max } \in(-\pi, \pi]$ is the value such that $f\left(\psi_{\max }\right)=\max f(\psi)$ (the $\arg \max$ ). We observe that $K_{c} \propto \Delta \omega$, which is expected if $K_{c} \propto \sigma$ as in the standard Kuramoto model (since for two oscillators $\sigma \propto \Delta \omega)$ and is observed in our numerical experiments even for $N \gg 2$.

We also observe that $K_{c}$ scales with the maximum value of the interaction function $f$, which in our numerical experiments depends on the parameter $a$. Similar dependence is also evident if we consider the $N=3$ case.

For $N=3$, we take the natural frequencies (without loss of generality) to be $0,-\sigma / 3, \sigma / 3$ respectively. As before, we convert to difference coordinates $\psi_{1}=\theta_{2}-\theta_{1}$ and $\psi_{2}=\theta_{3}-\theta_{2}$, and arrive at two conditions for existence of equilibria:

$$
\begin{equation*}
\frac{\sigma}{K}=f\left(\psi_{2}-\psi_{1}\right)-f\left(\psi_{2}\right)-2 f\left(\psi_{1}\right) \tag{A.20}
\end{equation*}
$$

$$
\begin{equation*}
\frac{\sigma}{K}=f\left(\psi_{2}-\psi_{1}\right)+2 f\left(\psi_{2}\right)+f\left(\psi_{1}\right) \tag{A.21}
\end{equation*}
$$

which simplify to

$$
\begin{equation*}
\frac{\sigma}{K}=f\left(\psi_{2}-\psi_{1}\right)+f\left(\psi_{2}\right) \tag{A.22}
\end{equation*}
$$

$$
\begin{equation*}
f\left(\psi_{1}\right)=-f\left(\psi_{2}\right) \tag{A.23}
\end{equation*}
$$

Hence, a necessary condition $K$ must satisfy for the existence of equilibria is

$$
\begin{equation*}
\frac{\sigma}{K} \leq 2 f\left(\psi_{\max }\right) \tag{A.24}
\end{equation*}
$$

So, just as in the $N=2$ case, we see that the critical coupling strength $K_{c}$ is proportional to the oscillator heterogeneity $\sigma$ and inversely proportional to the maximum of the interaction function $f$.

We hypothesize that similar scaling laws hold for $N \gg 1$, and find that such a hypothesis is consistent with data from numerical experiments shown in Fig. A. 6 .

## APPENDIX B

## Modeling advertising competition: appendix

## B.1. Case with sigmoidal advertising payoff

In the Section 3.3, we carried out our analysis assuming that the quantity demanded at zero price, $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$, took the form of a piecewise linear function given by Eq. (6). Here, we assume $Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)$ is given by a sigmoid:

$$
\begin{equation*}
Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)=\frac{\Delta Q_{\mathrm{ad}}}{2}\left\{\tanh \left[\frac{a_{i}-\bar{a}}{\lambda}\right]+1\right\}+Q_{\mathrm{min}} \tag{B.1}
\end{equation*}
$$

As in Section 3.3, we first consider the case where production and advertising costs scale linearly, i.e, $\mu=\nu=1$. We substitute Eqns. (9), (10), and (B.1) into (8), giving

$$
\tau \frac{d a_{i}}{d t}=\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{8 \lambda k_{\mathrm{P}}}\left[Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{Q}} k_{\mathrm{P}}\right] \operatorname{sech}^{2}\left\{\frac{a_{i}-\bar{a}}{\lambda}\right\}-k_{\mathrm{a}}
$$

$$
\begin{equation*}
=\mathrm{B}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{a}} \tag{B.2}
\end{equation*}
$$

where

$$
\begin{equation*}
\mathrm{B}\left(a_{i} \mid \vec{a}\right)=\frac{N-1}{N} \frac{\Delta Q_{\mathrm{ad}}}{8 \lambda k_{\mathrm{P}}}\left[Q_{\mathrm{free}}\left(a_{i} \mid \vec{a}\right)-k_{\mathrm{Q}} k_{\mathrm{P}}\right] \operatorname{sech}^{2}\left\{\frac{a_{i}-\bar{a}}{\lambda}\right\} . \tag{B.3}
\end{equation*}
$$

Similarly to Section 3.3, $\tau$ sets the time scale for equilibration, and we set $\tau=1$ without loss of generality.

Fixed points occur when $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ intersects $k_{\mathrm{a}}$. Figure 4 shows how firms would change their advertising given mean $\bar{a}$, according to (B.2). Of course, since $\bar{a}$ changes in time due to the distribution of advertising $\vec{a}$ evolving in time, $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ is variable in time as well. For a given $\vec{a}^{1}$ there are three fixed points for the system: two from the intersection of $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$ with $k_{\mathrm{a}}$ and one implied at zero as negative advertising is not physically meaningful in this problem. One can see from Fig. 4 that given $\bar{a}$ an equilibrium would involve a distribution of companies at zero and at $a=a_{\text {name }}>0$. Furthermore, the equilibrium must be distributed such that $\bar{a}=0 x+a_{\text {name }}(1-x)$, where $0<x<1$ gives the fraction of generic firms.

## B.1.1. Existence of the differentiated state

In Section 3.3, for the piecewise-linear advertising payoff function, we argued that the differentiated state may only exist when $k_{\mathrm{a}}$ intersects $\mathrm{B}\left(a_{i} \mid \vec{a}\right)$. This holds in the current sigmoidal case as well, so the following is a necessary condition for the existence of a differentiated state:

$$
\begin{equation*}
k_{\mathrm{a}}<k^{c r i t}\left(\lambda, N, Q_{\min }, \Delta Q_{\mathrm{ad}}, k_{\mathrm{Q}}, k_{\mathrm{P}}\right)=\max _{a} \mathrm{~B}(a \mid \vec{a}) \tag{B.4}
\end{equation*}
$$

Here $\max _{a} \mathrm{~B}(a \mid \vec{a})$ can be expressed algebraically as

$$
\max _{a} \mathrm{~B}(a \mid \vec{a})=\frac{4(N-1)}{27 N \lambda \Delta Q_{\mathrm{ad}} k_{\mathrm{P}}}\left(Q_{i}^{*}(\bar{a}) \sqrt{\Delta Q_{\mathrm{ad}}^{2}+\left[\min Q_{i}^{*}\right] \Delta Q_{\mathrm{ad}}+\left[\min Q_{i}^{*}\right]^{2}}+\frac{1}{2} \Delta Q_{\mathrm{ad}}^{2}\right.
$$

$$
\begin{equation*}
\left.-\left[\min Q_{i}^{*}\right] \Delta Q_{\mathrm{ad}}-\left[\min Q_{i}^{*}\right]^{2}\right)\left(Q_{i}^{*}(\bar{a})+\frac{1}{2} \sqrt{\Delta Q_{\mathrm{ad}}^{2}+\left[\min Q_{i}^{*}\right] \Delta Q_{\mathrm{ad}}+\left[\min Q_{i}^{*}\right]^{2}}\right) \tag{B.5}
\end{equation*}
$$

[^7]where $\min Q_{i}^{*}=Q_{\text {min }}-k_{\mathrm{Q}} k_{\mathrm{P}}$, and $Q_{i}^{*}(\bar{a})=\Delta Q_{\mathrm{ad}} / 2+Q_{\text {min }}-k_{\mathrm{Q}} k_{\mathrm{P}}$. This condition appears to be sufficient as well as necessary (see Section B.1.3).

## B.1.2. Stability of the differentiated state

We investigate the stability of the differentiated state by examining the functional form of $d a_{i} / d t$. After substituting Eqns. (B.1) and (10) into Eq. (B.2) we can reduce $d a_{i} / d t$ to the following form :

$$
\begin{equation*}
\frac{d a_{i}}{d t}=-\alpha \tanh ^{3}\left(\frac{a_{i}-\bar{a}}{\lambda}\right)-\beta \tanh ^{2}\left(\frac{a_{i}-\bar{a}}{\lambda}\right)+\alpha \tanh \left(\frac{a_{i}-\bar{a}}{\lambda}\right)+\beta-k_{\mathrm{a}} \tag{B.6}
\end{equation*}
$$

where $\alpha=(N-1) \Delta Q_{\mathrm{ad}}^{2} /\left(8 \lambda N k_{\mathrm{P}}\right)$ and $\beta=(N-1) \Delta Q_{\mathrm{ad}}\left(0.5 \Delta Q_{\mathrm{ad}}-k_{\mathrm{Q}} k_{\mathrm{P}}\right) /\left(8 \lambda N k_{\mathrm{P}}\right)$. Factoring out $\alpha$ from (B.6) gives

$$
\begin{equation*}
\frac{d a_{i}}{d t}=\alpha\left(-u^{3}-\gamma u^{2}+u+\gamma-k_{a}^{*}\right) \tag{B.7}
\end{equation*}
$$

where $u=u\left(a_{i} \mid \vec{a}\right)=\tanh \left[\left(a_{i}-\bar{a}\right) / \lambda\right], \gamma=\beta / \alpha$, and $k_{a}^{*}=k_{\mathrm{a}} / \alpha$. Because we are only interested in fixed points, we can ignore the $\alpha$ outside of the parenthesis as this just sets a time scale for equilibration.

We define $g\left[u\left(a_{i} \mid \vec{a}\right)\right]=-u^{3}-\gamma u^{2}+u+\gamma-k_{a}^{*}$. To assess linear stability, we show that the eigenvalues of the Jacobian matrix of our dynamical system are negative in a similar fashion to the method used in reference [33]. One can show that $\partial / \partial a_{j}\left(d a_{i} / d t\right)$ scales like $1 / N$ when $i \neq j$ and hence that these terms can be ignored when determining stability for $N \rightarrow \infty$. Now, consider the main diagonal of the Jacobian of our system as $N \rightarrow \infty$ :

$$
\begin{equation*}
\frac{\partial}{\partial a_{i}}\left(\frac{d a_{i}}{d t}\right)=g^{\prime}\left[u\left(a_{i} \mid \vec{a}\right)\right] \frac{\partial u}{\partial a_{i}} \tag{B.8}
\end{equation*}
$$

where

$$
\begin{equation*}
\frac{\partial u}{\partial a_{i}}=\operatorname{sech}^{2}\left[\frac{a_{i}-\bar{a}}{\lambda}\right]>0 . \tag{B.9}
\end{equation*}
$$

Hence, the sign of $\partial / \partial a_{i}\left(d a_{i} / d t\right)$ is determined by the sign of $g^{\prime}\left[u\left(a_{i} \mid \vec{a}\right)\right]$. Since stability must alternate in one dimensional flows, the only stable nonzero fixed point must be the far right intersection between B and $k_{\mathrm{a}}$ (see Fig. 4) located at $a=a_{\text {name }}$. We can write $g^{\prime}\left[u\left(a_{\text {name }} \mid \vec{a}\right)\right]$ as

$$
\begin{equation*}
g^{\prime}\left[u\left(a_{\text {name }} \mid \vec{a}\right)\right]=-\frac{1}{12} w^{2 / 3}-\frac{1}{3} v-\frac{4}{3} \frac{v^{2}}{w^{2 / 3}}, \tag{B.10}
\end{equation*}
$$

where
(B.11) $\quad v=\gamma^{2}+3$,
(B.12) $\quad w=72 \gamma-108 k_{\mathrm{a}}^{*}-8 \gamma^{3}+12 \sqrt{12 k_{\mathrm{a}}^{*} \gamma^{3}-12 \gamma^{4}+81\left(k_{\mathrm{a}}^{*}\right)^{2}-108 k_{\mathrm{a}}^{*} \gamma+24 \gamma^{2}-12}$.

It can be shown that the expression in (B.10) has a maximum of zero which occurs at one point, specifically,

$$
\begin{equation*}
k_{\mathrm{a}}^{*}=\frac{2}{27} \gamma\left(9-\gamma^{2}\right)+\left(\gamma^{2}+3\right)^{\frac{3}{2}}, \tag{B.13}
\end{equation*}
$$

or, expressed in terms of the original parameters,

$$
k_{\mathrm{a}}=\max _{a} \mathrm{~B}(a \mid \vec{a}) .
$$

This also happens to be the threshold for existence of the differentiated (bimodal) state. So, as long as the bimodal state exists, the Jacobian has negative values along its main diagonal and, therefore, it is a stable solution as $N \rightarrow \infty$.

## B.1.3. Stability of the undifferentiated state

To examine stability of the undifferentiated state, we set $a_{i}=\bar{a}=0$, for all $i$. We again assume $N \rightarrow \infty$, and thus any change in $\bar{a}$ is negligible under single firm perturbations. If there is an interval of positive perturbations neighboring zero for which $d a_{i} / d t<0$, then the undifferentiated state is stable under single firm perturbations for sufficiently small perturbations. If $a_{i}=\bar{a}=0$ then $u(0 \mid \overrightarrow{0})=\tanh \left[\left(a_{i}-\bar{a}\right) / \lambda\right]=\tanh (0)=0$. Hence, substituting $u=0$ in to Eq. (B.7) gives

$$
\begin{equation*}
\frac{d a_{i}}{d t}=\alpha\left(-u^{3}-\gamma u^{2}+u+\gamma-k_{\mathrm{a}}^{*}\right)=\alpha\left(\gamma-k_{\mathrm{a}}^{*}\right)<0 . \tag{B.14}
\end{equation*}
$$

By continuity of $d a_{i} / d t$, there must exist an interval $(0, \epsilon)$ such that $d a_{i} / d t<0$. Since $\alpha>$ 0 , the statement $k_{a}^{*}>\gamma$ is sufficient to ensure stability of the undifferentiated state. Taken together with Eq. (B.13) above, we observe that a region of bistability exists where both the undifferentiated state and the differentiated state are stable:

$$
\begin{equation*}
\gamma<k_{\mathrm{a}}^{*}<\frac{2}{27} \gamma\left(9-\gamma^{2}\right)+\left(\gamma^{2}+3\right)^{\frac{3}{2}} . \tag{B.15}
\end{equation*}
$$

In the original parameters with $N \rightarrow \infty$, Eq. (B.15) can be written as

$$
\begin{equation*}
\mathrm{B}(\bar{a} \mid \vec{a})=\frac{\Delta Q_{\mathrm{ad}}\left[Q_{\min }+\Delta Q_{\mathrm{ad}} / 2-k_{\mathrm{Q}} k_{\mathrm{P}}\right]}{4 \lambda k_{\mathrm{P}}}<k_{\mathrm{a}}<\max _{a} \mathrm{~B}(a \mid \vec{a}) . \tag{B.16}
\end{equation*}
$$

In more intuitive terms this can be expressed as

$$
\begin{equation*}
\left.\frac{\partial \pi_{i}}{\partial a_{i}}\right|_{a_{i}=\bar{a}}<0<\max _{a} \frac{\partial \pi_{i}}{\partial a_{i}} \tag{B.17}
\end{equation*}
$$

This result is identical to the one given in Eq. 3.20 .

## B.2. Welfare analysis

The existence and persistence of name brand goods is of interest in this section, but it is also of interest to determine if advertising is beneficial to society at large, and if advertising generates more profit for the industry as a whole. We address these questions by considering economic welfare, similarly to the approach taken in reference [112]. We define total welfare to be the sum of consumer welfare-the difference between the maximum price an individual is willing to pay for a good or service and the price they actually pay-and the total profit generated by the market. We use post-advertising preferences when calculating consumer welfare, effectively interpreting consumer demand after advertising to reveal the true preferences. Further consideration of this point is merited to better understand how advertising should factor into consumer welfare, but we leave that for future work.

Figure B.1 (a) displays the maximum possible consumer welfare (optimizing over all fractions generic in differentiated equilibria) for a given marginal advertising cost $k_{\mathrm{a}}$ (blue, solid) and the total welfare for the undifferentiated state (red, dashed). The black line indicates when $k_{\mathrm{a}}=\max \mathrm{B}\left(a_{i} \vec{a}\right)$ and thus, past that point the differentiated state ceases to be stable. The maximum possible consumer welfare for the differentiated state is greater than the undifferentiated state across all $k_{\mathrm{a}}$.


Figure B.1. Optimized consumer welfare, profit and total welfare given marginal advertising costs. In (a) we display the optimal total consumer welfare generated by the market across all possible differentiated equilibria for a given marginal advertising cost $k_{\mathrm{a}}$ (blue, solid) and the total consumer welfare generated by the undifferentiated state (red, dashed). The black line in all three panels indicates when $k_{\mathrm{a}}=\max \mathrm{B}\left(a_{i} \mid \vec{a}\right)$ and thus, past that point the differentiated state ceases to be stable. In (b) we display the optimal total profit generated by the market across all possible differentiated equilibria for a given marginal advertising $\operatorname{cost} k_{\mathrm{a}}$ (blue) and the total profit generated by the undifferentiated state (red, dashed). In (c) we display the optimal total welfare generated by the market across all possible differentiated equilibria for a given marginal advertising cost $k_{\mathrm{a}}$ (blue, solid) and the total welfare generated by the undifferentiated state (red, dashed). We set $N=100, Q_{\text {min }}=30, \Delta Q_{\text {ad }}=5$, and $\lambda=\mu=\nu=k_{\mathrm{Q}}=k_{\mathrm{P}}=1$, with $k_{\mathrm{a}}$ ranging over a range 0 to 45 .

Figure B.1 (b) shows the maximum possible total profit in the market (optimized over all stable fractions generic in differentiated equilibria) for a given marginal advertising cost $k_{\mathrm{a}}$ (blue, solid) and the total profit for the undifferentiated state (red, dashed). The differentiated state is more lucrative (at the industry-wide level) when marginal advertising costs are low, and the undifferentiated case is more lucrative for high marginal advertising costs.

Figure B.1(c) shows the maximum possible total welfare in the market. Note this is not simply the sum of the numbers in panels (a) and (b), as they may correspond to different stable
fractionations $x$. The differentiated state (blue, solid) is more beneficial when considering the interests of consumers and producers when marginal advertising costs are low, and the undifferentiated case (red, dashed) is more beneficial when marginal advertising costs are high. In this numerical experiment we set $N=100, Q_{\min }=30, \Delta Q_{\mathrm{ad}}=5$, and $\lambda=\mu=\nu=k_{\mathrm{Q}}=k_{\mathrm{P}}=1$, with $k_{\mathrm{a}}$ ranging over a range 0 to 45 .

According to Figure B.1, the differentiated state produces more total welfare than the undifferentiated state when the marginal cost of advertising is low. Conversely, the undifferentiated state generates greater total welfare when the marginal cost is relatively high. We leave analytical exploration of the effect of advertising on welfare and profit as a possible extension to this work.

## B.3. Comparing model predictions and data

In Section 3.3.6, we describe statistical tests used to assess how our model's predictions compare to real-world data. We use the Kolmogorov-Smirnov test (KS test) to determine if the sample data and the model predictions are likely to have come from same distribution. We use Hartigan's Dip Test to see if we could reject unimodality of the sample data. We display the resulting distributions from the application of these statistical tests in Fig. B.2.

Hartigan's Dip Test appears to be "excessively conservative and insensitive" at small sample sizes according to [33]. Clifton et al. used an altered test, the Least Unimodal Unimodal (LUU), to develop a more sensitive bootstrapped dip statistic to test for unimodality.

According to the LUU Test $53 \%$ of price distributions are inconsistent with unimodality at a significance level 0.05 compared to $46 \%$ of price distributions inconsistent with unimodality when applying Hartigan's Dip Test. For $58 \%$ of our model fits to the top 500 products, we fail


Figure B.2. Histograms of Fitting Statisitics. Distributions of $p$ values from tests for unimodality ( $\mathrm{a}-\mathrm{b}$ ) and consistency of price data with model predictions (c). The red dashed line indicates the significance level of .05 . Top and middle row: LUU and Hartigan's Dip Test (respectively). Rejection (low $p$-value) means the price distribution is not consistent with unimodal null hypothesis. Bottom row: KS test. Rejection (low $p$-value) means the price distribution is not consistent with same-distribution null hypothesis (i.e., model and data not from same distribution).
to reject the null hypothesis at a significance level of 0.05 : the data and the model prediction may come from the same distribution.

## B.4. Seeding the minimization algorithm

To seed the initial choice of parameters for fitting, we attempt to match the modes of the model distribution, $f(p)$, to the modes of the empirical distribution $g(p)$ inferred from the price data. Since our model only produces unimodal or bimodal distributions, we produce kernel density estimates (KDEs) of the data with increasing bandwidths until we detect at most two peaks at prices $P_{1}$ and $P_{2}$, where $P_{2}>P_{1}$. Numerically, we accomplish this using KDE bandwidths increasing progressively from 0.1 to 10 in increments of 0.1 . We apply a threshold to peak-detection such that only peaks of height greater than 0.1 are counted.

If the KDE is bimodal then we wish to choose seed values such that the two modes of the differentiated state (the generic brand price and name brand price) match $P_{1}$ and $P_{2}$. We cannot easily predict the equilibrium advertising value for name brands $a_{\text {name }}$, so $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ (and hence $P_{i}^{*}$ ) cannot be set exactly to match $P_{1}$ and $P_{2}$. Instead, we make the assumption that the average advertising level $\bar{a}$ is far away enough from 0 and $a_{\text {name }}$ that $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)$ is near its saturation values at $\pm \infty$ for generics and name brands-meaning that name brands receive the maximum demand increase due to advertising and generics receive no increase in demand due to advertising.

We minimize the magnitude of profit's derivative with respect to price when $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)=$ $\Delta Q_{\mathrm{ad}}+Q_{\text {min }}$ and $Q_{\text {free }}\left(a_{i} \mid \vec{a}\right)=Q_{\text {min }}$. This leads to the objective function

$$
\begin{equation*}
R\left(\Delta Q_{\mathrm{ad}}, k_{\mathrm{P}}, k_{\mathrm{Q}}, Q_{\min }, \mu\right)=\left|\frac{\partial \pi_{i}}{\partial P_{i}}\right|_{P_{i}=P_{1}, a_{i} \rightarrow-\infty}+\left|\frac{\partial \pi_{i}}{\partial P_{i}}\right|_{P_{i}=P_{2}, a_{i} \rightarrow \infty} . \tag{B.18}
\end{equation*}
$$

$R\left(\Delta Q_{\mathrm{ad}}, k_{\mathrm{P}}, k_{\mathrm{Q}}, Q_{\mathrm{min}}, \mu\right)$ is minimized when $\partial \pi_{i} /\left.\partial P_{i}\right|_{P_{i}=P_{1}, a_{i} \rightarrow-\infty}$ and $\partial \pi_{i} /\left.\partial P_{i}\right|_{P_{i}=P_{2}, a_{i} \rightarrow \infty}$ are both zero. When $R\left(\Delta Q_{\mathrm{ad}}, k_{\mathrm{P}}, k_{\mathrm{Q}}, Q_{\min }, \mu\right)$ is minimized the optimal price for name brands is $P=P_{2}$ (in the limit $a \rightarrow+\infty$ ) and the optimal price for generics is $P=P_{1}$ (in the limit $a \rightarrow-\infty)$. We minimize $R$ using the Nelder-Mead algorithm [110] and the resulting values for $\Delta Q_{\mathrm{ad}}, k_{\mathrm{P}}, k_{\mathrm{Q}}, Q_{\text {min }}$, and $\mu$ are used as seeds for fitting our model to the data. We set $\Delta Q_{\mathrm{ad}}=1.1 \times 10^{7}, k_{\mathrm{P}}=4.4 \times 10^{6}, k_{\mathrm{Q}}=100, Q_{\min }=2.5 \times 10^{7}$, and $\mu=0.4$ to seed the $R$-minimization problem (where these values are derived from a more complete exploration of the parameter space in the case of a single product from the data set).

The initial proportion of generic firms $x$ is set by measuring the percentage of price entries that are less than the quantity $P_{1}+w$, where $w$ is the peak width at half-prominence for the peak
at $P_{1}$. The initial advertising distribution is given by $x \delta\left(a_{i}-1\right)+(1-x) \delta\left(a_{i}-100\right)$, where $\delta$ represents the Dirac delta function.

If the KDE is unimodal (i.e., $P_{1}=P_{2}$ ), the we take a different approach. There is a critical cost of advertising $k_{\mathrm{a}}^{\text {crit }}$ defined by $k_{\mathrm{a}}^{\text {crit }}=\nu^{-1}\left(a_{\max }\right)^{1-\nu} \mathrm{B}\left(a_{\max } \mid \vec{a}\right)$, where for $k_{\mathrm{a}}>k_{\mathrm{a}}^{\text {crit }}$, $d a_{i} / d t<0$ for all $i$ and no bimodal state exists, since there will be no intersection between $B$ and the advertising cost curve $C_{a}^{\prime}(a)=\nu k_{\mathrm{a}} a^{\nu-1}$ (we assume $0<\nu \leq 1$ ). Here $a_{\text {max }}$ is defined as the advertising value such that $\mathrm{B}\left(a_{\max } \mid \vec{a}\right)=\max _{a} \mathrm{~B}\left(a_{i} \mid \vec{a}\right)$, i.e., the $\arg \max$ of the function $B$.

Note that this critical value $k_{\mathrm{a}}^{\text {crit }}$ is state-dependent. We set the initial advertising distribution for this undifferentiated case to $\delta\left(a_{i}-100\right)$, so $a_{i}=\bar{a}=100$ for $i=1, \ldots, N$. Given this choice, we found that setting $k_{\mathrm{a}}=\nu^{-1}\left(a^{*}\right)^{1-\nu} \mathrm{B}\left(a_{\text {max }} \mid \vec{a}\right)$ was sufficient to ensure that the system evolved to a unimodal price distribution, where $a^{*}$ is any sufficiently large value above $a_{\text {max }}$ (the threshold for "sufficiently large" can be calculated from model parameters). We used $a^{*}=300$ in our simulations.

Additionally, for the undifferentiated state, we redefine $R$ as

$$
\begin{equation*}
R\left(\Delta Q_{\mathrm{ad}}, k_{\mathrm{P}}, k_{\mathrm{Q}}, Q_{\min }, \mu\right)=\left|\frac{\partial \pi_{i}}{\partial P_{i}}\right|_{P_{i}=P_{1}, a_{i}=\bar{a}=0} \tag{B.19}
\end{equation*}
$$

We then minimize $R$ to set the parameters so the price in the undifferentiated state matches $P_{1}$. We implement the Nelder-Mead algorithm with the same seeds as in the previously described differentiated case.

There are three parameters we don't explicitly fit, namely, $k_{\mathrm{a}}, \lambda$ and $\nu$ (two in the undifferentiated case, since $k_{\mathrm{a}}$ is specified). For these we use the same values employed to seed $R$; the values result from a more thorough exploration of the parameter space in the case of a single
product from the data set. They are: $k_{\mathrm{a}} \approx 9100, \lambda \approx 9.7$ and $\nu \approx 0.57$. We chose not to fit these parameters for every product because price distributions appeared to be insensitive to their values. Given the inherent limitations of our "toy" model, our goal was to demonstrate that agreement between model and data is plausible, not to exhaustively discover the best possible agreement.

## B.5. Data adjustment

On examining the data, it becomes apparent that there is bias present toward prices ending in certain digits. Prior research has highlighted this phenomenon in pricing of goods, known as psychological pricing [13-116]. Since our model does not take into account this type of bias, we develop a method to remove this bias from the price distribution.


Figure B.3. Raw and Smoothed Price Data. Example of the raw price data with spikes situated near certain ending digits (blue) and the smoothed, debiased data that retains the quantitative properties of the underlying distribution while removing the spikes (orange). Red arrows point to example spikes located at $\$ 2.99$, $\$ 3.49$, and $\$ 3.99$.

This bias towards certain ending digits creates spikes in each product's price distributions, as displayed in Fig. B.3 (blue). One might consider using a kernel density estimate (KDE) [117] or a filter to remove these spikes [ $\mathbf{1 1 8}]$. However, with a KDE approach, the amount of smoothing required to effectively remove these spikes significantly alters the underlying distribution. With the filtering approach, the "probability mass" of the spike is redistributed nonlocally across the entire distribution, again altering it in an undesirable way. We develop a new method that attempts to remove spikes while keeping the underlying distribution largely unchanged. An example applying our method to data can be seen in Fig. B.3(orange).

Details of our "debiasing" method are reported in an upcoming paper [119], however we also explain the idea briefly here. We begin by defining a metric, $B_{1}(p)$, for how biased a distribution is towards a price value $p$. We calculate $B_{1}(p)$ as follows:

$$
\begin{equation*}
B_{1}(p)=\frac{\rho_{0}(p)}{\rho_{1}(p)} \tag{B.20}
\end{equation*}
$$

where $\rho_{0}(p)$ is the value of the normalized histogram at $p$ and $\rho_{1}(p)$ is a kernel density estimate of $\rho_{0}$ using a Gaussian kernel with a narrow bandwidth (e.g., here we use $h=0.01$ ). Conceptually similar to generation of a KDE, we then replace the data point at each price $p_{i}$ with a Gaussian that has variance $s B_{1}\left(p_{i}\right)^{r}$. This generates a new price distribution $\rho_{2}(p)$, where the parameters $s>0$ and $r>0$ determine the natural amount of smoothing when there is no bias and how the smoothing scales with bias-linearly, sublinearly, or superlinearly-respectively.

When we apply this method, we set $r=0.5$ and set $s=\sigma(4 / 3 n)^{-0.2}$ based on Silverman's "rule of thumb" $[\mathbf{1 2 0}]$, with $n$ being the number of data entries and $\sigma$ defined as the standard
deviation of the data. We iterate this process with $B_{i}(p)$ being defined as

$$
\begin{equation*}
B_{i}(p)=\frac{\rho_{0}(p)}{\rho_{i}(p)} \tag{B.21}
\end{equation*}
$$

until we arrive at a distribution $\rho_{\infty}$, such that it is fixed under this iteration.
As the sample size goes to infinity, the method is equivalent to the following integral transform:

$$
\begin{equation*}
\rho_{i+1}(p)=\int_{-\infty}^{\infty} \rho_{0}\left(p^{\prime}\right) \mathcal{N}\left(p^{\prime}, s\left(B_{i}\left(p-p^{\prime}\right)\right)^{r}\right) d p^{\prime} \tag{B.22}
\end{equation*}
$$

Letting $i \rightarrow \infty$ yields the smoothed distribution

$$
\begin{equation*}
\rho_{\infty}(p)=\int_{-\infty}^{\infty} \rho_{0}\left(p^{\prime}\right) \mathcal{N}\left(p^{\prime}, s\left(B_{\infty}\left(p-p^{\prime}\right)\right)^{r}\right) d p^{\prime} \tag{B.23}
\end{equation*}
$$

The distribution $\rho_{\infty}$ is fixed under this integral transform. As mentioned above, tests of how effectively this method performs "bias" removal are reported in an upcoming manuscript [119]. We move forward assuming that this method works sufficiently well; this assumption is supported by qualitative visual examination of distributions before and after this "debiasing" procedure ${ }^{2}$

## B.6. Data Availability

The data that support the findings of this study are available from The Nielsen Company (US), LLC but restrictions apply to the availability of these data, which were used under license

[^8]for the current study, and so are not publicly available. Data are however generally available for scientific research with an institutional or individual subscription [98].

Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

## APPENDIX C

## Size dimorphism: appendix

## C.1. Sex ratios



Figure C.1. Possible sex ratios. The solid black curve shows the threshold for existence of the anisogamous state given by Eq. (C.1). The anisogamous equilibrium exists below the threshold (blue shaded region) and ceases to exist above the threshold (red shaded region). Here the fraction "male" refers to the fraction with small gametes.

For the bimodal gamete size distribution to exist, $s^{*}$ must be real-valued, and hence $\beta=$ $\alpha^{2} x^{2}-6 \alpha x+1$ in Eq. (4.8) must be positive. This holds when

$$
\begin{equation*}
x<\frac{3-2 \sqrt{2}}{\alpha} . \tag{C.1}
\end{equation*}
$$

Thus there is an implied range of stable sex ratios for a given value of $\alpha$. Figure C. 1 illustrates the relationship between the power law exponent $\alpha$ and the fraction of the population with
small gametes (the fraction "male") given by Eq. (C.1). As $\alpha$ increases in magnitude the range of possible fractionations decreases.

Interestingly, an approximate 1:1 sex ratio is not attainable for some "reasonable" exponents of $N_{g}$ (e.g., 1, 2, and 3, each of which would correspond to a distinct simple measure of gamete "size"). This is likely a result of our specific choices of $\varphi_{\mathrm{g}}$ and $N_{g}$, as well as the restricted nature of the model. When we modify gamete reproductive potential to depend on both relative and absolute gamete size, in numerical simulation we observe stable anisogamous states with arbitrary sex ratios. Also note that our model purposefully omits frequency-dependent selection effects that would likely drive sex ratios toward 1:1 (the reproductive potential of a single "male" gamete in a community of mostly "female" gametes would be much higher than in a community of mostly "male" gametes because the likelihood of fusion would be higher and the likelihood of zygote survival would be higher, i.e., Fisher's principle [121]).

## C.2. Numerical simulations

We test predictions of our model via numerical simulation. Figure C. 2 shows the evolution of a population from a state that is isogamous to a one that is anisogamous. The gametes (yellow) move along the landscape (blue) in the direction that increases their reproductive potential. For this simulation, we set $\alpha=1, N=100$, and $w=1 / 10$ with a final fraction of individuals producing small gametes $x=0.1$. The initial isogamous distribution was sampled from the uniform distribution $\mathcal{U}(1,3)$.


Figure C.2. Simulation of the system. Panels (a)-(d) show the evolution of the system from an isogamous state to an anisogamous state. Here, the blue curve shows the reproductive potential landscape given by Eq. (4.4), the red circles indicate gamete sizes, and the yellow crosses give positions of gametes along the reproductive potential landscape. Panel (a) captures the isogamous initial condition $\mathcal{U}(1,3)$. Panel (b) shows the individuals moving along the landscape in the direction that increases reproductive potential. Panel (c) shows the beginning of two groups forming. In Panel (d), the simulation has arrived at an anisogamous equilibrium, with gamete sizes converging to zero or $s^{*}$ as given by Eq. (4.8). The final fraction of organisms that produce small gametes is $x=0.1$. For this numerical experiment, we set $\alpha=1, N=100$, and $w=1 / 10$.

## C.3. Stability tests

## C.3.1. Linear stability

Section 4.4.2, we outlined a restricted stability test of the anisogamous equilibrium where only single-gamete perturbations were allowed. A more rigorous test of stability is difficult because the Dirac delta functions that comprise the equilibrium gamete size distribution are actually generalized functions and thus must be treated carefully when perturbed. One straightforward
way to avoid this difficulty is to look at the linear stability of the equilibrium for finite $N$, then take the limit as $N \rightarrow \infty$.

One can show that, for finite $N$, the off-diagonal elements of the Jacobian matrix take the form

$$
\begin{equation*}
J_{i j}=\left(s_{i}\right)^{-\alpha-1}\left[2 \frac{N-1}{N^{2}} \frac{w \operatorname{sgn}\left(s_{i}-\bar{s}\right)+}{\left(w+\left|s_{i}-\bar{s}\right|\right)^{3}}-\frac{1}{N} \frac{\alpha\left(w+\left|s_{i}-\bar{s}\right|\right)}{\left(w+\left|s_{i}-\bar{s}\right|\right)^{3}}\right] \tag{C.2}
\end{equation*}
$$

It follows that these off diagonal elements approach zero as $N \rightarrow \infty$.
The diagonal elements of the Jacobian matrix take the form

$$
\begin{align*}
J_{i i} & =\frac{1}{\left(w+\left|s_{i}-\bar{s}\right|\right)^{3}}\left\{\left[\frac{N-1}{N}\right]^{2}\left[s_{i}-\bar{s}-2 \operatorname{sgn}\left(s_{i}-\bar{s}\right)\left(w+\left|s_{i}-\bar{s}\right|\right)\right]\right. \\
& \left.+\frac{\alpha(1+\alpha)}{s_{i}^{2}}\left(w+\left|s_{i}-\bar{s}\right|\right)^{2}\left(s_{i}-\bar{s}+\left|s_{i}-\bar{s}\right|\right)-w\left[\frac{N-1}{N}\right]\left[\frac{w+\left|s_{i}-\bar{s}\right|}{s_{i}}\right]\right\} . \tag{C.3}
\end{align*}
$$

One can show that these are all negative when $s_{i}=s^{*}$ or $s_{i} \rightarrow 0^{+}$and $\bar{s}=(1-x) s^{*}$ as $N \rightarrow \infty$ given that Eq. C. 1 is satisfied.

Since off-diagonal elements become infinitesimal, the eigenvalues of the Jacobian matrix are determined by the diagonal elements as $N \rightarrow \infty$, and thus all eigenvalues are negative, implying linear stability of the anisogamous equilibrium.

## C.3.2. Stable size distributions

We perform two numerical experiments to test the stability of the anisogamous state. First, we perturb the large group from its equilibrium value $s^{*}$ given in Eq. (4.8) by an amount $\xi_{i}$ drawn from $\mathcal{N}\left(0, \delta^{2}\right)$. Figure C.3a displays the result of the perturbation. Gamete sizes that were perturbed return their equilibrium value. Second, we perturb the small group $\xi_{i}$ drawn from


Figure C.3. Numerical test of the stability of anisogamy. We test the stability of the anisogamous state. Panel (a) shows the large gamete group being perturbed and then returning to its equilibrium value, $s^{*}$. Panel (b) shows the small gamete group being perturbed and then returning to its equilibrium value zero. Panel (a) and (b) demonstrate the stability of the anisogamous equilibrium and are consistent with the asymptotic theory from Eqns (4.9) and (4.11). In both panels, we set $\alpha=1 / 3, N=1000$, and $w=2$, with an initial fractionation $x=1 / 2$.
$\mathcal{U}(0, \gamma){ }^{17}$ Similar to the first test, Figure C.3b demonstrates that the perturbed group returns to its equilibrium value. We set $\alpha=1 / 3, N=1000, w=2$, and $x=1 / 2$ in both simulations.

## C.4. Nonidentical individuals

Our results appear to be robust to the inclusion of natural variation among the simulated individuals. In various numerical experiments, we considered variations in the width of the sigmoidal gamete reproductive potential function $\varphi_{\mathrm{g}}$ (see Eq. 4.3), as well as in its mean, minimum, and maximum values. We also varied the multiplicative factor in the gamete production function (see Eq. (4.2)). In all cases, the equilibrium gamete size distribution remained

[^9]qualitatively the same as in the case with identical individuals: the only change was the appearance of some variation around the expected delta function peaks (primarily the peak at $s^{*}$ ) at equilibrium. See Figure C. 4 .


Figure C.4. Simulation with a heterogeneous population. We display the evolution from isogamy to anisogamy population for a heterogeneous population. The large gamete group widens out when adding noise to the width of the sigmoid in Eq. (4.3). The final fraction of organisms that produce small gametes is $x=0.1$. For this simulation, we set $\alpha=1, N=100$, and $w$ was sampled from the distribution $\mathcal{N}\left(1 / 10,1 / 50^{2}\right)$.

## C.5. Nonzero size for small gamete group

Because reproduction requires the transfer of some minimal amount of physical material, the number of gametes cannot realistically diverge as $s \rightarrow 0^{+}$. Our results, however, appear to be robust to the inclusion of a minimal viable gamete size. In simulation, we incorporated a minimal size by multiplying the individual reproductive potential by $e^{-k / s}$, where $k>0$. This eliminated the singularity at zero and generated a point $0<s_{\text {small }}^{*}<s^{*}$ such that reproductive potential is maximized. In such simulations, the resulting equilibrium distribution was $\rho(s)=$ $x \delta\left(s-s_{\text {small }}^{*}\right)+(1-x) \delta\left(s-s^{*}\right)$, as expected. See Figure C.5.


Figure C.5. Simulation with a nonzero minimum gamete size. We display a simulation where the individual reproductive potential is multiplied by $e^{-k / s}$, where $k>0$. The initial isogamous population moves to an anisogamous population $\rho(s)=x \delta\left(s-s_{\text {small }}^{*}\right)+(1-x) \delta\left(s-s^{*}\right), 0<s_{\text {small }}^{*}<s^{*}$. Here, $N=100$, $w=1, \alpha=1, k=1$, and the final fraction of small gametes is $x=0.16$. The initial isogamous population was drawn from $\mathcal{U}(1,3)$.

## C.6. Absolute gamete fitness

In our model we assume that the reproductive potential of a gamete depends on its size relative to others in the population. In reality, there are likely some absolute size effects that also play a role. In Figure C.6, we numerically simulate our model with the inclusion of both absolute and relative gamete potential terms, with the results appearing to remain qualitatively unchanged. Figure C. 6 does show that a wider range of fractionations is possible at equilibrium with the final fraction of small gametes $x=0.4>(3-\sqrt{2}) / \alpha \approx 0.17$-the threshold given by Eq. (C.1).


Figure C.6. Simulation with both absolute and relative reproductive potential. In this simulation the individual reproductive potential was a weighted sum of two sigmoidal functions, one as in the Eq. (4.3) (i.e., centered at $\bar{s}$ ), and the other identical but centered at $c=1$. Weight was $90 \%$ absolute, $10 \%$ relative. The population converges to an anisogamous state with $40 \%$ small gametes. Here, $N=100, w=0.1+\mathcal{N}\left(0,0.01^{2}\right), \alpha=1$, and the initial population was drawn from $\mathcal{U}(0,1.5)$.

## APPENDIX D

## Debiasing data: appendix

## D.1. Derivation of the partial differential equation

We begin with Eq. (5.2). Writing out this equation explicitly yields
(D.1)

$$
\begin{aligned}
\frac{d y_{i}}{d t} & =\beta \frac{1-k}{2}\left[\left(y_{i+1}+y_{i-1}-2 y_{i}\right)-\left(y_{i+1,0}+y_{i-1,0}-2 y_{i, 0}\right)\right] \\
& -\alpha \frac{k}{4 \Delta x^{2}}\left(y_{i+2}+y_{i-2}-4 y_{i+2}-4 y_{i-2}+6 y_{i}\right)
\end{aligned}
$$

Multiplying Eq. (D.1) by $\Delta x^{2} / \Delta x^{2}$ yields

$$
\begin{align*}
\frac{d y_{i}}{d t} & =\beta \Delta x^{2} \frac{1-k}{2 \Delta x^{2}}\left[\left(y_{i+1}+y_{i-1}-2 y_{i}\right)-\left(y_{i+1,0}+y_{i-1,0}-2 y_{i, 0}\right)\right] \\
& -\alpha \Delta x^{2} \frac{k}{4 \Delta x^{4}}\left(y_{i+2}+y_{i-2}-4 y_{i+2}-4 y_{i-2}+6 y_{i}\right) \tag{D.2}
\end{align*}
$$

Setting $\alpha=\alpha^{*} / \Delta x^{2}$ and $\beta=\beta^{*} / \Delta x^{2}$ gives
(D.3)

$$
\begin{aligned}
\frac{d y_{i}}{d t} & =\beta^{*} \frac{1-k}{2 \Delta x^{2}}\left[\left(y_{i+1}+y_{i-1}-2 y_{i}\right)-\left(y_{i+1,0}+y_{i-1,0}-2 y_{i, 0}\right)\right] \\
& -\alpha^{*} \frac{k}{4 \Delta x^{4}}\left(y_{i+2}+y_{i-2}-4 y_{i+2}-4 y_{i-2}+6 y_{i}\right)
\end{aligned}
$$

The discrete operator for the second and fourth spatial derivative is given by
(D.4)

$$
\frac{d^{2} y_{i}}{d x^{2}}=\frac{y_{i+1}+y_{i-1}-2 y_{i}}{\Delta x^{2}}+\mathcal{O}\left(\Delta x^{2}\right)
$$

(D.5)

$$
\frac{d^{4} y_{i}}{d x^{4}}=\frac{y_{i+2}+y_{i-2}-4 y_{i+2}-4 y_{i-2}+6 y_{i}}{\Delta x^{4}}+\mathcal{O}\left(\Delta x^{2}\right)
$$

Now, sending $\Delta x \rightarrow 0$ gives

$$
\frac{\partial \rho}{\partial t}=\beta^{*}\left(\frac{1-k}{2}\right)\left(\frac{\partial^{2} \rho}{\partial x^{2}}-\frac{d^{2} \rho_{0}}{d x^{2}}\right)-\alpha^{*}\left(\frac{k}{4}\right) \frac{\partial^{4} \rho}{\partial x^{4}}
$$

matching Eq. (5.3).


[^0]:    ${ }^{1}$ Stability when $N$ is small is left for future work.

[^1]:    ${ }^{2}$ For identical oscillators, the magnitude of $K$ is not important, but for the nonidentical case see Section A. 4 for discussion of coupling strength effects

[^2]:    ${ }^{1}$ In our model we consider zero to be the minimal advertising state; note that this represents a state with zero excess advertising, since companies inevitably must package and sell their products under a brand name, and that could itself be considered a minimal form of advertising.
    ${ }^{2}$ By "indistinguishable" we mean that the product without branding is indistinguishable, but the brand label is always known to the consumer.
    ${ }^{3}$ This could be quantified, e.g., by clicks on a website ad banner, inserts in a newspaper, views of an ad on TV, or supermarket placement costs.

[^3]:    ${ }^{4}$ Power laws are common in both natural and engineered systems $\mathbf{1 0 7} \mathbf{1 0 8}$, and there is evidence that production costs can indeed be approximated by power law scaling [109] .

[^4]:    ${ }^{5}$ Numerical experiments suggest that stability conditions derived in this chapter hold for small $N$.

[^5]:    ${ }^{6}$ We suspect that an extension of this model to allow stronger within-segment competition (i.e., name-brands compete more strongly with each other than with generics.) would lead to additional modes.

[^6]:    ${ }^{1}$ See Section C. 5 in appendix for the case when the small gamete group has finite, nonzero equilibrium gamete size.

[^7]:    ${ }^{1}$ But of course $\bar{a}$ is itself a dynamical variable, and a self-consistent solution must be sought where the equilibrium distribution of firm advertising implied by $\bar{a}$ is consistent with $\bar{a}$.

[^8]:    ${ }^{2}$ Note that the KDE performs a uniformly diffusive smoothing operation, whereas our approach assumes that psychological bias in pricing is better corrected for by a locally varying redistribution of probability mass

[^9]:    ${ }^{1}$ We choose to perturb by the uniform distribution in order to avoid negative values

