Arithmetic Is All You Need

The Human Story of Computer Intelligence

Erich Elsen

Contents

Introduction

Predicting the future is not impossible, it's how we survive. Every living thing predicts the future — doing it better than others means survival and procreation. The game will be in that valley, the lions are the other way, the safe water hole is over yonder, next year will be drought, and she will like when I dance; someone who could predict all of these things would have no trouble surviving and passing on their prophetic abilities.

Our brains are evolved prediction machines. We make predictions in fractions of seconds about what other people want, if they can be trusted, and where that thrown spear will land; and we make predictions about longer time scales as well: used car prices are going down, I'll wait to buy; my parents want me to be a lawyer, but I think I'll be happier trying to be an actor; my chances of succeeding will increase if I go West to grow up with my country. We don't make these predictions in the language of mathematics, they happen through a process — thinking — that is still almost a complete mystery to us despite its ubiquitous nature.

But for thousands of years, humanity has also been learning to make predictions with the language of mathematics and arithmetic. Scientific laws are a way of saying "here's how to do arithmetic to predict this particular thing." Newton's laws of motion allow us to do arithmetic to predict how an object will move in space. The laws of thermodynamics allow us to do arithmetic to predict how heat will move. Combining fluid dynamics, thermodynamics and the laws of motion allow us to predict something as complicated as the weather.

Designing something like a bridge, or an airplane, is more than making a prediction; you have to have something to make a prediction about in the first place. Given a bridge, static mechanics allows us to predict how much load it can handle, but that's a textbook problem. Real bridges must satisfy a long list of requirements: span this distance, handle this much weight and wind, cost less than ten million dollars, and look nice. Before computers, humans would take all these requirements into account and come up with a proposed design that in their experience would work. Then they calculated if it would. If it didn't they made some tweaks and calculated again. One reason truss bridges were so popular before computers is that the calculations involved were tractable for a single human.

Then, in the mid-twentieth century, something fundamental changed. Our computational ability had advanced so far that we could do enough arithmetic to work backward from the requirements to the thing that would satisfy them. A computer could calculate the loads for one design, and then automatically propose the next one based on the weaknesses of the first. Then calculate the loads of that design and so on, until a final design satisfying all the requirements was reached. Optimization reshaped engineering — it is how we design artifacts too complex for intuition alone, like the shape of stealth airplanes and the antennas in our phones.

Optimization works just as well on virtual machines as physical ones; for example, optimizing virtual machines that predict which words come next, called language models, is the basis of what is colloquially known as artificial intelligence, or AI. Trying to formalize language the way we had formalized physics had not worked, but optimizing these machines, these language models *did*. When optimization began to produce machines that operate in domains we inhabit, producing language we recognized as familiar, we applied a different label. We called it intelligence. Computer intelligence.

Richard Feynman once wrote: "What I cannot create, I do not understand." It is a demanding ideal, and one we have never fully lived up to. We have long created things we did not understand — compasses guided navigation before magnetism was explained, and steam engines powered industry before thermodynamics existed. And we understand phenomena we cannot create, from stars to black holes, but the relationship between creation and understanding is now changing in a new way. We build meta-machines that produce other machines through optimization, a kind of automated search. What emerges is a form of creation in which we understand the method but not always the artifact: we can reproduce the result without being able to fully explain it. The gap between creation and understanding is not a failure; but we need to accept that a complete understanding of everything we create is no longer possible. If a divinity is granted through creation, then in a small sense, we have become lowercase gods.

My own path into this story came from a lifelong passion — *citius* — faster. Whether running, on horseback, or behind the wheel, I like going fast. As a graduate student at Stanford I worked on unlocking the computational power of graphics processors (GPUs), hardware built for video games but capable of far more arithmetic per second than conventional processors, to make scientific simulations orders of magnitude faster.

A few years later, researchers used that same hardware to train, that is to optimize, an artificial neural network called AlexNet that could identify objects in images as accurately as humans. I saw this and knew it was going to be big; bigger, as it turned out, than I imagined. I joined one of the first labs dedicated to using GPUs to train neural networks, and over the next decade, including work at DeepMind, I led the training of some of the largest neural networks ever built and contributed to quantifying the relationship at the heart of this book: that prediction quality improves predictably as computation increases.

Language had defeated every previous attempt at prediction by hand-designed rules, but optimization succeeded where human ingenuity alone could not. What we were doing when we trained these large language models was creating something that could predict the future — something that scientists and engineers had been doing for thousands of years; they had just been making predictions about systems whose behavior followed discoverable laws. Now we just needed to scale it up, make the networks bigger and use more arithmetic to optimize them. And if we were creating something that can predict language, that can *generate* language; is that not some kind of intelligence?

Rarely are there moments in science where the path to progress is laid out so clearly; the caricature of lone geniuses undergoing years of frustration followed by Eureka moments and fortuitous discoveries is not completely wrong; but here, now, there was a crank and all we had to do was figure out how to turn it faster. Luckily, I had just spent my PhD working out how to turn a slightly different crank.

As part of that work, I began to realize that if I changed the axes slightly, the same relationship has been true for thousands of years. On one axis, the amount of arithmetic humanity can bring to bear; on the other, the complexity of what we can predict. The plot is a line going up and to the right. I became curious about the path humanity had taken along this line. What did we learn to predict, and how? What were the bottlenecks that held us back, and how were they broken? What might we predict next — and what happens when the thing being predicted is us?

This book follows that line and visits the people and ideas along the way that made arithmetic faster and better prediction possible. At the end are the open questions that no history of computation can answer on its own: questions about what we choose to build and whether we will navigate the undiscovered country with fear or hope and wonder.

Chapter 1

As We May Think

> "Consider a future device … in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility." — Vannevar Bush, "As We May Think" (1945)

> "Every one knew how laborious the usual method is of attaining to arts and sciences; whereas, by this contrivance, the most ignorant person, at a reasonable charge, and with a little bodily labour, might write books in philosophy, poetry, politics, laws, mathematics, and theology, without the least assistance from genius or study." — Jonathan Swift, *Gulliver's Travels* (1726)

Computer intelligence is the culmination of a long human project: the collection and organization of our knowledge. It has always been distributed and the bottleneck has always been how effectively any individual could access what was already known. Knowledge of copper smelting took a millennium to travel the silk road to China and the secrets of silk farming took even longer to go the other way.

Writing was a step function in the speed that information could travel and the first technology to make our knowledge less distributed. But solving one problem always leads to another. Even in antiquity, human knowledge had become a collective enterprise spanning cultures and generations — civilization advanced by making accumulated knowledge accessible, but no individual could absorb it in full.

The Roman architect Vitruvius, writing in the first century BC, understood the debt this implied. "It was a wise and useful provision of the ancients to transmit their thoughts to posterity by recording them in treatises," he wrote, "so that they should not be lost … the ancients deserve no ordinary, but unending thanks, because they did not pass on in envious silence, but took care that their ideas of every kind should be transmitted to the future in their writings." The act of writing things down and refusing envious silence was the beginning. The Library of Alexandria is said to have held 200,000 papyrus scrolls, perhaps more, far more written knowledge than any one person could hope to read in a lifetime. That solution led to a new problem: once knowledge was recorded at a scale beyond any individual's capacity to read, let alone remember, how would anyone find the specific piece they needed?

Consider the problem as it would have presented itself in practice thousands of years ago. A small collection of a few dozen scrolls needs no organizational system at all because the librarian knows every work personally — where it is stored, and what it contains. A scholar arriving with a question can simply ask, and the librarian can retrieve the relevant text and its location in moments.

But this doesn't scale very far; at a hundred scrolls, a diligent librarian might still manage, though they may need to pause and think. At a thousand, memory alone is no longer sufficient. No one person can remember the contents of a thousand works with enough detail to reliably match a specific question to a specific passage. The librarian may remember that a certain subject was treated somewhere, but not in which scroll, or may confuse two works that treated similar topics differently. At ten thousand scrolls, human memory is an island in a vast ocean. Without some external system of organization, knowledge that exists in the library is the tree in the proverbial forest — it might as well not exist at all.

The early librarians at Alexandria solved this by organizing the library's holdings physically: scrolls were stored in rooms or alcoves grouped by subject, with tags attached to the end of each scroll identifying the author and title. Within a subject area, works were sometimes arranged alphabetically by author, which was much better than memory alone, but far from perfect. A work that touched on multiple subjects could only be stored in one place. And not all questions mapped obviously to a particular category. As the collection grew, the subject groupings themselves became unwieldy: "philosophy" might encompass thousands of scrolls spanning logic, ethics, natural philosophy, and metaphysics, with no further subdivision to guide the searcher.

Callimachus, a scholar and poet who worked at the library in the third century BC, recognized that physical organization alone could never solve the problem at this scale. Rather than relying on the arrangement of the scrolls themselves, he created a separate work, the Pinakes, a catalog of the entire collection. The Pinakes, reportedly 120 scrolls in its own right, listed the library's holdings by subject and author, noted the opening lines of works, provided biographical information about writers, and grouped everything into categories such as rhetoric, law, epic poetry, and medicine. It was humanity's first large-scale bibliographic database: a system of metadata that existed apart from the collection it described. A scholar no longer needed to browse the shelves or rely on a librarian's memory; they could consult the Pinakes to learn what the library held on a given topic and where to find it.

It was an extraordinary achievement, but the core problem that would persist for two millennia remained: even if knowledge existed somewhere, finding a specific piece of it was slow and uncertain. The Pinakes could tell you that a work existed and roughly where to look for it, but it could not tell you whether that work contained the particular fact or argument you needed. One would navigate to a subject area, browse promising titles, retrieve scrolls, and read substantial portions to see whether they contained what one was looking for. The process was measured in hours and days.

Difficulty in searching was a serious problem, but not without its own benefits. By browsing, one is exposed to ideas not explicitly sought. Discovering a new world is as exciting in a library as sighting land from a ship at sea.

For nearly two thousand years, this basic pattern barely changed. Unable to figure out how to make searching scale further, we found a different way to cope with ever accumulating knowledge. Each generation summarized and condensed the knowledge of the last in encyclopedias and textbooks that compressed centuries of learning into single volumes. And much was simply forgotten — works that were wrong, or superseded, or no longer relevant quietly dropped out of circulation. The number of books in a typical library stayed roughly comparable to the number of scrolls in Alexandria not because knowledge stopped growing, but because it was continually being re-summarized and culled.

Across two millennia tools had improved only modestly with the introduction of card catalogs and the Dewey Decimal System. Information finding remained slow and manual. As a child in the 1980s, I was still taught to use card catalogs to find the books and information I needed — at the time, it was the primary interface to human knowledge. You approached a wall of small wooden drawers, each labeled with a range of letters, pulled open the drawer and flipped through index cards scanning titles and subject headings. If you were lucky, you found a potentially useful book and noted the call number. You then walked to the shelves, found the right section, and browsed spines until you found your book. If it had been checked out, or didn't have what you needed, you tried browsing nearby books or started all over. The entire process, from question to answer, might take an hour for a simple search. For a serious research question, it could take days or weeks, and success was never guaranteed.

The frustration with the sheer friction of finding what you knew must exist somewhere led U.S. scientist Vannevar Bush, in 1945, to imagine the device described in this chapter's epigraph. Bush, who headed the U.S. Office of Scientific Research and Development at the time, called it the "Memex:" a desk-sized machine in which a researcher could store all of his books, records, and correspondence on microfilm, and retrieve any item by a system of codes and cross-references. The Memex was never built, but Bush had identified a problem that was about to get much worse: the volume of knowledge was beginning to grow so fast that summarization and curation could no longer hold the tide.

The real inflection came with computers and the internet. Early web directories, typified by Yahoo's hierarchical taxonomy, were recognizably Callimachean. Yahoo did not search the web so much as organize it: human editors placed websites into nested categories such as Science → Astronomy → Observatories or Arts → Literature → Poetry. To find information, users navigated these branches, browsing downward in the hope that the right page had been filed in the right place.

Using Yahoo in, say, 1996 was an experience closer to browsing a library shelf than to what we now think of as searching the internet. You arrived at a page of broad categories. You clicked on one — say, "Science" — and were presented with subcategories: Biology, Chemistry, Physics, Astronomy. You clicked again, and again, each time narrowing the scope but also committing to a path. If what you were looking for had been filed under a different branch, you might never find it. A page about the chemistry of interstellar dust might appear under Astronomy or under Chemistry, but probably not both, and the choice was made not by any algorithm but by a human editor applying judgment. This worked when the web was new — when a team of editors could plausibly visit and categorize most sites of consequence. But quickly, the system collapsed under its own success, and for exactly the reason Callimachus would have recognized: the knowledge grew faster than the catalog.

Full-text search engines such as AltaVista introduced a radically different idea: you could search directly for words. This was revolutionary, but brittle. Queries required choosing the right terms carefully and in advance; ambiguity was punishing. Searching for "apple" returned orchards and computers indiscriminately while "java" yielded islands and programming languages alike. A medical researcher searching for "treatment resistant depression" might find clinical papers, pharmaceutical advertisements, patient forums, and self-help blogs, all jumbled together with no way to distinguish authority from anecdote. The system retrieved pages that contained the right words, but it had no understanding of what those words meant or which sources were trustworthy. The serendipity of library browsing, of stumbling on related ideas nearby on the shelf, was replaced by something much less useful: stumbling over anything that happened to share a common word.

Advertising and search engine optimization made the problem worse. As incentives shifted, pages were increasingly written not for humans but for algorithms, polluting results with keyword-stuffed irrelevance. The web was becoming a library in which many of the books were written to trick the catalog into placing them on the most prominent shelf.

Google exploited hyperlinks, a unique property of the web, to mostly solve this problem by the end of the millennium. By analyzing how pages referenced one another, PageRank inferred which sources were likely to be valuable. A page that was linked to by many other pages was probably important itself. This applied the idea of academic citations, that highly cited papers are more likely to be relevant, at a much larger scale. But it was more complex than just vote-counting, it was vote-counting in a loop.

A link from a highly ranked page counted for more than a link from an obscure one, creating a recursive definition of authority that could only be resolved by computing across the web's entire link graph simultaneously. Every page starts with the same "vote", but then each page's vote is updated by adding the incoming votes and normalizing by the number of outgoing links; the result is that pages with lots of incoming and few outgoing links get more votes. Then the same procedure it done again and new votes are obtained. After 10-30 rounds of this, the "vote" that each page is assigned stops changing and the the final "vote" values are the page ranks. It is a measure of how likely you are to end up at the page if you start at a random website and just start following links at random.

The result was a ranking system that, for the first time, could distinguish between a personal homepage mentioning a topic in passing and an authoritative reference work treating it in depth — even when both pages contained the same keywords.

For two decades, this approach defined the state of the art. If relevant information existed somewhere on the internet, Google could usually point you toward it. The combination of full-text search with link-based ranking created a system that could handle both the scale of the web and the ambiguity of natural language queries with reasonable success. Two generations learned to begin almost any intellectual task by typing a few words into a search box; the interface was so successful that it became invisible, you simply expected it to work. It became a default and we stopped noticing what was still missing.

What was missing was the answer itself. Search could point you in the right direction, but the information you needed was rarely presented exactly in the form you needed it. You still had to read and interpret material spread across multiple sources. If you were trying to understand how a system worked or diagnose a problem search could direct you to relevant pages, but the work of actually using that information remained manual. Someone trying to get an old printer working with a new laptop would find a manufacturer's support page that hadn't been updated in three years, a forum thread where the solution was marked "resolved" with no explanation, a driver download page for a different operating system, and a YouTube video that covered a similar but not identical model. Bringing those fragments together into an actual solution remained entirely the user's responsibility.

Large language models represent the next, and possibly final, evolution of information finding. They can still retrieve documents, but the important difference is what happens next: rather than handing the user a list of sources, they construct answers. We compress an enormous fraction of humanity's written knowledge into these language models and equip them to search for what they might be missing. The resulting computer intelligence can synthesize knowledge on demand, tailored to your context and intent. Ask a question about playing the guitar, or ancient astronomical instruments, or the best way to debug a memory leak — provide the specifics of your situation — and the system generates an answer tailored to your circumstances, drawing on patterns learned from the billions of texts on which it was trained.

This is a better way to find information, but one that fundamentally breaks the tradition of knowing where knowledge came from. When answering factual questions, responses arrive without the sources that let us evaluate them, and can be confidently wrong in ways a list of references cannot. But it is also a fundamentally different interface, one that opens possibilities search never could; you can ask the system to write code, draft a document, or even delegate tasks entirely, letting the system act on your behalf. None of this was conceivable when the interface was a search box returning a list of links. For much of what these systems now do such as writing, coding, and reasoning through problems, the concept of a source doesn't even apply. Their power comes from the feedback loops they create between humans and computer intelligences; those loops expand what is possible for us to accomplish, but the computer intelligences also now shape us in return. Every force has an equal and opposite reaction.

How were these systems built? The first generation learned by reading the existing internet. The models were fed vast quantities of text[1] — books, articles, forums, encyclopedias, technical documentation, conversations — and learned patterns of language by predicting, word by word, what came next. The resulting systems could generate fluent, knowledgeable text across a wide range of subjects, because they had absorbed the statistical structure of how humans write about those subjects.

But very quickly it became clear, based on user interactions, that much valuable knowledge was still missing. It was embedded in interactions that had never before been worth recording in a widely searchable medium.

General principles were recorded in textbooks, but a senior engineer's instinct for debugging, a teacher's sense of which explanation will land with a struggling student, a doctor's clinical judgment built from thousands of patient encounters — actual transcripts of the application of general principles in specific situations were not. So they were created. Companies hired large numbers of people to produce exactly the kinds of documents that hadn't existed before: written demonstrations of expert judgment applied to specific situations.

Building the most capable models became as much a problem of data collection as of raw computation; the data collection required the coordinated work of millions of people and the expenditure of billions of dollars, making it one of the largest deliberate attempts ever undertaken to render tacit human knowledge explicit and reusable by machines. It was, in a sense, the largest bibliographic project since Alexandria — except that instead of cataloging existing texts, it was creating new ones, extracting knowledge that had previously existed only in human heads and human habits.

The result is something that feels like intelligence, but it's strange and savant-like: immensely impressive, shockingly fallible, possessing vast knowledge and instant recall, but little grounding. An example of this is the question, "I want to wash my car. The car wash is 100 meters away. Should I walk or drive?" Humans realize the entire exercise behind going to the car wash is to get the *car* washed, so you must drive. When the question first appeared in early 2026, even the best computer intelligence would sometimes recommend walking for various quasi-plausible but ultimately non-sensical reasons like avoiding the

[1]The text used to train these models was not always obtained with the consent of its authors. Large-scale data collection swept up (and even intentionally targeted) copyrighted works often without permission or compensation. The legal and ethical questions surrounding this practice remain unresolved at the time of writing. It remains unclear how to properly compensate sources whose knowledge is disintermediated. The standard advertising model of the internet will not work here — perhaps we can find something better for both producers and consumers of content.

hassle of parking. They are extraordinarily good at reorganizing what humanity already knows, and conspicuously bad in ways humans are not.

A crucial step towards computer intelligence was to make words, and eventually ideas, amenable to arithmetic. Once language could be represented numerically, systems built to manipulate numbers could begin to operate over meaning as well as symbols.

This was not the first attempt to formalize language. Beginning in the 1950s, researchers in artificial intelligence tried to understand language with rules and logic — an approach that later came to be associated with so-called "expert systems". Consider an airline reservation system that needs to understand the sentence, "Find me the cheapest nonstop flight from San Francisco to Boston next Tuesday after 3 p.m." To answer, it must extract the flight origin, destination, date, time and other constraints.

One simple scheme would be to find the word "from" and then assign whatever comes next as the origin. But immediately there's a problem: San Francisco would break that rule as it's two words. Instead there would need to be a long list of all acceptable place names for origins and a rule for checking if any of them appear after "from." If "nonstop" appears anywhere in the sentence, it is assumed to be a request for a nonstop flight. Dealing with time requires even more hand crafted rules; you have to deal with combinations of days, months, numbers, and modifier words like "next" and "after."

And after all this work, someone asks the same thing in a different way: "I need to be in Boston next Tuesday; I'm in San Francisco now, and I'd rather not stop anywhere if I can leave after midafternoon." And all your rules break.

These approaches could be made to work in narrow domains like airline reservation systems but they were brittle. Every new domain required new rules, and the rules interacted in ways that were difficult to predict. Natural language, with its ambiguity and context-dependence defeated every attempt to contain it in a rule-based framework.

The breakthrough came from a different direction entirely. The insight, anticipated by the linguist J.R. Firth's famous observation that "you shall know a word by the company it keeps," was that meaning could be inferred from context. Words that appeared in similar contexts probably meant similar things. If "dog" and "cat" appeared in many of the same sentences, near words like "pet," "veterinarian," and "fed", then they were probably related in meaning. This was not a definition of meaning in the philosophical sense, but it was something more useful: a way to measure it.

If we associate words with numbers, then we can have a computer program work through large volumes of text and optimize the assignment of numbers to words so that nearby words can be predicted from each other. (We'll discuss this process in more detail later.) At first we might try to associate each word with a single number: "dog" → 3.1, "leopard" → 2.7, "wolf" → 1.6.

But this immediately breaks down. Words are related to one another in too many different ways for a single number to capture. A dog may be closest to a wolf in an evolutionary sense, while a leopard may be closer to a wolf in terms of behavior or habitat. Dogs are domesticated; wolves and leopards are not. Leopards and wolves may be perceived as more dangerous than dogs. Dog can also be a verb, not just a noun. These relationships shift across contexts, uses, and even languages.

To represent all of this, we need more than one number per word.

The solution is to represent each word not as a number, but as a point — a position in a space with many dimensions. Distance in that space encodes similarity: words that are used in similar ways, or that share related meanings, end up near one another. Words that are similar in some ways, like tomato and apple (red and round), might be close in some dimensions but far apart in others (like taste).

To visualize the idea, imagine placing words within a three-dimensional space, like rooms in a house. Already we can do much more than with a single number. "Dog" and "wolf" might be close vertically, while "wolf" and "leopard" might be close horizontally. Notice that in two and three dimensions there are already more than two and three types of relationships that are possible.

To see why, consider a simplification: along each dimension, two words can be either "near" or "far". In two dimensions that gives four possible relationships: (near, near), (near, far), (far, near) and (far, far). In three dimensions this expands to eight. As the number of dimensions increases, the number of possible relationships expands *very* fast. Modern computer intelligences use thousands or even tens of thousands of dimensions, allowing an enormous number of relationships to coexist simultaneously.

Our intuition is tempted to label these dimensions: this one for size, that one for danger, another for cuteness and so on, but the instinct is misguided. The dimensions themselves do not correspond neatly to human concepts, they are simply degrees of freedom the system uses to arrange words so that useful relationships emerge. Meaning lives in the geometry of the whole which is not at all intuitive, and difficult for us to understand even after years of dedicated research.

Once words are represented this way, as points, language generation becomes a matter of movement through the space. Starting from a point corresponding to a word or phrase, the system repeatedly transforms that point — nudging it in directions suggested by context, grammar, and prior usage. Each transformation is millions, perhaps billions of numerical operations: additions and multiplications applied across many dimensions at once.

After enough of these steps, the point arrives in a region of the space near words that would plausibly come next. The system then selects among those nearby words, favoring closer ones but retaining some variability. For computer intelligence, language emerges as a result of navigating a high-dimensional landscape shaped by prior text.

Representing words, ideas, and meaning as points in space was a truly revolutionary shift. Rarely in intellectual history has a single abstraction collapsed many previously separate phenomena into one: Newton showed that the motion of the heavens, revealed by Kepler, and the motion of falling objects, measured by Galileo, obeyed the same laws; Darwin showed that the diversity of life could be understood as variations of a single process unfolding over time. The same is true here: a single abstraction unifies a wide range of problems that once appeared unrelated.

Concepts that share no words in common could now be recognized as related. The point in meaning space representing "food keeps getting stuck in my throat and it feels like it won't go down" would find itself near a cluster of related conditions — esophageal stricture, eosinophilic esophagitis, and achalasia; in the geometry of meaning, they live in the same neighborhood. This is what it means to represent knowledge geometrically: relationships that were previously implicit, requiring a trained human mind to recognize, become explicit in the structure of the space itself.

Translation, for example, emerges naturally once words from different languages are embedded within a common meaning space. A sentence is mapped into that space, where its content, its meaning, is expressed independently of any particular language. From there, it can be rendered back into words under the constraints of another language. What had been an enormously difficult problem for computers becomes a geometric operation: move from one region of the space (the source language) to a nearby region (the target language) where the same meaning is expressed differently.

Images can be treated in much the same way. Visual patterns are mapped into the same shared meaning space as language. Describing an image then becomes a matter of referencing the points representing the image, and rendering them into words. Generating an image reverses the process: take points representing the words, move to the points describing the image, and finally generate the pixels.

The same principle extends to speech, audio, video, and other modalities. What once required distinct systems, features, and hand-crafted pipelines now reduces to a single operation of navigating a shared space of meaning. But navigating this meaning space is computationally demanding; each step of generation is an expensive transformation applied across thousands of dimensions, and even a single such transformation involves millions of additions and multiplications. Producing a word requires repeating this process dozens, even hundreds of times. The price of unification is compute.

We can pay this price only because of the long accumulation of computational capacity humanity has built over millennia. For most of history, arithmetic was performed by individuals and was therefore slow and expensive. Then mechanical calculators, electronic computers, and industrial-scale infrastructure gradually compressed more and more computation into smaller amounts of time. That a machine can now perform, in a fraction of a second, the amount of arithmetic required to traverse a space of meaning at this scale is the culmination of a civilizational project to make computation abundant — one whose full history, and full cost, the following chapters will outline.

And yet an open question remains: can a system that navigates a space shaped by existing knowledge move beyond retrieval to discovery? Some of the greatest scientific advances have not required any new data, but only a reorganization of what was already known. Given Tycho Brahe's observations of planetary motion, Johannes Kepler discovered his empirical laws that predicted the heavens with unprecedented accuracy. Then, less than a century later, Newton required no new data beyond that available to Kepler to create a new framework, his laws of motion and universal gravitation, that explained how Kepler's empirical laws worked.

What Kepler and Newton did was extrapolation, but what experts mostly do and what generally makes expertise valuable is interpolation: combining known facts and methods in ways suited to a new situation. A doctor diagnosing a patient and an engineer sizing a beam, each draws on established knowledge applied to particular circumstances. Nearly every act of professional judgment is an interpolation: recognizing which known patterns apply here, and adapting them to the specifics at hand. Acts of true extrapolation — a doctor ignoring protocol and inventing a treatment from first principles, or a software engineer creating a new cryptographic scheme instead of using a standard one — are much more likely to be judged malpractice than genius.

A system that could reliably interpolate across the full breadth of human knowledge by combining insights from medicine with engineering, or from law with economics, at a speed and scale no individual could match would already be transformative, even if it never produced a single genuinely new idea.

Whether computer intelligence can make a genuinely extrapolative leap remains an open question, one we'll return to later. What we can answer is a prior question: how did we get here? The arithmetic required to produce a single sentence of machine-generated text would have consumed the working lives of a city of human calculators. That it now happens in a fraction of a second is not magic, but history.

Chapter 2

Antiquity Computes

> "I conceive that these things, King Gelon, will appear incredible to the great majority of people who have not studied mathematics, but that to those who are conversant therewith and have given thought to the question of the distances and sizes of the earth, the sun and moon and the whole universe, the proof will carry conviction." — Archimedes, *The Sand Reckoner* (c. 250 BC), translated by Thomas L. Heath

> "No motion impressed by natural causes, or by human agency, is ever obliterated. The momentary waves raised by the passing breeze, apparently born but to die on the spot which saw their birth, leave behind them an endless progeny, which, reviving with diminished energy in other seas, visiting a thousand shores, reflected from each and perhaps again partially concentrated, will pursue their ceaseless course till ocean be itself annihilated." — Charles Babbage, *The Ninth Bridgewater Treatise* (1837)

It is, in many ways, surprising that humanity's first successful systematic predictions about our world were actually predictions of other ones. The solar system and the stars are incongruous with anything encountered in daily life — yet there we made our first steps towards prediction and not closer to home.

Their size, distance and relatively small number are exactly the things which made them targets for our desire to predict. Absent the influence of the heavens themselves, things on Earth rarely repeat. The wind never blows quite the same way twice. And even relatively repeatable things are hard to quantify. A stone falls faster than our ability to time it with any accuracy without modern timepieces[1]. But the sun, the moon, the stars and the planets are always there

[1] Galileo's insight to put falling things on a very shallow ramp so that they fall slowly enough he could time them with his heartbeat was a brilliant insight thousands of years in the making. We should not be hard on the ancients for not thinking of it.

on a leisurely walk across the sky waiting to be measured.

Perhaps equally surprising to us is that for thousands of years the prediction of the movements of celestial bodies was intertwined with the prediction of Earthly events through horoscopes and fortunes. Astronomy and astrology were born together in ancient Babylon, one in service of the other. We used the one thing we *could* predict to make predictions about the things we *wanted* to predict, but could not.

The Babylonians were the first people to systematically observe and record their observations of the night sky. And they needed arithmetic to make sense of their observations. In Europe, before the adoption of Arabic numerals, Roman numerals (I-one, V-five, X-ten, L-fifty, …) were the standard way to write numbers. One problem with them is that writing nearby numbers can require a greatly different number of letters — it is hard to parse the scale of the number from its representation: 3888 = MMMDCCCLXXXVIII whereas 3900 is MMMCM. Another problem with them is that they make calculation clunky. Addition is not too bad, you can mostly just merge the symbols from the two numbers together.

`XLVII + XXIX = 47 + 29`
~~X~~`LVII + `~~X~~`XIX  cancel the subtracted X in 47 with an X in 29`
`LVII + XIX = 57 + 19`
`LVI`~~I~~` + X`~~I~~`X  do the same for the subtracted one in 29`
`LVI + XX = 56 + 20  merge all the remaining symbols together`
`LXXVI = 76`

But multiplication is really cumbersome; it's not impossible to use something like long multiplication with Roman numerals[2], but it's error prone and inefficient. The Romans worked around this by not writing down their arithmetic like we do today; they used an abacus or other counting device, did their sums

[2]Here is worked example of long multiplication with Roman numerals. You need to expand everything and then figure out how to simplify. Knowing a "Roman multiplication table" would make things a bit faster (e.g. V*X*=*L*, *V*C=D, …).

```
XIII   13
 VII    7
=========
   XIII
+  XIII
+ XXXXXVVV simplify XXXXX -> L and VV -> X
==========
   XIII 13
   XIII 13
+ LXV   65
===========
LXXXVIIIIII  simplify IIIIII -> VI
LXXXVVI      simplify VV -> X
LXXXXI       simplify LXXXX -> XC
XCI     91
```

and then just wrote the answer down using Roman numerals.

An abacus has some advantages — it is certainly faster for adding many numbers compared with writing things out, but its advantage decreases as the operations become more complex. One big disadvantage is that you have no record of your work; writing things down allows you to see all of your intermediate calculations and verify them individually and if there was a mistake, to correct it and only redo the portion that comes afterwards. An abacus in contrast is destructive — intermediate results are erased by the process of using it.

It would be natural to think that as the Babylonians were working a millennium before the Romans, they were using an even more primitive number system, but the opposite was true. The Babylonian number system was much closer to our own than to the Romans — it was a place value system that facilitated multiplication; the main differences were that they lacked a decimal point and used base 60 instead of base 10. The lack of a decimal point meant that a number like 1234 could be interpreted as 1.234 or 12.34 or 123.4 and so on; which one needed to be inferred from the context. Their base of 60 is what gives their number system its name, the sexagesimal system. Unlike our 10 digits, they had sixty different symbols to represent each number from 1 to 59[3]. And their multiplication table had 1,770 entries instead of 45.

With a multiplication table so large, memorization would be difficult, so it is not surprising that multiplication tables are a common occurrence on the Babylonian clay tablets that archaeologists have found. They have also found tables for reciprocals, squares, square roots, and even Pythagorean triples; lookup tables were used extensively to reduce the amount of work in solving common problems. Babylonian tables and their number system made serious arithmetic possible, and that arithmetic, applied to data recorded over decades and even centuries, is what enabled prediction.

The basis of all Babylonian prediction was discovering repetitive patterns in the positions of the Sun, Moon, and planets. Consider the Sun — its position in the Zodiac, that is, which stars appear behind it[4], repeats after one year; in fact, the Sun's position repeating is essentially the *definition* of a year. The planets, in contrast, all take much longer than a year to repeat their positions in the Zodiac. From a modern point of view: if the Earth orbits the Sun with a period of 365

[3] The symbols themselves were actually constructed a bit like Roman numerals, so it wasn't quite as bad as having to remember sixty completely different glyphs.

[4] You might wonder how to measure the Sun's position against the stars when the stars aren't visible while the sun is shining. One answer is to observe what stars are visible right on the horizon just before the Sun becomes visible in the same spot. Or to avoid the problem of needing to look towards the Sun (which you should NEVER do!), you can note what stars are exactly *opposite* of where the Sun rises. Then from your knowledge of the night sky, you can figure out later which stars were behind the Sun.

days and Venus orbits the Sun with a period of 225 days, then after 8 years Earth will have completed 8 revolutions and Venus will have done just shy of 13. That means that Venus will appear in almost the exact same place in the sky every 8 years. Even without understanding the mechanism, it is possible to observe these repetitions.

Although the Babylonians did not make plots as they dealt entirely with tabulated numbers, plotting their data is illustrative of what they were doing. If we plot the position of the Sun in the Zodiac against days, then it forms a nearly straight line going from 0 degrees to 360 degrees over the course of 365 days[5]. Then it starts all over again. This happens because Earth's orbit is *almost* circular, so that each day traverses nearly the same angle around the Sun. The planet Mercury which orbits close to the Sun roughly follows this straight line, but undulates around it due to its orbit, sometimes running ahead and sometimes falling behind. However, whereas the Sun returns to the same place after one year, the undulations of Mercury don't — it takes them 46 years to start repeating. After 365 days, the Sun has returned to the same place, but Mercury is in a different place in its orbit; only after 46 years do the Sun *and* Mercury return to approximately the same place.

{Note, there needs to be a plot there illustrating what was just described.}

Determining these periods of repetition takes decades, even centuries of systematic record keeping across generations. But if you can determine the period of the cycle, and you have observations of the previous cycle, then you can make predictions about the next one. The Babylonians discovered relationships that repeated over very long periods of time. Jupiter returns to nearly the same place in the sky every *71 years* and Mars every *79 years*. Even though the Babylonians had no idea *why* this happened, they knew from the data that it did, and that was enough to enable prediction.

This idea was taken to an extreme by a 3rd century BCE Babylonian priest named Berossus. As related by Seneca, he described the ultimate period: the point when all the planetary cycles simultaneously complete and reset; not just when Mars is in the same place in the sky, but when every planet is back in the same place. As the starting and ending point of this cycle, called The Great Year, he chose the time when all the planets were exactly in alignment. He predicted that if this happened with the planets in Cancer the world would end in fire and if in Capricorn, a great flood. Such an alignment hasn't happened yet, so the jury remains out on this doomsday prophecy[6].

These periods, while close, were not exact repetitions. For example the Babylonians knew that after 8 years Venus returned to almost the same place in the

[5]The real value is much closer to 365.2422 days. The fractional part being close to .25 or $\frac{1}{4}$ is why we have leap years every four years. And because it is actually a bit less, we also skip leap years when they are multiples of 100 but not 400.

[6]If such an event were to happen, it would not happen for millions of years. Long enough to not be too concerned about this potential end of the world.

sky, but it would be off by about 2.5 degrees. Almost but not quite. Predictions that just copied what happened the previous cycle would not be accurate enough, so it was important to note the small but real difference from an exact repetition.

This worked — figure out the period of repetition, note the slight differences from one cycle to the next, and use previous records to predict the future — but it must have been quite cumbersome to keep 79 years of observational records around on clay tablets! Reducing the number of numbers needed was the next step, but it came at the cost of doing more arithmetic.

The Babylonians would've realized from looking at the tables of observations that the amount by which the position changed from one observation to the next, that is, the differences between successive entries in their tables, was nearly constant across large stretches of the Zodiac. Their idea was to reduce the massive set of position observations to a much smaller set of differences and then to use those differences to reconstruct the positions on-demand.

They had two ways of approximating these differences — known by the aggressively unhelpful names System A and System B. The historian Neugebauer missed a great naming opportunity — I'll take it. System A, which we'll call Staircase, changed the difference value infrequently but in large increments when it did change. System B, Sawtooth, varied the difference amount up and down like teeth of a saw blade — it was always changed, but by smaller amounts. In either case the update procedure was much the same — take the current position, look up the difference, and add them together to get the next position. Sawtooth is theoretically more sophisticated than Staircase, but it was harder to deal with in practice, so both systems ended up being used side-by-side for thousands of years.

In all the tablets we've found of the Babylonian astronomer priests, we haven't found one that posits a mechanism, a model, for *why* the heavens behaved the way they did. The heavens moved, the priests observed, and from those observations, using arithmetic, they were able to make accurate predictions about the future. It would be left to the Greeks to ask the question — what system of motion could produce our observations?

A quick interlude to observe that our current approach to computer intelligence is far more Babylonian than Greek. We create computer intelligence "model-free" — our techniques are the ultimate black box capable of learning to predict language but also any other sequence, say for example DNA, equally well. There is nothing in our models that should make them good at language, or vision, or survival over any other task. We are still waiting for our Greeks to arrive.

Claudius Ptolemy lived in the 2nd century AD, perhaps in Alexandria, and with his astronomical treatise, the *Almagest*, had the final say on the matter for the next 1500 years.

His system placed the Earth at the center of the Universe, which, with hindsight, was quite a large blunder. But even today it is not so easy for a layperson to refute his two arguments for why the Earth must not move: the lack of stellar parallax and the fact that objects not attached to the Earth stay motionless relative to the ground. Parallax is best understood by placing your thumb at arms length, covering one eye and pointing it at an object a few arm lengths away; swap which eye is covered and note that your thumb has moved and is no longer pointing at the object. Stellar parallax means that we should observe the same thing when looking at stars with our "eyes" being the Earth at opposite points in its orbit around the Sun. But Ptolemy could not see it; indeed nobody was able to directly observe it until the 19th century. The reason is both simple and incomprehensible — the stars are absurdly far away. Given the choice between the Earth not moving and the stars being nonsensically far, it was an easy choice for Ptolemy and his contemporaries[7].

The system he describes in his *Almagest* is far more complicated than the Sun, Moon and planets moving in circles around the Earth. *That* system would have produced horrendous predictions about the positions of celestial bodies, certainly less accurate than the Babylonian system based purely on observation and arithmetic. The actual system Ptolemy devised required sophisticated mathematics and arithmetic to fit the model to observations and then use the model to make predictions.

To fit the observations, in particular "retrograde" motion, when the planets move backward opposite their normal direction across the sky, Ptolemy hypothesized the planets move on epicycles, circles whose center itself moves around a bigger circle which encircles the Earth. Then, to account for the nonuniform speed throughout the Zodiac that the Babylonians also observed, he had to place the center of the larger circle away from the Earth itself. This offset, called an equant, meant that sometimes the planets were closer to Earth and appeared to move faster than when they were farther away. The equant was not easily accepted by all; to some, it violated the notion that the Cosmos must consist only of perfectly uniform circular motions. Fifteen hundred years later Copernicus's distaste for the idea drove him to seek a new model of the Cosmos with the Sun at the center.

[7] Well, except Aristarchus.

The equant also made the arithmetic particularly complicated, Ptolemy needed trigonometry. Trigonometry allows the astronomer to measure what is tractable and then, based on that measurement, calculate that which cannot be measured. Measuring angles was the astronomer's bread and butter; with trigonometry and angles alone, the positions of the heavenly bodies could be predicted.

Euclid, based in Alexandria a few hundred years earlier, had provided the theorems with his *Elements*. In it he laid out all the known results of plane geometry and number theory, proving ever more general results from a small set of basic axioms. The basic ingredients necessary for solving triangles are there, but a very practical piece was missing: how to deal with arbitrary angles like 37 or 52 degrees that came from real astronomical observations.

A quick example of the power of trigonometry. Suppose we want to measure the height of a mountain and we are some distance away from it, but we don't know how far. We can measure the angle formed by the ground and a line point the mountaintop, say 27 degrees; then walk a kilometer and take another measurement of the angle, now 23 degrees. Knowing the two angles and the exact distance we walked is all we need to calculate the height of the mountain to be 2546 meters. We can calculate that which cannot be measured[8], but to do it we needed to know the value of trigonometric functions for 23 and 27 degrees, which Euclid and his Geometry does not give us.

Ptolemy's solution was a table of chords[9], a function closely related to our modern *sine* function. Imagine a circle with a radius going from the center to the circle's edge. Now imagine a line perpendicular to this radius going from one side of the circle across the radius to the other — that perpendicular line is a chord. The two ends of the chord form a triangle with the circle's center. If the chord is drawn very near the edge of the circle it will be quite small and the angle of the triangle at the circle's center will also be quite small. If we slide the chord along the radius until it's very close to the center, it will almost be as big as the diameter of the circle and the angle will be almost 180 degrees.

He tabulated the relationship between the angle of this triangle and the length of the resulting chord for every angle from $\frac{1}{2}$ of a degree all the way up to 180, a fine enough division to be used for precise astronomical calculations. This table, combined with Euclid's plane geometry and Menelaus's spherical trigonometry, would allow him to create a model of the movement of the Heavens and use it to generate predictions of the future. It was also the largest computational undertaking man had yet performed.

To build this table, Ptolemy started from what geometry *could* provide: exact

[8]The height is given by this formula: $h = \frac{dtan\alpha tan\beta}{tan\alpha - tan\beta}$. d is the distance between the two measurements, α is the first angle and β is the second.

[9]We know from Theon of Alexandria's *Commentary on the Almagest* that before Ptolemy, Menelaus composed a book titled *On the calculation of the chords in a circle* and that Hipparchus also composed a similar work even earlier, but both are lost. Hipparchus's table contained entries only every $7\frac{1}{2}$ degrees compared with Ptolemy's $\frac{1}{2}2$.

chord lengths for a handful of special angles. Values of the chord function were known, due to Euclid, for special angles related to regular polygons: equilateral triangles, squares, pentagons and hexagons. They gave you values for the chords of 120, 90, 72, and 60 degrees, but what about the rest? Getting the rest relied on a remarkable theorem, today known eponymously as Ptolemy's Theorem[10], which is proved in the *Almagest* and then used to make practical geometry and trigonometry possible. The theorem relates the lengths of the sides and diagonals of a quadrilateral inscribed in a circle. It allowed Ptolemy to derive two critical results. The first was the relationship between two angles with known chords, and chords for the sum and difference of those angles. For example, if you knew the chords for 72 and 60, then you get could get the chords for 132 and 12. The second was a way to generate the chord value for half of a given angle with a known chord. If you knew the chord for 60, then you could get the chord for 30.

With just a few exact starting values and these chord / angle relationships Ptolemy had a clear, if laborious, path to a table with fine spacing. Starting from the hexagon's 60 degrees, he could halve his way down: 30, 15, $7\frac{1}{2}$, $3\frac{3}{4}$ degrees, and so on, getting finer with each step. Once he had a fine enough value, he could use the first relationship to build back up one step at a time; this would eventually populate a table.

But Ptolemy was more clever than this. The regular pentagon gives the chord of 72 degrees and from his addition formulas, he could calculate the chord of 75 degrees (since $75 = 45 + 30$, and he could obtain both via halving 90 and 60). The subtraction formula then gives $75 - 72 = 3$ degrees. Now he only needed to halve twice — from 3 to $1\frac{1}{2}$ to $\frac{3}{4}$ of a degree — rather than grinding down from 60 through many more halvings.

However, there was one problem left. Halving can take you from 3 degrees to $\frac{3}{4}$ of a degree, but it cannot give you exactly 1 degree — for that, you would need to trisect an angle, which the Greeks had proven to be impossible by geometric construction alone and did not know how to do with algebraic tools. To reach the chord of 1 degree, and from there to fill in his table at half-degree intervals, Ptolemy needed a different approach. He proved a pair of inequalities that bounded the chord of 1 degree between two values he *could* calculate, and showed that the bounds were tight enough, agreeing to the precision he needed, that the estimate was as good as exact for practical purposes.

[10] Despite Ptolemy's theorem being perhaps the most important theorem for actually *doing* things with geometry, it is rarely taught in high school geometry anymore. On the rare occasions when it is taught, it happens without any context to how it was used historically and why it actually matters. The author remembers it being in the back of his high school textbook where it was presented as curious relationship about lengths of the sides and diagonals of an inscribed quadrilateral.

Finding the value for the chord of $\frac{1}{2}$ was just the beginning — most of the work still remained to be done. Each entry in the table required many multiplications and divisions and was dependent on the ones before it[11]; and any introduced errors would propagate and render the table useless. It would have taken about a month of doing arithmetic all day long to produce the final table which occupied several pages of the *Almagest*. It gave chord values for every half degree from $\frac{1}{2}$ to 180 degrees, accurate to the equivalent of about five decimal places, all expressed, of course, in the sexagesimal system inherited from the Babylonians.

With this table it was now possible to use triangles and trigonometry to make sense of the Heavens. Ptolemy could take the geometric scaffolding of his model, the epicycles, the eccentric circles, the offsets, and compute its parameters: how big each circle was, how fast things moved and with those compute actual predicted positions for the Sun, Moon, and planets at any given time. Every prediction required trigonometry and a fair amount of arithmetic — multiple lookups in the chord table, combined with multiplications and additions.

Ptolemy's system needed 20-30 observations to accurately determine the parameters of each planet. Once those parameters were established, you could just run the model forward and predict the future as far forward as you wanted to go — if you were willing to do the math. Contrast this with the Babylonian system: without a model of the underlying system they needed records of the entire 79 year cycle of Mars, and ideally multiple such cycles, thousands of observations in all, to be able to predict its motion.

Having solved one problem — the prediction of the Heavens — Ptolemy created another. Astronomy was still done significantly in the service of astrology: predicting the heavens to predict earthly events. But if the heavens could be predicted completely, what did that mean for us? If the positions of the planets shaped your life, and those positions are determined by a mechanical system that grinds inexorably forward, then the future is already written. What room is left for choice?

Ptolemy confronted the problem directly in his *Tetrabiblos*, a companion work to the *Almagest* that applied his astronomical system to astrology. His answer had two parts. The first was that celestial influences are tendencies, not commands. Just as the Sun drives the seasons and the Moon governs the tides, the planets shape conditions — but a farmer who knows a drought is coming can store water. Foreknowledge does not eliminate agency; it enhances it. The stars incline, Ptolemy argued, but do not compel.

[11] To minimize errors, you wouldn't get the value of 73 by starting from 1 and calculating 1.5, 2, 2.5, 3, … Instead you would start from 72 (which you know from the pentagon) and combine it with 1. Same thing for 4, you would start from 3 and add 1. Accuracy would be maximized by minimizing the number of steps between the final result and an exact result.

The second was complexity. Even if celestial influences were in principle deterministic, they interact with local conditions — geography, climate, upbringing, the circumstances of the moment — in ways too intricate to predict exactly. The prediction is real but approximate; certainty is unattainable in practice.

This same problem would arise again, more seriously, when Newton gave precise laws to describe how *everything* moved — not just the heavens but objects on Earth. If the universe is a machine whose future state follows inevitably from its present one, then free will becomes difficult to defend. Laplace laid out the following thought experiment: he imagined an intelligence vast enough to know the position and velocity of every particle in the universe, and concluded that for such an intelligence "nothing would be uncertain and the future, as the past, would be present to its eyes." In this view, our sense of choice is merely ignorance; we feel free only because we cannot see the machinery that determines us.

Ptolemy's complexity defense turned out to be durable. Even if Newton's laws are deterministic, the equations governing three or more bodies interacting gravitationally cannot be solved exactly. Many systems exhibit what we now call chaos — an extreme sensitivity to initial conditions in which tiny differences in starting points lead to wildly different outcomes. To predict the future of such a system you would need to know its present state with *infinite* precision, and no measurement can provide that. Laplace's demon would need not just vast computational power but perfect knowledge of the present, down to the last decimal place, and there is no last decimal place.

But, even a completely deterministic system with known initial conditions can be unpredictable. It depends not just on the thing itself, but on the computational resources of the one doing the predicting[12]. Consider a simple example: a computer program that takes a single number as input and produces a long sequence of numbers as output. The program is entirely deterministic; given the same input, it always produces the same output. But the sequence it produces *looks* random; it passes every statistical test for randomness you might think to apply and no pattern can be found in it.

Such programs, called pseudorandom number generators, are so effective that they are the foundation of modern cryptography. We will return to this in a later chapter — arithmetic's ability to create sequences that look random is, it turns out, how it keeps secrets. An observer who managed to work backward from the output to the input, who "cracked" the code so to speak, would be able to predict every element in the sequence: the randomness would be gone. But

[12]https://arxiv.org/abs/2601.03220 *From Entropy to Epiplexity: Rethinking Information for Computationally Bounded Intelligence* This is a new enough idea, that readers, especially those with a math or science background, might find a direct reference useful.

an observer who has only the output, and not enough computational power to work backward to the input, cannot distinguish it from a sequence produced by flipping a coin. The "randomness" is not in the sequence; it is in the gap between the sequence's complexity and the observer's capacity to analyze it.

Determinism that cannot be exploited, whether because chaos demands impossible precision in the initial conditions or because the computation required exceeds what any real observer can perform, is operationally indistinguishable from genuine freedom. Whether the universe is "truly" random at its foundations or merely intractable to predict may be a distinction without a difference for any intelligence that actually exists within it.

With the nature of free will still on the scales, the practical work of predicting the heavens continued. Ptolemy's model was so successful it superseded everything that came before it. It was disseminated by Arab astronomers across the Islamic world and even farther East. While each civilization it touched would refine and extend its ideas, the core would remain the same until it made its way back to Western Europe.

Arithmetic Shines Light on the Heavens

> "I also ask you, my friends, not to condemn me entirely to the mill of mathematical calculations, and allow me time for philosophical speculations, my only pleasures." — Johannes Kepler, letter to Vincenzo Bianchi (1619)

> "Since nothing is more tedious, fellow mathematicians, in the practice of the mathematical arts, than the great delays suffered in the tedium of lengthy multiplications and divisions…" — John Napier, *Mirifici Logarithmorum Canonis Description* (1614)

One and a half millennia after Ptolemy, in the 16th century, the prevailing view of the heavens had not changed from the static geocentric world of Aristotle and Ptolemy. But then, in 1543, Copernicus reignited Aristarchus's ancient claim that the Earth moved around the Sun by publishing "On the Revolutions of the Heavenly Spheres." Scholars vigorously defended geocentrism and the ancients from attacks by the upstart heliocentrists inspired by Copernicus. We know now that Copernicus was right. But at the time, his theory, while elegant, did not actually offer more *predictive* power than Ptolemy: he still used circles and epicycles, circles going round circles, and surprisingly, had little more data than Ptolemy despite the passage of 1500 years.

Tycho Brahe, a Danish nobleman born in 1546, grew up in this environment of intellectual discord. He also grew up in an era of great astronomical luck — he would observe a solar eclipse, great conjunction, supernova, and great comet within the span of 17 years. That's not likely to happen again for another approximately 5,000 years.

The first event, a solar eclipse in 1560, is an uncanny event for anyone to observe — day turns into night over the course of a single minute, not the slow, majestic fade of sunset. But it particularly impressed the young astronomer because it had been *predicted*.[1] The prediction being off by a day did not seem to lessen his excitement or wonder that we could predict the heavens. Suitably awed, he set about learning to make the same kinds of predictions and promptly acquired all the available material he could on Ptolemy's techniques from the experts of the time.

It was with his new predictive powers and a discerning eye that he observed the great conjunction of 1563, an event where Jupiter and Saturn appeared very near to each in the sky. It too, had been predicted by the almanacs and ephemerides of the time, but by his judgment — poorly. It was this event that convinced him that more and better *data* was needed to make better predictions, a truly novel thought at the time.

In November of 1572 a truly rare astronomical event occurred: a new star appeared in the sky. A star in the constellation Cassiopeia had exploded, i.e. gone supernova, an event that released so much energy that the star's dying moment 8,000 light years away was one of the brightest objects in the sky. This event shook Brahe to his core. It is hard for us moderns to appreciate what a constant presence the night sky was, equanimous in its ability to be unperturbed by human events. It was there, the same, every night, slowly rotating across the sky; some stars were only visible at certain times of year, but they always came back and new ones never arrived.

Aristotelian philosophy turned this observation into dogma: the heavens were perfect and unchanging. Yet this nova ("new star"), a term Brahe coined in the title of his 1573 publication, *De nova stella*, was brighter than Venus, the brightest object in the night sky, and visible for months[2]. According to his observations it lay beyond the Moon, in the supposedly immutable sphere of the fixed stars. Something about the old theories was wrong.

Aristotle had reasoned about the heavens without measurement. Ptolemy had

[1] The author has been unable to find records of exactly *who* made this prediction. It is relatively easy to predict that an eclipse will occur every 18 years, due to the so-called Saros cycle of the sun, moon and earth. This was known to the ancients. But predicting exactly *where* a solar eclipse will happen is far harder and was not mastered until well into the 17th century. Perhaps whoever made this prediction just got lucky that the solar eclipse happened to be observable in western europe. Or maybe they were far ahead of their time.

[2] There are conflicting accounts as to whether it was brighter than Venus (the otherwise brightest object in a moonless night sky), or just as bright. Tycho claims brighter, other sources, the same. It would only have been at maximum brightness for 1-2 weeks and faded in and out.

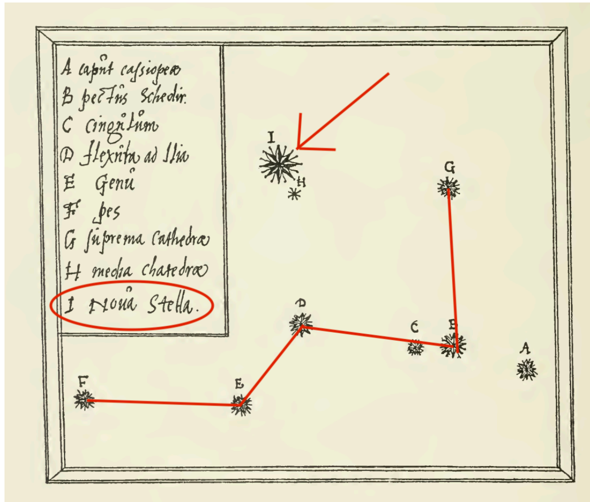


Figure 3.1: Tycho Brahe *De Stella Nova* Location of New Star

made just enough observations to constrain his models. Medieval scholars debated cosmology largely without data at all. Tycho, in contrast, collected far more information than he, or anyone else, could immediately explain[3]. It isn't hyperbole to say that Tycho initiated the entire scientific revolution with his approach of using observation to decide which system of the world was correct.

In 1577, his measurements of a great comet that appeared in the sky for 3 months convinced him further that the old Aristotelean wisdom must be wrong. The data implied that it too must lie beyond the moon. The answers lay in data and already his new observatory, Uraniborg, was being built on a small island off the coast of Denmark; it was to be the largest scientific data collection operation the world had yet seen.

Turning each observation of the night sky into a usable measurement involved a staggering amount of arithmetic. Observations measured positions relative to the horizon; but to be useful, they needed to be transformed into absolute coordinates on the celestial sphere — a process called reduction. Tycho and his assistants made their observations at exactly the right time to minimize the amount of work involved and it still required about ten multiplications, twenty trig table lookups and a few divisions.

The accuracy at Uraniborg was the equivalent of measuring the width of a hu-

[3]Or at least, to measure *in addition* to speculating.

man hair at arm's length. That level of precision meant the trigonometric tables needed 6 or more digits of accuracy, and each multiplication needed to operate on six digits as well. Ptolemy's table of chords had been improved upon during the fifteen intervening centuries, and the 16th century tables of Regiomontanus, Rheticus and Viète now possessed sufficient accuracy, both in terms of spacing and digits, to match.

To a modern reader, the difficulty of multiplication is easy to underestimate. We memorize the multiplication table for single-digit numbers and learn a set of rules for extending it to larger ones. Beyond a few classroom exercises, perhaps extending to three-digit problems worked once and forgotten, multiplication rarely intrudes into everyday life. When it does, we are surrounded by machines that make it effectively free. Typing the inputs now takes far longer than getting the result.

To give you, the reader, an idea of the difficulty, I timed myself adding and then multiplying the same two six digit numbers[4]. It took me 10 seconds to do the addition and 3 minutes and 30 seconds to do the multiplication, or over twenty times as long. I also made an error, which took a further 5 minutes to discover and correct.

Now imagine needing to do this ten times *per* observation. It could take an hour to produce one error-free reduced observation. Dozens of observations could be recorded each evening; the amount of effort to reduce them all the next day would require many scribes working dawn to dusk. Tycho had built the world's first data center and for the first time the speed of arithmetic was the limiting factor in scientific progress.

In 1580, a mathematician named Paul Wittich arrived with a solution – he possessed a way to calculate faster than everyone else. He did not fully understand where it came from; he had found it buried in the work of a little-known mathematician from the early sixteenth century, Johannes Werner, and could not even prove that it worked. But he knew that it did and more importantly, he knew where this discovery would be most valuable. Tycho recognized its worth and immediately put it to work[5].

[4] I used the numbers 931605 and 284917. I had originally intended to use 10 digit numbers, but scaled back my ambition someway through the multiplication when I was sure I had made errors due to misaligning some columns. Certainly, with more practice and some basic learnings (like start the multiplication problems on the right edge of the paper, so there is room for the values to grow to the left and leave spacing between columns so that they remain *exactly* aligned), I would get faster. But the main point remains – multiplication is a lot slower than addition and only becomes more so the bigger the numbers become.

[5] It is likely that Wittich took something in return – the designs of Tycho's astronomical instruments and some of his methods. It was the beginning of a long-running dispute over who deserved

What was Wittich's secret? He could multiply numbers by doing only addition[6]. More specifically, he could multiply the sines of two angles by adding and subtracting the angles and then looking up the cosines of the sum and difference in a trigonometric table. Since the tables already existed this was an order of magnitude reduction in work and time for the practical problems they were solving at the observatory. An additional benefit is that since addition is so much less work than multiplication, it is also much less error prone. Speed went up and errors went down.

However, the reduction process required multiplying sines together *and* cosines together. Wittich's method only handled the former; products of cosines could not be simplified with what Wittich brought to Uraniborg. Wittich had done well to find this obscure bit of knowledge and bring it to the right place and the right time, but further progress required a mathematical brilliance which neither Wittich nor Tycho possessed.

The solution to the cosine problem and a much more general one would come from a Swiss clockmaker, Jost Bürgi. Tycho maintained a correspondence with Bürgi's aristocratic patron who called Bürgi a "second Archimedes" for his mechanical and mathematical brilliance. It is likely from this relationship that Bürgi learned of Wittich's technique and the cosine problem. Reportedly, it took him little time to produce a proof of the sine formula and from this proof only a small modification was required to derive the formula for the product of cosines via addition. Armed with these two formulas, Bürgi was able to remove most of the arithmetic labor from the process of reduction[7].

These formulas were actually waiting there patiently since Ptolemy wrote down his theorem in the Almagest. He used them to *construct* his tables, but the recognition that once the tables had been painstakingly constructed, the formulas could be used backwards together with the tables to avoid multiplication took 1500 years. The leap required both need, brilliance and better tables. The tables of Ptolemy with a spacing of half a degree were nowhere near fine enough to be used for multiplication; but the tables of Regiomontanus, first computed in 1467, had a spacing of .017 degrees — thirty times finer than Ptolemy and accurate enough for Uraniborg.

---

credit for what.

[6]The technique was later named prosthaphaeresis, from the Greek for addition and subtraction.

[7]Much of the work in the reduction process comes from applying the spherical law of cosines — which requires three multiplications. Prosthaphaeresis eliminates all of them:

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha \tag{3.1}$$

$$= \cos b \cos c + \tfrac{1}{2}\left[\cos(b-c) - \cos(b+c)\right] \cos \alpha \tag{3.2}$$

$$= \cos b \cos c + \tfrac{1}{2}\cos(b-c)\cos\alpha - \tfrac{1}{2}\cos(b+c)\cos\alpha \tag{3.3}$$

$$= \tfrac{1}{2}\left[\cos(b-c) + \cos(b+c)\right] + \tfrac{1}{4}\left[\cos(b-c-\alpha) + \cos(b-c+\alpha) - \cos(b+c-\alpha) - \cos(b+c+\alpha)\right] \tag{3.4}$$

Some readers will recognize these as the "product-to-sum" formulas from high school trigonometry. In the author's experience they were memorized and promptly forgotten without context as to why they were once useful or that they were once among the most prized tools of European science. The author might have been more likely to remember if our math and science curriculum told stories about our collective intellectual journey rather than just leaving random artifacts out in a field all jumbled about.

It is possible to generalize this technique to the multiplication of arbitrary numbers, even those not arising from the reduction calculations. And while it's certainly better than doing a brute-force calculation most of the time, it's actually still rather cumbersome for arbitrary numbers.

The solution for the general case was actually foreshadowed by Archimedes himself in his work *The Sand Reckoner*. In his time, it was thought that there was not a number big enough to count all the grains of sand on Earth. Archimedes goes further — not only will he name a number large enough to count the number of grains on the Earth, he will name the number required to fill the Earth, and then the number of grains required to fill the *Universe*. This incredible work is one of the only sources we have for the existence of Aristarchus's heliocentric theory — that the Earth goes round the Sun. Archimedes assumes Aristarchus's theory because the heliocentric universe is significantly larger than one with the earth at the center, which makes his feat of counting the grains of sand required to fill it all the more impressive. He estimates the size of the Sun, Moon, and Earth, then uses those to estimate the size of the universe itself.

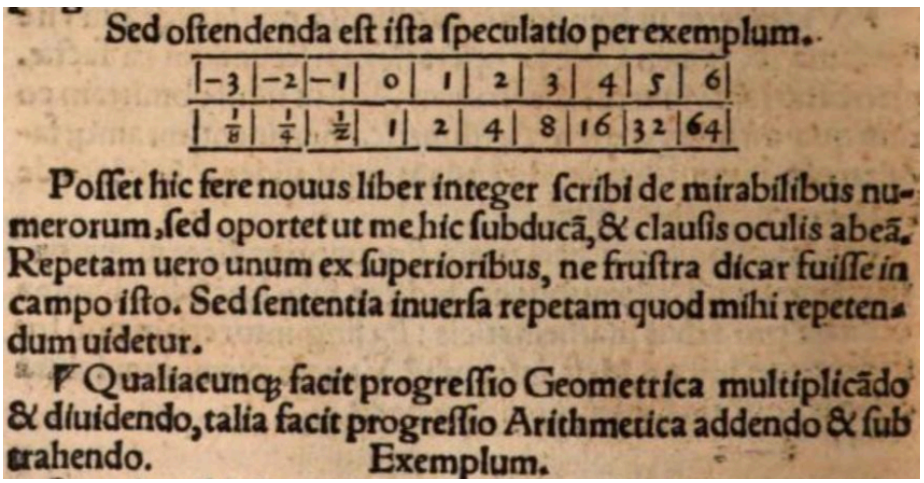
Archimedes also had to solve another problem. In ancient Greece, the largest named number was a myriad — 10,000; they did not have words or notation to systematically express bigger numbers and the number he needed to express was far, far larger. His solution was to invent what we would call exponents.

An exponent tells you how many times a number is multiplied by itself. For example, $10^2 = 10 * 10 = 100$ and $2^3 = 2 * 2 * 2 = 8$. The number being multiplied by itself is called the *base*; in the first example the base is 10 and the exponent 2 and in the second the base is 2 and the exponent is 3. In order to manipulate these exponentiated numbers, Archimedes proved that when you multiply two exponentiated numbers together, you *add* their exponents to get the result. Archimedes's first big number was a myriad-myriad, or myriad multiplied by itself, $10^4 * 10^4 = 10^8$, which is one hundred million. His estimate for the size of the universe was about 10^{63} grains of sand.

It is this result, that multiplying exponentiated numbers requires only *adding* their exponents, which connects sand reckoning back to simplifying multiplication. But it would take nearly two thousand years for anyone to say so explicitly. In 1544, the German mathematician Michael Stifel published *Arithmetica integra*. Eleven years earlier Stifel had achieved some notoriety for using numerology on a verse from the Revelation of John to predict the end of the world would happen on October 19, 1533; when it failed to arrive, he required protective custody from those he had deceived: by spending some time in prison[8]. But his *Arithmetica integra* includes this remarkable passage (the original Latin is in fig. 3.2):

"Here almost a whole new book could be written about the wonders of numbers, but it is necessary that I withdraw myself here, and go away with closed eyes. I will repeat indeed one thing from the above, lest I be said to have been in this field in vain…

Whatever things a Geometric progression does by multiplying and dividing, such things an Arithmetic progression does by adding and subtracting."



Sed oſtendenda eſt iſta ſpeculatio per exemplum.

−3	−2	−1	0	1	2	3	4	5	6
1/8	1/4	1/2	1	2	4	8	16	32	64

Poſſet hic fere nouus liber integer ſcribi de mirabilibus numerorum, ſed oportet ut me hic ſubducā, & clauſis oculis abeā. Repetam uero unum ex ſuperioribus, ne fruſtra dicar fuiſſe in campo iſto. Sed ſententia inuerſa repetam quod mihi repetendum uidetur.

Qualiacunq; facit progreſſio Geometrica multiplicādo & diuidendo, talia facit progreſſio Arithmetica addendo & ſubtrahendo. Exemplum.

Figure 3.2: Stifel's *Arithmetica Integra* folio 249

Stifel clearly knew that this connection would allow for the replacement of multiplication by addition and even gave some examples of it, but he didn't (or couldn't) extend it to work for arbitrary numbers. In Stifel's table, each number in the bottom row is double the one before it, and each number in the top row increases by one. To multiply $1/8$ by 64, you find their exponents in the top row: -3 and 6. Add those to get 3, then look up 3 in the top row to find 8 in the bottom row. Multiplication has become addition.

[8] The German expression *einen Stiefel rechnen* (to calculate a boot), meaning to calculate nonsense, is said to derive from this incident. Like Plato and his tale of the world's creation, it is a *likely* story.

But what do you do if you want to multiply 7×3.5? Stifel's table is no help here. In Stifel's original table each number on the bottom row was the previous one multiplied by 2. If instead, he replaced it with 1.5, then the first few numbers would be 1, 1.5, 2.25 — closer together. If 1.1, then 1, 1.1, 1.21 — even closer! The closer the ratio is to 1, the finer the divisions — and the more likely we are to find any number we need in the table. But it also means creating such a table would require a Sisyphean number of multiplications — perhaps this is what Stifel meant when he wrote that "a whole new book could be written about the wonders of numbers" before closing his eyes and walking away.

Jost Bürgi not only understood the principle but had the brilliance to make the calculation tractable. He constructed it entirely through addition, exploiting a multiplication trick sometimes taught to grade school students. To multiply a number by eleven, you just shift it one digit right and then sum the original and new number. For example, 11×13:

```
13
 13
---
143
```

But he needed a number very close to 1, so he chose 1.0001. Instead of shifting by one digit, he needed to shift by four digits. The first three numbers in his sequence would be: 1, 1.0001, 1.00020001.

```
10001
    10001
---
100020001
```

That's it. He just needed to perform additions like that over and over and over again; in the end 23,027 times until his sequence ended at ten[9]. That's a non-trivial amount of work, certainly, but modern estimates suggest it might have only taken him a few months of calculating. If multiplication was needed at each step instead of addition it would have taken a lifetime. The resulting table was more valuable than its weight in gold, and Bürgi treated it that way — like buried treasure he proceeded to share it with no one; it was a closely guarded secret for two decades.

Johannes Kepler though, knew what Bürgi had, if not the location. He arrived in Prague in 1600 to work as an assistant to Tycho Brahe. Tycho had been forced out of Denmark when the new king, Christian IV, whose gaze never turned

[9] He could stop at 10 because once he had all the "bottom" numbers between 1 and 10, he could scale numbers bigger than 10 down and numbers smaller than 1 up to still use the table.

skyward, cut his funding, and he had resettled at the court of the Holy Roman Emperor Rudolf II, bringing with him the most valuable scientific dataset in the world — decades of planetary observations of unprecedented precision. Tycho's first task for Kepler was to figure out how to predict the orbit of Mars. His unassailable measurements disagreed with the predictions of Ptolemy which meant the true nature of the solar system was yet to be discovered.

He died just a year later with the task far from complete. On his deathbed he is reported to have said, "Let me not seem to have lived in vain," to Kepler before entrusting him with the data[10] and tasking him with shining light not just on Mars but the Heavens. Tycho's favored theory of the solar system was not correct in the end, but his approach of measuring to find better theories and truth laid the foundations of the scientific revolution. Few men can claim to have led lives *less* in vain.

Kepler's initial attack was to assume that the Ptolemaic model was essentially correct, but just that its parameters, like the size of the circles and how fast they moved, hadn't been determined accurately enough. He modified Ptolemy's model slightly, allowing more freedom in the orbital parameters to account for why previous models didn't fully predict Mars's motion.

To determine the parameters for his modified Ptolemaic model he relied on a special kind of observation called oppositions. They are taken when the Sun, Earth and Mars all lie on the same line; if you flew in a straight line from the Sun to the Earth and kept going you would get to Mars[11]. An Earth-Mars opposition occurs every approximately 2 years and Tycho had 12 recorded observations. Kepler made an initial guess for the values informed by Ptolemy and then would compare that prediction against 4 of the 12 measured values; based on the differences, he would update the guess and start again. Each guess involved a chain of trigonometric calculations — multiplying sines and cosines, solving triangles on the celestial sphere.

In chapter 16 of his *Astronomia Nova*, Kepler addressed the reader directly regarding this effort: "If this wearisome method has filled you with loathing, it should more properly fill you with compassion for me, as I have gone through it at least seventy times at the expense of a great deal of time." Kepler and Bürgi worked in the same building at Prague Castle, and Kepler knew how much effort Bürgi's tables could save. But Bürgi would not share them. Years later, Kepler would write of him in the preface to his *Rudolphine Tables*: "Justus Byrgius was led to these very logarithms many years before Napier's system appeared; but he, a hesitant man and a guardian of secrets, abandoned the child at birth, and did not raise it for the common benefit."

[10]Kepler would be embroiled in a dispute with Brahe's heirs for years over who really owned the data. The skeptical might think it convenient that Kepler, the only narrator of Brahe's final wishes, was also their main benefactor.

[11]If you imagine a typical top-down 2D representation of the solar system. In three dimensions this isn't quite true: a straight line from the sun to the earth almost never actually goes through Mars.

All this effort was not wasted. He found a set of parameters that matched Tycho's observations within his observational error of 2 arc-minutes. But Kepler did not stop at this apparent vindication of Ptolemy; he went on to check whether the distances implied by this theory also agreed with the observations — and they didn't. Correcting the parameters to make the distances agree increased the error of the predictions to eight arc-minutes.

Eight arc-minutes is less than a quarter of the apparent width of the Moon. Previous astronomers could have shrugged it off; their data was not precise enough to detect so small an error. But Kepler knew that Tycho's observations were accurate to within two minutes, and he refused to ignore the discrepancy. "These eight minutes alone," he wrote, "will lead us along a path to the reform of the whole of Astronomy."

And so they did — though not without further years of the same brutal arithmetic. Kepler tried curve after curve, each attempt requiring fresh rounds of iterative calculation, until finally he tried an ellipse. It fit. The orbit of Mars was not a circle, but an ellipse with the Sun at one focus. It swept out equal areas in equal times, moving faster when closer to the Sun and slower when farther away. These first two laws appeared in his *Astronomia Nova* of 1609, wrested from Tycho's data with prosthaphaeresis and years of sheer computational force.

But to extend that discovery to the rest of the solar system and build the prediction tables that astronomers and navigators actually needed would require the very tool that Bürgi had hoarded. Kepler and Bürgi had worked side by side at Prague Castle for nearly a decade, so Kepler knew about the tables and had almost certainly seen them. But fifty-eight pages of nine-digit numbers — twenty-three thousand entries — were not something one could borrow for an afternoon and copy. There was, almost certainly, only one manuscript, and it was Bürgi's.

It was a Scottish nobleman, John Napier, who gave it to the world. Unlike Bürgi, Napier believed, as he wrote in his preface, that "the secret is best made common to all, as all good things are." Napier's formulation was different from Bürgi's: it was an elegant thought experiment of a point moving along a line with a speed proportional to how far it had left to travel. Imagine Zeno's arrow, trying to cross the remaining distance but at each moment covering only a fraction of what remains — moving slower and slower, never quite arriving without an infinitude of time.

His tables took nearly two decades to compute, compared with the months Bürgi had needed. But Napier was willing to give the world the artifact produced by those two decades of his life. Bürgi produced a solution to the multiplication bottleneck decades before Napier published his, yet it is Napier who changed the world while Bürgi's works collected dust.

Kepler received a copy of Napier's *Mirifici Logarithmorum* in 1617 and immediately grasped their power. He, however, disapproved of Napier's formulation so much that he later wrote his own book, the *Chilias Logarithmorum*, replacing the kinematic foundation with one built solidly on the Greek geometrical tradition even though the practical result was much the same. Whether grounded in motion or in Euclid, Laplace would later say that logarithms "doubled the life of the astronomer." Every astronomer, for the next four centuries.

Armed with these tables, Kepler could finish the work that the ellipse had begun. His third law, that the square of each planet's orbital period is proportional to the cube of its distance from the Sun, appeared in 1619, an insight likely inspired by logarithms. And in 1627, eighteen years after *Astronomia Nova*, he published the *Rudolphine Tables*: comprehensive predictions for every planet, computed using logarithms, with Napier's tables bound into the volume as an essential tool. In the preface, he reflected on encountering logarithms a decade earlier; he called it a "happy calamity" — it had made predicting the motions of all the planets possible.