# The **BEST WRITING** on **MATHEMATICS**

# 2020

Copyrighted material

Copyright © 2020 by Princeton University Press

Princeton University Press is committed to the protection of copyright and the intellectual property our authors entrust to us. Copyright promotes the progress and integrity of knowledge. Thank you for supporting free speech and the global exchange of ideas by purchasing an authorized edition of this book. If you wish to reproduce or distribute any part of it in any form, please obtain permission.

Requests for permission to reproduce material from this work should be sent to permissions@press.princeton.edu

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

All Rights Reserved

ISBN 9780691207575
ISBN (pbk.) 9780691207568
ISBN (e-book) 9780691213651

British Library Cataloging-in-Publication Data is available

Editorial: Susannah Shoemaker and Kristen Hop
Production Editorial: Nathan Carr
Text Design: Carmina Alvarez
Jacket/Cover Design: Chris Ferrante
Production: Jacquie Poirier
Publicity: Matthew Taylor and Katie Lewis
Copyeditor: Paula Bérard

This book has been composed in Perpetua

Printed on acid-free paper. ∞

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

Copyrighted material

# *Contents*

*Color illustration insert follows page 128*



Copyrighted material

Copyrighted material

# *Introduction*

Mircea Pitici

This anthology is the 11th in the annual series of *The Best Writing on Mathematics*. It contains pieces originally published in late 2018 and throughout 2019 in various venues, including specialized print and online magazines, research journals, newspapers, books, and collections of conference proceedings. The volume should be considered by the readers in conjunction with the other ten previously published in the series.

## *Overview of the Volume*

In a piece eerily reminding us of the current coronavirus health crisis, Steven Strogatz recounts the little-known contribution of differential equations to virology during the HIV crisis and makes the case for considering calculus among the heroes of modern life.

Peter Denning and Ted Lewis examine the genealogy, the progress, and the limitations of complexity theory—a set of principles developed by mathematicians and physicists who attempt to tame the uncertainty of social and natural processes.

In yet another example of fusion between ideas from mathematics and physics, Bruce Boghosian describes how a series of simulations carried out to model the long-term outcome of economic interactions based on free-market exchanges inexorably leads to extreme inequality and to the oligarchical concentration of wealth.

Stan Wagon points out the harmonic-average intricacies, the practical paradoxes, and the policy implications that result from using the miles-per-gallon measure for the fuel economy of hybrid cars.

Jørgen Veisdal details some of the comparative reasoning supposed to take place in majoritarian democracies—resulting in electoral strategies that lead candidates toward the center of the political spectrum.

Copyrighted material

In an autobiographical piece, John Baez narrates the convoluted professional path that took him, over many years, closer and closer to algebraic geometry—a branch of mathematics that offers insights into the relationship between the classical mechanics and quantum physics.

Erica Klarreich explains how Hao Huang used the combinatorics of cube nodes to give a succinct proof to a long-standing computer science conjecture that remained open for several decades, despite many repeated attempts to settle it.

A graph-based explanation, combined with a stereographic projection, also helped Richard Montgomery solve one of the questions posed by the dynamical system formed by three masses moving under the reciprocal influences of their gravitational pulls, also known as the three-body problem.

Chris King, who created valuable online resources freely available to everyone, describes the algebraic iterations that lead to families of fractal-like, visually stunning geometric configurations and stand at the confluence of multiple research areas in mathematics.

In the next contribution to our volume, Jim Henle presents several paper-and-pencil games selected from the vast collection invented by Sid Sackson.

Dave Linkletter breaks the classic Rubik's cube apart and, using the mechanics of the cube's skeleton, counts for us the total number of possible configurations; then he reviews a collection of mathematical questions posed by the toy—some answered and some still open.

Colin Adams introduces with examples, defines, and discusses several important properties of the hyperbolic 3-manifold, a geometric notion both common to our physical environment and difficult to understand in its full generality.

In a similar geometric vein, with yet more examples, physical models, and definitions, followed by applications, Boris Odehnal presents an overview of higher dimensional geometries.

With linguistic flourishes recalling Fermat's cryptic style, James Propp traces the history of two apparently disconnected results in the theory of numbers—which, surprisingly, turned out to be strongly related—and tells us how an amateur mathematician used the parallelism to prove one of them.

Patrick Honner works out in several different ways a simple multiplication example to compare the computational efforts required by the

Copyrighted material

algorithms used in each case and to illustrate the significant benefits that result when the most efficient method is scaled up to multiply big numbers.

Ben Orlin combines his drawing and teaching talents to prove that ignorance of widely known mathematics can be both hilariously ridiculous and academically rewarding!

Donald Teets's piece is entirely concerned with the young Karl Friedrich Gauss's contribution to the history of the Christian calendar.

Paul Thagard proposes five conjectures (and many more puzzling questions) on the working of mathematics in mind and society and formulates an eclectic metaphysics that affirms both realistic and fictional qualities for mathematics.

Mark Colyvan asserts that explanation in mathematics—unlike explanation in sciences and in general—is neither causal nor deductive; instead, depending on the context, mathematical explanation provides either local insights that connect similar mathematical situations or global answers that arise from non-mathematical phenomena.

Gerry Hahn, Necip Doganaksoy, and Bill Meeker call (as they have done over a long period of time) for improving statistical inquiry and analysis by using new tools—such as tolerance and prediction intervals, as well as a refined analysis of the role of sample size in experiments.

## *More Writings on Mathematics*

Readers of this series of anthologies know that in each volume I offer many other reading suggestions from the recent literature on mathematics: book titles in the Introduction and articles in the section on Notable Writings, toward the end of the volume. As a matter of principle, I never included in these lists materials I have not seen; thus, my ability to keep up with the literature has been considerable affected by the health crisis that closed university campuses and libraries during the spring of 2020. I thank the authors and the publishers who sent me books over the last year; complete references are at the end of this introduction.

To start my book recommendation list, special mention deserves—for exceptional illustrations and insightful contributions—the collective volume published by the Bodleian Library with the title *Thinking 3D*, edited by Daryl Green and Laura Moretti. Also—for visual aspect,

Copyrighted material

inspired humor, and teaching insights—Ben Orlin's books *Math with Bad Drawings* and *Change Is the Only Constant*.

Excellent expository introductions to specific topics are Julian Havil's *Curves for the Mathematically Curious*, Steven Strogatz's *Infinite Powers*, and (slightly more technical) David Feldman's *Chaos and Dynamical Systems*.

In applied mathematics and connections to other domains, we have *The Mathematics of Politics* by Arthur Robinson and Daniel Ullman, *Modelling Nature* by Edward and Michael Gillman, *Data Analysis for the Social Sciences* by Douglas Bors, *Islands of Order* by Stephen Lansing and Murray P. Cox, *Producers, Consumers, and Partial Equilibrium* by David Mandy, and *Ranking: The Unwritten Rules of the Social Game We All Play* by Péter Érdi. Featuring mathematics in astronomy are *Finding Our Place in the Solar System* by Todd Timberlake and Paul Wallace and *Our Universe* by Jo Dunkley; on mathematics in military affairs, *The (Real) Revolution in Military Affairs* by Andrei Martyanov. Two expository statistics books are *The Art of Statistics* by David Spiegelhalter and *Statistics in Social Work* by Amy Batchelor; and an excursion into computer science is *Computational Thinking* by Peter Denning and Matti Tedre.

Interdisciplinary with historical elements but also with ramifications in contemporary affairs are *Proof! How the World Became Geometrical* by Amir Alexander and *How Charts Lie* by Alberto Cairo. Several books last year were dedicated to the increasing role of algorithms in daily social affairs, including *Algorithmic Regulation* edited by Karen Yeung and Martin Lodge, *The Ethical Algorithm* by Michael Kearns and Aaron Roth, and *The Information Manifold* by Antonio Badia.

A few recent books on the history of mathematics are *Power in Numbers* by Talithia Williams, *Bernard Bolzano* by Paul Rusnock and Jan Šebestík, and *David Hume on Miracles, Evidence, and Probability* by William Vanderburgh. Also historical, with strong reciprocal influences between mathematics and the cultural, social, and linguistics contexts are *Disharmony of the Spheres* by Jennifer Nelson, *Republic of Numbers* by David Lindsay Roberts, and *Roads to Reference* by Mario Gómez-Torrente. In logic and philosophy of mathematics is *Reflections on the Foundations of Mathematics* edited by Stefania Centrone, Deborah Kant, and Deniz Sarikaya. Mathematical notions in practical philosophy appear in *Measurement and Meaning* by Ferenc Csatári and in *Conscious Action Theory* by Wolfgang Baer.

Copyrighted material

# The **BEST** **WRITING** on **MATHEMATICS**

# 2020

Copyrighted material

Copyrighted material

# *Outsmarting a Virus with Math*

Steven Strogatz

In the 1980s, a mysterious disease began killing tens of thousands of people a year in the United States and hundreds of thousands worldwide. No one knew what it was, where it came from, or what was causing it, but its effects were clear—it weakened patients' immune systems so severely that they became vulnerable to rare kinds of cancer, pneumonia, and opportunistic infections. Death from the disease was slow, painful, and disfiguring. Doctors named it acquired immunodeficiency syndrome (AIDS). No cure was in sight.

Basic research demonstrated that a retrovirus was the culprit. Its mechanism was insidious: The virus attacked and infected white blood cells called helper T cells, a key component of the immune system. Once inside, the virus hijacked the cell's genetic machinery and coopted it into making more viruses. Those new virus particles then escaped from the cell, hitched a ride in the bloodstream and other bodily fluids, and looked for more T cells to infect. The body's immune system responded to this invasion by trying to flush out the virus particles from the blood and kill as many infected T cells as it could find. In so doing, the immune system was killing an important part of itself.

The first antiretroviral drug approved to treat HIV appeared in 1987. It slowed the virus down by interfering with the hijacking process, but it was not as effective as hoped, and HIV often became resistant to it. A different class of drugs called protease inhibitors appeared in 1994. They thwarted HIV by interfering with the newly produced virus particles, keeping them from maturing and rendering them noninfectious. Though also not a cure, protease inhibitors were a godsend.

Soon after protease inhibitors became available, a team of researchers led by David Ho (a former physics major at the California Institute of Technology and so, presumably, someone comfortable with calculus)

Copyrighted material

and a mathematical immunologist named Alan Perelson collaborated on a study that changed how doctors thought about HIV and revolutionized how they treated it. Before the work of Ho and Perelson, it was known that untreated HIV infection typically progressed through three stages: an acute primary stage of a few weeks, a chronic and paradoxically asymptomatic stage of up to 10 years, and a terminal stage of AIDS.

In the first stage, soon after a person becomes infected with HIV, he or she displays flulike symptoms of fever, rash, and headaches, and the number of helper T cells (also known as CD4 cells) in the bloodstream plummets. A normal T cell count is about 1,000 cells per cubic millimeter of blood; after a primary HIV infection, the T cell count drops to the low hundreds. Because T cells help the body fight infections, their depletion severely weakens the immune system. Meanwhile, the number of virus particles in the blood, known as the viral load, spikes and then drops as the immune system begins to combat the HIV infection. The flulike symptoms disappear, and the patient feels better.

At the end of this first stage, the viral load stabilizes at a level that can, puzzlingly, last for many years. Doctors refer to this level as the *set point*. A patient who is untreated may survive for a decade with no HIV-related symptoms and no lab findings other than a persistent viral load and a low and slowly declining T cell count. Eventually, however, the asymptomatic stage ends and AIDS sets in, marked by a further decrease in the T cell count and a sharp rise in the viral load. Once an untreated patient has full-blown AIDS, opportunistic infections, cancers, and other complications usually cause the patient's death within two to three years.

The key to the mystery was in the decade-long asymptomatic stage. What was going on then? Was HIV lying dormant in the body? Other viruses were known to hibernate like that. The genital herpesvirus, for example, hunkers down in nerve ganglia to evade the immune system. The chicken pox virus also does this, hiding out in nerve cells for years and sometimes awakening to cause shingles. For HIV, the reason for the latency was unknown.

In a 1995 study, Ho and Perelson gave patients a protease inhibitor, not as a treatment but as a probe. Doing so nudged a patient's body off its set point and allowed the researchers—for the first time ever—to track the dynamics of the immune system as it battled HIV. They found

Copyrighted material

that after each patient took the protease inhibitor, the number of virus particles in the bloodstream dropped exponentially fast. The rate of decay was incredible: half of all the virus particles in the bloodstream were cleared by the immune system every *two days*.

## *Finding the Clearance Rate*

Calculus enabled Perelson and Ho to model this exponential decay and extract its surprising implications. First, they represented the changing concentration of virus in the blood as an unknown function, $V(t)$, where $t$ denotes the elapsed time since the protease inhibitor was administered. Then they hypothesized how much the concentration of virus would change, $dV$, in an infinitesimally short time interval, $dt$. Their data indicated that a constant fraction of the virus in the blood was cleared each day, so perhaps the same constancy would hold when extrapolated down to $dt$. Because $dV/V$ represented the fractional change in the virus concentration, their model could be translated into symbols as the following equation:

$$dV/V = -c\,dt$$

Here the constant of proportionality, $c$, is the clearance rate, a measure of how fast the body flushes out the virus.

The equation above is an example of a differential equation. It relates the infinitesimal change of $V$ (which is called the differential of $V$ and denoted $dV$) to $V$ itself and to the differential $dt$ of the elapsed time. By applying the techniques of calculus to this equation, Perelson and Ho solved for $V(t)$ and found it satisfied:

$$\ln\,[V(t)/V_0] = -ct$$

Here $V_0$ is the initial viral load, and ln denotes a function called the natural logarithm. Inverting this function then implied:

$$V(t) = V_0 e^{-ct}$$

In this equation, $e$ is the base of the natural logarithm, thus confirming that the viral load did indeed decay exponentially fast in the model. Finally, by fitting an exponential decay curve to their experimental data, Ho and Perelson estimated the previously unknown value of $c$.

Copyrighted material

The discovery that HIV replication was so astonishingly rapid changed the way that doctors treated their HIV-positive patients. Previously physicians waited until HIV emerged from its supposed hibernation before they prescribed antiviral drugs. The idea was to conserve forces until the patient's immune system really needed help because the virus would often become resistant to the drugs. So it was generally thought wiser to wait until patients were far along in their illness.

Ho and Perelson turned this picture upside down. There was no hibernation. HIV and the body were locked in a pitched struggle every second of every day, and the immune system needed all the help it could get and as soon as possible after the critical early period of infection. And now it was obvious why no single medication worked for very long. The virus replicated so rapidly and mutated so quickly, it could find a way to escape almost any therapeutic drug.

Perelson's mathematics gave a quantitative estimate of how many drugs had to be used in combination to beat HIV down and keep it down. By taking into account the measured mutation rate of HIV, the size of its genome, and the newly estimated number of virus particles that were produced daily, he demonstrated mathematically that HIV was generating every possible mutation at every base in its genome many times a day. Because even a single mutation could confer drug resistance, there was little hope of success with single-drug therapy. Two drugs given at the same time would stand a better chance of working, but Perelson's calculations showed that a sizable fraction of all possible double mutations also occurred each day. Three drugs in combination, however, would be hard for the HIV virus to overcome. The math suggested that the odds were something like 10 million to one against HIV being able to undergo the necessary three simultaneous mutations to escape triple-combination therapy.

When Ho and his colleagues tested a three-drug cocktail on HIV-infected patients in clinical studies in 1996, the results were remarkable. The level of virus in the blood dropped about 100-fold in two weeks. Over the next month, it became undetectable.

This is not to say that HIV was eradicated. Studies soon afterward showed that the virus can rebound aggressively if patients take a break from therapy. The problem is that HIV can hide out. It can lie low in sanctuary sites in the body that the drugs cannot readily penetrate or lurk in latently infected cells and rest without replicating, a sneaky way

Copyrighted material

of evading treatment. At any time, these dormant cells can wake up and start making new viruses, which is why it is so important for HIV-positive people to keep taking their medications, even when their viral loads are undetectable.

In 1996, Ho was named *Time* magazine's Man of the Year. In 2017, Perelson received a major prize for his "profound contributions to theoretical immunology." Both are still saving lives by applying calculus to medicine: Ho is analyzing viral dynamics, and some of Perelson's latest work helped to create treatments for hepatitis C that cure the infection in nearly every patient.

The calculus that led to triple-combination therapy did not cure HIV. But it changed a deadly virus into a chronic condition that could be managed—at least for those with access to treatment. It gave hope where almost none had existed before.

## *More to Explore*

**Rapid Turnover of Plasma Virions and CD4 Lymphocytes in HIV-1 Infection.** David D. Ho et al. in *Nature*, Vol. 373, pp. 123–126, January 12, 1995.

**Modelling Viral and Immune System Dynamics.** Alan S. Perelson in *Nature Reviews Immunology*, Vol. 2, pp. 28–36, January 2002.

Copyrighted material

# *Uncertainty*

Peter J. Denning and Ted G. Lewis

In a famous episode in the "I Love Lucy" television series—"Job Switching," better known as the chocolate factory episode—Lucy and her best-friend coworker Ethel are tasked to wrap chocolates flowing by on a conveyor belt in front of them. Each time they get better at the task, the conveyor belt speeds up. Eventually they cannot keep up, and the whole scene collapses into chaos.

The threshold between order and chaos seems thin. A small perturbation—such as a slight increase in the speed of Lucy's conveyor belt—can either do nothing or it can trigger an avalanche of disorder. The speed of events within an avalanche overwhelms us, sweeps away structures that preserve order, and robs our ability to function. Quite a number of disasters, natural or human-made, have an avalanche character—earthquakes, snow cascades, infrastructure collapse during a hurricane, or building collapse in a terror attack. Disaster-recovery planners would dearly love to predict the onset of these events so that people can safely flee and first responders can restore order with recovery resources standing in reserve.

Disruptive innovation is also a form of avalanche. Businesses hope their new products will "go viral" and sweep away competitors. Competitors want to anticipate market avalanches and side-step them. Leaders and planners would love to predict when an avalanche might occur and how extensive it might be.

In recent years, complexity theory has given us a mathematics to deal with systems where avalanches are possible. Can this theory make the needed predictions where classical statistics cannot? Sadly, complexity theory cannot do this. The theory is very good at explaining avalanches after they have happened, but generally useless for predicting when they will occur.

Copyrighted material

## *Complexity Theory*

In 1984, a group of scientists founded the Santa Fe Institute to see if they could apply their knowledge of physics and mathematics to give a theory of chaotic behavior that would enable professionals and managers to move productively amid uncertainty. Over the years, the best mathematical minds developed a beautiful, rich theory of complex systems.

Traditional probability theory provides mathematical tools for dealing with uncertainty. It assumes that the uncertainty arises from random variables that have probability distributions over their possible values. It typically predicts the future values of the variable by computing a mean of the distribution and a confidence interval based on its standard deviation. For example, in 1962 Everett Rogers studied the adoption times of the members of a community in response to a proposed innovation (5). He found they follow a normal (bell) curve that has a mean and a standard deviation. A prediction of adoption time is the mean time bracketed by a confidence interval: for example, 68% of the adoption times are within one standard deviation of the mean, and 95% are within two standard deviations.

In 1987, researchers Per Bak, Chao Tang, and Kurt Wiesenfeld published the results of a simple experiment that demonstrated the essence of complexity theory (4). They observed a sand pile as it formed by dropping grains of sand on a flat surface. Most of the time, each new grain would settle into a stable position on the growing cone of sand. But at unpredictable moments, a grain would set off an avalanche of unpredictable size that cascaded down the side of the sand pile. The researchers measured the time intervals between avalanche starts and the sizes of avalanches. To their surprise, these two random variables did not fit any classical probability distribution, such as the normal or Poisson distributions. Instead, their distributions followed a "power law," meaning the probability of a sample of length $x$ is proportional to $x^{-k}$, where $k$ is a fixed parameter of the random process. Power law distributions have a finite mean only if $k > 2$ and variance only if $k > 3$. This means that a power law with $k \leq 2$ has no mean or variance. Its future is unpredictable. When $2 < k \leq 3$, the mean is finite but not the confidence interval. Bak et al. had discovered something different—a random process whose future could not be predicted with any confidence.

Copyrighted material

This was not an isolated finding. Most of the random processes tied to chaotic situations obey a power law with $k < 3$. For example, the appearance of new connections among web pages is chaotic. The number of web pages with $x$ connections to other pages is proportional to $1/x^2$—the random process of accumulating links produces 1/4 as many pages with $2x$ connections as with $x$ connections. This was taken as both bad and good news for the Internet. The bad news is that because there are a very few "hubs"—servers hosting a very large number of connections—an attacker could shatter the network into isolated pieces by bringing down the hubs. The good news is that the vast majority of servers host few connections and thus random server failures are unlikely to shatter the network. What makes this happen is "preferential attachment"—when a new web page joins the network, it tends to connect with the most highly connected nodes already in the network. Start-up company founders try to plot strategies to bring about rapid adoption of their technologies and transform their new services into hubs.

Hundreds of processes in science and engineering follow power laws, and their key variables are unpredictable. Innovation experts believe that innovations follow a power law—the number of innovations adopted by communities of size $x$ is proportional to $x^{-2}$—not good news for start-up companies hoping to predict that their innovations will take over the market.

Later Bak (1) developed a theory of unpredictability that has subsequently been copied by popular writers like Nassim Nicholas Taleb (6) and others. Bak called it *punctuated equilibrium*, a concept first proposed by Stephen Jay Gould and Niles Eldredge in 1972 (3). The idea is that new members can join a complex system by fitting into the existing structure; but occasionally, the structure passes a critical point and collapses and the process starts over. The community order that has worked for a long time can become brittle. *Avalanche* is an apt term for the moment of collapse. In the sand pile, for example, most new grains lodge firmly into a place on the pile, but occasionally one sets off an avalanche that changes the structure. On the Internet, malware can quickly travel via a hub to many nodes and cause a large-scale avalanche of disruption. In an economy, a new technology can suddenly trigger an avalanche that sweeps away an old structure of jobs and professions and establishes a new order, leaving many people stranded. Complexity

Copyrighted material

location than further away. The preparedness strategies include rapid mobilization of law enforcement just after an attack to counter the tendency for a new attack and identifying optimal geographic locations for positioning recovery resources and supplies. Resilience strategies include rapidly mobilizing technicians and artisans to restore broken communications and facilities. Adaptiveness strategies include scenarios and war games.

## *Uncertainty in Professional Work*

What can we do when we find ourselves in chaotic situations and must still navigate through the uncertainty to achieve our goals?

One of the most difficult environments to navigate is the social space in which we perform our work. This space is dominated by choices that other people make beyond our control. When we propose innovations, we are likely to encounter resistance from some sectors of our community that do not want the innovation; they can be quite inventive in finding ways to block our proposals (2). When we start new projects or even companies, we do not know whether our plans are going to take off or just wither away. Even in normal everyday working environments, conflicts and contingencies suddenly arise and we must resolve them to keep moving forward.

The analogy of a surfer is useful in approaching these situations. A surfer aims to ride the waves to the shore without losing balance and being swept under. The waves can be turbulent and unpredictable. The surfer must maintain balance, ride the crests moving toward the shore, and dodge side waves and cross currents. The surfer may need to jump to a new wave when the time is right, or quickly tack to avoid an unfavorable current or wind. Thus, the surfer generates a path through the fast-changing waves.

In the social space, waves manifest as groups of people disposed to move in certain directions and not in others—sometimes the waves appear as fads or "memes," and they have a momentum that is difficult to divert. As professionals, we become aware of these waves and try to harness them to carry us toward our goals. As each surprise pops up, we instinctively look for openings into which we can move—and, more importantly, we create openings by starting conversations that assuage the concerns of those whose resistance threatens to block us.

Copyrighted material

These little deals cut a path through the potential resistance and get us to our goal.

The lesson here is that we listen for the waves, ride their momentum toward our goals, and make adjustments by creating openings in our conversations with other people. At its best, the complexity theory helps us understand when a process is susceptible to unpredictable avalanches. We move beyond the limitations of the theory by generating openings in our conversations with other people.

## *References*

1. Bak, P. 1996. *How Nature Works: The Science of Self-Organized Criticality*. Springer-Verlag, New York.
2. Denning, P. 2018. Winning at innovation. *IEEE Computer* (Oct.), 51:10, 32–39.
3. Eldredge, N., and Gould, S. J. 1972. Punctuated Equilibria: An Alternative to Phyletic Gradualism. In *Models in Paleobiology*, T. J. M. Schopf, Ed. Freeman Cooper, San Francisco, 82–115.
4. Lewis, T. G. 2011. *Bak's Sand Pile: Strategies for a Catastrophic World*. Agile Press, Williams, CA, 382.
5. Rogers, E. M. 2003. *Diffusion of Innovations,* 5th ed. Free Press, New York.
6. Taleb, N. N. 2007, 2010. *The Black Swan: The Impact of the Highly Improbable*. Random House, New York.

Copyrighted material

# *The Inescapable Casino*

Bruce M. Boghosian

Wealth inequality is escalating at an alarming rate, not only within the United States, but also in countries as diverse as Russia, India, and Brazil. According to investment bank Credit Suisse, the fraction of global household wealth held by the richest 1% of the world's population increased from 42.5 to 47.2% between the financial crisis of 2008 and 2018. To put it another way, as of 2010, 388 individuals possessed as much household wealth as the lower half of the world's population combined—about 3.5 billion people; today Oxfam estimates that number as 26 individuals. Statistics from almost all nations that measure wealth in their household surveys indicate that wealth is becoming increasingly concentrated.

Although the origins of inequality are hotly debated, an approach developed by physicists and mathematicians, including my group at Tufts University, suggests that they have long been hiding in plain sight—in a well-known quirk of arithmetic. This method uses models of wealth distribution collectively known as agent-based, which begin with an individual transaction between two "agents" or actors, each trying to optimize his or her own financial outcome. In the modern world, nothing could seem more fair or natural than two people deciding to exchange goods, agreeing on a price and shaking hands. Indeed, the seeming stability of an economic system arising from this balance of supply and demand among individual actors is regarded as a pinnacle of Enlightenment thinking—to the extent that many people have come to conflate the free market with the notion of freedom itself. Our deceptively simple mathematical models, which are based on voluntary transactions, suggest, however, that it is time for a serious reexamination of this idea.

Copyrighted material

In particular, the *affine wealth model* (called thus because of its mathematical properties) can describe wealth distribution among households in diverse developed countries with exquisite precision while revealing a subtle asymmetry that tends to concentrate wealth. We believe that this purely analytical approach, which resembles an x-ray in that it is used not so much to represent the messiness of the real world as to strip it away and reveal the underlying skeleton, provides deep insight into the forces acting to increase poverty and inequality today.

## *Oligarchy*

In 1986, social scientist John Angle first described the movement and distribution of wealth as arising from pairwise transactions among a collection of "economic agents," which could be individuals, households, companies, funds, or other entities. By the turn of the century, physicists Slava Ispolatov, Pavel L. Krapivsky, and Sidney Redner, then all working together at Boston University, as wel as Adrian Drăgulescu, now at Constellation Energy Group, and Victor Yakovenko of the University of Maryland, had demonstrated that these agent-based models could be analyzed with the tools of statistical physics, leading to rapid advances in our understanding of their behavior. As it turns out, many such models find wealth moving inexorably from one agent to another—even if they are based on fair exchanges between equal actors. In 2002, Anirban Chakraborti, then at the Saha Institute of Nuclear Physics in Kolkata, India, introduced what came to be known as the "yard sale model," called thus because it has certain features of real one-on-one economic transactions. He also used numerical simulations to demonstrate that it inexorably concentrated wealth, resulting in oligarchy.

To understand how this happens, suppose you are in a casino and are invited to play a game. You must place some ante—say, $100—on a table, and a fair coin will be flipped. If the coin comes up heads, the house will pay you 20% of what you have on the table, resulting in $120 on the table. If the coin comes up tails, the house will take 17% of what you have on the table, resulting in $83 left on the table. You can keep your money on the table for as many flips of the coin as you would like (without ever adding to or subtracting from it). Each time you play, you will win 20% of what is on the table if the coin comes up heads, and

Copyrighted material

you will lose 17% of it if the coin comes up tails. Should you agree to play this game?

You might construct two arguments, both rather persuasive, to help you decide what to do. You may think, "I have a probability of $\frac{1}{2}$ of gaining \$20 and a probability of $\frac{1}{2}$ of losing \$17. My expected gain is therefore:

$$\tfrac{1}{2} \times (+\$20) + \tfrac{1}{2} \times (-\$17) = \$1.50$$

which is positive. In other words, my odds of winning and losing are even, but my gain if I win will be greater than my loss if I lose." From this perspective, it seems advantageous to play this game.

Or, like a chess player, you might think further: "What if I stay for 10 flips of the coin? An extension of the above argument indicates that my expected winning will be $(1 + 0.015)^{10} \times \$100 = \$116.05$. This is correct, and it seems promising until I realize that I would need at least six wins to avoid a loss. Five wins and five losses will not be good enough, since the amount of money remaining on the table in that case would be

$$1.2 \times 1.2 \times 1.2 \times 1.2 \times 1.2 \times 0.83 \times 0.83 \times 0.83 \times 0.83 \times 0.83 \times \$100 = \$98.02$$

so I will have lost about \$2 of my original \$100 ante. The trouble is that, of the $2^{10} = 1{,}024$ possible outcomes of 10 coin flips, only 386 of them result in my winning six or more times, so the probability of that happening is only 386/1024 = 0.377. Hence, while my reward for winning is increasing, the probability of my winning is simultaneously decreasing." As the number of coin flips increases, this problem only worsens as the following table makes clear:

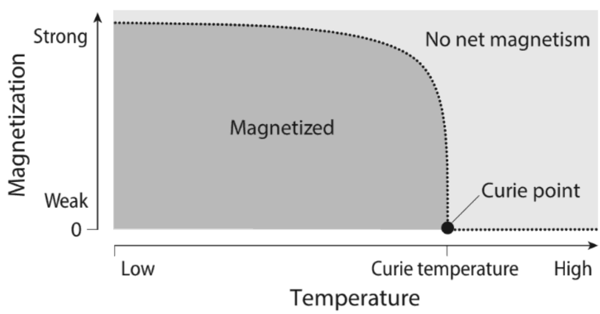
| Number of Coin Flips | Expected Gain | Number of Wins Required to Avoid Loss | Probability of Avoiding a Loss |
|---|---|---|---|
| 10 | 0.1605 | 6 | 0.377 |
| 100 | 3.432 | 51 | 0.460 |
| 1,000 | $2.924 \times 10^{6}$ | 506 | 0.364 |
| 10,000 | $4.575 \times 10^{64}$ | 5,055 | 0.138 |
| 100,000 | $4.020 \times 10^{646}$ | 50,544 | 0.000294 |

Copyrighted material

## The Physics of Inequality

When water boils at 100 degrees Celsius and turns into water vapor, it undergoes a phase transition—a sudden and dramatic change. For example, the volume it occupies (at a given pressure) increases discontinuously with temperature. Similarly, the strength of a ferromagnet falls to zero (*line in Figure A*) as its temperature increases to a point called the Curie temperature, $T_c$. At temperatures above $T_c$, the substance has no net magnetism. The fall to zero magnetism is continuous as the temperature approaches $T_c$ from below, but the graph of magnetization versus temperature has a sharp kink at $T_c$.

**A Phase Change in a Ferromagnet**

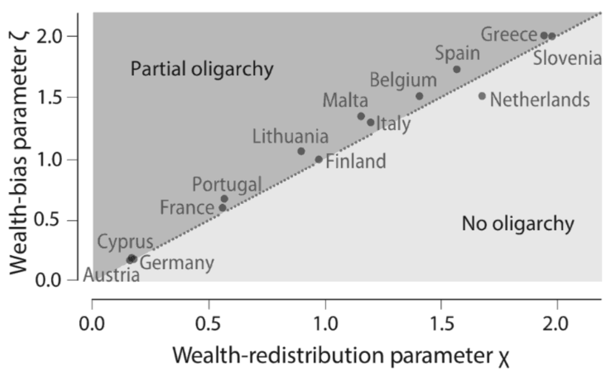



Conversely, when the temperature of a ferromagnet is reduced from above to below $T_c$, magnetization spontaneously appears where there had been none. Magnetization has an inherent spatial orientation—the direction from the south pole of the magnet to the north pole—and one might wonder how it develops. In the absence of any external magnetic field that might indicate a preferred direction, the breaking of the rotational symmetry is "spontaneous." (Rotational symmetry is the property of being identical in every orientation, which the system has at temperatures above $T_c$.) That is, magnetization shows up suddenly, and the direction of the magnetization is random (or, more precisely, dependent on microscopic fluctuations beyond our idealization of the ferromagnet as a continuous macroscopic system).

Copyrighted material

Economic systems can also exhibit phase transitions. When the wealthbias parameter ζ of the affine wealth model is less than the redistribution parameter χ, the wealth distribution is not even partially oligarchical (*area on the right in Figure B*). When ζ exceeds χ, however, a finite fraction of the wealth of the entire population "condenses" into the hands of an infinitesimal fraction of the wealthiest agents. The role of temperature is played by the ratio χ/ζ, and wealth condensation shows up when this quantity falls below 1. (See also color insert.)

**B Phase Transition in Economic Systems**




Another subtle symmetry exhibited by complex macroscopic systems is "duality," which describes a one-to-one correspondence between states of a substance above and below the critical temperature, at which the phase transition occurs. For ferromagnetism, it relates an ordered, magnetized system at temperature $T$ below $T_c$ to its "dual"—a disordered, unmagnetized system at the socalled inverse temperature, $(T_c)^2/T$, which is above $T_c$. The critical temperature is where the system's temperature and the inverse temperature cross (that is, $T = (T_c)^2/T$). Duality theory plays an increasingly important role in theoretical physics, including in quantum gravity.

Like ferromagnetism, the affine wealth model exhibits duality, as proved by Jie Li and me in 2018. A state with $\zeta < \chi$ is not a partial oligarchy, whereas a corresponding state with this

Copyrighted material

relation reversed—that is, with the "temperature" $\chi/\zeta$ inverted to $\zeta/\chi$—is. Interestingly, these two dual states have exactly the same wealth distribution if the oligarch is removed from the wealth-condensed economy (and the total wealth is recalculated to account for this loss).

Significantly, most countries are very close to criticality. A plot of 14 of the countries served by the European Central Bank in the $\chi - \zeta$ plane in Figure B shows that most countries lie near the diagonal. All except one (the Netherlands) lie just above the diagonal, indicating that they are just slightly oligarchical. It may be that inequality naturally increases until oligarchies begin to form, at which point political pressures set in, preventing further reduction of equality.

casino—you win some and you lose some, but the longer you stay in the casino, the more likely you are to lose. The free market is essentially a casino that you can never leave. When the trickle of wealth described earlier, flowing from poor to rich in each transaction, is multiplied by 7.7 billion people in the world conducting countless transactions every year, the trickle becomes a torrent. Inequality inevitably grows more pronounced because of the collective effects of enormous numbers of seemingly innocuous but subtly biased transactions.

## *The Condensation of Wealth*

You might, of course, wonder how this model, even if mathematically accurate, has anything to do with reality. After all, it describes an entirely unstable economy that inevitably degenerates to complete oligarchy, and there are no complete oligarchies in the world. It is true that, by itself, the yard sale model is unable to explain empirical wealth distributions. To address this deficiency, my group has refined it in three ways to make it more realistic.

In 2017, Adrian Devitt-Lee, Merek Johnson, Jie Li, Jeremy Marcq, Hongyan Wang, and I, all at Tufts, incorporated the redistribution of wealth. In keeping with the simplicity desirable in applied mathematics models, we did this by having each agent take a step toward the mean

Copyrighted material

wealth in the society after each transaction. The size of the step was some fraction $\chi$ (or "chi") of his or her distance from the mean. This is equivalent to a flat wealth tax for the wealthy (with tax rate $\chi$ per unit time) and a complementary subsidy for the poor. In effect, it transfers wealth from those above the mean to those below it. We found that this simple modification stabilized the wealth distribution so that oligarchy no longer resulted. And astonishingly, it enabled our model to match empirical data on U.S. and European wealth distribution between 1989 and 2016 to better than 2%. The single parameter $\chi$ seems to subsume a host of real-world taxes and subsidies that would be too messy to include separately in a skeletal model such as this one.

In addition, it is well documented that the wealthy enjoy systemic economic advantages, such as lower interest rates on loans and better financial advice, whereas the poor suffer systemic economic disadvantages, such as payday lenders and a lack of time to shop for the best prices. As James Baldwin once observed, "Anyone who has ever struggled with poverty knows how extremely expensive it is to be poor." Accordingly, in the same paper mentioned above, we factored in what we call wealth-attained advantage. We biased the coin flip in favor of the wealthier individual by an amount proportional to a new parameter, $\zeta$, (or "zeta"), times the wealth difference divided by the mean wealth. This rather simple refinement, which serves as a proxy for a multitude of biases favoring the wealthy, improved agreement between the model and the upper tail (representing very wealthy people) of actual wealth distributions.

The inclusion of wealth-attained advantage also yields—and gives a precise mathematical definition to—the phenomenon of partial oligarchy. Whenever the influence of wealth-attained advantage exceeds that of redistribution (more precisely, whenever $\zeta$ exceeds $\chi$), a vanishingly small fraction of people will possess a finite fraction, $1 - \chi/\zeta$, of societal wealth. The onset of partial oligarchy is in fact a phase transition for another model of economic transactions, as first described in 2000 by physicists Jean-Philippe Bouchaud, now at École Polytechnique, and Marc Mézard of the École Normale Supérieure. In our model, when $\zeta$ is less than $\chi$, the system has only one stable state with no oligarchy; when $\zeta$ exceeds $\chi$, a new, oligarchical state appears and becomes the stable state [*see* box "Winners, Losers"]. The two-parameter ($\chi$ and $\zeta$) extended yard sale model thus obtained can match empirical data on U.S.

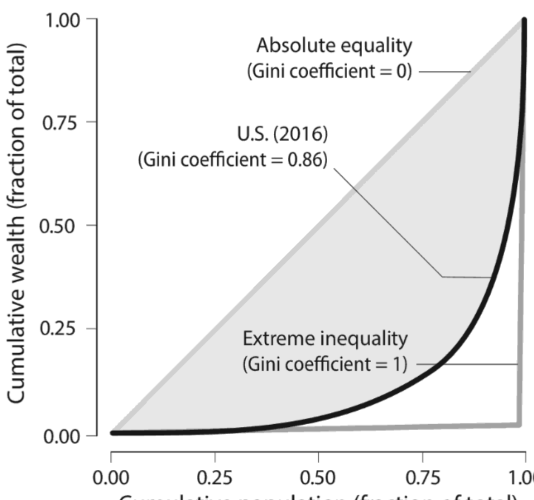
Copyrighted material

## Measuring Inequality

In the early twentieth century, American economist Max O. Lorenz designed a useful way to quantify wealth inequality. He proposed plotting the fraction of wealth held by individuals with wealth less than *w* against the fraction of individuals with wealth less than *w*. Because both quantities are fractions ranging from 0 to 1, the plot fits neatly into the unit square. Twice the area between Lorenz's curve and the diagonal is called the Gini coefficient, a commonly used measure of inequality.

Let us first consider the egalitarian case. If every individual has exactly the same wealth, any given fraction of the population has precisely that fraction of the total wealth. Hence, the Lorenz curve is the diagonal (*green line in Figure A*), and the Gini coefficient is 0. In contrast, if one oligarch has all the wealth and everybody else has nothing, the poorest fraction *f* of the population has no wealth at all for any value of *f* that is less than 1, so the Lorenz curve is pegged to 0. But when *f* equals 1, the oligarch is included, and the curve suddenly jumps up to 1. The area between this Lorenz curve (*orange line on the right of the x-y axis*) and the diagonal is half the area of the square, or ½, and hence the Gini coefficient is 1. (See also color insert.)

**A Lorenz Curves**




Copyrighted material