

**KELLY CLANCY**

# **Playing**

**How Games**

# **With**

**Shape Our World**

# **Reality**

Copyrighted Material

allen lane

# *Playing With Reality*

**Copyrighted Material**

**Copyrighted Material**

# *Playing With Reality*

*How Games Shape Our World*

KELLY CLANCY



ALLEN LANE  
an imprint of  
PENGUIN BOOKS

ALLEN LANE

UK | USA | Canada | Ireland | Australia  
India | New Zealand | South Africa

Allen Lane is part of the Penguin Random House group of companies  
whose addresses can be found at [global.penguinrandomhouse.com](http://global.penguinrandomhouse.com).



Penguin  
Random House  
UK

First published in the United States of America by Riverhead Books 2024  
First published in Great Britain by Allen Lane 2024  
001

Copyright © Kelly Clancy, 2024

The moral right of the author has been asserted

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

The authorized representative in the EEA is Penguin Random House Ireland,  
Morrison Chambers, 32 Nassau Street, Dublin D02 YH68

A CIP catalogue record for this book is available from the British Library

ISBN: 978-0-241-54550-8

[www.greenpenguin.co.uk](http://www.greenpenguin.co.uk)



Penguin Random House is committed to a sustainable business, our readers and our planet. This book is made from Forest Stewardship Council® certified paper.

*To the next generation—especially to my son,  
Cillian, and my niblings Fiona, Kira, and Ronan.  
If play is the engine of creation,  
I cannot wait to see what new worlds you build.*

**Copyrighted Material**

# Contents

## *Part I* How to Know the Unknown

1 THE PLAY OF CREATION	3
2 HOW HEAVEN WORKS	16
3 DICE PLAYING GOD	41

## *Part II* Naming the Game

4 KRIEGSSPIEL, THE SCIENCE OF WAR	71
5 RATIONAL FOOLS	85
6 THE CLOTHES HAVE NO EMPEROR	112
7 A MAP THAT WARPS THE TERRITORY	137

## *Part III* Building Better Players

8 CHESS, THE <i>DROSOPHILA</i> OF INTELLIGENCE	167
9 THE END OF EVOLUTION	177
10 NOUS EX MACHINA	200
11 COGITO ERGO ZERO SUM	222

## *Part IV* Building Better Games

12 <i>SIMCITY</i>	249
13 MORAL GEOMETRY: PLAYING UTOPIA	257
14 MECHANISM DESIGN: BUILDING GAMES WHERE EVERYONE WINS	278
<i>Epilogue</i>	301

<i>Acknowledgments</i>	307
------------------------	-----

<i>Notes</i>	309
--------------	-----

<i>Bibliography</i>	325
---------------------	-----

<i>Index</i>	349
--------------	-----

Copyrighted Material

**Copyrighted Material**

*Part I*

# How to Know the Unknown

---

**Copyrighted Material**

# 1

## The Play of Creation

*Lila* is the play of creation. To awakened consciousness, the entire universe, with all its joys and sorrows, pleasures and pains, appears as a divine game, sport, or drama. It is a play in which the one Consciousness performs all the roles.

KENDRA GROSSEN BURROUGHS

The long-forgotten game rithmomachia was once medieval Europe's most popular educational technology. It remained a standard in the monastic curriculum for nearly five hundred years. Ovid hailed it as "the leaf, flower, and fruit of Arithmetic, and its glory, laud, and honor." Thomas More depicted the virtuous citizens of his *Utopia* playing rithmomachia instead of "ruinous" games like dice. Elites played it, in part, to demonstrate their erudition. Church leaders believed that the game was not only an educational force but also an enlightening one: it instructed players in the greater harmony of the universe and even tamed violent tempers. Rithmomachia, a "banquet proper of the mind," wordlessly introduced students to divine truths through the beauty of play.

Also known as the Philosopher's Game, rithmomachia was a capture game related to chess. Two players moved black and white pieces across a checkered board according to rules prescribed by their respective

shapes. Unlike in chess, pieces were engraved with numbers meant to represent ideas in number theory. A player could capture an opponent's piece only if the numbers inscribed on the battling pieces had some geometric relation. A player won by arranging three or four pieces on their opponent's side of the board into a "harmony," or geometric progression.

The game has been traditionally credited to the fifth-century BCE Greek philosopher Pythagoras, though the earliest written evidence of its existence dates only to 1030 CE. Nevertheless, it was clearly designed to reflect his teachings. Rithmomachia was considered the apotheosis of the quadrivium, a philosophy of higher education that embodied Pythagorean ideals, centered around his worship of numbers. The quadrivium consisted of arithmetic (pure numbers), geometry (numbers in space), music (numbers in time), and astronomy (numbers in space and time). In rithmomachia, numbers and shapes danced across the board, organized into spatial patterns unfolding over time. These ideas, harmless as they seem, laid the foundations for one of the most absurd cover-ups in intellectual history.

Though none of Pythagoras's writings survived the march of history, he left an outsize influence on European thought. Born in Greece, he was rumored to have spent twenty years studying with religious mystics in Egypt before establishing a cult in the Grecian colony of Croton. His was an ascetic religion: he preached vegetarianism and temperance in an era of decadence and growing political instability. His followers shared all their property and ate their meals together. Pythagoras forbade his disciples from eating beans, which he fretted might house the souls of men, given their resemblance to human fetuses. Above all, his teachings held numbers sacred: they were the building blocks for the entire cosmos. Numbers were beyond divine—they were the very cause of gods.

"All is number," Pythagoras professed. Yet not all numbers were

created equal, in his view. Central to his cosmology was the idea of harmony. He had discovered that musical pitch varies depending on a plucked string's length, and musical harmony arises from plucking strings that are precise multiples of one another. For this reason, he held ratios sacred, because he believed they expressed the harmony of the universe. He worshipped what are now known as rational numbers, including whole numbers (like one, two, and three) and their ratios (like one-third and five-sixths). These fundamental elements, he believed, comprise all other phenomena in the universe.

His growing cult gained political clout over the two decades that it resided in Croton. Eventually, local leaders deemed the bizarre vegetarian a threat to their power. They organized an angry mob to drive the cult away, burning the commune to the ground. Different legends have Pythagoras meeting his end by exile, fire, or suicide. One story has him fleeing for his life, only to be stopped in his tracks by a field of beans. Unwilling to trample across the sacred crop, Pythagoras was caught and executed by his pursuers. Despite this fall from grace, his cult persisted for another three hundred years, and his ideas dominated European philosophy and education for nearly two millennia.

Dogma is anathema to intellectual progress. Unfortunately, Pythagoras's devotion to rational numbers motivated a scholarly subterfuge. Hippasus, a member of Pythagoras's cult, is thought to have accidentally discovered irrational numbers in the fourth century BCE. Whereas rational numbers can be expressed as the ratio of two numbers, irrational numbers can't be. Rational numbers can be precisely determined: One-half is equal to 0.5. The value of one-third is 0.3333 . . . , where the three repeats forever. We automatically know that its seven-billionth digit is three. The square root of two, on the other hand, is irrational. It equals 1.41421356237 . . . , a decimal without repeating or terminating digits. Irrational numbers can't be precisely

determined. To know the seven-billionth digit of the square root of two, we must calculate it out. Irrational numbers, Hippasus realized, are alogos, or “inexpressible.” Their existence directly contradicted the Pythagorean doctrine that all phenomena are composed of whole numbers and their ratios. They contravened all the certainty and comprehensibility that Pythagoras believed numbers endowed upon the universe. Hippasus disappeared into history shortly after sharing his discovery; he was reportedly executed at sea. Legend has it that the cult justified his murder as an offering to Poseidon, rendering it an act outside legal jurisdiction.

Some medieval historians blame the quadrivial curriculum for holding European scholarship back hundreds of years. Its focus on harmony blinkered scientists to fundamental advances, such as the discovery of irrational numbers and celestial mechanics. The quadrivial curriculum was largely abandoned by the seventeenth century, as new mathematical techniques from Persia and India led to breakthroughs in calculus and probability theory. Rithmomachia similarly fell by the wayside. The harmonies that Pythagoras and his followers believed made up the material cosmos were flimsy ideals. Planets move along elliptical, not spherical, orbits, defying models of cosmic conformity. Not only do irrational numbers exist, but the size of their set is so huge it utterly swamps that of rational numbers. If one were to write down all rational and irrational numbers on slips of paper and place these into an immense hat, one would have almost no chance of picking out a rational number—they’re vanishingly rare in the universe of mathematics.

Games are a kind of untrue truth. What is internally consistent within a game need not reflect anything about reality. Yet games have been increasingly adopted as models of the world. Games are a method by which we can learn **Copyrighted Material** the behavior of objects in rule-based systems. At their best, games have generated a deeper under-

standing of mathematics. At the same time, ideologues have used games to smooth reality's rougher edges.

To be fair, the problem is never a game itself but the dogma it's co-opted to serve. It was the doctrine of harmony, not rithmomachia, that held thinkers back. Rithmomachia was the doctrine's emissary, imprinting generations of scholars with its aesthetic ideals. A game rewards its players for adopting its precepts. A pacifist cannot win *Call of Duty* without shooting in-game enemies; a socialist cannot win Monopoly without adopting capitalist behaviors. Similarly, rithmomachia seduced its players with the beauty of Pythagorean ideals. Games are more than models of the world—they're models that reward us for believing in them. They can influence how we think of and see the world, for better or worse. As such, games are also a window into the beliefs and habits of the people who play them. By examining our relationship with games throughout history, we can better understand the beliefs of people long past and more clearly see our own.

Games are older than written language. Games like Go, chess, backgammon, and mancala are living artifacts; they've outlasted empires and conquered cultures. They transcend even language: through play, we can interact with other people's minds, regardless of their tongues. One can imagine an ancient Mesopotamian tavern populated by far-flung merchants and travelers, unable to converse but still able to enjoy an evening together engrossed in a board game. Throughout history, immigrants carried their treasured games with them and kept them alive for thousands of years, as cherished as any cultural memento. Today, multiple subcultures spanning continents and generations define themselves through the games they play. Games endure because they engage a learning system common to all people. They're a product of both culture and biology—a stimulus the brain devised, over time, to serve itself for pleasure. Although often dismissed as "trivial," games have captivated human attention for millennia.

People are so entranced by games that they've routinely used them to anesthetize the anxiety of troubled times. In his account of the Greco-Persian Wars, Herodotus contends that gaming saved the Lydian people in a time of deprivation. The Lydians were much like the Hellenes, he writes. They were the first people to mint coins and, by their own account, the first to invent several now ubiquitous games. Around the second millennium BCE, they were devastated by a decades-long famine but found solace in play:

When they saw the evil still continuing, they sought for remedies, and some devised one thing, some another; and at that time the games of dice, knucklebones, ball, and all other kinds of games except draughts, were invented, for the Lydians do not claim the invention of this last; and having made these inventions to alleviate the famine, they employed them as follows: they used to play one whole day that they might not be in want of food; and on the next, they ate and abstained from play; thus they passed eighteen years.

Though the Lydians couldn't have invented the games they claimed to have invented—knucklebones and dice far predate them—it's conceivable that they took refuge in games when food was scarce. Games are, by design, utterly absorbing. History is littered with stories of famous figures ruined by their love of games or gambling. To this day, a handful of people die every year while gaming—usually from exhaustion, though some are murdered in the heat of competition. This is not cited to demonize games but to demonstrate their power. For thousands of years, games have fascinated people. Today, games are arguably the dominant cultural media. They are a superstimulus, a psychotropic technology, having been tweaked over time to maximally engage our reward systems. **Copyrighting Material** Copying or distributing in print or electronic forms without written permission of the copyright owner is prohibited. What's more, we can better understand humans. What's more, we can better under-

stand understanding: games have shaped how we generate knowledge and reason about the unknown.

One difficulty in discussing games is their elasticity as a metaphor. Is make-believe a game? Is an exam a game? Is a puzzle? There are zero-player games, massively multiplayer online games, and games with negative-, zero-, or positive-sum rewards. To avoid lengthy classifications, I'll roughly define a game as a system furnished with a goal. Players make decisions given some uncertainty (whether the roll of a die or the strategy of an opponent) in pursuit of that goal. Play is all about the unknown and learning how to navigate it. "Fun is just another word for learning," writes game designer Raph Koster. Dice and chess have little in common except that players work to predict the world in both cases. In gambling, players hope to anticipate the environment. In chess, they hope to anticipate their opponents. A game might look like an auction (with the goal of acquiring a coveted good at a reasonable price), social media (with the goal of attracting attention), or *SimCity* (with the goal of designing a sustainable digital metropolis).

Chance is central to many games because randomness is nature's foundational search algorithm. Random mutations are the driving engine of evolution. Modern scientific experiments incorporate randomization to eliminate the influence of scientists' personal choices and biases. Machine-learning engineers use randomness in their models to jolt the system out of local minima and explore the solution space more thoroughly. It's also at the heart of our ancient drive to play. Before humans, before mammals, this drive emerged with the earliest animals.

Play expanded nature's search strategy of randomness into the behavioral realm. It's been crucial to the emergence of intelligence. Evolution responds to changes in the environment on relatively slow timescales. Innovation happens only as fast as genetic mutations can

spread through a population. The development of the nervous system allowed animals to rapidly respond to their environments—to migrate when the climate changes, say, or learn how to avoid an invasive poisonous species. Hardwired reflexes can be maladaptive, however: take, for instance, many cats' absurd overreaction when they suddenly see a cucumber, triggered by their instinctual fear of snakes. Play uncouples behavior from rote instinct, facilitating flexibility. It draws chance into the realm of experience; it's a safe platform to test the unknown. Play opens new realms for animals, allowing them to build an inventory of adaptable, robust behavioral programs. In play, animals develop a range of strategies through random exploration: If condition A, try B. If condition C, try D, then E. In place of the reactive response of a hardwired reflex, play enables animals to explore a suite of options, some more adaptive than others. It is the crucible of invention and a learning system that mimics the genius of evolution. In natural selection, mutations resulting in bodies that meet the demands of their environments are rewarded with survival. In play, behaviors that correctly anticipate the demands of the environment persist. Play is to intelligence as mutation is to evolution.

This is also why play so often involves being out of control in some way. Otters sled down muddy hills; birds surf on the wind; wrestling children tumble. Play is practice for the unexpected. Playing animals get themselves into purposeful but safe trouble, a strategy that helps them learn how to best escape real trouble. Play allows our brains to build more robust models of the world, sampling from experiences we wouldn't otherwise encounter. It's also how we develop and stress-test social relationships. Play is an activity that brings the wilderness of reality into the realm of the understood.

While play is how animals explore their bodily agency, games are how people explore their mental agency. Games have helped humans improve their powers of reasoning and interactive decision-making

for millennia. The first evidence of game boards dates to about ten thousand years ago, just as cats were beginning to be domesticated and when agriculture was still an emerging technology. Limestone boards carved with rows of depressions were common in Neolithic households of the Near East, potentially used for a counting game similar to mancala. The game may have helped players understand basic mathematics. By manipulating pebbles for fun, they would have grown more familiar with the abstraction we now call numbers.

Beyond the pleasure they bring us, games have reshaped how we think. Games are mental practice, empowering players to develop cognitive skills like strategy, numeracy, and theory of mind in a safe domain. Chess, Go, and related games develop social and strategic thinking. Language and knowledge games like trivia, Scrabble, and puns develop mnemonic abilities and literacy. Games of chance help people develop counting skills and probabilistic estimation. Ninth-century Chinese printmakers invented the first playing cards, founding games of incomplete information. Players, unable to see one another's cards, have to reason about their opponents' psychology and intentions. These games sharpen our understanding of each other. Games, says AI researcher Julian Togelius, are like casts of the mind. Each is fit to a different cognitive capability, and new games emerge as we identify new mental niches and functions. Games like *Pokémon* scratch our itch to collect things, while games like *Tetris* satisfy our compulsion to organize things. Game design is a cognitive science. Games themselves are like a "homebrew neuroscience," says game designer Frank Lantz, "a little digital drug you can use to run experiments on your own brain." We see ourselves—our biases, weaknesses, strengths—more clearly through play.

In life, we often learn the rules of the world by observing the consequences of our actions. For example, if we touch a hot stove, we learn that it's painful and that we should avoid touching it in the future. We infer rules ("don't touch hot stoves") from consequences

(“ouch, that hurt”). In games, consequences are determined by rules. Players must reason forward about the ramifications of the decisions they make. In order to infer the effects of their choices, chess players must understand the rules behind how pieces interact. Games train us to explicitly reason about effects—a crucial mental skill for humans living in complex societies mediated by rules.

In fact, as we’ll continue to explore in this book, games are at the very heart of how we conceive of understanding itself. The late physicist Richard Feynman expounded on this in his famous Caltech course, which he taught in the early 1960s. “What do we mean by ‘understanding’ something?” he asked of the crowded lecture hall. We can say we understand a system if we know the rules governing it:

We can imagine that this complicated array of moving things which constitutes “the world” is something like a great chess game being played by the gods, and we are observers of the game. We do not know what the rules of the game are; all we are allowed to do is to watch the playing. Of course, if we watch long enough, we may eventually catch on to a few of the rules. The rules of the game are what we mean by fundamental physics. Even if we knew every rule, however, we might not be able to understand why a particular move is made in the game, merely because it is too complicated and our minds are limited. If you play chess, you must know that it is easy to learn all the rules, and yet it is often very hard to select the best move or to understand why a player moves as he does. [. . .] We must, therefore, limit ourselves to the more basic question of the rules of the game. If we know the rules, we consider that we “understand” the world.

The rules of a system (or game) are its most efficient representation. We can’t hope to know every possible move in a game. The

enormous space of all legally possible Go games (somewhere between  $10^{800}$  and a googolplex, or  $10^{10^{100}}$ , unfolds as a direct consequence of the game's three fundamental rules, playing out over a nineteen-by-nineteen grid. Once we know the rules well enough to predict or explain a move, we can claim to understand the game, even without elaborating all its possible outcomes. Prediction, as we'll see, is the main currency of the brain. It feels rewarding to have a predictive model of the world. Yet knowing the rules of a system doesn't always mean we can anticipate their outcomes. Incredibly complex dynamics can arise out of deceptively simple rules. Game models can manufacture the illusion that we understand a system better than we do, echoing Pythagoras's obsession with rational numbers obscuring the complexities of material reality.

Games have endured because they represent a model of how our minds work. They're the outgrowth of a learning system that has been central to the evolution of intelligence. Play is a tool the brain uses to generate data on which to train itself, a way of building better models of the world to make better predictions. Perhaps this is also why games have been traditionally associated with divination: throughout history, people have harbored an enduring intuition that games are somehow linked to knowing the future. Cards, dice, and lots have long been used as decision-making tools—an instinct that, as we'll discover, helped to de-bias human choices. The notion of rules and their consequences sits at the root of what we mean by "understanding." None of this is an accident. Games are more than an invention; they are an instinct.

Games have been a powerful generator of knowledge. Though rithmomachia was saddled with Pythagorean dogma, other games yielded profound mathematical insights. Their contemplation gave rise to probability theory and modern economic thought, as well as new ideas in moral philosophy and artificial intelligence. In contrast to

Pythagoras and his cover-up of irrational numbers, the mathematician John Conway discovered an immense new universe of numbers, now known as the surreal numbers, in 1974, after contemplating the end-games of Go. This was the largest magnitude of infinity that mathematicians had discovered in a century. Conway was stunned by the immensity of his revelation. It put him in a stupor for weeks—he felt as though he’d stumbled upon a new continent.

Even so, games are ultimately mathematical objects. We can use them to generate knowledge about simulated worlds, but this knowledge doesn’t necessarily translate to reality. The orderly randomness of dice is a poor model for the untamed randomness of the real world. This discrepancy contributed to the 2008 global financial crash because traders failed to incorporate appropriate measures of risk into their bets. Game theory, once an arcane corner of pure mathematics, now serves as the foundation of modern economics—despite being a poor model of real people. This hasn’t stopped companies and academics from employing game theory in the design of economic and political systems that underpin our everyday lives.

We should be especially careful about using games as a lens for understanding people, because the metaphors we use to understand ourselves matter. In defending his controversial theories about human behaviors, the twentieth-century psychologist B. F. Skinner claims that “no theory changes what it is a theory about; man remains what he has always been.” Refuting this claim lies at the heart of this book. It’s certainly true in realms like physics: atoms survived millennia unscathed by our early, incorrect theories about them. An electron’s orbit is entirely determined by the attractive and repulsive physical forces buffeting it, untouched by our models. But people aren’t passive objects of physics. People, unlike atoms, learn. They are independent agents who make choices dependent on their beliefs about the world and how it works—and their beliefs about themselves. As the engineer

Edsger W. Dijkstra writes, “The tools we use have a profound (and devious!) influence on our thinking habits, and, therefore, on our thinking abilities.”

At the individual level, a person who believes humans to be nasty and brutish will make different choices in life than one who believes that human nature is gentle and generous. Their beliefs affect their politics, their priorities, their prejudices. At the societal level, an incorrect model of people, put into practice in the economic spheres they inhabit, can warp their behaviors and radically change their experience of life. A game rewards its players for adopting the precepts from which it's built. Its rules dictate whether players should cooperate or compete, play honestly or cheat. Game theorists and game designers increasingly shape our shared civic circles, and we're rewarded for adopting their models of humanity in games we cannot escape. It's never been more critical to examine how games have come to dominate modern thought, and how we might clarify our thinking from their rule.

## How Heaven Works

To learn is a natural pleasure, not confined to philosophers, but common to all men.

ARISTOTLE

In the early twentieth century, a fever swept through the world, hastened by armies fanning out across Europe during World War I. Previously unknown to science, it was popularly deemed “sleeping sickness.” People across the world fell into lethargy, sometimes instantaneously. A British doctor recalled the first case he encountered: a healthy girl walking home from a concert simply folded in half and fell to the ground. Within half an hour, her sleep was so deep that she could no longer be roused; within twelve days, she was dead.

Encephalitis lethargica, as it’s now known, is a disease without a fixed character. Its elastic nature is reflected in the diversity of its symptoms. With its most common form, patients are overwhelmed by the desire to sleep, though even in this pseudosleep they remain dimly aware of what is happening around them. Some patients recover. Others progress to a chronic form of the disease, characterized by a motley collection of symptoms: euphoria, ~~increased sexual drive~~ “excessive” pun making, tremors, muscle rigidity, hallucinations, and self-mutilation. An eight-

year-old girl pulled out all her teeth and plucked out both of her eyes. A seventeen-year-old boy became obsessed with horrible odors. He sought out armpits and feces and hoarded trash in his room. Some patients progress to paralysis, coma, or death. Patients languish in long-term care for decades, still as statues, stone-faced and trapped in perpetual slumber. The cause of the disease remains unknown to this day. Its effects, doctors later found, could be mitigated with a then obscure chemical now known as dopamine.

The neurotransmitter dopamine suffers one of the most persistent branding problems of any chemical. This is much to its own credit: it simply does so much. Today, it's commonly mischaracterized as hedonism's biological counterpart, the "pleasure molecule." But for decades after its discovery, scientists thought little of it. It was first synthesized in 1910 as an intermediary in the production of adrenaline, which was then used as an asthma medication. Dopamine continued to be found, here and there, in various bodily tissues. Still, it was dismissed as nothing more than a way station for the chemicals, like adrenaline, that were thought to matter. A traditional Ayurvedic medicine would soon upend this.

In India, high blood pressure, fevers, and insanity had been treated for centuries with the flowering shrub sarpagandha. Mahatma Gandhi, who suffered from hypertension, devoutly took six drops of sarpagandha tincture in his tea daily. Sarpagandha was later "discovered" in the early 1950s by the American doctor Robert Wilkins, after having been studied in clinical trials by Indian scientists for over a decade. Chemists isolated its active compound, reserpine, which became a popular blood-pressure medication and antipsychotic. It was also sometimes used as an animal tranquilizer: in high doses, reserpine rendered animals catatonic. No one knew precisely how.

In 1957, the Swedish researcher Arvid Carlsson and some of his colleagues injected mice with reserpine. They found that it depleted

the animals of several chemicals, including dopamine and its downstream products, such as noradrenaline. Noradrenaline was then known for its effect of arousing the body to action. The scientists therefore expected that injecting noradrenaline back into the catatonic animals would restore their ability to move. It did not, but injecting them with L-dopa, a precursor to dopamine, did. In some cases, the animals even grew hyperactive. This, along with Katharine Montagu's 1957 discovery of dopamine in brain tissue, definitively established dopamine's role as a neurotransmitter—a finding that earned Carlsson a Nobel Prize. The notion that neurons could talk to one another through bulk chemical transmission was still baffling. At the time, electrical signaling, in which individual cells send messages to select partners, was the more familiar mode of neural communication. Here, a single message encoded in the concentration of dopamine was being broadcast to entire brain areas. But what was it saying?

Oleh Hornykiewicz, a Viennese neurologist, noticed that the effects of dopamine depletion in animals echoed the symptoms of the common neurodegenerative disorder Parkinson's disease. Parkinson's was first described in 1817 by the surgeon James Parkinson, whose patients displayed "involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forward, and to pass from a walking to a running pace: the senses and intellects being uninjured." Parkinson, an avid fossilist and naturalist, described it as a new "species of disease," one that he'd drawn out from the jungle of nebulous neurological symptoms, like a botanist classifying a newly discovered flower.

Though Parkinson's is most common in elderly patients, it was also frequently seen in people with chronic sleeping sickness. In the years following the original sleeping sickness epidemic, seemingly catatonic patients flooded care facilities. In the 1960s, more than thirty years after the first epidemic, Hornykiewicz began collecting the brains of

recently deceased patients. He discovered that the brains of patients with parkinsonism were low in dopamine. What if he could reverse parkinsonism with L-dopa, just as Carlsson had revived catatonic mice? He wasted no time and immediately delivered his store of the drug to a colleague in charge of an elder-care facility in Vienna. The medical staff administered L-dopa to Parkinson's patients. The results seemed miraculous. Patients who had been immobile for decades, imprisoned by sleeping sickness, stood up and walked, their voices and personalities returned to them.

Dopamine replacement therapy became the standard of care for Parkinson's patients, as it remains to this day. It's not a cure, unfortunately; it becomes less effective over time. In his breakout work, *Awakenings*, the neurologist Oliver Sacks recounted the brief but luminous revival of encephalitis lethargica patients, likening them to "extinct volcanoes" erupting back into life. With this spectacular success, dopamine rocketed into the scientific spotlight, becoming one of the most researched neurotransmitters. Scientists have since discovered dopamine in almost every animal with a nervous system, making it evolutionarily ancient; it's involved in the movement of flatworms, fireflies, flounder, and falcons alike. All of this seemed to cement dopamine's place as the neurotransmitter responsible for movement, but nature is never so simple.

- - - - -

No sooner had the computer been conceived than people wondered whether it could think. The world's earliest programmer, Ada Lovelace, was the first to realize that Charles Babbage's proto-computer, the Analytical Engine, could do more than crunch numbers. It might one day

make music, prove math theorems, play games. Yet, she argues, it could only carry out instructions; it had “no pretensions whatever to *originate* anything. It can do whatever we *know how to order it to perform*.” A century later, Alan Turing, whose work led to the creation of more powerful and flexible computers, believed that these machines would one day be capable of much more. He responded to “Lady Lovelace’s Objection”: “A better variant of the objection says that a machine can never ‘take us by surprise.’ [ . . . ] Machines take me by surprise with great frequency.” He predicted that computers would eventually be able to generate new knowledge and understanding. But first we’d have to teach them how to learn.

Designing a machine that captured all the complexity and knowledge of a fully formed adult, Turing anticipated, would be difficult. Its intelligence might simply be credited to its creators—it would not, as Lovelace proclaimed, originate anything. “Instead of trying to produce a programme to simulate the adult mind,” Turing writes, “why not rather try to produce one which simulates the child’s? If this were then subjected to an appropriate course of education, one would obtain the adult brain.” He imagined that this “child” machine could be, like children, educated through punishment and reward. This entailed solving two separate problems. First, researchers would have to build a computer program that reflected a child’s capacity to learn. Then they’d have to design its education. Today, we’d call the solution to the first problem a learning algorithm and the solution to the second its training data.

Turing suggested board games as an ideal training arena for bringing up these child machines. Games are miniature worlds, abstractions of interaction, and their discrete nature makes them beautifully suited to computers. In games like chess and checkers, a few simple rules unfold into campaigns of astronomical complexity. Games have long been thought to demonstrate their players’ intelligence. A player’s progress from novice to expert is conveniently assessed and tracked by metrics

like their ranking. Thus, games could serve as both a training regimen for learning agents and a benchmark for measuring their intelligence. Structured games standardize players' agency, evening the playing field. Games impose symmetry. Players work against each other toward a shared goal: winning. They're given identical pieces and bound by the same rules. To measure time, we use clocks; to measure space, rulers. Games came to be adopted as a meter of intelligence. They are an ancient form of rhetoric—a debate not of words but of choices, actions parried across time and space.

John McCarthy, the researcher who coined the term *artificial intelligence*, also provided one of its most enduring definitions.\* “Intelligence,” he writes, “is the computational part of the ability to achieve goals in the world.” We might assess a machine's intelligence through conversation, as in Turing's “imitation game,” wherein players—both computer programs and real people—vie to convince a human judge that they're human via typed conversation. Or we might call a program that can beat a human in chess intelligent.

Turing was, by his own admission, a mediocre chess player. He nurtured his dream of building an artificially intelligent chess program in conversations with his favorite chess partner and colleague, Donald Michie, who was equally mediocre at the game. Both Turing and Michie worked at Bletchley Park during World War II, devising methods to decode encrypted Axis communications. Michie had come to Bletchley Park almost by accident, having enrolled in a cryptography class with hopes of doing “something unspecified but romantic” for the war effort. Today, we know that his research was instrumental to the Allied win: Michie's insights helped crack the Lorenz cipher and vastly improved the Colossus II computer. Thanks to his work, communications that

---

### Copyrighted Material

\* I use the term *artificial intelligence* (AI) not because these programs are necessarily meaningfully intelligent themselves but because this is the aspiration of their creators.

had previously taken days to decode took mere hours, enabling Allied forces to dodge ambushes and anticipate enemy troop movements.

Michie was smitten with Turing's vision: "I resolved to make artificial intelligence my life as soon as it became feasible," he writes. But computers as a technology were vastly outpaced by researchers' ambitions for them. They were also impossibly expensive, and rare outside military circles. After the war ended, Michie returned to academia and shifted his focus to genetics, inspired by his boyhood love of mice. He was a middling biologist; his greatest contribution was to the research of his wife, Anne McLaren, whose work paved the way for in vitro fertilization. But Michie never gave up on the dream of artificial intelligence, despite having no access to computers.

In 1961, Michie took a bet with a colleague who was skeptical that machines could learn. Michie would win the bet using nothing more than three hundred-odd matchboxes and some tinted glass beads. He built a learning system that could play tic-tac-toe and named it MENACE: Matchbox Educable Noughts and Crosses Engine. Each matchbox represented a state of the tic-tac-toe board, and the boxes were stacked in piles representing sequential moves for all possible arrangements of X's and O's. The beads, nine colors in all, indicated every possible next move from the current state. Initially, Michie stashed an equal number of colored beads in each matchbox drawer. On every turn, he would randomly pull a bead from the matchbox, which determined MENACE's next move and the game board's next state. The color of the glass bead taken from the next box decided the next move, and so on. The drawers were left open as a record of which moves had been played. If MENACE lost at the end of the game, Michie didn't return the beads to the open drawers, lessening the probability that MENACE would take those actions in the future. If the game ended in a draw, he added **Copyrighted Material** one extra bead of the appropriate color to each matchbox. If the game resulted in a win, he returned three beads to

each box. MENACE learned through reinforcement, with Michie rewarding correct moves and punishing incorrect moves.

At first, MENACE was terrible: “Random games have an extremely idiotic character, as can readily be verified by playing through one or two examples,” Michie writes. But over the course of hundreds of games, the colored beads were redistributed within the boxes in a way that made winning moves more likely and losing moves less likely, much like deepening the ruts in the path leading to victory. Eventually, MENACE played tic-tac-toe perfectly. A mindless system had achieved expert performance through trial and error alone.

Michie had drawn inspiration from the trial-and-error theory of learning that had been the focus of psychology research earlier that century. The psychologist Edward Thorndike was determined to understand “animal stupidity”—how seemingly intelligent, purposeful behavior can be built from simple associations. He placed cats inside a puzzle box and put scraps of fish outside the box, just out of reach. The box had a trapdoor that opened only if its occupants pressed a lever. Once a cat’s accidental step on the lever resulted in the door opening and a fish reward, the animals quickly learned, in subsequent trials, that they could escape the box by pressing the lever again. Thorndike named this the “law of effect”: a behavior that leads to a pleasurable outcome will be selected and repeated. A behavior that leads to an unpleasant outcome will be extinguished. It’s a bit like evolution, where gene variants that confer fitness are rewarded with survival in the population. In trial-and-error learning, a random action that leads to reward “survives.” Instead of being recorded in DNA, it’s stored in memory. The strength of its association with reward dictates how likely that action will be reproduced in the future. In MENACE’s case, winning a game of tic-tac-toe ensured the survival of moves that led to the win. Victory **Copyrighted material** rewarded winning moves, whereas defeat removed copies of losing moves from the pool of available actions.

Michie later applied similar learning methods to the endgames of chess. He was fond of a saying attributed to the Soviet mathematician Aleksandr Kronrod: chess was to the study of artificial intelligence as *Drosophila*, the fruit fly, had been to genetics. As studies of the simple *Drosophila* genome, consisting of only four chromosomes, had paved the way for understanding more complex human genetics, Michie pronounced that “the use of chess now as a preliminary to the knowledge engineering and cognitive engineering of the future is exactly similar, in my opinion, to the work on *Drosophila*.”

Meanwhile, in the US, engineer Arthur Samuel had been working on a program that could play checkers. He’d originally conceived the project as a quick stunt to drum up funding to finish work on a computer he was building, not realizing that it would become a focus of his research for the next thirty years. Samuel wasn’t a particularly strong checkers player, and his early efforts to build a checkers engine only performed as well as he could program it. To create a system that could exceed his own limited talents, he took up Turing’s goal of building a program that could learn for itself, popularizing the phrase *machine learning* in 1959.

By the late 1950s, Samuel had struck on a game training regimen that would become central to the field of AI: self-play. In self-play, a program is trained against copies of itself. These copies tweak their parameters after each game to improve their win rates. Self-play has been so successful in part because players learn best from well-matched opponents. When a program is pitted against a copy of itself, it’s always met by an adversary playing at its own level. If paired with a much better player, it might never learn, because it will always get creamed. If paired with a weaker opponent, it might win so easily that it’s never challenged to improve.

However, training by self-play requires copious training time—a luxury few researchers had, given that computers were so rare. By

then, Samuel was working at IBM, whose executives were not particularly enthused by his checkers side hustle. After his colleagues had gone home and the company's computers stood idle, Samuel snuck into the lab to train his program every night from midnight to seven in the morning. By 1956, the program was good enough to contend with novice players. Impressed by Samuel's furtive progress, the president of IBM arranged a public demonstration of the program, and the price of IBM stock jumped fifteen points overnight.

- - - - -

At the same time that computer scientists were attempting to engineer intelligent systems, neuroscientists were working to untangle the biological basis of intelligence. In the 1980s, a young medical doctor named Wolfram Schultz established his research lab focusing on Parkinson's disease. He intended to record the electrical activity of dopamine neurons to better understand their role in movement. Though dopamine neurons comprise less than 1 percent of all neurons in the brain, they're relatively easy to find and record from because they're concentrated in a cluster of midbrain areas. Schultz and his colleagues implanted macaque monkeys with electrodes and measured the activity of dopamine neurons while the animals performed simple motor tasks. Based on their known involvement in movement, the dopamine neurons might be expected to fire whenever the animals moved. Instead, they fired when the animals were given a reward.

This wasn't entirely unexpected. Though dopamine was known to be involved in movement, what is movement ultimately for, save the pursuit of reward and avoidance of punishment? Biologists had been slowly piecing together the basis of this orienting system. Nineteenth-century

scientists, smitten with the newly discovered phenomenon of electricity and dimly aware of its involvement in the nervous system, stuck stimulating electrodes in the brains of humans and animals, guided more by exuberance than reason. By the early twentieth century, the neurosurgeon Wilder Penfield had refined this technique and used it to build an atlas of brain functions. He annotated a map of the brain with the effects of stimulation: the visual cortex was labeled “lights and shadows.” Stimulation of a region he called “memory” caused patients to become immersed in flashbacks so vivid that it seemed as if they were playing out in the present. The frontal cortex, involved with executive functions, was termed “silence” because it extinguished patients’ internal monologues. “Throughout my career,” Penfield writes, “I was driven by the central question that has obsessed both scientists and philosophers for hundreds of years. Are mind and body one?” It appeared—almost too tidily—as though mental functions could be directly mapped onto distinct brain regions.

A flurry of stimulation experiments ensued, linking brain areas to their evoked behaviors. Stimulating deep in the brains of rats caused them to become aggressive. Stimulation of another area made them fearful, and these rats avoided returning to the places where they’d been zapped. In 1953, a postdoc named James Olds made a fortuitous discovery while performing his first electrode placement surgery in a rat. His surgery had been a fraction of a millimeter off target. The electrode impulse didn’t make the rat fearful; rather, it seemed to be rewarding. The rat repeatedly visited the area where it had been zapped, instead of avoiding it. Like in a game of hot and cold, the animal could be “pulled” to any spot by delivering an electrical stimulus after each move in the right direction.

Olds quickly rigged up a lever system that allowed the animal to self-administer stimulation. The implanted rat pressed the lever incessantly. Later studies revealed that rats would choose to

stimulate this area rather than eat, drink, or mate. They'd press the lever even if it was paired with a painful electric shock. Like addicts, animals would press the lever all day, every day, often until collapse, convulsion, or sometimes death. Rats, cats, monkeys, and even a dolphin died this way. These results captivated popular media, and futurists predicted that all striving and desire would soon be replaced with electrophysiological contentment. Science-fiction writer Isaac Asimov concludes: "Evidently all the desirable things in life are desirable only insofar as they stimulate the pleasure center. To stimulate it directly makes all else unnecessary."

Stimulating this area, it was later discovered, elicited dopamine release. Dopamine's reputation as the "movement molecule" was soon eclipsed by its new identity as the "pleasure molecule." Yet this was always a dubious association. Humans implanted with electrodes in an analogous brain region didn't experience pleasure when that area was zapped. It was more like the feeling of *wanting*. Subjects reported feeling compelled to press the lever, as though scratching an itch. Absent the ability to talk to animals, we can't say whether what we call reward feels pleasurable to them. Rather, it reinforces the behaviors that lead to it, as with Thorndike's law of effect. Dopamine isn't a measure of pleasure. It's more like the "motivation molecule," driving subjects to take actions to get what they want.

Schultz's new data complicated this picture. His team had confirmed that dopamine neurons fired when the animals received a reward—but only when they didn't expect it. The scientists had trained the monkeys to tap a lever to get a juice reward after a light cue was flashed. The untrained animals behaved, at first, randomly. They pressed the lever erratically, and this behavior was occasionally reinforced with juice when they happened to push at the right time. The dopamine neurons of these untrained animals fired when the reward was delivered. But once the animals were experts and had made

an association between cue and reward, the dopamine neurons no longer fired in response to the juice delivery. Rather, they fired in response to the light cue that preceded it. Even more tellingly, if the light came on but the reward was withheld, the dopamine neurons reduced their firing. They were signaling what they expected to happen. The dopamine neurons weren't tracking movement, exactly, or reward; they were tracking *belief*. They even signaled when something they expected to happen didn't happen, as though surprised.

Psychologists already knew that surprise was an important part of learning. Animals don't always learn from simple repetition and reward. Surprise indicates that something remains to be learned—its attention-grabbing effect potentiates learning. Even infants pay more attention to surprising stimuli, such as a video of a ball rolling uphill. Surprise happens when experience violates expectations, and the brain uses this as a learning signal. This is why language programs often teach humorous sentences with unexpected associations: “Why is the banana wet?” and “My horses collect teeth.” When teachers employ surprise, students are more likely to remember the lesson.

Schultz's findings also resembled classic results in psychology, like Pavlov's dogs salivating in anticipation at the sound of a bell signaling that food was soon to come. Decades later, the psychologist B. F. Skinner continued this line of research and dubbed behavioral feedback “reinforcement.” Negative reinforcers, like an air puff, could abolish certain behaviors, while positive reinforcers, like food, strengthened them. He believed that this simple valence, the push and pull of desire and avoidance, was the axis along which all intelligent behavior could be built. In this way, he hoped to reduce all the complexity of animal behavior to a simple sort of physics, shaped by the attraction and repulsion of reward and punishment. Skinner believed that animals could be effectively “programmed” with reinforcers to behave in arbitrary ways. He thought that this was as true for humans as it was for