Numbers and the World
Essays on Math and Beyond

David Mumford

Professor Emeritus
Brown and Harvard Universities

Draft of August 16, 2022

To Erika, Jenifer and Alice

Contents

Cover: Euler meets the Human Face

I submitted four images to the AMS for the book's cover, with the accompanying text "Euler meets the Human Face", illustrating the max and min curvatures and the lines of curvature on a human face, a good example of "Numbers and the World". This theory goes back to Euler in 1760, [Eul60]. Here, he considers the intersection of a given surface S with planes containing its normal at a point P. He proves that the resulting plane curves have a maximum and minimum curvature, called the max and min *principal curvatures* at P, and that their tangent lines are perpendicular vectors on S at P.

In the late 1990s, one of the areas my students studied was the use and statistics of laser range images of the world and another area was face recognition. In those days, 3D range sensors were expensive and rare. One of my students, Gaile Gordon, found a startup in California, Cyberware Laboratories, who had designed a laser range scanner to create a 3D image of a person's head. They rotated a laser 360 degrees around the person's head scanning vertical slices, yielding a 512 x 256 range image $I(\theta, z)$. Their original business model was creating custom sculptures, but instead found the data was much more valuable to the Hollywood special effects industry – e.g. scanning the actors of Star Trek to enable more realistic and creative computer graphics. They generously agreed to help with Gaile's research by scanning her head (used here), as well as all of their employees, to create a small 3D face database.

The face proper amounted only to 66×73 pixels but we interpolated, smoothed a bit and got enough data to work out the main differential geometric features. Our work appeared in the joint book [**Bk-1999**], which studied the face from many perspectives, including a section *How to Sculpt a Face* and many curvature diagrams including "ridge curves" where one of the curvatures has a max or min along its line of curvature.

The top left is a plastic model of the face created from the data. The top right and bottom left images show level curves of the min and max principal curvatures respectively, with the zero value, the parabolic curves, thickened. For min curvature, the parabolic curves surround the convex parts of the face where both curvatures are positive, especially the tip of the nose; for max curvature, they surround the concave parts where both curvatures are negative, especially the eye sockets. For these figures, I chose a degree of
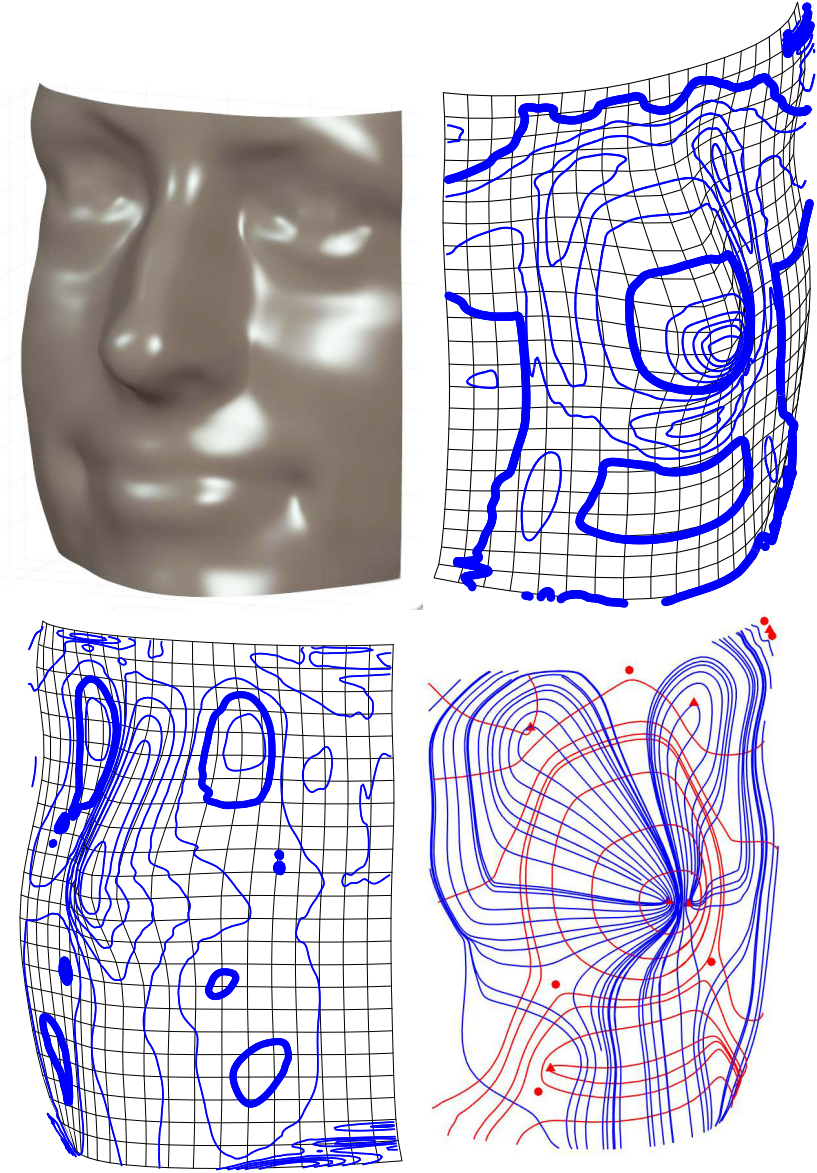
smoothing where the strong features are visible but the result is not cluttered with details.

The bottom right image shows samples of the lines of curvature. These form an orthogonal net as Euler showed but with singularities at the *umbilic points* where max and min curvatures are equal. These come in two types as the lines rotate $\pm\pi$ when you go around an umbilic, sometimes called "lemons" and "stars". These are denoted by little black triangles (the lemons) and circles (the stars). The lines of curvature move around lemon umbilics like comets as though attracted to the umbilic but, near a star umbilic, look like they are repelled. Note that the nose must have two lemons on it because the lines of curvature must rotate by 2π. Note too the star umbilic at the chakra on your brow.

Preface: Confessions of a Polymath

Firstly, the pdf below is the draft of my book that I sent to the AMS on August 16, 2022. It has almost no input from the AMS. It was TeXed as a plain latex book and the cover figures were produced entirely by me. The figure credits have been added and were almost all obtained by me in the fall of 2022. When I retired at age 70, I thought it would be a lot of fun to write a blog where I could sound off on anything and never worry about picky referees. A particular fact that is both my problem and the motivation for my blog, is that I keep getting excited about something new. This is sometimes a technical area, where I am neither known as a regular contributor nor do I know "the rules" which regulate publication there. Other times, it is some area of general interest where I get fired up and the AMS was helpful in constraining my impulses to stir up controversy. I'm afraid I'm addicted for wanting to learn the essential ideas in more and more fields as well as getting involved with more and more debatable issues.

My greed in this respect goes way back. I became fascinated in high school with designing a relay driven calculator and reading up on special relativity and on the foundations of math; I spent 2 summers in college working on simulating submarine atomic reactors with analog computers at Westinghouse; at Harvard, I tried to learn more physics, biology and astronomy as well as math, not to mention fliers taking art history and Anglo-Saxon. (Art history was a struggle: I never knew when I submitted a paper whether I would get an A or a C. I didn't dare take courses in music or philosophy because I doubted my competence.) I remember talking with Barry Mazur when we were grad students and both agreeing that we wanted to have a basic understanding of *all* fields of math. Why settle for less. Then however I fell under the spell of Oscar Zariski, John Tate and Alexander Grothendieck and began to focus. Zariski, in particular, had an infectious passion for Algebraic Geometry. When he said the words "Let V be a *variety*," you felt he had access to a secret garden in which this abstract mathematical construct called a "variety" was a species of beautiful flowers with wonderful exotic properties.

I wanted the key to his garden and I settled down with algebraic geometry. I loved the ideas of "infinitely near points" and "blowing up" but was especially attracted to moduli spaces, maps in an abstract world. Sometimes these spaces seemed almost tangible, as

in the bijection between suitably embellished abelian varieties and 2-adic "Gaussian-like" measures.[1] But after about 30 years, when Joe Harris and I [HM82] reached a milestone in this corner of the garden, wanderlust struck again. At an algebraic geometry meeting in Ravello, Jayant Shah and I got talking over some duty-free whiskey and discussed what was going on in Artificial Intelligence. Benoit Mandelbrot had visited Harvard a couple of years earlier and, if there ever was a successful polymath, who combined math with applications, he was it. Jayant and I threw ourselves into learning the AI-relevant computer science as well as neurobiology. David Marr [Mar82] had defined the area like this: there should be a unified "theory of the computation" in AI underlying its distinct implementations in silicon and in neural tissue and one should combine insights from math, statistics, computer science, engineering, psychology and biology to frame this evolving theory. He proposed, and we agreed with him on this, that it was prudent to start with a simpler instance of a cognitive skill, namely vision. We were encouraged that vision has been mastered by such diverse animals as octopuses and man. So, for about 20 years, we concentrated on vision.

Jayant and I had fun bringing some math into the field of computer vision [**V-1989**]. However, the field was really driven by engineers who vied for incremental improvements in various benchmarks at each annual get together. For me, one central problem was what was the best math to model our use of "shape" in understanding images of the world. When I met Peter Michor through the IMU in the late 90's, I learned that he had created wonderful machinery for doing infinite dimensional Riemannian geometry [KM97]. We worked together on the mathematics of shape for about a decade and I was delighted to gain a deeper understanding of Riemannian geometry and non-linear analysis, to learn, for instance, a bit about *a priori inequalities*. Then, in 2007, I retired from teaching.

Now I had time to pursue long standing interests. Another polymath, Freeman Dyson, was one of my heroes, always thinking about new things with an unorthodox perspective [Dys81] and I hoped to follow his example. One interest was math education. What was the root cause of the depressingly familiar comment made by a new acquaintance, "Sorry, math was my worst subject," after you admit to them that you are a life-long mathematician? I discuss some of my thoughts on this in Chapter 1. Another was the History of Math. I had recently met David Pingree and started a math course based on history of math for non-math majors. I take this up in Chapters 4-7. A third topic was physics. I had learned a lot of quantum mechanics from George Mackey and from von Neumann's book [vN55] but "Schrödinger's Cat" had always bothered me and I was eager to learn a bit about quantum field theory. I discuss this in Chapters 14-15. A fourth big area was the study of the foundations of math. I believed in Christopher Freiling's negative answer to the continuum hypothesis [Fre86] but wanted to dig deeper. I write about my thoughts here in Chapter 13. I've put a lot of time into all four and much of this appears in this book. I was not immune, however, to issues of more general interest, to wanting to learn and comment a bit of what is going on in the world these days, to learn a bit of philosophy

---

[1]See **A-1966a**, part II, §8,9 and **Bk-2010a**, pp.622-648.

and also to speculating about the future. Such issues are discussed in Chapters 16-19.

I'd like to add some words about an issue that comes up when you venture as "visiting mathematician" into another field and have some unorthodox ideas that you feel are reasonable and valid. Mathematicians are often distrusted as aggressive amateurs who force their abstractions on fields for which they don't have a good "feel," the depth of knowledge that comes from a lifetime of work. This has some validity but can also be very frustrating if you have worked hard precisely to get some feel. Let me give a couple of examples. When I studied neurobiology in the 80's, essentially all modeling was based on the idea that *feed-forward* pathways, with one way flow of information from senses to thought to action, was the basis of cortical function. The ubiquitous *feed-back* pathways in the brain were explained as merely representing attentional modulation. I thought this made no sense and wrote extensively on alternate models, see e.g. [**B-1991, B-1998, B-2003**] and Chapter 9. Related to this is the theory that Bayesian statistics or its cousin Grenander's *Pattern Theory* [GM07] should be used to model thinking. Engineers, like the neurobiologists, were focussed on feed-forward algorithms so algorithms using feedback from prior probabilities were anathema. But if I read the tea leaves well, I think both communities are coming around to the view that feedback is a crucial component of cortical thinking.

In the case of math education, if you think that your brilliant suggestion for modifying the math curriculum and exciting more students is going to be listened to, forget about it. The math education establishment is made up of teachers, school boards, textbook writers and publishers, examination factories, college admissions officers and outraged parents and the idea that, of all things, a research mathematician should have any say is a joke. Each of these groups has entrenched views and real power. (The CORE curriculum was a short lived exception but it has had its share of pushback.) I learned soon enough not to expect much from my ideas that I talk about in Chapter 1. My own angle is based on G. Harel's Principle "Students are most likely to learn when they see a need for what we intend to teach them..." [Har07] and I have found a small community agreeing with this including Heather Dallas at UCLA and Sol Garfunkel at COMAP. More power to them.

Math History is nearly as tough to break into. No matter how much you have read about some ancient culture, a research mathematician will always be accused of propagating "whig history," anachronistically misinterpreting ancient texts by comparing them with modern ideas. I find this weird: no one criticizes consulting metallurgists to understand the mineral content of an ancient sword! I give my favorite example in Chapter 4, where Archimedes is unmistakably calculating a Riemann sum of an integral, a comment that is not only something no historian apparently knows but that would bring down the wrath of referees for distorting Archimedes' thinking if you mentioned it in a paper submitted to one of their journals. I subscribe to Littlewood's description of Archimedes and his contemporaries as "Fellows in another college." Historians are really good at history but, when dealing with mathematical material, I believe they would benefit from partnering with research mathematicians (see, for example, the controversy over the Babylonian tablet Plimpton 322 in Chapter 4).

In this book, I've put down some of my ideas on education, history, AI, current and future issues and even a little actual math and physics partly taken from my blog. I have divided it into parts with various common themes plus an interlude that I trust will be understood as a spoof. I deeply enjoy the math and have written from my heart about all these issues. I know that some bits in every part of what I have written are controversial and will strike some readers as radical or wrong or misconceived. However I believe strongly that controversy is healthy and no one should be "cancelled" for their opinions, no matter how passionately one holds a different opinion. In any case, all opinions in the book are entirely my own and do not in any way represent opinions of the publisher, whom I thank for its tolerance.

For those who are skimming this densely written book, let me emphasize a few take-home points as a sort of executive summary:

1. Chapter 1: High school math ought to be taught so students believe it is useful and relevant to their lives,

2. Chapter 10/18: The experience of passing time is the essence of consciousness.

3. Chapter 13: Applied math suggests a major revision of set-theoretic foundations.

4. Chapter 14: DNA mutations may be creating "cat-states", high rank macroscopic density matrices.

5. Chapter 19: In a treacherous future, eugenics is likely to reappear.

Finally, I want to express my thanks first of all to the American Mathematical Society, especially to Sergei Gelfand, Catherine Roberts and Eriko Hironaka who have helped me put this volume together and allowed me to express my feelings on many things not usual in math books. But equally, I need to thank the many people who have given me helpful comments, suggestions and references and who have checked various parts for accuracy. In alphabetical order, these include Michael Artin, Alain Connes, Al Cuoco, Heather Dallas, P.P. Divakaran, Harvey Friedman, Stuart Geman, Sol Garfunkel, Gaile Gordon, Alice Gorman, Robin Hartshorne, Jens Høyrup, Curt McMullen, Peter Michor, John Myers, Jeremy Mumford, Linda Ness, Mark Nitzberg, Ulf Persson, Nick Trefethen, Hugh Woodin, Jakob Yngvason, Song-Chun Zhu and doubtless others over the many years in which I have written these essays.

Part I

Opening more Eyes to Mathematics

This first part concerns topics in mathematics that have involved non-mathematicians, students, life scientists and lay people, with mathematical issues.

Chapter 1 is a discussion of how it is that, in the K-12 sequence of classes, so many students "turn off" when it comes to math. I got involved in math pedagogy when Deborah Hughes-Hallett was working on a sequence of calculus books together with Andrew Gleason and Bill McCallum [HHM98]. As I recall, it started because I objected when a textbook asked for a gradient vector in a 2D plot involving different units of each axis, e.g. plotting temperature $T(x, t)$ as a function of a space coordinate and time. In such a case, the differential makes sense but not the gradient. About the same time, I was writing the book *Indra's Pearls* with Caroline Series and Dave Wright [**E-2002**] and we had to describe in the Preface what math background our readers needed. We came up with the phrase, "(if) you can handle high school algebra with confidence," then you can read our book. Only after the book was published and I gave copies to various friends did I realize how small this cohort is. This was very dismaying and I didn't know very many people who were trying to remedy this. I felt then and still do that the biggest part of the problem was trying to teach math in isolation instead of teaching it by solving problems important to students and adults alike. Sol Garfunkel and I wrote an op-ed piece published in the Times on this [**E-2011b**]. Sadly, we got deluged by objections from people who held onto a dream that "pure" math was the single most important thing taught in high school and thought teaching its applications was "dumbing it down"! Lynn Steen, who defined the goal of math education as "quantitative literacy," acidly remarked to me that changing the math curriculum is harder than getting the permits needed to move a cemetery.

The second chapter concerns an obituary for Alexander Grothendieck that John Tate and I wrote that was rejected by Nature magazine as too technical as we mentioned higher degree polynomials and complex numbers. It is a sad fact that many of those in the Life Sciences have forgotten much of the math they once knew. This is not a healthy situation. Grothendieck, in my book, is the most original mathematician in the second half of the 20th century and is the person whom I have no hesitation in describing as a genius. Surely, the unique and amazing way Grothendieck thought can somehow be told in a way that non-mathematicians can appreciate. His reputation outside the math community is growing so this is a challenge with some significance. I have come to realize how difficult this is, how many layers build one upon another underlying his key results. I write about some attempts along these lines in this chapter.

On a more positive note, the third chapter concerns *beauty* in mathematics and two projects springing from the belief that there are beautiful math formulas. This hardly does justice to the topic but I doubt that any consensus on what is beautiful in math can ever be reached. Working researchers in math all experience, from time to time, an epiphany over the beauty of something they see. But I think that, to codify this, a light-hearted approach is the best we can hope for if not offered a dive into an MRI tube as some mathematicians were in the second project described there.

Chapter 1

How to get Middle School Students to love Formulas & Triangles

All mathematicians are familiar with the usual reaction when they answer the question "And what do you do?" in a party. If I had a dollar for every awkward response, often a confession that math was the questioner's worst subject, well, you know the rest. Where does the education system go astray that this is how math is viewed? I think that, by and large, arithmetic is accepted by everybody as a key skill, useful even if many people forget the rules for long division or even how to add $\frac{1}{2} + \frac{1}{3}$. But in middle school, either in the 7th or 8th grade, they hit algebra. Suddenly it's all x's and y's and many students rapidly loose their bearings. It is pointless to drill students for three or four years in something most of them will forget as soon as they have taken their SATs. In a year or two, they hit geometry. This tends to be a bit more accessible but nonetheless irrelevant to their lives.[1]

i. Algebra

I'm sure everyone saw the viral image of a blackboard on which a student has written "Dear Algebra, Please stop asking us to find your X. She's never coming back and don't ask Y." Sigh. I believe there is a way to present algebra to middle schoolers that breaks the log jam of "what the hell are x and y?." To make middle and high school math work it is essential to get as many students as possible to see that formulas are useful and intuitive ways to

---

[1]This Chapter is partly based on my blog, dated October 22, 2014 and partly on an unpublished paper by Heather Dallas and me. This article was intended for a journal of the National Council of Teachers of Mathematics (NCTM) but it was rejected. It follows a New York Times OpEd piece Sol Garfunkel and me that appeared on August 21, 2011, [**E-2011b**]]. Other work on Math Education is on my website www.dam.brown.edu/people/mumford/beyond/education.html.

see how numbers in their real lives *are connected to each other*. Formulas are simply the natural language for talking about any quantitative relationship. Once this step is made, a great deal of science, economics and further math is open for exploration.

The first and most important thing is *not to use x and y until much later, but instead make formulas using whole words or abbreviations for words*. In a nutshell the reason for the usefulness of algebra is this: life is full of situations where several numbers are needed to describe a situation. These numbers vary from one situation to another but in each case the numbers usually have some fixed arithmetical relationship to each other that doesn't vary. Writing these relationships as equations gives you a clearer grasp of all these situations, much as having the right word in your vocabulary can help you grasp immediately new situations described by this word: in both cases, your mind learns a structure that will fit many situations in the future. An equation can be thought of as a class of quantitative situations. Those who never internalize this equation are condemned to dredge up isolated rules every time similar situations come their way.

Arguably, the simplest case of a useful formula is this: in any trip, distance travelled is the product of the time the trip takes by the speed of travel. Going by plane, 3000 miles from NYC to SF equals 6 hours times 500 miles per hour; a 2 mile walk is 40 minutes (2/3 of an hour) times a typical walking pace of 3 miles per hour. We can write this:

$$\text{distance} = \text{speed} \times \text{elapsed time}$$

or

$$\text{dst} = \text{spd} \times \text{tm}$$

or just

$$d = s \cdot t$$

Here and in all other real world numerical situations that cry out for a formula, *use simple abbreviations*. Do we cite Einstein's most famous equation by saying "if x is the energy of an object, y its mass and z the speed of light, then $x = yz^2$ "? No, we say $E = mc^2$ where E and m are obvious abbreviations for energy and mass and c was the universally used abbreviation for the speed of light[2].

But there's another easily explained advantage to thinking in terms of a formula. Take the travel case again. Clearly, if the speed s and the elapsed time t are known, the formula tells us that the distance travelled d is gotten by multiplying s and t. But algebra tells us that we can also play the game getting the value s or t from the other two numbers. This is because the formula can be rewritten:

$$s = d/t \text{ or } t = d/s$$

---

[2]In fact, Einstein wrote it first as "change in mass (in grams) equals change in energy (in ergs) divided by 9×10^{20}, or, as a formula: $\Delta(L) = \Delta(m) \cdot c^2$." He used L because he was talking about the energy of light(*licht*) and everyone knew the speed of light is close to 3×10^{10} in these particular metric units.

so that if we know d and t, we get out s by division, and if we know d and s, we get out t by division. The rules of algebra show how a numerical relationship of one kind can be used in multiple ways. Once you get the hang of thinking in terms of a formula, the formula becomes a much clearer way of describing a situation than an awkward long sentence. It becomes the natural way of grasping how sets of connected numbers *fit together*. But before this happens, you need to see a lot of meaningful instances and schools, all too often, just drill the student in abstract formulas with no real world meaning. *You must not start with the abstract formulation and afterwards illustrate it with examples.* No, you must start instead with multiple concrete instances, enough so the leap to a general abstract formulation is natural and easy.

Formulas of the "thing A is the product of thing B and thing C" abound. Converting quantities measured with one unit to their value in another unit is extremely common. Liquid measures like cups, pints, quarts, gallons convert to weight measures like ounces and pounds using the memorable verse "a pint a pound the world around." Traveling abroad on continental Europe, the essential tool is the simple formula:

$$\text{price in dollars} = (\text{price in euros}) \times (\text{rate: dollar per euro})$$

or

$$p_{\text{dollar}} = p_{\text{euro}} \times \text{rate} \left(\frac{\text{dollar}}{\text{euro}} \right).$$

The second version has the advantage that you can imagine it comes from cancelling the word 'euro' in the two right hand terms. Similar conversions between metric and English measurements and between the zillions of units used in cooking occur all the time and most of us face these to some degree. I confess it is not easy though, at a European gas station, to convert euros per liter into dollars per gallon on the fly, because you need two ratios: gallons per liter and dollars per euro! My favorite problem in converting between units is this: "how fast does your hair grow in *miles per hour*?." Now that's going to make students laugh as well as learn how to multiply big numbers.

A spreadsheet is a terrific stepping-stone for learning algebra and all middle and high school students should have access to one. To use these efficiently, you enter formulas into cells that calculate new values from values in other cells. It is all based on using variables for the number in each cell, e.g. E7 in a formula means the number in column E, row 7. "E7" plays the role of x. Symbols for variables can be anything you like. Ancient Indian mathematicians used color names for variables. Thus entering into a cell "=A1*E7+D3" will result in adding cell D3's value to the product of the numbers in cells A1 and E7 and then putting the result in the new cell. A spreadsheet is not merely a set of numbers but becomes much more useful and powerful when it contains formulas, hence contains a whole web of numerical relationships. Spreadsheets have numerous nifty tricks to do common things fast. For example, suppose you have a long column of figures that use one unit and you want to convert all the numbers to another unit, i.e. multiply them all by some

ratio. All you need to do is enter this once, click in some way on this result and *drag* the cursor down forming a new column. Bingo: all the corresponding multiplications are made automatically.

More broadly speaking, there are few topics that will get more student's attention and will give them useful tools for adult life than *money*. It's never too early to have students set up a business plan plus a daily history for a lemonade stand in a spreadsheet. It is in financial matters that most of us need to grasp numerical relationships more clearly and where formulas and spreadsheets can help a lot and give everyone the power not to have to accept blindly what is told to us by 'experts' (who are usually salesmen).

Finances and algebra really connect when it comes to exponents and compound interest. Every person handling money takes out loans, whether they are credit card loans, college tuition loans, loans for purchases like a car or house mortgages. And in Civic's class, they need to learn that credit are loans that banks or corporations make to you while bonds are loans people make to a corporation or the government. But understanding compound interest and what is involved in paying off loans really needs a little math and even the dreaded *polynomials*. Susan Forman and Lynn Steen [FS99] came up with an example of a typical difficult math problem that is likely to be faced by typical middle class adults navigating the financial world:

> The rent on your present apartment is $1,200 per month and is likely to in-crease 5% each year. You have enough saved to put a 25% down payment on a $180,000 townhouse with 50% more space, but those funds are invested in an aggressive mutual fund that has averaged 22% return for the last several years, most of which has been in long-term capital gains (which now have a lower tax rate). Current rates for a 30-year mortgage with 20% down are about 6.75%, with 2 points charged up front; with a 10% down payment the rate increases to 7.00%. The interest on a mortgage is tax deductible on both state and federal returns; in your income bracket, that will provide a 36% tax savings. You expect to stay at your current job for at least 5-7 years, but then may want to leave the area. What should you do?

The figures are completely out of date, terms like "points" and APR need to be defined, but the basic situation is as current as ever. This is not straightforward math as it involves rough estimates and weighing choices as well as math. But Forman and Steen's point is that High School math ought to prepare him/her for such problems.

Bringing this closer to your typical student, say our average high school senior wants a car and may be able to get one by taking out a loan. But, for instance, if they charge a mediocre credit risk teen-ager 1.33% interest *per month* (16% APR) on a 5 year loan, he would do well to know that his total cost works out to be about 50% more for the car than he would pay if he had the cash. I would suggest that high school math class ought to give every student the confidence to "do the math" him or herself and not rely on others with their own agendas. The first step is to assign simple abbreviations to the numbers involved.

Use C for the cost of the car, P for the monthly payment, r for the rate of interest (e.g. $r = 0.05$ for 5%.) The second step is to translate what interest means into a formula: one month's interest increases the loan from C to $C + r.C$ and one payment decreases it from C to $C - P$. So, after a month, the outstanding loan changes from C to $C.(1 + r) - P$. It's easier to see what's happening if you let $R = 1 + r$, the factor by which your balance increases every month. Then, after each month, your balance goes from C to $C.R - P$. I would argue that, broken into small steps, formulas will begin to make sense to all students. Repeating this for the second month, the balance owed becomes $(C.R - P).R - P$. This seems like a mess only a math nerd would love. But use the rules of algebra and it becomes a quadratic polynomial in the number R:

$$C.R^2 - P.R - P.$$

Aha, so polynomials actually occur in *real* life! If you go on for, say 4 months, the balance owed will be this polynomial:

$$C.R^4 - P.\left(R^3 + R^2 + R + 1\right).$$

Wow, more polynomials. We're not giving a lecture here, just hoping to show how algebra can be useful. So let's just say – if you use the stuff taught in every Algebra II class and pursue what we have started, you'll wind up, maybe not easily but eventually, seeing that if you need to pay off the loan in 5 years (60 months), your total payment P will be

$$P = C.\frac{R^{60}}{R^{60} - 1}.$$

In the example above, make $r = 0.0133$ and work out his total cost, $60P$, on a hand held calculator, and you get about 1.5 times the cost C of the car.

The formula above, though it might show up in a New Yorker cartoon with white-coated scientists, reveals, when you play with it, an essentially simple relationship between interest rates, loans and payments. The majority of real life scientists work on real problems like this and not on abstract stuff in ivory towers. A nation-wide discussion, verging on a political fight, has been going for the last decade concerning the Common Core State Standards in Math (CCSS-M), `http://www.corestandards.org/Math/`, with many voices, pro and con, including the many state boards of education. As I see it, the CCSS-M have considerably upped the ante in abstract math but have also opened the option of introducing "modeling" a code word for math that might relate to the real world as students know it. All K-12 math can be enlivened and made relevant to students, exciting even, by dipping into the vast array of applications that math has to real life. Our message: math, properly taught, can be relevant, interesting and maybe even memorable.

After the above was posted on my blog, I had a lot of correspondence, esp. with Bill McCallum, one of the principal authors of the Common Core in Math, and with Al Cuoco, a major author (see [Cuo10]) and advisor at the Educational Development Center (EDC).

They both said they agreed with much of what I wrote but that the big question was "How to get there." Bill goes on to say "An important part of the problem ... is grasping the complex relationship between fluency and conceptual understanding" and "If you start kids out with a flexible understanding of arithmetic, then they are more likely to appreciate your formula for the total amount paid on a loan." Here flexible means not just memorizing but seeing how useful rearranging terms can be, as in the example:

$$9 + 16 = 9 + (1 + 15) = (9 + 1) + 15 = 10 + 15 = 25.$$

I certainly agree there.

Unknown to me, Al had actually written almost the identical discussion of the interest payment problem to which he has given me the link [Cuo19]. He also wrote me when the blog was first posted these comments:

> You claim that "The first step is to assign abbreviations to the numbers involved." My colleagues at EDC and I have used this example for decades in our own CME high school curriculum and in our high school teaching before that. The step of writing down the relationships in precise algebraic language is somewhere near the midpoint of a long development that is preceded by carefully orchestrated numerical calculations, an introduction to functions and recursively defined functions, and experiments with a spreadsheet and later with a CAS. Once the basic algebraic relationships are in place, there are a host of other sophisticated ideas that need to be in place before one can get the closed form for the monthly payment.

Although I agree with much that Al says, I worry that recursion is best introduced in the context of teaching how to code computers. His recent paper [CG21] does exactly that. I feel teaching the basics of computer coding should be a central component of 21st century high school math curriculum and recursion is one of its key principles. But I want here to reiterate here a criticism that applies to much of Common Core as well: many pure mathematicians subscribe to the idea that you cannot understand an idea until you have a general definition for it and almost all their writings put the abstract concept first. I would put it backwards, especially when it comes to K-12 education: students will not understand a general idea such as that of a recursive function until they have seen some motivating examples. Fortunately, after working with numbers in spreadsheets, a recursive formula with abbreviations is not a big step. I do not think understanding that interest adds to your outstanding balance and a payment subtracts from it is going to be hard for students to understand and then to write as the formula above. I want to stick to my guns: show *real* examples, relevant to the students first, trusting that the concrete context allows the teacher to explain easily the arithmetic in the formula. As quoted in the Preface, G. Harel asserted the principle: "Students are most likely to learn when they see a need for what we intend to teach them...." Later, after enough examples are seen, one might introduce
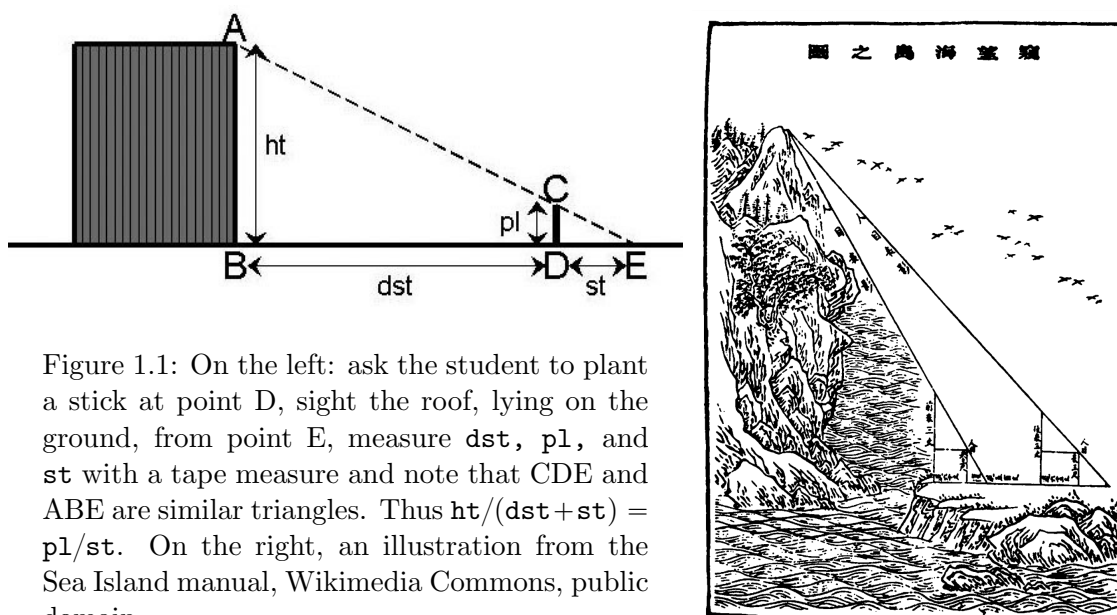
Figure 1.1: On the left: ask the student to plant a stick at point D, sight the roof, lying on the ground, from point E, measure `dst, pl,` and `st` with a tape measure and note that CDE and ABE are similar triangles. Thus $\texttt{ht}/(\texttt{dst}+\texttt{st}) = \texttt{pl}/\texttt{st}$. On the right, an illustration from the Sea Island manual, Wikimedia Commons, public domain.

the general concept of functions and of recursion rules. A confession: this is how my mind works and the standard approach of starting with abstract definitions has been a stumbling block for me in reading many math books.

ii. Geometry

When my children were taking geometry in High School and we had parent-teacher conferences, I repeatedly asked their math teacher whether they ever took their class outdoors and had them measure something, e.g. the height of a tree. They always treated me like a foolish math professor interfering with their job and their professional training. I guess nobody remembers that the very word "geometry" *means* measuring the earth? This is such a wonderful opportunity to show students how math is relevant to the real world! The last summer, I showed my 10 year old grandson how to measure the height of a tree. The idea is clearly explained by Figure 1 left for the case of the height of a building.

Actually, a refinement of this idea was invented by the Chinese mathematician Liu Hui in 263 CE in his book "The Sea Island Mathematical Manual" [Swe92]. The idea is made clear in Figure 1 right: The distance to the island `dst` is unknown but if you can sight the peak from two points instead of one, you can solve for `dst`! If the right sort of clouds are in the sky, one can use this technique to work out the height of some point on a cloud. Of course, the formula is very sensitive to how far apart are your two observations. Finding some bounds on the estimated distances using rough measures of how accurate your sitings

are is another valuable lesson.

There are so many ways of measuring the earth that can be taught in High School. Perhaps the most fun is using the curvature of the earth. So long as your school is near a moderately large body of water or a totally flat prairie, this is easy to include in a geometry class. The basic idea is given in the diagram of Figure 2. You apply Pythagoras's rule to the two right triangles:

$$R^2 + d_i^2 = (R + h_i)^2, \text{ for } i = 1, 2, . \text{ hence}$$
$$d_i^2 = 2Rh_i + h_i^2$$

Then you point out that $2R \gg h_i$, so you might as well just forget the h_i^2 terms. This, in itself, is a great lesson to teach: in the real world, approximate answers are usually just as useful as exact ones. For that matter, the earth is an oblate spheroid so there is no single value for R. OK, ignoring the h_i^2 terms, we get:

$$d_1 + d_2 = \sqrt{2Rh_1} + \sqrt{2Rh_2}$$

which can be solved for R. I used this with my Brown class for non-math majors, using a photo the Newport bridge taken 18 miles up Narragansett Bay and got decently accurate estimates for R. But one can use houses or trees or boats seen with binoculars across a lake at a distance of say 5 miles while sitting in a kayak. This effect is really obvious when boating off shore, say 8 miles out. Conversely, one can use this formula to find the distance to the horizon when standing on the shore. If you stand on the waterline so your eyes are $5'$ or $6'$ over the waterline – call this $1/1000^{\text{th}}$ of a mile and use the rough figure $R \approx 4000$ miles, then the horizon is a little less than 3 miles away. An Israeli friend of mine said that he used to lie on the beach at the waterline watching the sun go down and, at the last moment when the sun fully disappeared, stand up fast and count the seconds until it disappears again. Believe it or not, this should be around 4 seconds! This is easy to check: the sun moves 360 degrees in 24 hours, hence 15 degrees in an hour, 1/4 of a degree in a minute, 1/240 degrees in a second. Converting this to radians, it moves about 0.000073 radians in a second. When you stand up, your eyes, feet and the horizon make a triangle with sides 0.001 miles and 3 miles, hence a small angle of 0.00033 radians (denoted α in Figure 2), the amount the sun moves in about 4.5 seconds.

Triangles are everywhere, not only in geometry textbooks. Carpenters, architects, city planners and map makers use them all the time. Why aren't some of these applications used in teaching geometry? Hipped roofs are a source of fascinating problems and drawing plans is a lot of fun. Even better is to teach trig at the same time as geometry. What better place to talk about ratios of lengths in a right triangle than at the time when you introduce Pythagoras's rule? In chapter 4, where I discuss the origin of this rule, I discuss how it appears to have originated in the need to survey land as city states emerged and a primitive form of a trig table occurs on the famous tablet Plimpton 322.
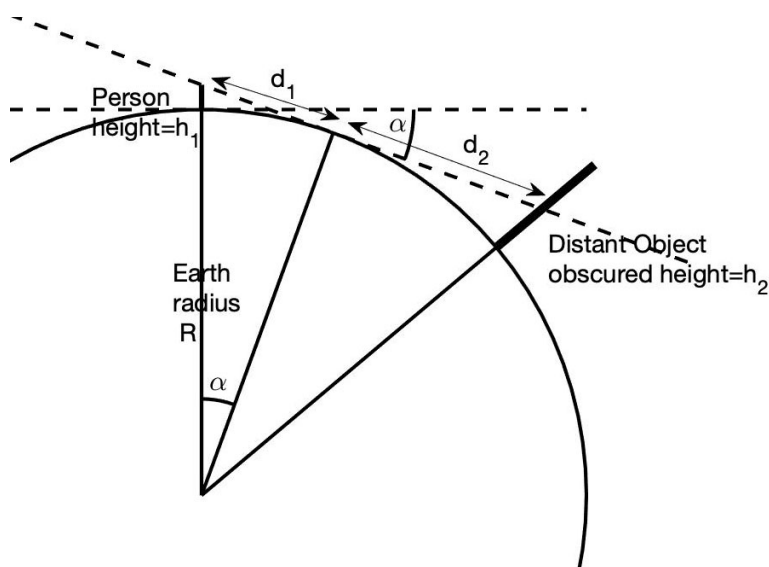
Figure 1.2: The generic diagram for finding the radius R of the earth from images of distant objects whose bases are obscured by the curvature of the earth. Given the height of the observer h_1 above the sea or flat ground, an estimate h_2 of how much of the object is obscured, and an estimate of $d_1 + d_2$, the distance to the object, R can be calculated.

Traditionally, geometry went into great length discussing angles, congruent triangles (e.g. the side-angle-side criterion) and simple proofs of properties of configurations of triangles, parallelograms, etc. In particular, proofs were constructed. using columns of statements and reasons. Much of this seems to have been dropped as being archaic throwbacks to Euclid not relevant to the 21st century. Actually, I believe a very important thing was taught with such exercises: how to be 100% certain of something if the need arises. It is natural to think loosely, analogically, metaphorically. But the law, for example, requires precise logic and meticulous dissection of circumstances. This is obviously still relevant today. But I think there is another context that demands it and is thoroughly 21st century useful: computer programming. Writing code that compiles and runs correctly requires 100% precision in your code. The least error and the code will fail. My suggestion for all High School students would be a semester learning simple programming. For example, code simple web pages using raw html and put your friends faces and ideas and sports scores on your page. The Euclidean algorithm is an example where algebra and coding intersect as described vividly by Al Cuoco and Paul Goldenberg in [CG21]. They make the case that this can be an eye-opening experience. Writing code is really very similar to formulating 2 column proofs: everything must be defined in the right place and references have to be consistent. As a retired professor, I cannot count how many jumbled, incoherent explanations of a formula I have read in the "blue books" of exam finals. A little practice

where nothing less than 100% accuracy is acceptable is a good preparation for life. I cannot resist mentioning what my colleague Phil Griffiths once said to a pre-med student who complained about being marked down severely for a "trivial" mistake: "You may be my surgeon some day and how much partial credit is deserved if you botch my operation?"

Chapter 2

Explaining Grothendieck to Non-Mathematicians

i.   Nature Magazine vs. rings & schemes

John Tate and I were asked by Nature magazine to write an obituary for Alexander Grothendieck[1]. Now he is a hero of mine, a person clearly deserving of the accolade "genius." I got to know him when he visited Harvard and John, Shurik (as he was known) and I ran a seminar on "Existence theorems." His devotion to math, his disdain for formality and convention, his openness and what John and others call his naiveté struck a chord with me.

Figure 2.1: Alexander Grothendieck, 1970, Wikimedia Commons by Konrad Jacobs.

_____

[1]This first section is based on a blog post with the same title, dated Dec.14, 2014

So John and I agreed and wrote the obituary below. Since the readership of Nature were more or less entirely made up of non-mathematicians, it seemed as though our challenge was to try to make some key parts of Grothendieck's work accessible to such an audience. Obviously the very definition of a scheme is central to nearly all his work, and we also wanted to say something genuine about categories and cohomology. Here's what we came up with:

<u>Alexander Grothendieck</u>

Although mathematics became more and more abstract and general throughout the 20th century, it was Alexander Grothendieck who was the greatest master of this trend. His unique skill was to eliminate all unnecessary hypotheses and burrow into an area so deeply that its inner patterns on the most abstract level revealed themselves – and then, like a magician, show how the solution of old problems fell out in straightforward ways now that their real nature had been revealed. His mathematical strength and intensity were legendary. He worked long hours, transforming totally the field of algebraic geometry and its connections with algebraic number theory. He was considered by many the greatest mathematician of the 20th century.

Grothendieck was born in Berlin on March 28, 1928 to an anarchist, politically activist couple – a Russian Jewish father, Alexander Shapiro, and a German Protestant mother Johanna (Hanka) Grothendieck, and had a turbulent childhood in Germany and France, evading the holocaust in the French village of Le Chambon, known for protecting refugees. It was here in the midst of the war, at the (secondary school) Collège Cévenol, that he seems to have first developed his fascination for mathematics. He lived as an adult in France but remained stateless (on a "Nansen passport") his whole life, doing most of his revolutionary work in the period 1956 - 1970, at the Institut des Hautes Études Scientifique (IHES) in a suburb of Paris after it was founded in 1958. He received the Fields Medal in 1966.

His first work, stimulated by Laurent Schwartz and Jean Dieudonné, added major ideas to the theory of function spaces, but he came into his own when he took up algebraic geometry. This is the field where one studies the locus of solutions of sets of polynomial equations by combining the algebraic properties of the rings of polynomials with the geometric properties of this locus, known as a *variety*. Traditionally, this had meant complex solutions of polynomials with complex coefficients but just prior to Grothendieck's work, André Weil and Oscar Zariski had realized that much more scope and insight was gained by considering solutions and polynomials over arbitrary fields, e.g. finite fields

or algebraic number fields.

The proper foundations of the enlarged view of algebraic geometry were, however, unclear and this is how Grothendieck made his first, hugely significant, innovation: he invented a class of geometric structures generalizing varieties that he called *schemes*. In simplest terms, he proposed attaching to *any* commutative ring (any set of things for which addition, subtraction and a commutative multiplication are defined, like the set of integers, or the set of polynomials in variables x, y, z with complex number coefficients) a geometric object, called the *Spec* of the ring (short for spectrum) or an affine scheme, and patching or gluing together these objects to form the scheme. The ring is to be thought of as the set of functions on its affine scheme.

To illustrate how revolutionary this was, a ring can be formed by starting with a field, say the field of real numbers, and adjoining a quantity ε satisfying $\varepsilon^2 = 0$. Think of ε this way: your instruments might allow you to measure a small number such as $\varepsilon = 0.001$ but then $\varepsilon^2 = 0.000001$ might be too small to measure, so there's no harm if we set it equal to zero. The numbers in this ring are $a + b \cdot \varepsilon$ with real a, b. The geometric object to which this ring corresponds is an infinitesimal vector, a point which can move infinitesimally but to second order only. In effect, he is going back to Leibniz and making infinitesimals into actual objects that can be manipulated. A related idea has recently been used in physics, for superstrings. To connect schemes to number theory, one takes the ring of integers. The corresponding Spec has one point for each prime, at which functions have values in the finite field of integers mod p and one classical point where functions have rational number values and that is 'fatter', having all the others in its closure. Once the machinery became familiar, very few doubted that he had found the right framework for algebraic geometry and it is now universally accepted.

Going further in abstraction, Grothendieck used the web of associated maps – called morphisms – from a variable scheme to a fixed one to describe schemes as *functors* and noted that many functors that were not obviously schemes at all arose in algebraic geometry. This is similar in science to having many experiments measuring some object from which the unknown real thing is pieced together or even finding something unexpected from its influence on known things. He applied this to construct new schemes, leading to new types of objects called *stacks* whose functors were precisely characterized later by Michael Artin.

His best known work is his attack on the geometry of schemes and varieties by finding ways to compute their most important topological invariant, their

cohomology. A simple example is the topology of a plane minus its origin. Using complex coordinates (z, w), a plane has four real dimensions and taking out a point, what's left is topologically a three dimensional sphere. Following the inspired suggestions of Grothendieck, Artin was able to show with algebra alone that a suitably defined third cohomology group of this space has one generator, that is the sphere lives algebraically too. Together they developed what is called *étale cohomology* at a famous IHES seminar. Grothendieck went on to solve various deep conjectures of Weil, develop *crystalline cohomology* and a meta-theory of cohomologies called *motives* with a brilliant group of collaborators whom he drew in at this time.

In 1969, for reasons not entirely clear to anyone, he left the IHES where he had done all this work and plunged into an ecological/political campaign that he called *Survivre*. With a breathtakingly naive spirit (that had served him well doing math) he believed he could start a movement that would change the world. But when he saw this was not succeeding, he returned to math, teaching at the University of Montpellier. There he formulated remarkable visions of yet deeper structures connecting algebra and geometry, e.g. the symmetry group of the set of all algebraic numbers (known as its Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$) and graphs drawn on compact surfaces that he called 'dessin d'enfants'. Despite his writing thousand page treatises on this, still unpublished, his research program was only meagerly funded by the CNRS (Centre Nationale de Recherche Scientifique) and he accused the math world of being totally corrupt. For the last two decades of his life he broke with the whole world and sought total solitude in the small village of Lasserre in the foothills of the Pyrenees. Here he lived alone in his own mental and spiritual world, writing remarkable self-analytic works. He died nearby on Nov. 13, 2014.

As a friend, Grothendieck could be very warm, yet the nightmares of his childhood had left him a very complex person. He was unique in almost every way. His intensity and naivety enabled him to recast the foundations of large parts of 21st century math using unique insights that still amaze today. The power and beauty of Grothendieck's work on schemes, functors, cohomology, etc. is such that these concepts have come to be the basis of much of math today. The dreams of his later work still stand as challenges to his successors.

The sad thing is that this was rejected as much too technical for their readership. Their editor wrote me that 'higher degree polynomials', 'infinitesimal vectors' and 'complex space' (even complex numbers) were things at least half their readership had never come across. The gap between the world I have lived in and that *even of scientists* has never seemed larger. I am prepared for lawyers and business people to say they hated math and not to remember any math beyond arithmetic, but this!? Nature is read only by people belonging

to the acronym 'STEM' (= Science, Technology, Engineering and Mathematics) and in the Common Core Standards, all such people are expected to master a heck of a lot of math. Very depressing.

Well, Nature magazine really wanted to publish *some* obit on Grothendieck and wore us out until we agreed with a severely stripped down re-edit. The obit came out in the Jan.15 issue, which is now free to download. The whole issue of trying to bridge the gap between the mathematician's world and that of other scientists or that of lay people is a serious one and I believe mathematicians could try harder to find bridges. An example is Gower's work on bases in Banach spaces: when he received the Fields Medal, no one to my knowledge used the example of musical notes to explain Fourier series and thus bases of function spaces to the general public.

In the case of our obit, I had hoped that the inclusion of the unit 3-sphere in $\mathbb{C}^2 - (0,0)$ would be fairly clear to most scientists and so could be used to explain the Mike Artin's breakthrough that $H^3_{\acute{e}tale}(\mathbb{A}^2 - (0,0)) \neq (0)$. No: excised by Nature. I had hoped that the "web of maps" was an excellent metaphor for the functor represented by an object in a category and gave one the gist. No: excised by Nature. I had hoped that the "symmetry group of the set of all algebraic numbers" might pass muster to define this Galois group. No: excised by Nature. To be fair, they did need to cut down the length and they didn't want to omit the personal details.

The essential minimum I thought for a Grothendieck obit was to make some attempt to explain schemes and say something about cohomology. To be honest, the central stumbling block for explaining schemes was the word "ring." If you haven't taken an intro to abstract algebra, where to begin? The final draft settled on mentioning in passing three examples – polynomials (leaving out the frightening phrase "higher degree"), the dual numbers and finite fields. We batted about Spec of the dual numbers until something approaching an honest description came out, using "very small" and "infinitesimal distance." As for finite fields, in spite of John's discomfort, I thought the numbers on a clock made a decent first exposure. OK, $\mathbb{Z}/12\mathbb{Z}$ is not a field but what faster way to introduce finite rings than saying "a type of number that is added like the hours on a clock – 7 hours after 9 o'clock, the clock reads 4 o'clock, not 16 o'clock." We then describe characteristic p as a "discrete" world, in contrast to the characteristic 0 classical/continuous world. Here is our final draft, omitting the beginning and end parts that were only lightly edited: .

Alexander Grothendieck (1928-2014)

Mathematician who rebuilt algebraic geometry.

Algebraic geometry is the field that studies the solutions of sets of polynomial equations by looking at their geometric properties. For instance a circle is the set of solutions of $x^2 + y^2 = 1$ and in general such a set of points is called a variety. Traditionally, algebraic geometry was limited to polynomials with real or complex coefficients, but just prior to Grothendieck's work, André Weil and Oscar Zariski had realized that it could be connected to number theory if you allowed the polynomials to have coefficients in a finite field. These are a type of number like the hours on a clock – 7 hours after 9 o'clock is not 16 o'clock, but 4 o'clock – and it creates a new discrete type of variety, one variant for each prime number p.

But the proper foundations of this enlarged view were unclear and this is where, inspired by the ideas of the French mathematician Jean-Pierre Serre, but generalizing them enormously, Grothendieck made his first, hugely significant innovation: he proposed that a geometric object called a scheme was associated to any commutative ring – that is, a set in which addition and multiplication are defined and multiplication is commutative, $a \times b = b \times a$. Before Grothendieck, mathematicians considered only the case in which the ring is the set of functions on the variety that are expressible as polynomials in the coordinates. In any geometry, local parts are glued together in some fashion to create global objects, and this worked for schemes too.

An example might help in illustrating how novel this idea was. A simple ring can be generated if we make a ring from expressions $a + b\varepsilon$ where a and b are ordinary real numbers but ε is a variable with only 'very small' values, so small that we decide to set $\varepsilon^2 = 0$. The scheme corresponding to this ring consists of only one point, and that point is allowed to move the infinitesimal distance ε but no further. The possibility of manipulating infinitesimals was one great success of schemes. But Grothendieck's ideas also had important implications in number theory. The ring of all integers, for example, defines a scheme that connects finite fields to real numbers, a bridge between the discrete and classical worlds, having one point for each prime number and one for the classical world.

Probably his best known work was discovering how all schemes have a topology. Topology had been thought to belong exclusively to real objects, like spheres and other surfaces in space. But Grothendieck found not one but two ways to endow all schemes, even the discrete ones, with a topology, and especially with the fundamental invariant called cohomology. With a brilliant group of collaborators, he gained deep insights into the theory of cohomology, and established

them as one of the most important tools in modern mathematics. Owing to the many connections that schemes turned out to have to various mathematical disciplines, from algebraic geometry to number theory to topology, there can be no doubt that Grothendieck's work recast the foundations of large parts of twenty-first century mathematics.

——omit last two paragraphs——

The whole thing is a compromise and I don't want to say Nature is foolish or stupid not to allow more math. The real problem is that such a huge and painful gap has opened up between mathematicians and the rest of the world. I think that Middle and High School math curricula are one large cause of this. If math was introduced as connected to the rest of the world instead of being an isolated exercise, if it was shown to connect to money, to measuring the real world, to physics, chemistry and biology, to optimizing decisions and to writing computer code, fewer students would be turned off. In fact, why not drop separate High School math classes and teach the math as needed in science, civics and business classes? If you think about it, I think you'll agree that this is not such a crazy idea.

I got a lot of feedback after posting this blog. My old friend at UCLA, David Gieseker, wrote to me about what is happening there:

> *We've been having a lot of trouble with scientists, in particular life scientists. They are teaching calculus by radically dumbing it down. E.g. no trig, a half page on the chain rule, ... and very weak exams. This is being pushed by the Dean of Life Science, ostensibly so that math phobic students are not turned off science. The people in charge seem to be ecologists and they don't believe in any math that's not what they use. I suspect these students will be in real trouble when they take physics. I also suspect the readers of Nature think they know all important math and get upset if it's hinted that there's important math they haven't even heard of.*

A sad story. Let's be honest: how much math do biologists need? I would argue first of all that oscillations are central part of *every* science plus engineering/economics/business (arguably excluding computer science) and one needs the basic tools for describing them – sines and cosines, all of trig of course, especially Euler's formula $e^{ix} = \cos(x) + i.\sin(x)$ and Fourier series. And, of course, modeling a system by the path of a state vector in some $\mathbb{R}^n$, often with a PDE, is also ubiquitous. For example, surely all ecologists have studied the Lotka-Volterra equation (wolf and rabbit population cycles). Algebra is more of a mixed bag. Splines are much more useful than polynomials for engineers, finite fields arise mostly in coding applications and I doubt that the abstract idea of a *ring* is ever needed. But polynomials and varieties have been used in Sturmfels' algebraic statistics and, as Lior Pachter noted (see below), is very effectively used in modeling genome mutation. But evolutionary genomics is one community within biology and John and I figured we needed to throw into the obit a rough definition of a ring.

I also received email from a computational biologist Steven Salzberg about the challenge of bridging the gap between math and biology, including a link to a fascinating blog on this gap by another mathematical biologist Lior Pachter: `http://liorpachter.wordpres s.com/2014/12/30/the-two-cultures-of-mathematics-and-biology`. Pachter details how varieties arise as sets of probabilities consistent with a class of models, an application I was only dimly aware of when writing the obit with John Tate. He then elaborates at length on the many ways in which the culture of mathematicians and of biologists differ, cultures that he straddles at UC Berkeley. As he goes on to say, "The extent to which the two cultures have drifted apart is astonishing" and worse, both sides seem happy to ignore each other. To illustrate this, he cites another side to the situation at UCLA mentioned by Gieseker – that the math dept is not one of 15 partner departments to UCLA's new "Institute for Quantitative and Computational Biosciences." This split is to their joint detriment and as Pachter says:

> *The laundry list of differences between biology and math that I aired above can be overwhelming. Real contact between the subjects will be difficult to foster, and it should be acknowledged that it is neither necessary nor sufficient for the science to progress. But wouldn't it be better if mathematicians proved they are serious about biology and biologists truly experimented with mathematics?*

But forgetting biologists, what would we really want to explain about Grothendieck's ideas? I had another opportunity quite recently:

ii.   A geologist vs. π_1 & topoi

I was asked by a good friend, the Bulgarian geologist Andrew Stancioff, a man with broad curiosity and interests, can you explain to me the result for which Shinichi Mochizuki is famous? Well, he is famous for proving a conjecture of Grothendieck and Andrew did not want me to talk about his proof, only what it meant. And he knows a lot more math than the editors of Nature. When I began thinking about this, I slowly realized that I had to go back quite a ways. The conjecture concerns the fundamental group of a curve over a number field. This is a group extension of the fundamental group (OK, the "pro-finite" completion) of the points of the curve over the complexes, a smooth real surface, by the Galois group of the field of all algebraic numbers (over the field where the curve is defined). The conjecture is that this extension determines the curve.

This complex of ideas results from the merger of topological (π_1) and algebraic (Galois theory) ideas. This merger has roots in the late 19$^{\text{th}}$ century. It came from the tight parallels discovered then between the algebra of the rings of algebraic numbers and that of the rings of polynomial functions on affine curves over $\mathbb{C}$; and between the finite field extensions of number fields and of the fields of rational functions on complex algebraic curves. This analogy is described in glowing terms in Felix Klein's book on the History of

Math in the 19$^{\text{th}}$ Century [Kle79]. On p.334, you find the remarkable passage where he is describing the ideas foreshadowed by papers in Kronecker's *Festschrift* of 1881:

> *Es handelt sich nicht nur um die reinen Zahlkörper oder Körper, die von einem Parameter z abhängen, oder die Analogisierung dieser Körper, sondern es handelt sich schliesslich darum, für Gebilde, die gleichzeitig arithmetisch und funktionentheoretisch sind, also von gegebenen algebraischen Zahlen und gegebenen algebraischen Funktionen irgendwelcher Parameter algebraisch abhängen, das selbe zu leisten, was mehr oder weniger vollständig in den einfachsten Fällen gelungen ist.*
> *Es bietet sich da ein ungeheuerer Ausblick auf ein rein theoretisches Gebiet, welches durch seine allgemeinen Gesetzmässigkeiten den grössten ästhetischen Reiz ausübt ...*
> (Free translation: This isn't only about number fields or fields that depend on one parameter, or the analogs of such fields. Ultimately, one wants to carry over what has been done, more or less, in those basic cases, to objects that are simultaneously arithmetic and function-theoretic, that is objects that depend on given algebraic numbers and algebraic functions of arbitrary parameters.
> This offers an enormous vision of a purely theoretical field, which through its general principles has the greatest aesthetic appeal.)

Is Klein channeling Grothendieck or what!!? His *Gebilde* are surely examples of what Grothendieck called schemes. I want to give a simple example that illustrates the synthesis he is talking about. This example uses for algebraic numbers the square roots of particular numbers and for algebraic functions the square roots of particular polynomials. Starting with integers a, b, the set of all numbers of the form $a + b\sqrt{2}$ is a ring of algebraic numbers. Next, we can form the two algebraic functions $s_1(x) = \sqrt{x + \sqrt{2}}$, $s_2(x) = \sqrt{x - \sqrt{2}}$. Now consider the collection of all expressions formed from 8 polynomials $a(x), \cdots, h(x)$ with integer coefficients:

$$a(x) + b(x)\sqrt{2} + \left(c(x) + d(x)\sqrt{2}\right) s_1 + \left(e(x) + f(x)\sqrt{2}\right) s_2 + \left(g(x) + h(x)\sqrt{2}\right) s_1 s_2$$

It's easy to see that the product of two such expressions is another such expression, i.e. the set of all such expressions forms a ring that *mixes algebraic numbers and algebraic functions*. Moreover, this ring has *symmetries*: you can flip the sign of s_1 or flip the sign of s_2 or leave them alone but replace $\sqrt{2}$ by $-\sqrt{2}$ (thus also interchanging s_1 and s_2. Technically, the first two are symmetries in $\pi_1(\mathbb{C} - \{\sqrt{2}, -\sqrt{2}\})$ and the third is a symmetry in a Galois group. So all the ingredients of Grothendieck's conjecture are here. The three symmetries generate the non-commutative Galois group of order 8 for the quotient field of the above ring over the field of rational functions in x with rational coefficients, a finite version of the groups in Mochizuki's theorem.

I'm not sure how many scientists or engineers have the patience to follow the above but it does contain the key point, that symmetries from numbers and from functions intertwine.

If one wants to delve a bit deeper, one needs to describe π_1. If you want to explain this to a non-mathematician, the hard point to accept is *not to make a definition*. Instead, one needs to to give an illustrative example or perhaps a simile or metaphor with a bit of vagueness. I would suggest a slinky. Collapsed, it is a cylinder and squooshed, just a circle. But expanded it is a very long wire. Imagine that the wire has no ends but winds infinitely often both above and below. Then the slinky is the *universal covering space* of the circle and π_1 is the set of its symmetries that shift the loops a discrete number of loops up or down, while keeping every point always at the same angle to the axis. Mathematically, this is described by the complex exponential function $e^{ix} = \cos(x) + i\sin(x)$. This takes the whole x-line and wraps it infinitely many times around the unit circle, identifying any two points that differ by a multiple of 2π. Adding a real part to ix, the full plane covers the plane minus the origin wrapping it around the origin infinitely often. So the idea is that the covering by the "log"-plane displays the topology of the punctured plane. The complex plane, for algebraic geometers, is just the points of the affine line in the canonical algebraically closed field $\mathbb{C}$. The log is approximated by taking higher and higher n^{th} roots. What we see is that, by taking roots, you are getting closer and closer to topologically trivial spaces.

For a surface, e.g. the surface of a pretzel, one can also unwind all the many circles on it and this was one of the main topics of Klein's research. He gave wonderful ways to visualize even these covering spaces, as described in my book with Caroline Series and Dave Wright [MSW02].

Now Grothendieck hated particular examples and always sought the most abstract essence of a problem. He was not content with the idea of schemes but saw a scheme as a special case of a *topos* (plural *topoi*). We can roughly explain this in two stages. The first stage involves what it means to break a space up into its simplest parts. Here too a real life illustration may explain something of what he did. Your body is a complicated shape (ignore the head) but one covers it when needed with a shirt, pants, two gloves and two socks. These 6 items cover the whole body with some overlap, giving what mathematicians call a covering of a topological space by what are called open subsets. The shirt and the pants both have circles on them so that can be "unwrapped" by variants of the log function. But now suppose the person adds a shawl or a wrap-around skirt. These overlap *themselves*, covering a single point of the body multiple times. Like unwrapping, these items multiply cover parts of the body. With the shawl and the wrap-around skirt, we have pieces of a covering that do not correspond to a *subset* of the body. Coverings are the bread-and-butter of topology but Grothendieck realized they need not be made up of subsets but can have multiple layers and this led him to define a *site* where all the objects come with a set of distinguished coverings called *sieves*.

The second stage has already been suggested when introducing schemes. Prior to his work, people thought of space as primarily a set of points and secondarily as having coordinate functions on these points. Grothendieck said the functions come first, the points second. Grothendieck's insight with schemes was to invert this: why shouldn't any ring be

imagined as the set of local functions on some sort of a geometric object. (Well, not quite. One better insist that $f \times g = g \times f$ or the analogy gets much subtler.) The points of the geometric object take second stage compared to the rings of functions on the pieces of the scheme. The concept of a topos now *ignores the points entirely* but focuses on ways to assign data to each open set that fit consistently together for each covering. The simplest case of this are the rings of algebraic functions on the covering pieces but one can even take finite discrete data. Such things are called *sheaves*. Phew. Topoi are now a cottage industry in math but were described in a very poetic, metaphoric way in Grothendieck's extraordinary reflective work *Recoltes et Semailles* [Gro86]. These passages were pointed out to me by Curt McMullen. In §2.13, Grothendieck writes:

> *Un lit si vaste en effet (telle une vaste et paisible rivière très profonde. . .), que*
> > *"tous les chevaux du roi*
> > *y pourraient boire ensemble. . . "*
> *- comme nous le dit un vieil air que sûrement tu as dû chanter toi aussi, ou du moins l'entendre chanter. Et celui qui a été le premier à le chanter a mieux senti la beauté secrète et la force paisible du* topos, *qu'aucun de mes savants l'élèves et amis d'antan. . .*
> (A bed so vast indeed (like a vast, peaceful and very deep river) such that "all the king's horses could drink together," – as in the old song that you must have sung or, at least, heard sung. And whoever was first to sing it sensed the secret beauty and the peaceful force of a *topos* like none of my learned students or old friends ... .)

A curious thing is that the full song is about a cobbler seducing a beautiful lady and the bed is where they consummate the relationship. A second passage to which Curt drew my attention is where he notes certain similarities between his reformulation of the idea of space using schemes and Einstein's using general relativity. In §2.20, he writes:

> *La comparaison entre ma contribution à la mathématique de mon temps, et celle d' Einstein à la physique, s'est imposée à moi pour deux raisons : l'une et l'autre oeuvre s'accomplit à la faveur d'une mutation de la conception que nous avons de "l'espace" (au sens mathématique dans un cas, au sens physique dans l'autre); et l'une et l'autre prend la forme d'une vision unificatrice, embrassant une vaste multitude de phénomènes et de situations qui jusque là apparaissaient comme sèparés les uns des autres.*
> (The comparison between my recent contribution to mathematics and Einstein's to physics occurred to me for two reasons: both works involve a mutation in our conception of "space" (one in mathematics, the other in physics); and both take the form of a unifying vision, embracing a vast multitude of phenomena and of situations that had previously been viewed as separate from each other.)

His legacy is indeed vast and unifying.

I cannot resist describing one final topic that resulted from Grothendieck's unification of topology and algebra even though more math is needed to follow this. For me, conversations with Barry Mazur and Michael Artin in the late 50's gave rise to an unexpected analogy, but the idea was likely wide-spread. Take the simplest of all schemes, $\mathrm{Spec}(\mathbb{Z})$. *The idea began to jell that this scheme was like a 3-sphere with all the primes being knots in it!!* More generally, from an étale cohomology point of view, the rings of algebraic integers were all like 3-manifolds. To my knowledge, the first theorems on this appear the notes from the 1964 Woods Hole Algebraic Geometry Symposium in the seminar by Michael Artin and Jean-Louis Verdier [AV64].

Where does this astonishing idea come from? First of all, for all finite fields $k = \mathrm{GF}(p^n)$, their absolute Galois group $\mathrm{Gal}(\bar{k}/k)$ is just $\hat{\mathbb{Z}}$, the pro-finite completion of the integers, so the schemes $\mathrm{Spec}(k)$ should be thought of as a simple circles contained in whatever sort of space $\mathrm{Spec}(\mathbb{Z})$ is. Secondly, each such finite field is the residue field of a unique complete local ring R_p of characteristic zero with p generating its maximal ideal. Its Spec has the same unramified extensions as the finite field and should be thought as a thickening of the circle. Thirdly, the Spec of the quotient field K_p of R_p should be thought of as the boundary of this thickening as we get it by throwing away the closed point. So what is its absolute Galois group $\mathrm{Gal}(\overline{K_p}/(K_p)$? It has a "tame" part obtained by n^{th} roots of p for integers n with $p \nmid n$ and a "wild" part[2] of extensions with degrees that are powers of p.The tame part is readily seen to be close to the completion of $\mathbb{Z}^2$ but not quite. It has two generators, the Frobenius map ϕ that lifts the $(p^n)^{\mathrm{th}}$ power in the residue field and ψ, multiplication by roots of unity for the roots of p. But they don't commute. Instead $\phi \circ \psi \circ \phi^{-1} = \psi^{p^n}$. The conclusion is that, geometrically, the boundary of the tube around the circle corresponding to a prime is like a twisted form of a 2-torus. Already, this suggests that $\mathrm{Spec}(\mathbb{Z})$ ought to be 3-dimensional.

The simplest way to bring H^3 into the picture is to cover $\mathrm{Spec}(\mathbb{Z})$ by two "open" sets and use the Mayer-Vietoris exact sequence. Actually, it's much better to start with the ring R of algebraic integers in some number field K that contains the ℓth roots of unity for some odd prime ℓ. Choose a closed point $x \in \mathrm{Spec}(R)$ (not of characteristic ℓ) and "cover" $X = \mathrm{Spec}(R)$ by:

1. $U_1 = X - \{x\}$

2. $U_2 = \mathrm{Spec}(\overline{R_x})$, the Spec of the completion of the local ring at x

I have put the words open and cover in scare quotes because U_2 is obviously *not* an open subset. But its cohomology is known to be the same as that of the *henselization* of R_x and the henselization is the direct limit of bona fide étale covers of X. Although I don't have a

---

[2]One should mention here the amazing theory of Peter Scholze's *perfectoids* [Sch12], that go beyond schemes and eliminate this awkward wild part.

reference, I believe the Mayer-Vietoris exact sequence is valid with $U_1 \cap U_2 = U_1 \times_X U_2 = \text{Spec}(\overline{K_x})$, with $\overline{K_x}$ the quotient field of $\overline{R_x}$ and gives the homomorphism:

$$H^2(\text{Spec}(\overline{K_x}), \mathbb{Z}/\ell\mathbb{Z}) \to H^3(X, \mathbb{Z}/\ell\mathbb{Z}).$$

More work is needed to check that the left side is $\mathbb{Z}/\ell\mathbb{Z}$ and that the arrow is bijective (using Hasse's theory of division algebras over R). But it is a theorem that the right hand side is indeed $\mathbb{Z}/\ell\mathbb{Z}$.

There are other striking analogies, e.g. the symmetry of the linking number of two circles connects to Gauss's quadratic reciprocity. None are exact but many are suggestive. A recent survey is [Mor12].

Chapter 3

Are Mathematical Formulas Beautiful?

i.   Equations as art

This Chapter has two parts, both dealing with the question: *what is a beautiful mathematical formula*? Mathematicians do like to talk of a "beautiful result" and often it can be condensed into a formula, but what does this mean? Strangely, at roughly the same time, two mathematicians, Dan Rockmore and Michael Atiyah, decided, in two different ways, to try to pin this down. The first part is about an astonishing project of Dan Rockmore and Bob Feldman that has, unbelievably, placed some such math formulae in art museums and collections around the world![1]A couple of years ago, my good friend Dan sent me by FedEx a remarkable invitation: write on copper plate "what you think is your most significant and elegant equation," for a limited edition of etchings. Even for Dan, whom I've known for unorthodox projects, this seemed off the wall. But OK: Yole Zariski, the wife of my PhD advisor Oscar Zariski, had her artist brother cast the symbols of some of Oscar's results on a necklace that she loved. Maybe the odd symbols that we put together might be viewed as a contemporary form of magic and, even if not understood, having a McLuhan-esque significance. The project is described in their website www.concinnitasproject.org.

Now I've always had a complex attitude towards (all caps) ART that started with my sister Daphne and brother-in-law (Charles Duback) being artists and watching them struggle with evolving tastes and fashions and expressing their own muse's visions. Then my oldest son Steve became an artist, my second son Peter became a photographer, I married an artist Jenifer, whose sister Mimo, second son Andrew and his wife Heather are all artists and finally Steve married the artist Inka Essenhigh – you get the picture. Of course we collect a lot of art – "friends and family" we call it, so I follow prices, galleries, and reviews a tiny bit. I'm aware, especially after reading the book *Seven Days in the Art*

---

[1]This is based on my post "Is it Art?" dated Feb.9, 2015.

World, [Tho08], of some of the bizarre aspects of the art scene. In another direction, I have found striking parallels between the history of art and the history of math going back to 1800 when abstraction begins to play a role in both (see Chapter 7). These are two fields that are not dependent on language and so can manifest the zeitgeist of the age more directly. In yet another direction, both the Paris group in computer vision led by Jean-Michel Morel and my own research on the statistics of images were led to stochastically synthesized images and we have noticed how naturally some kinds of abstract art emerge.

Dan's project had emerged from a serendipitous meeting on a trans-continental flight with an unorthodox publisher, Bob Feldman of Parasol Press, who has created beautiful portfolios of many great artists, so we were in very exalted company. Sol Lewitt is perhaps the best point of reference. Bob had always wondered if math could be made into art and Dan had likewise wondered if art could be made out of math. So the mailing tube Dan sent me was full of many types of paper and drawing instruments (no copper plate) and Jenifer and I spread them out on the dining room table. Thank god that she knows art materials and after I play with charcoal a bit, I find I can make believe I am talking to a class and writing on a blackboard. My contribution was a startling identity that arose studying the geometry of moduli space. Besides being a lovely, basic fact about moduli spaces, this identity is most peculiar in having the number 13 appear in it! As I said in the accompanying blurb, the only numbers bigger than 2 that are likely to appear in a math article are usually page numbers.[2] '13' was, to say the least, really unexpected. This identity also has the merit that it has been used by string theorists.

Now the plot thickens. Together with 9 other mathematicians, physicists and computer scientists, the portfolio of formulae was put together using *aquatints* that inverted the colors, now white on black like chalk on a blackboard. This is apparently an awfully hard process to master, especially with thin lines scratched on the paper. But Harlan and Weaver succeeded and the lot is being sent around the world with the title *Concinnitas* from art gallery to art gallery: Zurich, Seattle, Portland, Yale and even the Metropolitan Art Museum in NYC[3]. A panel discussion was arranged at the Yale Art Gallery where I met much of the cast of characters. Amazingly, a couple of hundred people showed up to hear the discussion. That's where I heard how challenging the aquatint process was and had the pleasure of meeting Bob Feldman. And we also learned from Yale professor Asher Auel that, like artists with different favorite paints, mathematicians can avail themselves of three types of chalk that make quite different sorts of lines, something new to me.

Most of the panel discussion, however, centered on the question – "Is it Art?." In fact, this precise question was even discussed in the Scientific American `http://blogs.scient ificamerican.com/sa-visual/2015/01/27/math-can-be-beautiful-but-is-it-art`! I had just seen upstairs in the museum a quite wonderful wall done by Sol Lewitt made from panels with all permutations of two curved arcs, butting up to each other. What

---

[2]This isn't true of physics: see the compendium of numbers assembled by Nick Trefethen, see `people.maths.ox.ac.uk/trefethen/5farmelo.pdf.`

[3]OMG, something I drew sat in the Met for a month!!

he found was that the serendipitous pairings on adjacent panels created a spider web of contours that, for me anyway, "worked" as an entry point for math into art. I was not so sure that any such unanticipated magic emerged from our scrawls, that would raise them above the status of fetishistic objects for the layman to worship. Still confused about what is art, my wife and I went the next day to stay in NYC with two people in the thick of it – my son Steve and his wife Inka. Steve told me: "read Tom Wolfe's *The Painted Word*," [Wol08]. And I did. What an eye opener. The whole history of 20th century art began to make sense. If you haven't cracked this slim volume, let me reproduce the quote that sets him off, from Hilton Kramer in the April 28, 1972 Times (reviewing a show at Yale in fact):

> Realism does not lack its partisans, but it does rather conspicuously lack a persuasive theory. And given the nature of our intellectual commerce with works of art, to lack a persuasive theory is to lack something crucial – the means by which our experience of individual works is joined to our understanding of the values they signify.

Wolfe, not persuaded himself, goes on to detail the many theories shilled by art critics that supported all the isms of 20th century art. Now formulas began to seem more plausible as grist for this mill. Minimalism? Conceptual Art? Urban graffiti? Surely there's a place there somewhere for formulas. All it needs is its own unique *persuasive theory*! Once this is found, Dan and Bob's project will have legs. Another school says that great art is what people still enjoy a century or two later. This is a more usable definition than the presence of a persuasive theory though it does require a lot more patience than critics have at hand.

On the next page are thumbnail reproductions of these ten aquatints from the Concinnitas portfolio.. There is no room for the caption so I want to add that these are reproduced by permission of the publisher Parasol Press, Ltd.

ii.   Equations reflected in MRI scans and mathematical tribes

Now the second part of the Chapter. Here the question is: is there a special part of cortex which is highly active when mathematicians do math and see beautiful formulas?[4] Recently Professors Michael Atiyah and Semir Zeki have addressed this question, collaborating on an astonishing *experimental* investigation of these questions culminating in a paper entitled "The experience of mathematical beauty and its neural correlates," [ZRBA14]. Fifteen mathematicians were scanned using fMRI (functional magnetic resonance imaging) while viewing 60 mathematical formulas and rating them as ugly, neutral or beautiful. The first 15 are shown below in a table following the ten favorite formulas Dan and Bob solicited. Their main result is that activity in the mOFC = medial (near the centerline) orbital (in

---

[4]This is based on my post "Math & Beauty & Brain Areas" on Oct.11, 2015.

the inward curl of the cortex just above the eyes) frontal cortex *correlates* to some extent with their judgement of beauty (though strangely activity in mOFC relative to baseline diminishes). My aim in the second part of this chapter is to argue for the view that the subjective nature and attendant excitement during mathematical activity, including a sense of its beauty, varies greatly from mathematician to mathematician and that that would make it plausible for quite different parts of the brain to be active during mathematical reflection. I do not claim any scientific basis for this as my only evidence comes from opportunities to talk with colleagues and being struck with the remarkably diverse ways they seem to have of "doing math".

A word of apology before I get started: much of what i want to say is understandable to non-mathematicians, but, in order to make my case, I need to cite many specific mathematicians and mathematical results that are only clear to fellow mathematicians. I have included some background to make the ideas clearer to non-mathematicians but this is an uneasy compromise.

1.	$e^{i\pi} = -1$	Euler's identity relating e, i and π
2.	$\cos^2\theta + \sin^2\theta = 1$	The Pythagorean identity via trig functions
3.	$V - E + F = 2$	The Euler characteristic for a spherical polyhedron
4.	$\int_M K dA + \int_{\partial M} k_g ds = 2\pi\chi(M)$	The Gauss-Bonnet formula connecting curvature and topology
5.	$e^{ix} = \cos(x) + i\sin(x)$	The complex exponential
6.	$\int_{-\infty}^{\infty} e^{-x^2} = \sqrt{\pi}$	The definite Gaussian integral, key in stat and physics
7.	$\frac{1}{\zeta(s)} = \sum_{n=0}^{\infty} \frac{\mu(n)}{n^s}$	Dirichlet series for inverse zeta function
8.	$\exp(X) = \sum_{n=0}^{\infty} \frac{X^n}{n!}$	Series expansion for exponential
9.	$\mathcal{F}_x\left[e^{-ax^2}\right](k) = \sqrt{\frac{\pi}{a}} e^{-\pi^2 k^2/a^2}$	Fourier transform of a Gaussian is Gaussian
10.	$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n$	Compound interest definition of e
11.	$2^{|S|} > |S|$	A generalization of the cardinality of the reals being greater than the cardinality of the integers
12.	$z_{n+1} = z_n^2 + c$	The iteration leading to the Mandelbrot set
13.	$f(x) = \int_{-\infty}^{\infty} \delta(x - y) f(y) dy$	The definition of the delta function
14.	$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k!)(1103 + 26390k)}{(k!)^4 396^{4k}}$	An utterly insane baroque formula for the inverse of π
15.	$1729 = 1^3 + 12^3 = 9^3 + 10^3$	An odd fact famous from Ramanujan's quoting it to Hardy while he was on his sickbed

I think one can make a case for dividing mathematicians into several *tribes* depending on

what most strongly drives them in their esoteric world. I like to call these tribes *explorers, alchemists, wrestlers and detectives.* Of course, many mathematicians move between tribes and some results are not cleanly part the property of one tribe.

- Explorers are people who ask – are there objects with such and such properties and if so, how many? They feel they are discovering what lies in some distant mathematical continent and, by dint of pure thought, shining a light and reporting back what lies out there. The most beautiful things for them are the wholly new objects that they discover (the phrase 'bright shiny objects' has been in vogue recently) and these are especially sought by a sub-tribe that I call Gem Collectors. Explorers have another sub-tribe that I call Mappers who want to describe these new continents by making some sort of map as opposed to a simple list of 'sehenswürdigkeiten'.

- Alchemists, on the other hand, are those whose greatest excitement comes from finding connections between two areas of math that no one had previously seen as having anything to do with each other. This is like pouring the contents of one flask into another and – something amazing occurs – like an explosion!

- Wrestlers are those who are focussed on relative sizes and strengths of this or that object. They thrive not on equalities between numbers but on inequalities, what quantity can be estimated or bounded by what other quantity, and on asymptotic estimates of size or rate of growth. This tribe consists chiefly of analysts and uses integrals that measure the size of functions, but people in every field get drawn in.

- Finally Detectives are those who doggedly pursue the most difficult, deep questions, seeking clues here and there, sure there is a trail somewhere, often searching for years or decades. These too have a sub-tribe that I call Strip Miners: these mathematicians are convinced that underneath the visible superficial layer, there is a whole hidden layer and that the superficial layer must be stripped off to solve the problem. The hidden layer is typically more abstract, not unlike the 'deep structure' pursued by syntactical linguists. Another sub-tribe are the Baptizers, people who name something new, making explicit a key object that has often been implicit earlier but whose significance is clearly seen only when it is formally defined and given a name.

I want to give examples for each tribe of specific beautiful results and specific people I have known and interacted with in this tribe.

Explorers:

Arguably the archetypal discovery by explorers was the ancient Greek list of the five Platonic solids: the only 'regular' convex polyhedra (meaning that any face and vertex on that face can be carried to any other such face, vertex pair by a rotation of the polyhedron). This discovery is sometimes attributed to Theaetetus, is described by Plato in

the *Timaeus* dialog and worked out in detail in Euclid's *Elements*. I find it curious that nowhere, to my knowledge, is an icosahedron or a dodecahedron ever described in Indian or Chinese writings prior to the 17th century merging of their mathematical traditions with those of the West. Enlarging the mathematical universe from three dimensions to higher dimensions started a gold rush for explorers. In the 19th century, the Swiss mathematician Ludwig Schläfli extended the Greek list to regular polytopes in n dimensions, finding that there were 6 in four dimensional space but only 3 in all higher dimensional spaces. In the 20th century, exploring all possible low dimensional manifolds (both homeomorphic, piecewise-linear and differentiable types of manifolds) has been a major focus. I knew my contemporary Bill Thurston fairly well and he seems to me to have been clearly a member of the explorer tribe. He was a fantastic topologist and it was especially intriguing to me that he was born cross eyed, thus his understanding the 3D world was forced to depend more on parietal brain areas and hand-eye coordination than occipital cortex, stereo based learning. I never met anyone with anything close to his skill in visualization (except perhaps for H. S. M. Coxeter).

But explorers are not all geometers: the list of finite simple groups is surely one of the most beautiful and striking discoveries of the 20th century. Although he is not a card-carrying explorer, having devoted much of his career to detective work, in the second half of his career, Michael Artin discovered an amazing rich world of non-commutative rings lying in the middle ground between the almost commutative area and the truly huge free rings. "Rings" are sets of things that can be added and multiplied, but here he allows $x \cdot y \neq y \cdot x$. He really set foot on a continent where no one had a clue what might be found: this exploration is ongoing. And then there is that most peculiar, almost theological world of 'higher infinities' that the explorations of set theorists have revealed.

My own career has been centered in the mapper sub-tribe. My maps are called moduli spaces of varieties (finite-dimensional objects) and moduli spaces of sub-manifolds of Euclidean spaces (infinite-dimensional objects). But one can make the case that the earliest members of the explorer tribe, even the earliest mathematicians, were literally mappers. I have in mind the story told by cuneiform surveying tablets. The earliest organized states in the world confronted the tasks of keeping track of land ownership and of taxing farmers. We are lucky to have a vast collection of Mesopotamian tablets from the late third millennium to the mid first millennium BCE. Many of these tablets contain idealized maps of land or of geometric constructions stimulated by surveying tasks. It seems fairly clear that the scribes who wrote these tablets went on to discover much of the geometric algebra, Pythagoras's rule and the quadratic equation, as a result of being presented with practical land use and accounting challenges. They had no interest in questions of proof, only in algorithms related to measuring the earth, its distances and areas, (which they called the wisdom of the goddess Nisaba with her rope and measuring reed).

The Atiyah-Zeki list has very few results of explorers, perhaps because their results are not usually expressed by formulas. However, it contains three gems: #12 shown in the table above, the function whose iterations lead to the Mandelbrot set; #15 also in the

table, an integer expressible two ways as a sum of two cubes, famous because Ramanujan told it to Hardy when Hardy mentionned he had arrived by a taxi numbered 1729; and #28, not shown, is $3^2 + 4^2 = 5^2$, the formula that shows there is a right triangle with sides $(3, 4, 5)$. Among the formulas in the Rockmore-Feldman project described earlier in this Chapter, one finds a gem from the short list of finite simple groups, here the groups discovered by Rimhak Ree. I would like to add that some of the things that gave me the most pleasure in my own research were discovering unusual previously unknown geometric objects: one was a negatively curved algebraic surface whose homology was the same as that of the positively curved $\mathbb{P}^2$.

Alchemists:

For many people, the most wonderful results in mathematics are those that reveal a deep relationship between two very distant subjects, for instance a link between algebra and geometry, algebra and analysis or geometry and analysis. Such links suggest that the world has a hidden unity, previously concealed from our mortal eyes but blindingly beautiful if we stumble upon it. An early example of such a link is the connection of the geometric problem of trisecting an angle and the algebraic problem of solving cubic polynomial equations. The first was one of the major unsolved problems of the ancient Greek tradition. In the Renaissance, Italian algebraists found a mysterious formula for the roots of a cubic polynomial. But in the case where all three roots are real, their formula led to complex numbers and cube roots of such numbers. The French mathematician Viète was the 'alchemist' who made the link c. 1593: he showed how, if you can trisect angles, you can solve these cubic equations and vice versa. It wasn't until the early18th century, however, that another Frenchman, Abraham De Moivre, really explained the result with his formula

$$(\cos(\theta) + i.\sin(\theta))^n = \cos(n\theta) + i.\sin(n\theta).$$

This is surely alchemy. But I would classify the leading mathematicians of the 18th and early 19th century, Leonard Euler from Switzerland and Carl Fredrich Gauss from Germany as the "strip miners" who showed how two dimensional geometry lay behind the algebra of complex numbers. Euler's form of De Moivre's formula appears as #5 (and #1) of our table of the Atiyah-Zeki list.

My PhD advisor Oscar Zariski was surely an alchemist. His deepest work was showing how the tools of commutative algebra, which had been developed by straight algebraists, had major geometric meaning and could be used to solve some of the most vexing issues of the Italian school of algebraic geometry. More specifically, the algebraic notions of integral closure and of valuation rings were shown to relate to geometry in Zariski's 'Main theorem' and in his work on resolving singularities. He used to say that the best work was not proving new theorems but creating new techniques that could be used again and again.

The famous Riemann-Roch theorem has been an especially rich source of alchemy. It was from the beginning a link between complex analysis and the geometry of algebraic

curves. It was extended by pure algebra to characteristic p, then generalized to higher dimensions by Fritz Hirzebruch using the latest tools of algebraic topology. Then Michael Atiyah and Isadore Singer linked it to general systems of elliptic partial differential equations, thus connecting analysis, topology and geometry at one fell swoop. Out of modesty, Atiyah did not include this in his list but he did put in its special case, the Hirzebruch signature formula, in his aquatint in the Feldman-Rockmore project. These aquatints also include the Dyson-MacDonald combinatorial formula for $\tau(n)$, numbers which come from complex analysis: surely alchemy. Finally, a most bizarre formula for $1/\pi$ appears as formula #14 in the Atiyah-Zeki list. I suspect this was included by the authors because they suspected that many would think it ugly. I have no idea where it comes from but whoever found it belongs to the sub-tribe of Baroque Alchemists. It stands in contrast to the much simpler but nonetheless alchemical formula #30: $\pi = 1 - \frac{1}{2} + \frac{1}{5} - \frac{1}{7} + \cdots$ for π.

Wrestlers:

Wrestling goes back to Archimedes: he loved estimating π and concocting gigantic numbers. The very large and very small have always had a fascination for wrestlers. Calculus stems from the work of Newton and Leibniz and in Leibniz's approach depends on distinguishing the size of infinitesimals from the size of their squares which are infinitely smaller. A laissez-faire attitude towards infinities and infinitesimals dominated the 18th century, resulting in alchemy gone amuk as in Euler's really strange formulas:

$$\frac{1}{2} = 1 - 1 + 1 - 1 + 1 - \cdots, \quad \frac{1}{4} = 1 - 2 + 3 - 4 + 5 - \cdots$$

Of course Euler knew these only made sense when viewed in a very special way and he himself had not gone crazy. In fact, many might say the above are very beautiful formulas. A notable much more understandable achievement of wrestlers in this century was Stirling's formula $n! = (n/e)^n \sqrt{2\pi n}(1 + o(n))$ for the approximate size of $n!$ (#41 in the Atiyah-Zeki list). The modern father of the wrestling tribe in the 19th century should be the Frenchman Augustin-Louis Cauchy who finally made calculus rigorous. His eponymous inequality, that the absolute value of the dot product of 2 vectors is less than the product their lengths,

$$|(\vec{x} \cdot \vec{y})| \leqslant \|\vec{x}\| \cdot \|\vec{y}\|$$

remains the single most important inequality in math. Atiyah-Zeki include the related triangle inequality $\|x + y\| \leqslant \|x\| + \|y\|$ as #25.

I was not trained as a wrestler but I, at least, had a small education later because of my work in applied math. I did fall in love with the wonderful inequalities of the Russian analyst Sergei Sobolev. The simplest of these illustrates what many contemporary wrestlers deal with: say $f(x)$ is a smooth function on the real line. Then for all a,b, one has the simple Corollary of Cauchy's inequality:

$$|f(b) - f(a)|^2 \leqslant |b - a| \cdot \int(\tfrac{df}{dx})^2 dx.$$

Thus one says that a square integral bound on the derivative "*controls*" its point wise values. When I was teaching algebraic geometry at Harvard, we used to think of the NYU Courant Institute analysts as the macho guys on the scene, all wrestlers. I have heard that conversely they used the phrase 'French pastry' to describe the abstract approach that had leapt the Atlantic from Paris to Harvard.

Besides the Courant crowd, Shing-Tung Yau is the most amazing wrestler I have talked to. At one time, he showed me a quick derivation of inequalities I had sweated blood over and has told me that mastering this skill was one of the big steps in his graduate education. It's crucial to realize that outside pure math, inequalities are central in economics, computer science, statistics, game theory, and operations research. Perhaps the obsession with equalities is an aberration unique to pure math while most of the real world runs on inequalities.

Other examples of wrestler's work in the Atiyah-Zeki list are #11 (Cantor's inequality); #26 (the prime number theorem – (number primes $\leqslant n$) $\approx \frac{n}{\log(n)}$); and #38 (inequality of geometric and arithmetic means):

$$\frac{1}{n} \sum_{k=1}^{n} a_k \geqslant \left(\prod_{k=1}^{n} a_k \right)^{\frac{1}{n}}$$

Detectives:

Andrew Wiles said he worked on Fermat's claim that $x^n + y^n = z^n$ has no positive integer solutions if $n \geqslant 3$ obsessively for eight years, describing the work as follows (in a PBS interview `http://www.pbs.org/wgbh/nova/physics/andrew-wiles-fermat.html`):

> I used to come up to my study, and start trying to find patterns. I tried doing calculations which explain some little piece of mathematics. I tried to fit it in with some previous broad conceptual understanding of some part of mathematics that would clarify the particular problem I was thinking about. Sometimes that would involve going and looking it up in a book to see how it's done there. Sometimes it was a question of modifying things a bit, doing a little extra calculation. And sometimes I realized that nothing that had ever been done before was any use at all. Then I just had to find something completely new; it's a mystery where that comes from. I carried this problem around in my head basically the whole time. I would wake up with it first thing in the morning, I would be thinking about it all day, and I would be thinking about it when I went to sleep. Without distraction, I would have the same thing going round and round in my mind. The only way I could relax was when I was with my children. Young children simply aren't interested in Fermat. They just want to hear a story and they're not going to let you do anything else.

Although this is extreme, this sort of pursuit is well known to all mathematicians. The English mathematical physicist Roger Penrose once described his way of working similarly: "My own way of thinking is to ponder long and, I hope, deeply on problems and for a long time ... and I never really let them go." In many ways this is the public's standard idea of what a mathematician does: seek clues, pursue a trail, often hitting dead ends, all in pursuit of a proof of the big theorem. But I think it's more correct to say this is one way of doing math, one style. Many are leery of getting trapped in a quest that they may never fulfill. Peter Sarnak at the Princeton Institute for Advanced Study has described what it feels like to be a research mathematician by the sentence "The steady state of a mathematician is to be blocked." Arguably Landon Clay may have done maths no service by singling out seven of the deepest, most difficult math problems and putting a million dollar bounty on each. Putting a dollar value on a proof is quite bizarre and the prize was declined by Grigori Perelman, the only winner in this contest so far. In any case, I believe it is more common among mathematicians to become intimately familiar with a range of related problems while not necessarily actively working on any of them. But these problems are not far from their consciousness and from time to time, a clue will show up, a hint of some connection, and then it all rushes back and hopefully some progress is made on one of the problems.

Among those who attack major problems, a very small number are able to imagine a deeper more abstract layer of meaning in the problems of the day, that others never imagined. They are detectives who feel the answer is deeply hidden, so you need to strip away all the features of the situation that are accidental and thus irrelevant to understanding it. Underneath you find its true mechanisms, what makes it tick. It seems only logical to call such people strip miners, though not in a pejorative sense. The greatest contemporary practitioner of this philosophy in the 20th century was Alexander Grothendieck. Of all the mathematicians that I have met, he was the one whom I would unreservedly call a "genius." But there have been others before him.

I consider Eudoxus and his spiritual successor Archimedes to be strip miners. The level they reached was essentially that of a rigorous theory of real numbers with which they are able to calculate many specific integrals. Book V in Euclid's *Elements* and Archimedes *The Method of Mechanical Theorems* testify to how deeply they dug. Some centuries later and quite independently, Aryabhata in India reached a similar level, now finding what are essentially derivatives, fitting them into specific differential equations. But it is impossible to fully document the achievements of either of these mathematicians as only fragments of their work survive and there is no way to reconstruct much of the mathematical world in which they worked, the context for their discoveries. Grothendieck's ideas, however, and the world both before and after his work are very clearly documented. He considered that the real work in solving a mathematical problem was to find *le niveau juste* in which one finds the right statement of the problem at its proper level of generality. And indeed, his radical abstractions of schemes, functors, K-groups, etc. proved their worth by solving a raft of old problems and transforming the whole face of algebraic geometry. Mike Artin,

John Tate and I and many others have documented his greatest successes in the Notices of the AMS [**E-2014,iii**]. Pretty wonderful French pastry.

Many of the formulas in the Atiyah-Zeki list seem to me to come from the Baptismal subtribe. #10 defines e; #13 defines the δ-function; #21 defines π; #24 defines eigenvectors; #47 defines Möbius maps; #48 defines Clifford algebras. I have not mentioned many of the remaining equations in the Atiyah-Zeki list. It seems to me that many are intermediate results in a developing theory, found by detectives doing great work. It is hard for me to judge which are more beautiful: their attraction comes from their bringing to mind a whole beautiful theory of which they are one part. For instance, #36, $\langle B, B \rangle_t = t$, the variance of Brownian motion, is hugely important and beautiful but I would think of it as a natural consequence of the more basic fact that, when you add independent random variables x and y, their standard deviations follow the stochastic version of Pythagoras's rule.:

$$\text{St.Dev.}(x + y) = \sqrt{(\text{St.Dev.}(x))^2 + (\text{St.Dev.}(y))^2}$$

Brain areas for the different forms of beauty?:

It is clear that members of each tribe will make different judgements on the relative beauty of specific mathematical formulas or theorems. I want to take up each one in turn and ask what cortical activity they might produce. Explorers clearly find a tremendous thrill in the *Systema Naturae*, the flora and fauna and gazetteers produced by their explorer colleagues. Exotic creatures like non-standard differential structures on Euclidean 4-space continue to amaze and to defy visualization. But I suspect that geometers have mental tricks that allow them to piggy-back a sense for high dimensional constructions on top of their 3-dimensional skills. Thus constructions like surgery and suspension can be visualized in the simplest cases and the mind builds the skills that allow the general case to be grasped as an analog of these. I remember Zariski, getting stuck at a certain point in his lectures, drawing a bit of an algebraic plane curve (a cubic with a double point) in the corner of blackboard to kickstart his intuition. Steve Kosslyn and others have studied cortical activity with fMRI while a subject is forming a visual mental image of some object. One reference is `http://www.ncbi.nlm.nih.gov/pubmed/15183394`. There seems to be a complex pattern of widespread activity – frontal as well as parietal and temporal – as well suppression of activity in what I guess is pretty close to Zeki's mOFC (see the blue area in the top row in the figure on p.231 of the cited paper). But people who are not geometers may never use visualization in their research. There's a probably apocryphal story about the algebraist Irving Kaplansky: asked what he saw when you asked him to think about a *ring* replied "I see the letter 'R' ."

The most common "beautiful" formulas are alchemical. The famous:

$$e^{i\pi} = -1$$

brings together exponential growth with the geometry of the circle. When a formula

connects two concepts that would seem to have absolutely nothing to do with each other, you get a chill running down your back. It feels as though the universe wasn't forced to be this way so it is not unreasonable to ask God "why did you decide to make this happen?." In other words, it is hard to dispel a sense of mystery that clings to them. Is there an area of the brain which is active when you can't figure out why something happened, when you are mystified by some event? It would seem hard to devise fMRI experiments to find such a "mystery-center." But I believe that alchemists find the greatest beauty in such mysteries.

What is going on in the minds of wrestlers? My guess is that estimating size and relative power of math things is connected to our social behavior, to Darwinian selection of the fittest. Animal life is all about being strong enough to get the stuff you need. A large number of species exists in a hierarchical social setting, with each individual learning rapidly whom to defer to, whom to dominate. And Robin Dunbar has shown that the size of your working social group goes up exponentially with your brain size, thus humans must have large cortical areas devoted to deeply understanding the interactions of their large groups – he estimates that on average each person lives in group of some 150 people whom "you wouldn't feel embarrassed about joining for a drink if you happened to bump into them in a bar." Although I have not seen any experiments with this focus, I feel there must be cortical areas specialized for learning social structures and the complex web of pair relationships. (Perhaps anterior cingulate cortex and/or insula?) Given how central this is in our brains and lives, it feels to me that when structuring math objects, especially functions, by size (rate of growth, degree of smoothness, etc.), you would utilize this machinery built in for creating social hierarchy. I don't mean that you personify these math sizes, but only that making a partially ordered graph like structure is a skill you already have because of evolution.

Solving a puzzle is the basic drive for the detective tribe and the goal that gives them the greatest pleasure. In this case, there need not be a beautiful formula that encapsulates the solution. Rather, the proof itself is wonderful and beautiful. (Confession: I personally find quite stupid puzzles like Sudoku rather addictive.) This is surely a central aspect of pre-frontal lobe activity: planning your activities is finding a path in a world satisfying many constraints that leads to some desired goal. Math is, however, a bit different from the world: if you are trying to prove a theorem, you have to be prepared to reverse course and prove its negation. Never put all your money on one result. Perhaps, way out the imaginary axis, the Riemann zeta function does have a zero with real part not equal to 1/2.

Summarizing, I see visualizing an alien abstract world, finding new mysteries, creating vast hierarchies or solving the hardest puzzles as four aspects of what mathematicians find most beautiful. But each has its characteristic form of beauty that connects it to distinct parts of our mental life. Can we expect to nail each down to a specific part of the brain? Recall that most of the qualities localized in 19th century phrenology have long since not been dropped as labels for specific cortical areas. The perception of mathematical beauty may also turn out to be a higher order derivative phenomenon characterized by patterns

of activity widely distributed over the brain.

Part II

The History of Mathematics

I first got involved with the History of Math when I volunteered to teach a class for non-math majors in the Division of Applied Math at Brown. I knew that others used particular topics like voting systems or knot theory that wouldn't scare students by seeming too abstract and that such classes were derisively called "math for poets." I felt that I needed a better "hook" to get students interested, one that both connected to their lives and to the other topics they were studying. Explaining math by actually doing some serious calculations with *spreadsheets* and plotting results (things like the spectrum of a singing voice) was one hook. The other was teaching it through its history. Brown has a rich heritage in the History of Math and I discovered the work of Otto Neugebauer and the amazing math in Mesopotamia around 2000 BCE that he uncovered. So Mesopotamia was where my course started. Of course I went on to Newton but the big challenge I decided to talk about was Fourier series. I wanted to show the students how the singing voice and other musical instruments can be decomposed into a superposition of frequencies. Incidentally, I stumbled on the fact that the true inventor (discoverer?) of Fourier series was the mathematical astronomer Alexis Clairaut in 1754. I feel this is an important little known fact, so to convince the reader, I include the following slide from a lecture of mine where he gives both the expansion and the formula for the coefficients:
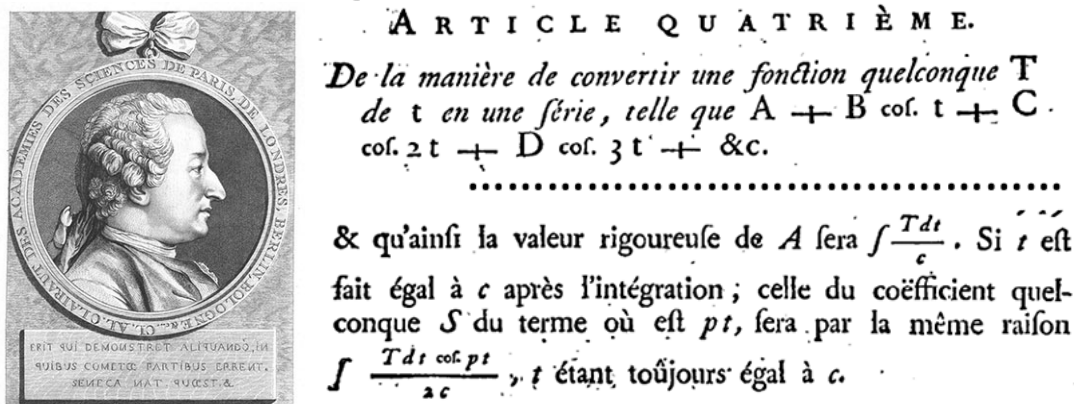
Figure 3.1: Left: a portrait of Alexis-Claude Clairaut, from Wikimedia Commons, public domain, and right a reproduction from his original 1754 memoire [Cla54], from Gallica, Bibliothéque Numèrique de France.

(Translation of the text reads: *Concerning the manner of converting any function $T(t)$ into a series such as $A + B\cos(t) + C\cos(2t) + D\cos(3t) + \cdots$* and then: *Thus the rigorous value of A will be $\int T dt/c$ if t equals c after integration* (meaning c is the period); *that of any coefficient S of the term with pt will, for the same reason, be $\int T dt \cos(pt)/2c$, t* (integrated between 0 and) c.

This part contains four chapters. The first concerns Pythagoras's rule that the sum of the squares of the lengths of the two short sides of a right triangle equals the square

of the length of the hypotenuse. This is the key fact that reduces geometry to algebra and I consider it to be "Theorem One" in the field of mathematics.[5] Amazingly, it occurs in Mesopotamia around 2000 BCE, in India around 800 BCE and in China (earliest date unknown but also likely before Pythagoras). This raises a huge question: how was this ever discovered and did it spread or was it discovered independently in these places?

The next chapter concerns the history of algebra, the long struggle to define and manipulate algebraic formulas and the bizarre problems that were concocted at each stage to illustrate the latest methods. Not unexpectedly, money was usually the focus of the problems but, equally often, the problems set seem to be just play. One might compare the history here with the struggles of contemporary students addressed in Chapter One. In both cases, I think the ease with which it is used by those who have mastered the idea of algebra, makes it very hard for them to see why it was (resp., is) so hard to discover (resp., learn) this technique.

The third chapter is extracted from a lecture I gave in 2013 at the IMA in Minneapolis. One of the chief things that fascinated me in studying history was the fact that the fact that, in historical situations, when one specific event repeats itself, too many related things have changed. So how can you know what factor caused the event, when it might have been one of many things. It seems to me that the History of Math is unique in that sometimes identical discoveries are made in different countries and you can get closer to seeing what causes what. I wanted to come to grips with the question of how much math has been unique to one or another country and how much resulted from ideas crossing from one culture to the another. In the first case, you can see how math is or isn't affected by differences in the culture of the respective countries. The bottom line is that the truth is mixed, each culture has its own personality and unique ideas but also a conqueror (like Alexander) or an inquisitive ruler (like Al-Mamun) or a wandering trader (like Fibonacci) can carry the spark of an idea quite far afield. The chapter is based on 5 *space-time* charts in which more and more mathematicians are added as time moves on.

The fourth chapter concerns the remarkable parallels between the History of Art and the History of Math from early in the 19th century to the present. Both have undergone a huge turn to abstraction, involving an analysis of the basics on which each is based. This led to both being called "modern" in the 20th century. Some instances of their parallelism are so synchronized that it is hard not to believe that this trend was driven by a world-wide zeitgeist, an intangible expression of the focus of the intellectual/artistic community. Both math and art are relatively free from purely national trends, hence express aspects of the international zeitgeist more clearly.

In all this, I need to confess that I am not a card-carrying Historian. I'm sure some of my ideas are off-base and may even look totally wrong to some professional historians. But this is what I came up with, diving into this immense field.

---

[5]Pursuing the metaphor, Theorem Two should surely be the formula for the volume of the sphere and Theorem Three Euler's result $e^{ix} = \cos(x) + i\sin(x)$. This is a fun thing to argue about at a mathematician's dinner party.

Chapter 4

Pythagoras's Rule

This chapter discusses the origin of the rule that, in a right triangle, the square of the length of the hypotenuse equals the sum of the squares of the lengths of the two shorter sides. The rule is not just an odd fact about triangles but rather it is key that *connects geometry and algebra.* More precisely, if you start from a pair of perpendicular lines in a plane, then distances in this plane can be calculated by the rule as shown in Figure 1. For exactly this reason, the rule was extremely useful in early city-states both for construction, city planning and especially calculating area of fields for taxation purposes. Its extension to three, n-dimensions and infinite dimensions have made it the square root of the sum of squares the key tool for measuring size in much of higher math.

$$d(p, q)^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$
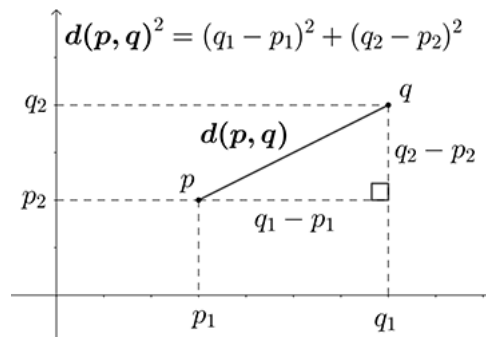
Figure 4.1: Pythagoras's Rule allows one to compute the distances using a pair of perpendicular lines. Wikimedia Commons courtesy of Kmhkmh

Although traditionally named for Pythagoras, the earliest extant documents that show knowledge of the rule are Babylonian tablets dating from the centuries around Hammurabi's time, c. 1800 BCE. I am calling it a rule, not a theorem, following Jens Høyrup's suggestion, because it appears as a rule for connecting these lengths, not a theorem, in most of its early history. In any case, we don't know if Pythagoras proved it or not. After the Babylonians

it next appears in extant records in Indian Vedic altar construction manuals, composed and transmitted orally as early as 800 BCE. Due to the wholesale destruction of documents in China in the Qin dynasty (221-206 BCE), the earliest records we have for the rule from China date from the second century BCE though it was likely known in China much earlier. This is a sparse set of sources indeed. But because this rule may be described in math-talk as the first "non-trivial" mathematical theorem to be discovered, there has been extensive debate about when and where it was first found, whether it was discovered independently in several places and how it was found. All this work belongs to what André Weil called "protohistory," an attempt to be scholarly when surviving documents are not only sparse but also possibly unrepresentative of a tradition, and totally absent from other cultures. The full history of Pythagoras's rule is a perfect example of a problem about which we mostly speculate. But that's what I want to do in this chapter. However, all is not speculation and, for great help in all that a real scholar of the History of Math might study, I want to thank Jens Høyrup for all his help.[1]

How should one view such speculation? My view of history in general, not just proto-history, is that it is always an exercise in *Bayesian inference.* We never have full knowledge of any past part of space-time. Even in our own lifetimes, we rely on faulty and selective memories in reconstructing events. Scholars have the illusion when they are relying only on primary sources that they are not making significant inferences, but I believe they are mistaken. Of course primary sources are much better than secondary ones, but everyone has built up their personal *prior* on human behavior and human culture and uses this to expand the meager sources that survive into a full blown reconstruction of some events. Indeed, Salman Rushdie quotes his Cambridge Professor Hibbert saying "You must never write history until you can hear the people speak." Of course this is also the fundamental reason why histories of the same event written at various times in later centuries typically differ so much.

My personal experience reading Archimedes for the first time illustrates my bias: after getting past his specific words and the idiosyncrasies of the mathematical culture he worked in, I felt an amazing certainty that I could follow his thought process. I knew how my mathematical contemporaries reasoned and his whole way of doing math fit hand-in-glove with my own experience. I was reconstructing a rich picture of Archimedes based on my prior. Here he was working out a Riemann sum for an integral[2], here he was making the irritating estimates needed to establish convergence. I am aware that historians would say I am not reading him for what he says but am distorting his words using my modern understanding of math. I cannot disprove this but I disagree. I take math to be a fixed set of problems and results, independent of culture just as metallurgy is a fixed set of facts that must be used to analyze ancient swords. When, in the same situation, I read in his manuscript things that people would write today (adjusting for notation), I feel justified

---

[1]This chapter is an edited version of a blog post on Jan.9, 2015.

[2]I listened to a major historian of ancient math, who had apparently never heard of Riemann sums of integrals, referring to this as an obscure technical digression.

in believing I can hear him "speak."

i.   Its discovery

Getting back to the Pythagorean rule, I think the first task is to ask why ancient peoples were led to study right triangles. I think there are two interconnected and quite convincing reasons. One is that the value of a field depends on its area and for buying and selling and inheriting and taxing farms, the numerical value of this area is indispensable. Another is that as towns grew and became cities, the most convenient shape for buildings and for the street plan was a rectangle. In the first case, the natural method is to break the field up into approximate rectangles or right triangles. A right triangle is half a rectangle and a rectangle can divided into two right triangles by its diagonal. So you need to be able to lay out perpendicular lines and recognize when one corner of a triangle is a right angle, when a quadrilateral is a rectangle. In other words, the rulers of all ancient kingdoms needed skilled land measurers and master builders who knew some basic facts from geometry. This does not mean they *required* the Pythagorean rule, but it suggests how useful it would be.

In Mesopotamia we are unbelievably lucky that records made in clay tablets, unlike records made on paper, papyrus, birch bark or string, are nearly permanent. Fire, for instance, makes clay *more* permanent instead of destroying it. We have a nearly three millennium record of clay tablets (and tokens) from Mesopotamia from which its cultural history can be reconstructed. Denise Schmandt-Besserat [SB92] has used this data to construct a very convincing story of the origin of writing in third millennium BCE Mesopotamia starting from clay tokens, then clay envelopes containing tokens and finally cuneiform on solid clay tablets. Essentially, her theory says it all started from needing to say "Mr. so-and-so owes me such-and-such." Their highly sophisticated place-value base 60 arithmetic seems to have originated from the need for a unified central accounting (perhaps in Ur III) including goods and labor which had been measured with many units often related by multiples such as 4,5,6,10,12 etc. Remarkable accounting tablets survive with detailed entries of labor and goods: see the book by Richard Mattessich on "The Beginnings of Accounting" [Mat00].

How about the measurement of land? The following wonderful paean to the Goddess Nisaba, who received literacy and numeracy as a wedding present from Enlil and passed it down to human beings, is found on one Babylonian tablet:

> Nisaba, woman sparkling with joy,
> Righteous woman, scribe, lady who knows everything:
> She leads your fingers on the clay,
> She makes them put beautiful wedges on the tablets,
> She makes them sparkle with a golden stylus,
> A 1-rod reed and a measuring rope of lapis lazuli,
> A yardstick, and a writing board which gives wisdom:

Nisaba generously bestowed them on you.

The "1-rod reed" and the "measuring rope" are the basic tools of the surveyor, here praised on a par with writing. Many "deed" tablets survive with plans of fields and measurements. A recent study by Daniel Mansfield of two such tablets [Man20], YOS 1,22 and Si.427, describes in detail how the area of two fields with a rather complicated shapes were calculated, subdividing it into approximately right triangles, especially how right triangles with "regular" sides (meaning their length and its inverse are finite sexagesimals) were used (referred to as "Pythagorean triples").

Pythagoras's rule is ostensibly a theorem about triangles – but really it describes distances in Cartesian coordinates in 2 dimensions as shown in Figure 1. Iterating it, one gets the distance in $\mathbb{R}^n$ as the square root of the sum of the squares of each coordinate difference:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The great importance of Pythagoras's rule is this Corollary.

And here from Uruk in Babylon, sometime in the 17th century BCE, we find *this rule used in 3-space.* This most impressive demonstration of their knowledge of Pythagoras's rule is on the tablet MS 3049 in the Schøyen collection. In this tablet, the authors calculate the diagonal distance in a gateway through a thick wall from e.g. the distance from the inner left bottom corner to the outer right top corner, going straight in/out, left/right and bottom/top all at the same time. Below is a rough translation of the calculation following Joran Friborg's book [Fri07], pp. 181-2. All Mesopotamia ran on base 60 but without a "decimal" point, indicating the division between whole numbers and fractions (this was always inferred from the context by the reader). In Mesopotamia, lengths were measured in "nindas" (or rods) about 21 feet, "cubits" each 1/12 of a ninda, hence about 1'9", and "fingers" which are 1/30 of a cubit (about 0.7"). Except at the top, the tablet uses the unit ninda throughout and uses sexagesimal fractions are written here as ;$xyz\cdots$ meaning $x/60 + y/3600 + z/216000 + \cdots$ ninda (the semi-colon is inserted for my readers and nothing like is on the tablet).

> If the inner cross-over of a gate he shall do
> 5 cubits and 10 fingers,
> the height of the gate
> ;8 53 20 (= decimal 4/27) ninda the width
>   (this comes from some missing tables)
> and ;6 40 (= decimal 1/9) the thickness of the wall, you see.
> ;26 40 (= decimal 4/9) the height of the gate, let eat itself
>   (This means square it), then
> ;11 51 06 40 you see.

;8 53 20, the width of the gate, let eat itself, then
;1 19 (missing number) 44 26 40 you see.
;6 40, the thickness of the wall, let eat itself, then
;0 44 26 40 you see.
Heap them (meaning add them), ;13 54 34 14 26 40 you see.
Its likeside (meaning square root) let come up, then :28 53 20
    (=decimal 13/27) you see
(for) the date that (has) ;26 40 (as its) height
So you do

They have added the squares of the gate's dimensions in all three dimensions and then taken its square root! The attentive reader will notice that the Babylonians contrived this so that the base of the thick gate with its diagonal is similar to a (3,4,5) triangle and the vertical side together with the diagonal on the base forms a (5,12,13) triangle – the two simplest rational right triangles. Besides Pythagoras, the tablet shows a remarkable skill in base 60 arithmetic.

An aside: another tablet, Plimpton 322, is often used as evidence of the Mesopotamians' knowledge of the Pythagorean rule. This contains a list of pairs (s, d) where $d^2 - s^2$ is a square of a regular sexagesimal number ℓ – namely Pythagorean triples (s, ℓ, d). As the tablet lists these for triangles with angles steadily decreasing from about 44 degrees to 32 degrees, it has been thought to be an *equivalent* of a table of sines (without any angle measurements) or perhaps a manual for earthworks giving simple distances that could be laid out by surveyors. However, Eleanor Robson has proposed instead [Rob02] that it was simply a table of reciprocal pairs $(x, 1/x)$ (now missing because the tablet broke) together with their sums and differences reduced to sexagesimally simple forms to simplify the work of setting problems, i.e. a teacher's manual. Nonetheless, the heading on Plimpton 322 contains a particular word for "diagonal" that refers to the diagonal of a rectangle. Daniel Mansfield [Man21] has therefore made the alternative proposal that the tablet could be a manual for surveyors who used these regular rectangles and the right triangles obtained by dividing them in half to subdivide fields into pieces of known area. For my money though, I like MS 3049 the most as it *explicitly* uses the Pythagorean rule twice, making their knowledge of the rule and *its relevance to measuring Euclidean distances* indisputable.

Who were the people who came up with this – arguably the first "non-trivial" fact in mathematics? We know that there were scribal schools in Mesopotamia where apprentices were trained in the three 'R''s, reading, (w)riting and (a)rithmetic, all highly skilled professions at the time. (Aside: besides the base 60 arithmetic being quite a challenge, the script, like contemporary Japanese, was a mixture, in this case of Sumerian logograms and the Akkadian syllabic alphabet, hence another major challenge.) Bins of hundreds of discarded student tablets, many with errors, survive! Students in these schools became scribes working as bureaucrats, accountants, surveyors or teachers. But I contend that some scribes must have been mathematical geniuses too or the Pythagorean rule could not
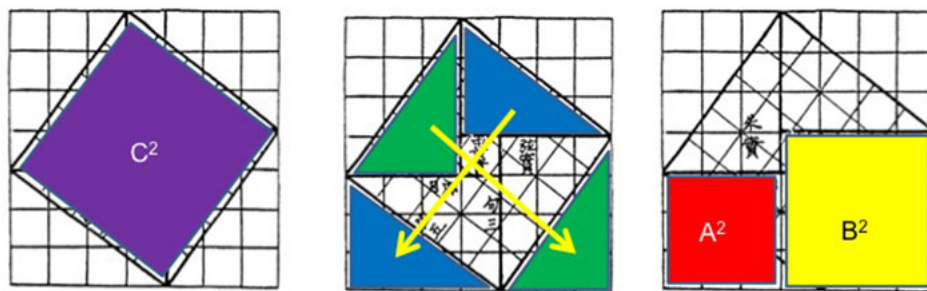
Figure 4.2: How to cut, shift and reassemble the squares on the three sides of a right triangle. This is the simplest proof of Pythagoras's rule that I know.

have been discovered. Should we think of them as the world's first mathematicians? There is some controversy here. For Eleanor Robson, all this work was oriented to engineering, administrative and instructional needs – measuring and designing canals, earthworks, etc. and she asserts that thinking of them as mathematicians is a misguided anachronism that ignores the society in which they lived.

Perhaps this is just a reflection of the age-old tension between pure and applied mathematics. Many engineers have been mathematical geniuses. You don't have to be a *professional* mathematician to be a mathematical genius and it does seem a stretch to call anyone from that time a mathematician. Following Hibbert's dictum, let's imagine a brilliant civil servant whose day job was measuring fields or construction sites and writing tablets with associated plans but whose imagination was caught by these geometric diagrams and who then played with how these diagrams constrained lengths and areas (one might think of Einstein in the Swiss patent office).

But how was the rule found, what led them to this strange looking rule? This is the real mystery. Jens Høyrup in his book "Length, Width, Surfaces: A Portrait of Old Babylonian Algebra and its kin" proposes, in connection his analysis of tablet Db_2146, that the Babylonians discovered a version of the famous Xian Tu diagram that appears in Chinese manuscripts of the Early Han dynasty (see Figure 5 below). The key to this diagram is to inscribe one square inside another at the angle that makes the gaps in the four corners all equal to the given triangle. Unfortunately, no trace of such a diagram has been found on a tablet. However, the case where the inner square is oriented at 45° is found on tablet BM 15285 shown in Figure 4 left. And once you conceive of this diagram, there are many ways to prove the rule. Høyrup, analyzing very carefully the exact words on tablet Db_2146, proposes one in his book, p.259, figure 67. Figure 2 shows my favorite derivation of the Pythagorean rule using the Chinese diagram with A, B, C denoting the sides of the white triangles in the four corners.

To my mind, it seems more likely that the rule was discovered from working with *similar* triangles. There are a number of tablets showing a set of similar triangles formed
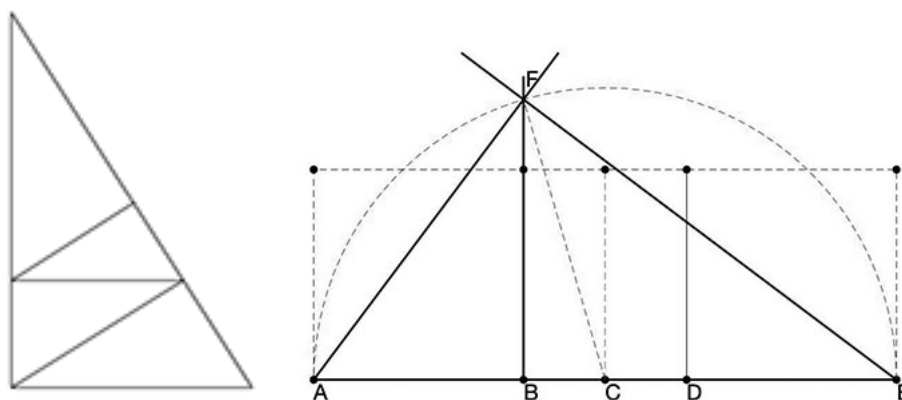
Figure 4.3: Left: The diagram appearing on IM 55357, right: the diagram leading to the Pythagorean rule, and, with faint lines, the well-known construction of a square with area equal to that of a rectangle, see text.

by intersecting a wedge with various parallel lines. A good example is IM 55357 working with the lengths and areas of various parts of the diagram in Figure 3 left. In the right side of that figure, I show how readily the Pythagorean rule can be deduced from a pair of similar triangles. This diagram also has similarities with the diagram on tablet TMS 1. In my figure, the similar triangles are (i) AEF and FEB gotten by flipping and shrinking the first around the vertex E and sharing the angle $\angle$AEF; and (ii) EAF and FAB gotten by flipping and shrinking around vertex A and sharing the angle $\angle$FAE. The similarity tells us that (i) AE/FE = FE/BE; and (ii) AE/AF = AF/AB. Therefore,

$$FE^2 + AF^2 = AE.BE + AE.AB = AE^2.$$

I have drawn the dashed line FC to show how we are dealing with the well-known diagram used to construct a square with the same area as a rectangle: define D as the point on the line AE such that $\overline{DE} = \overline{AB}$ assume we want to square the rectangle with dashed lines over $\overline{BE}$. The standard construction begins by halving BD and constructing F as the intersection of the circle with center C, radius $\overline{AC}$ and the extension of the vertical line through B. The desired square has side $\overline{BF}$. Note that once you know FEB is similar to AEF and EAF is similar to FAB, you also know FEB and EAF are similar, hence AB/BF = BF/BE, so BF does square the rectangle AB × BE.

Given the familiarity of this construction as well as the study of similar triangles, it feels as if this could be a plausible route to the discovery of Pythagoras's rule. Though this is a considerable speculative leap, MS 3049 makes it unmistakable that somehow they found the rule, so I think we have to entertain such a speculation.

ii. How did it spread and was it rediscovered?

But then did other cultures discover the result independently? Not necessarily: if we accept that Pythagoras's rule and the accompanying geometry were very useful for taxes and building, it is only natural that its knowledge would spread to nearby civilizations with which Mesopotamia had regular trade. Master builders and surveyors would be in demand and some would likely migrate. Thus both the Egyptian and the Indus Valley cultures flourished at overlapping times and so might learn of the latest technology from Babylon. Sadly, in both cases, we have much sparser remains from which to deduce what they knew. From Egypt, the so-called "Scorpion Macehead" shows the pharaoh seeding the fields adjacent to the Nile after its flood and is dated c.3000 BCE. To reconstruct the fields, "rope stretchers" were employed and paintings testify that knotted ropes were their principal tools. It is widely believed that they used the 3-4-5 triangle to lay out right angles for construction purposes. But the only evidence for this is problem 1 in the Berlin Papyrus 6619 where the equation $x^2 + y^2 = 100$, $y/x = 3/4$ is solved. According to a recent review [Imh09], judging from the mathematical papyri that have survived, it is doubtful that the Egyptians knew the statement of the Pythagorean rule in general. Moreover, structures such as the great pyramid of Giza were built about 800 years before the above tablets were written. My guess is that, in the Old Kingdom, squares were laid out by using ropes to ensure that all sides were equal and both diagonals were equal. It's also plausible that the technique of laying out right triangles by a rope with knots at spaces 3, 4 and 5 could have been transmitted from Babylon during the Middle Kingdom while its theoretical background was not.

As for the Indus Valley culture, we have about 3700 inscriptions containing about 400 symbols but this is no help as they are still untranslated. But there are Sumerian descriptions of trade to a place in the East called "Meluhha," often identified with the Indus Valley, and identical clay seals are found in the Indus Valley and in Mesopotamia. Their cities were laid out with very regular rectangular street plans indicating their need for skilled surveying (as does the universal concern with fields). What makes the possibility of transmission of the full Pythagorean rule to the Indus valley a bit more plausible, however, is how the rule crops up very explicitly in the Indian Vedic period, in the Sulba Sutra of Baudhayana, usually dated c. 800 BCE. Here the rule is used not for laying out fields, streets or buildings but for laying out sacrificial fire altars. The Vedic invaders of Northwest India are thought to have occupied the Indus Valley during the late periods of the Indus Valley culture and then to have spread East. How they interacted or interbred with the natives in this land and what, if anything, they picked up from them are the subjects of great controversy. A strong case for significant interaction is laid out in Wendy Doniger's book [Don09] and in the article of Hyla Stuntz Converse [Con74].

Regardless of where you stand on these sensitive issues, it is startling to find in Vedic Sutras not only the Pythagorean rule but the basic geometric constructions with ropes used in Mesopotamia and Egypt (and likely the Indus Valley): see Figure 4 middle. If you
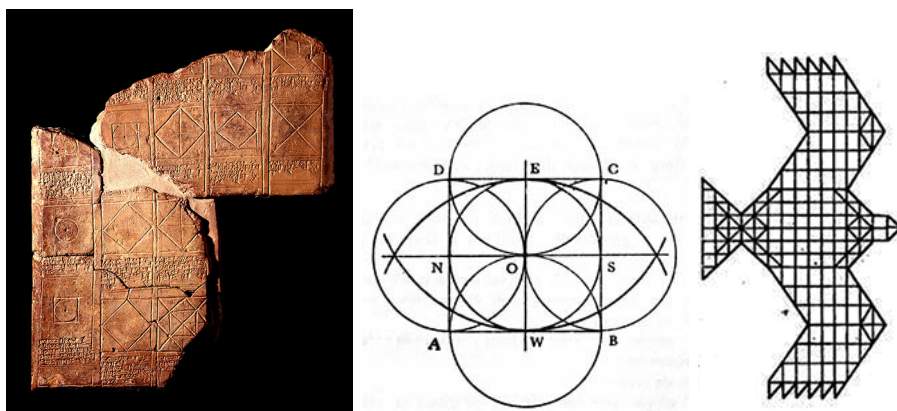
Figure 4.4: On the left, a photograph of the Babylonian tablet BM 15285, replete with many elementary geometric diagrams, by permission of the British Museum. Note the square within a square, rotated 45°, a possible precursor to the Xian Tu construction. In the middle, a drawing of the circles laid out via ropes and aligned to NSEW prior to building a Vedic sacrificial fire altar, as per the prescriptions in the Baudhayana Sulbasutra and identical to later Euclidean constructions, from [Amm99], p.30, by permission from Ravi Jain, *Motilal Banarsidass*. On the right, the bottom layer of brick tiles for the falcon altar of that type, as described in [SB83], by permission of the Indian National Science Academy. The sulbasutras describe many startling shapes for their altars, always made by multiple layers of clay bricks of standard rectangular size (or halved).

put the Sulba Sutras next to a book on the geometry in the Mesopotamian tablets, the similarities are stunning. You might wonder why *area* was important to the Vedic peoples? There is a simple ritual reason: if a sacrifice did not achieve its aim, it was repeated after doubling, tripling etc. the area of the altar until is worked its magic. If you use Pythagoras's rule, this is easy to do with ropes. We also find, a bit later, very sophisticated accounting used in the Maurya empire. All in all, it seems a reasonable speculation that a good deal of math was transmitted from Mesopotamia, via the Indus Valley people, to the Vedic peoples.

How about China? A key problem with the history of Chinese math is that mathematics and mathematicians never held an important place in Chinese culture. Math was a tool for low level bureaucrats and, in many dynasties, was not even part of the imperial exams. Astronomy and its sister, Astrology, held a somewhat higher place. But these were not esteemed nearly as much as writing poetry and essays on Confucian ideals. After the massive burning of ancient documents and the burying alive of recalcitrant mandarins in the Qin dynasty, the Han dynasty scholars were able to reconstruct much of the ancient dynastic histories and Confucian manuscripts but only the final state of the math, not its history. Nonetheless, in what they reconstructed the Pythagorean rule emerges full blown.
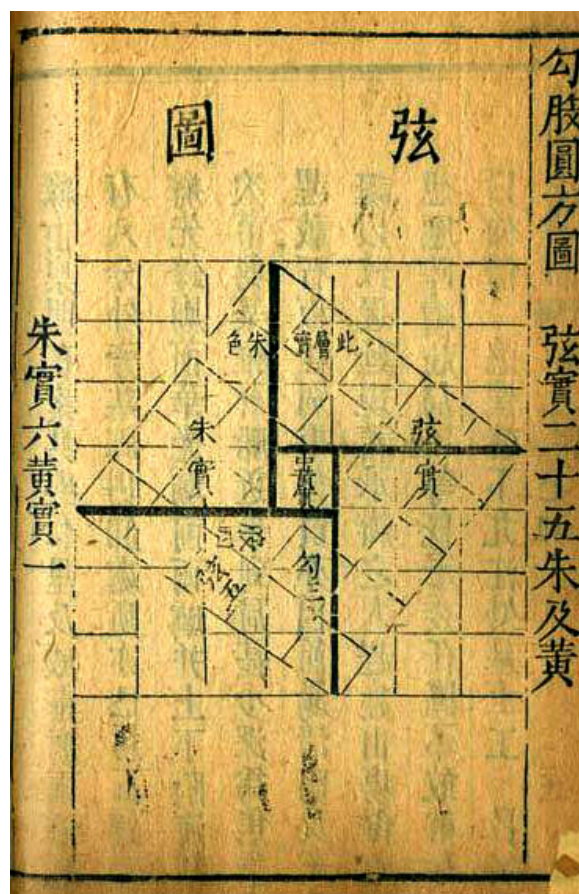
Figure 4.5: The famous Xian Tu diagram from which the Chinese deduced Pythagoras's theorem from a 1603 manuscript of the Zhou Bi Suan Jing. Photocopy of illustration from Swetz & Katz's Math Association of America collection *Mathematical Treasures*

It occupies a full chapter in the main Han dynasty treatise, the "Nine Chapters on the Mathematical Art" (Jiu Zhang Suan Shu) and the proof using the famous diagram Xian Tu (figure 5) appears in somewhat garbled form in the surviving late Zhou manuscript "Zhou Bi Suan Jing" (sometimes translated as the "Arithmetical Classic of the Gnomon").

Was this rule, as well as the use of Gaussian elimination and negative numbers to solve systems of linear equations, all discovered in the burst of creative activity in the Han dynasty? Chinese culture had expanded and built sophisticated societies with elaborate governments, earthworks etc. for over a thousand years preceding the Qin. Confucius had lived three centuries earlier as had scientifically inclined philosophers like Mo Tzu. Although there is no direct evidence, it seems much more likely that Pythagoras's rule had been discovered sometime in the Zhou dynasty (1046-256 BCE, often subdivided into the

Zhou proper, then the Spring and Autumn period and finally the Warring States period). It also seems unlikely that its statement might have been transmitted from the Middle East in these early times. The culture of the Middle Kingdom has its own very distinct writing and founding myths. It seems most likely to me that another unsung mathematical genius discovered it in China in the early first millennium BCE.

Enough speculation. My central point is first that early math was applied math, embedded in practical tasks, especially accounting and surveying. Secondly, the algorithms in these fields can be transmitted to other cultures by their practitioners – bureaucrats, scribes and master builders – just as well as by the experts who first formulated them. But thirdly, for a few of these experts, the math they uncovered took on a life of its own, they pushed things to a deeper level and their discoveries, such as the Pythagorean rule, should be celebrated as much as the discovery of metals and of wheels. I think it is not anachronistic to call those experts mathematicians and I suspect they felt not unlike how my colleagues feel today when they find something new.

Chapter 5

The Checkered History of Algebra

The history of algebra is completely different from the history of geometry or the history of analysis. Geometry arose from measuring areas and laying out constructions, like buildings and streets. Analysis arose from the modeling of machines like pulleys and clocks and from the beginnings of calculus. But algebra lagged behind, engaged only with solving arithmetic problems, both prosaic and elaborate ones, until it came into its own in the 20th century with groups, rings and fields. During much of this history, people struggled to find good *notation*, adequate symbolism for unknown numbers and especially for expressing the relationships of unknown numbers linked in some context. This is one thread that connects algebra in multiple times and places and that I will try to sketch. Inventing the needed notation is an example of *reification*, making a manipulable tangible thing out of something you previously knew only indirectly as an abstraction. This is essential for several reasons. Firstly, it allows you convert prose phrases into formulas. Secondly, having symbols for unknowns allows you to formulate the rules for manipulating and simplifying formulas. Thirdly, it allows you to substitute entire expressions for the unknowns. We shall see how each of these benefits first appears and transforms the solution of even simple arithmetic problems.

The other theme I want to discuss is the curious fact that, with few exceptions, every advance in algebra was illustrated by *meaningless* problems, frequently challenges in the form of "word problems" having no importance in the real world. And I find it odd that no book on the History of Math points out how many algebra problems in every era are crazy concoctions whose main point is to show how smart their creator was and perhaps to torture the student. It's a fascinating, not well-known side of math history.[1]

---

[1]This Chapter is based on my blog post "Ridiculous Math Problems," April fool's day, 2020, and on a lecture, "The Invention of Algebra as Reification," delivered in Calicut, Kerala, India, Sept.1, 2010.

i.   Babylon

Curiously, the creation of meaningless math problems goes back to the earliest known mathematical documents. A truly ridiculous question was posed four thousand years ago on a Babylonian tablet inscribed with cuneiform and concerns solving for numbers involved with building a wall. You are given the *sum* of the number of laborers, number of days needed and the number of loads of bricks used and must work out the number in each! Of course, such a sum has no significance whatsoever and no overseer would ever need to solve any problem like this. Never mind: the problem was probably devised to test the poor student's knowledge of the quadratic formula. Or could it have been a brain-teaser for scribes in their leisure time? Here's what the actual cuneiform says:

> *I added the bricks, the laborers and the days so that it was 140. The days were 2/3rd's of my workers.* (Note: It was also assumed known that a worker can carry $3/20^{\text{th}}$ of a load each day). *Find (the number of) bricks, laborers and days for me.*

If you figure out that there were 30 laborers, working 20 days and carrying 90 loads of bricks, "you are get a gold star" as we did in K-5. You'll need to solve a quadratic equation, something the Babylonians did by completing the square.

For me, as an applied math guy, disregarding the units of measurement when carrying out arithmetic operations is one of the cardinal sins. Days are units of time, loads are units of weight and only workers is a number without a scale. So simply posing such a problem shows they are playing with algebra, not doing anything remotely useful. Secondly, note that there are no symbols here, everything is stated as a pure "word problem." Thirdly, in the full tablet, the solution was described not by writing the requirements as formulas but simply by giving the steps of the algorithm that solves it, as by computer code: add this to this, multiply by this, take the square root of this etc., etc. The scribe memorized the steps, perhaps understanding the logic, perhaps not. But having to write out the steps in this way is ultimately a consequence of the fact that they had no notation for formulas or variables.

This Babylonian problem even sounds like a lot of the so-called "word problems" posed in high school algebra today. It reminds me of the chestnut: "If Jim can dig this ditch in 2 days and Bob can dig it in 3 days, how long would it take them if they dig together?" Actually, I think that problem is a pretty good one to master and problems like it might actually be useful. It requires the student to realize that Jim digs 1/2 the ditch in one day, Bob 1/3 of the ditch, because the number of days and the fraction dug in one day are inverses of each other.

By the way, some word problems in textbooks are also really ridiculous. Here's a problem coming from Richard Feynman's autobiography [Fey85] describing his work reviewing textbooks for the California Board of Education:

> *Finally I come to a book that says, "Mathematics is used in science in many ways. We will give you an example from astronomy, which is the science of stars." I turn the page, and it says, "Red stars have a temperature of four thousand degrees, yellow stars have a temperature of five thousand degrees . . ." – so far, so good. It continues: "Green stars have a temperature of seven thousand degrees, blue stars have a temperature of ten thousand degrees, and violet stars have a temperature of . . . (some big number)." There are no green or violet stars, but the figures for the others are roughly correct. It's vaguely right – but already, trouble! ….*
>
> *Anyway, I'm happy with this book, because it's the first example of applying arithmetic to science. I'm a bit unhappy when I read about the stars' temperatures, but I'm not very unhappy because it's more or less right – it's just an example of error. Then comes the list of problems. It says, "John and his father go out to look at the stars. John sees two blue stars and a red star. His father sees a green star, a violet star, and two yellow stars. What is the total temperature of the stars seen by John and his father?" – and I would explode in horror. My wife would talk about the volcano downstairs.*

It's always makes me laugh – adding temperatures of some set of objects is such a nutty meaningless idea. Imagine if, in the course of the Covid pandemic, a hospital were to post the total temperature of all its patients!

ii.   Greece

Ancient Greek math was not known for its algebra with the exception of the work of Diophantus. What do his problems look like? Here's a typical one:

> *IV.39: To find three numbers such that the difference of the greatest and the middle has to the difference of the middle and the least a given ratio, and further such that the sum of any two is a square.*

What is always implicit in his book is that he wants all his numbers to be positive rational fractions. In this specific case, he goes on to specialize the problem to ask for the given ratio to be 3 and comes up with expressions for the three numbers depending on a fourth rational number that you can choose as any fraction between 0 and 2. He then gives this representative answer: 29/242, 939/242 and 3669/242! Really? Is this significant? The most exciting thing is that he now had *formulas*, shown in Figure 1 for IV.39 both as he wrote it, in a transliteration and in modern form.

His variable, that he calls the *arithmos*, is hugely useful but his biggest problem is that he has only *one* symbol for an unknown. In his solution, he makes a substitution, an *ansatz*, taking the square in the formula equal to $(3 - u.x)$ where u is a new variable. But

$$\Delta^{\Upsilon} \bar{\gamma} \, \mathsf{s} \, \overline{\iota\beta} \, \mathring{\mathsf{M}} \, \bar{\theta} \; \ddot{\iota}\sigma. \; \Box^{\varphi}$$
$$x^2 3x12[\text{cnst}]9 = [\text{square}]$$
$$3x^2 + 12x + 9 = y^2$$

Figure 5.1: An equation Diophantus is led to in solving IV.39. ς is the unknown x (short for *arithmos*,) Δ^{Υ} is its square (short for *dynamis*, $\mathring{\mathsf{M}}$ is a constant, $\iota\sigma$ means equals (*isos*) and the alphabetic characters with a bar over them are the consecutive numerals in decimal notation (so $\overline{\iota\beta}$ is 12). You can see his conventions from the middle line above and the same formula as we write it on the last line.

he has no symbol handy for u! Therefore, he has to resort to awkward circumlocutions. Here's literally what he says in solving the above problem:

> So I am led to make the 3 dynameis (x^2) (and) 12 arithmous (x) (and) 9
> units equal to a square (number). I form the square from 3 units wanting **some
> (number of)** arithmous (x); and the arithmos (x) comes from **some number**
> taken six times and augmented by 12 (units), that is, the (quantity) of the 12
> units of the equalization, and divided by the excess of the square formed from **the
> number** on the (quantity) 3 of the dynameis (x^2) in the equalization. Therefore
> I am led to find **a number** which when taken six times and augmented by 12
> units, and divided by the excess that the square on it exceeds the 3 units, makes
> the quotient (parabolê) less than 2 units. (many thanks to Jean Christianidis
> for this translation. "**some number**" is the new variable u)

After this the bold "number" now becomes a new arithmos. Using our notation, his manipulations are easy to check. Note also that "equalization" means he is rearranging the equation the same way we do it. This is easy for him using nice formulas.

He had a few tricks for coming up with such bizarre solutions and he spun this out to hundreds of such problems. He would appear to be randomly flailing in a sea of irrelevant games, though André Weil does claim to see an underlying logic. Weil, in his retirement, studied history of math extensively and, in particular, analyzed Diophantus using contemporary math, algebraic geometry and number theory, to reveal structure behind his choice of problems. So the fact that the study of integer and rational solutions of polynomial equations is known today as "Diophantine Analysis" may not be unreasonable. On the other hand, his specific problems must have appeared pretty meaningless to his contemporaries.

iii.   China

Let's skip across the world now and look at what the Chinese were doing, as appears in their major Han dynasty treatise, the "Nine Chapters on the Mathematical Art" (Jiu Zhang

Suan Shu), a compilation that was assembled around 100 BCE from earlier manuscripts. There is a whole lot of algebra here *but not a single formula*! For example, there is a solution of the problem:

> *Now given 3 bundles top grade paddy, 2 bundles medium grade, 1 bundle low grade. Yield: 39 dou of grain. 2 bundles top, 3 bundles medium, 1 bundle low. Yield 34 dou. 1 bundle top, 2 bundles medium, 3 bundles low. Yield 26 dou. Tell: how much paddy does one bundle of each grade yield?*

This is clearly a set of 3 linear equations in 3 unknowns.

$$
\begin{array}{ccccccc}
3T & + & 2M & + & L & = & 39 \\
2T & + & 3M & + & L & = & 34 \\
T & + & 2M & + & 3L & = & 26
\end{array}
$$

How could they ever solve this without some notation? By analog computation! They laid out a 4×3 grid of squares on a flat surface, made a whole lot of short red and black sticks (known as counting rods), and they made the whole 4×3 matrix of integers by placing sticks in each square. Red was for positive numbers (because red is auspicious), black for negative and numbers were made with 0,1,2,3,4 rods and 5 added as a roof if needed. Place value was given by alternating horizontal and vertical orientations. Once this is done, they then implemented Gaussian elimination *exactly the way we still do it* (if forced to do this by hand!). I think this was an amazing tour-de-force. But note the paddy problem is not ridiculous. It is a practical, useful problem that might be encountered by the supervisor of paddy market, seeking to assess the prices for each quality of the product. All this is so unlike the Western tradition: no formulas but useful applications. Notice also how the Chinese procedure is identical to what happens in a computer. Here also there is no need for symbols for the variables: the *location* of each number (or bit) gives it a name and labelling the coefficients with symbols T, M, L is what programmers call "syntactic sugar", useful for humans with poor memory but wholly unnecessary for a machine.

This approach got even more remarkable when, in the Song dynasty, Zhu Shijie (c.1300 CE) carried out polynomial arithmetic in several variables with counting rods. Now the coefficient of $x^n y^m$ is placed in the $(n, m)^{\text{th}}$ grid square. An example is shown in Figure 2. Zhu went on to create elimination theory, computing a polynomial $f(x)$ in the ideal generated by two polynomials in two variables, $g(x, y), h(x, y)$. This anticipated Bezout's work by about 500 years. This remarkable work was not followed up in China but it spread to Korea and Japan. This and much more were developed especially by the Samurai mathematicians Seki (his family name) Takakazu (1642-1708), who also introduced determinants, and his pupil Takebe (family name) Katahiro (1664-1739), see [OM19] for many papers on this work.

Curiously, this approach to math reflects the low estimation of math in Chinese culture. Mandarins were expected to know the classics and write poetry but not to do math. Math

y^j	$\cdots$	2	-8	28	
$\cdot$			0	-1	6	-2
$\cdot$			0	0	0	-1
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$x^i y^j$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	x^i

$$2y^3 - xy^2 - 8y^2 + 6xy - x^2 + 28y - 2x$$
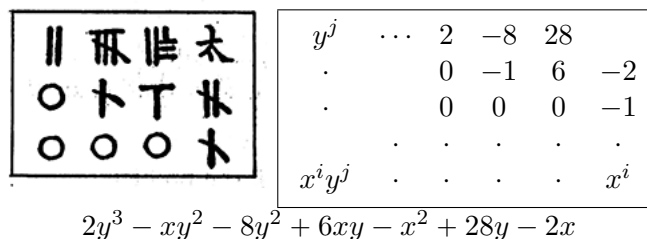
Figure 5.2: Zhu Shijie's analog representation of a polynomial in two variables, from the *Siyuan yujian*, 1303 CE. The original figure is on the left using rod counting sticks, slash for negative, reversing orientation for place value and for the number 5. Modern equivalent on the right and below. Coefficients are read from top down for powers of x and right to left for coefficients of y.

was done by lower level technicians. The main exception to this was a consequence of the need to predict eclipses. These predictions were essential in demonstrating that the emperor enjoyed the "mandate of heaven." This made the bureau of astronomy very important and, of course, it required mathematical skills. Astronomy also underpinned map making since they found latitude from the height of the north star (or the height of the sun at solstices). I have written 2 papers on this [**E-2012b, E-2016**].

iv.   India

Symbolic notation in India goes back to the famous Sanskrit grammar, the *Astadhyayi* of Panini, c. 500 BCE. This paved the way for introducing variables in math. Every important document in India in those days was transmitted by memory and, either for this reason or just because it was his preferred style, Panini wrote extremely compactly and cryptically, using abbreviations and lists to internally reference one verse to another. The whole work is a tightly woven nest of cross references. A simple example is sutra 1.4.14:

suptiṅantam padam

What does this mean? Firstly, the suffixes of nouns have been put in a long list starting in *su* and ending in *p*. Thus the prefix *sup* in this sutra refers to all nouns. Secondly, the suffixes of verbs have also been listed from *ti* to *ṅ*. This *tiṅ* refers to all verbs. Since *padam* means "word," the sutra simply says that a word is what ends in something in the *sup* list or in the *tiṅ* list, i.e. is a noun or a verb.

Skipping over Pingala who studied binary notation and Pascal's triangle – all very much in the above cryptic, highly compressed, symbolic fashion – we come to the famous Bakhshali manuscript where we find full fledged formulas quite similar to those of Diophantus. This manuscript is a long rolled up piece of birchbark unearthed (and badly

damaged) at the time of its modern discovery underneath a farmer's plow. It's impossible to date exactly but is likely from the early to mid-first millennium CE. An excerpt with a formula is shown in Figure 3. The double brackets distinguish the self-contained formula, transcribed literally underneath and, on the right, as a pair of modern style formulas. The unknown is now indicated by a small black filled circle, known as śūnya sthāna, the empty place. Note that solving this pair of equations is an exercise of no particular significance, yet another meaningless problem. A little thought shows that x must equal 11.

$$x + 5 = \square$$
$$x - 7 = \square$$
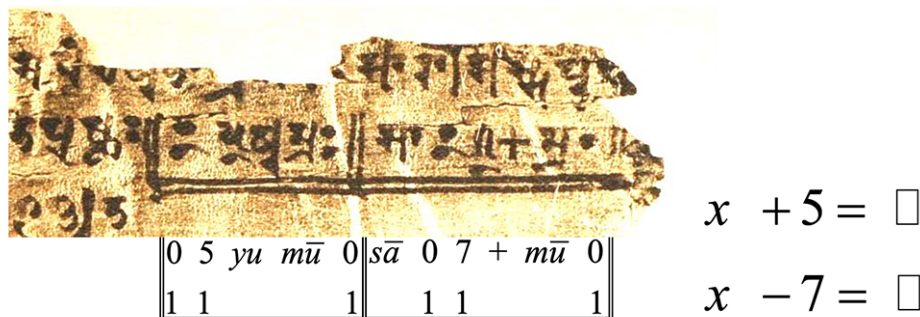
Figure 5.3: On top, a scan of a snippet of the Bakhshali manuscript. Below, a transcription of the bracketed formula. The symbols with a twiddle (transcribed as 1) underneath are numbers and variables, the other letters are operations. The filled dots (transcribed as 0) are variables. The notation is postfix; yu is a contraction of $yuta$, " joined together" and means add; $+$, oddly, means subtract the first on the left from the second; $m\bar{u}$ is a contraction of $m\bar{u}la$, root, and indicates that an integer on the right is the square root of what is written on the left. I assume $s\bar{a}$ means continue with the same variable. On the right, there is a modern version where the squares mean some square of a whole number.

A large part of the Bakhshali manuscript deals with summing arithmetic progressions, e.g. $\sum_{k=0}^{k=t}(a + bk)$ for specific numbers a, b. The sum is a quadratic function of t so, most curiously, they interpolated this sum for t not a whole number! For example, one problem asks you to solve for t in the formula:

$$\sum_{k=0}^{k=t-1} (5 + 6k) = \sum_{k=0}^{k=t-1} (10 + 3k)$$

and the author finds $t = 4\frac{1}{3}$.

But it was Brahmagupta (c.598–c.668 CE) who is the true father of algebra in India, and who invented a full fledged system for writing equations. He invented what seems to be the first complete system of algebraic notation, using multiple colors for extra variables. Below, I give a table of his notations that is subsequently illustrated by an excerpt from Bhāskara II.

yā→yāvat-tāvat="as many as so many"; ⎫
cā→ cālaca=black;   ⎬ variables
nī →nīlaca=blue;   ⎭
rū →rūpa=number; (for a constant)
dot over symbol → negative;
new line → equals;
v →square;
c →square root;

Brahmagupta's great achievement in algebra was his discovery of the algebra of real quadratic number fields. Otherwise put, this is the study of the integer solutions of equations $Nx^2 + C = y^2$ or $(y - \sqrt{N}x)(y + \sqrt{N}x) = C$. Here is an instance of a genius playing with algebra and finding major ideas that are rediscovered in 19th century Europe. His excitement led him, at one point, to declare "The person who can solve this problem within a year is a mathematician".

Moreover, Brahmagupta's writings apparently made their way to the caliphate in Baghdad where they likely inspired the Persian Muhammed ibn Musa al-Khwarizmi, c.780-c.850 CE. Though often called the father of algebra, his book, after explaining basic arithmetic and the solution of quadratic equations, consists almost entirely in working out legacies according to islamic law and involving slaves and dowries but with little historical significance.

Algebra reached its high point in medieval India with Bhāskarāchārya (or Bhāskara II, 1114-1185 CE). Like many others, he could not resist the temptation to show how powerful were his ideas with a meaningless problem:

> *If thou be conversant with operations of algebra, tell the number of which the biquadrate* (4th power) *less double the sum of the square and 400 times the simple number is a myriad* (10,000) *less one."* (Vija-Ganita, V.138)

$$\text{ya v v 1 ya v } \overset{\bullet}{2} \text{ ya } \overset{\bullet}{400} \text{ ru 0}$$

$$\text{ya v v 0 ya v 0 ya 0 ru 9999}$$

$$x^4 - 2x^2 - 400x = 9999$$

Well, this is a bizarre 4th degree polynomial equation. He suggests the natural idea is to add $400x + 1$ to make the LHS a square, but this is a dead end! "Hence ingenuity is called for" he says. Instead add $4x^2 + 400x + 1$, getting $(x^2 + 1)^2 = (2x + 100)^2$, hence $x^2 + 1 = \pm(2x + 100)$, hence $x = 11$. He ignores the possible minus sign. Honestly, I would not have had a clue how to solve it.

v.   Early Modern Europe

Algebra began in early modern Europe with Fibonacci of Pisa (c.1170-c.1240 CE). The son of a world trader who took him along to Africa and Asia, Leonardo de Pisa (his proper name) wrote a remarkable book *Liber Abaci* that introduced Europe not only to Arabic numerals but also to algebra and its rules. Though living only a generation after Bhāskarāchārya (and a generation before Marco Polo), it doesn't seem likely that he went as far as India. He must have learned his algebra in the Middle East. After chapters on the basics, his book is mostly a huge collection of concocted problems of which I want to give an example belonging to a class of traditional but wildly unrealistic money puzzles (Chapter 12, p.415 in [Sig03]):

> *On Three Men with Sterling*
> *Three men had pounds of sterling, I know not how many, of which one half was the first's, one third was the second's and one sixth's was the thirds; as they wished to have it in a place of security, every one of them took from the sterling some amount, and of the amount that the first took he put in common one half, and of it that the second took, he put in common a third part, and of that which the third took, he put in common a sixth part, and from that which they put in common every one received a third part, and thus each had his portion.* If you are confused, below are the equations he has in mind.

$$\frac{x_1}{2} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{1}{2}\left(x_1 + x_2 + x_3\right)$$

$$\frac{2x_2}{3} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{1}{3}\left(x_1 + x_2 + x_3\right)$$

$$\frac{5x_3}{6} + \frac{1}{3}\left(\frac{x_1}{2} + \frac{x_2}{3} + \frac{x_3}{6}\right) = \frac{1}{6}\left(x_1 + x_2 + x_3\right)$$

This is 'just' a simple set of three linear equation in three unknowns. But even with modern methods, I struggled not to make arithmetic mistakes solving them. Gold star if you find 33:13:1 for wealth of the three men. His book has much text and a few illustrations, but I have not been able to see clearly how Fibonacci solved the problem.

A most extraordinary competition occurred in Northern Italy in the first half of the sixteenth century *over formulas for solving polynomial equations of degree 3 and 4*! From the time of the Babylonians, it was known how to solve quadratic equations. Why the problem of higher degree polynomials obsessed Renaissance Italians is unknown, at least to me, but the story apparently started with one Scipione del Ferro in Bologna discovering the formula for one type of third degree polynomials early in the sixteenth century *but*

keeping the rule a secret! However, he told the rule to his student Antonio Fior. Meanwhile, Niccolo Tartaglia found a formula for another type of cubic and challenged Fior, not to the customary duel, but to solve 30 cubic equations that each sent to the other! During the night of Feb. 12-13, 1535, Tartaglia had an inspiration and rapidly solved all of Fior's equations. The story continues: Gerolamo Cardano inveigles the formula out of Tartaglia and then, with the help of Ludovico Ferrari, worked out the formula for 4th degree polynomials as well. When, against his sworn word, Cardano published both, Tartaglia was incensed and challenged them to a debate in Milan, 1548. He lost, sued, lost again and retreated in disgrace to Venice. Seldom has math led to such public clashes.

Cardano (1501-1576), however, had published his book, *Ars Magna*, that immortalized all this joint work. In this book, the unknown is *rem ignotam, quam vocamus positionem*, which he abbreviates to *pos*. Its square is *quad* and he writes for example "6.m̄.1.pos.m̄.R.v.4.m̄.1.quad" for $6 - x - \sqrt{4 - x^2}$. Here, in translation, is an excerpt from this sixteenth century best seller:

> For example,
> $$x^3 + 6x = 20.$$
> Cube 2, one-third of 6, making 8; square 10, one-half the constant; 100 results. Add 100 and 8, making 108, the square root of which is $\sqrt{108}$. This you will duplicate: to one add 10, one-half the constant, and from the other subtract the same. Thus you will obtain the *binomium* $\sqrt{108} + 10$ and its *apotome* $\sqrt{108} - 10$. Take the cube roots of these. Subtract [the cube root of the] *apotome* from that of the *binomium* and you will have the value of *x*:
> $$\sqrt[3]{\sqrt{108} + 10} \; - \; \sqrt[3]{\sqrt{108} - 10}$$

He chooses a cubic equation with coefficients 6 and 20, apparently more or less at random, to show how his formula works. Oddly, he doesn't mention that the specific cube roots above can be evaluated and are equal to $\sqrt{3} \pm 1$, so that $x = 2$ is the solution. Of course, this is particular to his choice of coefficients. Anyway, can you imagine a crowd turning up today to hear two math guys argue over solving oddball equations? But I should add: like Diophantus, Cardano's work led to something really big, in his case Galois theory. But in addition, his formulas led him to both negative numbers and square roots of negative numbers. He viewed both of these with great suspicion but still made initial steps in setting up the algebra of complex numbers. Thus he says, correctly from our perspective, that if you want to divide 10 into two parts whose product is 40, the answer is $10 = (5 + \sqrt{-15}) + (5 - \sqrt{-15})$. The full story of cubics, complex arithmetic and the trisection of angles took another two centuries to work out.

Nearly a century later, with Descartes (1596-1650), we find a nearly modern algebraic notation, *though he still assumed his variables were positive numbers*. Thus he takes y as his positive horizontal coordinate, x the vertical, and he describes a hyperbolic arc in the positive quadrant over a segment $0 < c \leqslant y \leqslant a$ with asymptotes $y = 0$ and $y = a + c - \frac{c}{b}x$

on p.54 (original p.322) of *La Geometrie* [Des54] by:

$$yy \; \infty \; cy \; \text{--} \; \tfrac{cx}{b}y + ay \; \text{--} \; ae$$

Note his odd equals sign and minus sign. I have no idea where these came from. It took nearly another generation before Wallis (1616-1703) made negative numbers full partners of positive ones. I have written about this peculiar resistance in early modern Europe to negative numbers [**E-2010c**].

vi.   Today

These days, in K-12 school and in popular pseudo-math, almost anything can be used as a symbol for an unknown. There is a small industry of odd ball algebra problems on the web. I was challenged by a neighbor to solve one of these and thought it utterly trivial, only to find I had missed a detail and was quite wrong. I couldn't use that image as, in spite of its going viral, the *copyright* was unknown and the AMS nixed it. But a friendly problem guru, Rajesh Kumar, very kindly drew me a variant that is in Figure 4, just as crazy.

   I need to admit that math puzzles can be a lot of fun. A whole cult followed the puzzles and games that Martin Gardner wrote up in his Scientific American columns. And KenKen is addictive. I grew up with variants of Alcuin's famous wolf and river problem that dates from about 1200 years ago:

> A man had to take a wolf, a goat and a bunch of cabbages across a river. The only boat he could find could only take one passenger or baggage at a time. But he had been ordered to transfer all of these to the other side in good condition (i.e. the goat cannot be left alone either with the cabbages or the wolf). How could this be done?

Suffice it to say that the solution requires you to bring various things *back* after ferrying other things across. (There's also an X-rated variant with condoms that I will not reproduce.)
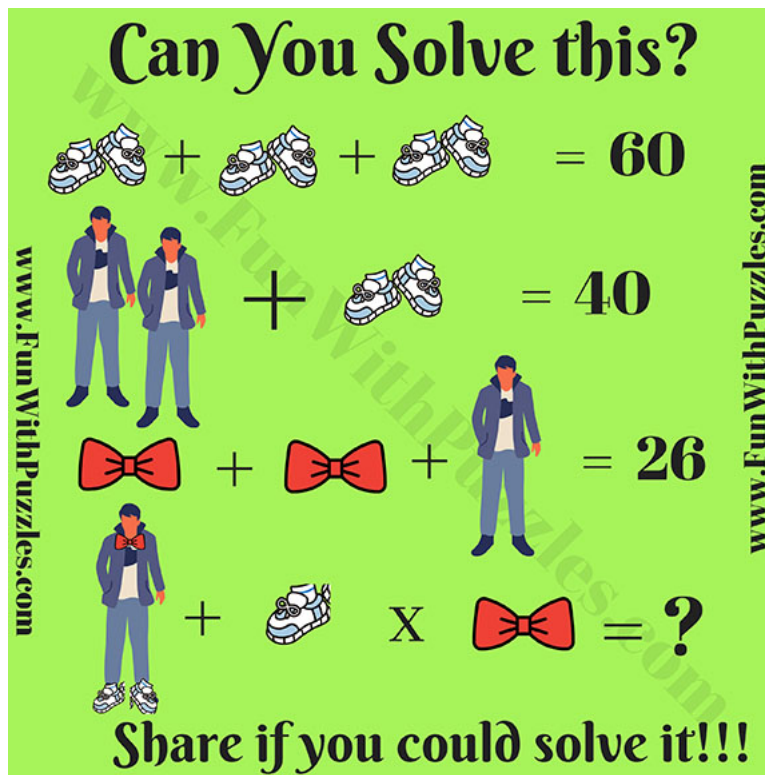
Figure 5.4: An array of equations, like the Chinese paddy problem or Fibonacci's problem. Do the numbers on the right mean prices? You can see both shoes, figures and bowties. Apparently, the objects are unknowns but weirdly, prices get multiplied, units being mixed up as in the Babylonian problem! This figure drawn by and used by permission of Rajesh Kumar, www.FunWithPuzzles.com.

Chapter 6

Multi-cultural Math History in 5 Slides
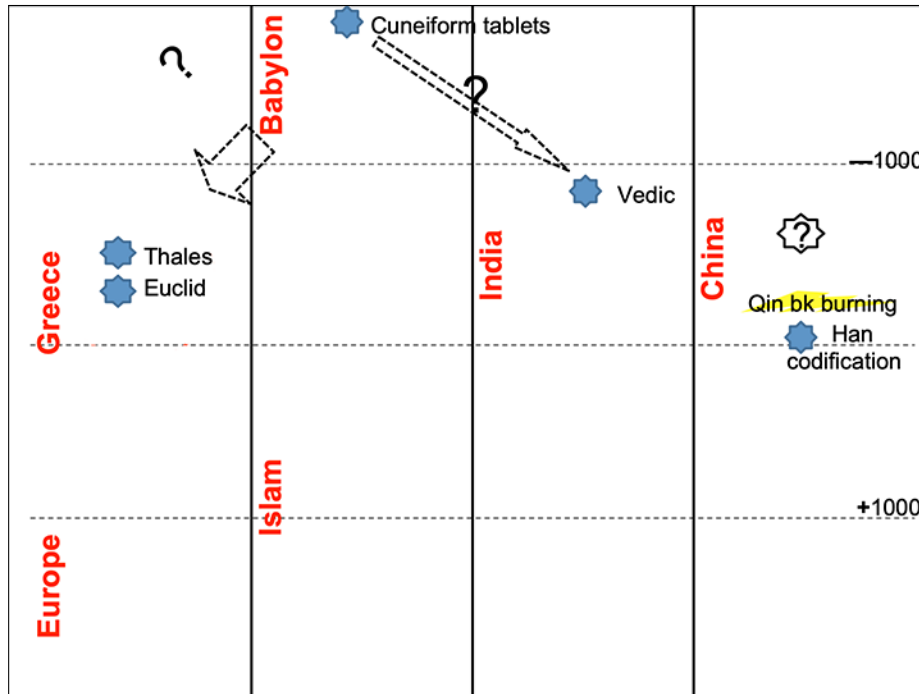
We begin with a recap of the history of the Pythagorean rule. This is an example of a big mathematical idea springing up all over the world, maybe independently, maybe not. It started in the city-states of Ur, Babylon etc. in Mesopotamia around 2000 BCE. Arguably, it might have been transmitted first to the Indus Valley kingdom, thence to the Vedic peoples where it appears explicitly in the rules for constructing sacrificial fire altars. Presumably independently, it is discovered in China but all traces of its discovery were erased by the nearly complete Qin dynasty destruction of ancient documents. At around this time, Grecian mathematicians, whom some believe (see Joran Friberg [Fri07]) absorbed the basic ideas of "geometric algebra" from Mesopotamia, incorporated it into their thinking, e.g. into Euclid's *Elements*. It is not too much to say that this rule came into its own in modern times when the size of an n-dimensional vector came to be defined as the "root-mean-squared" $\sqrt{\sum_i x_i^2}$, not to mention Gauss's statistical use of it in defining the variance of approximate observations when he recovered the position of the asteroid Ceres.
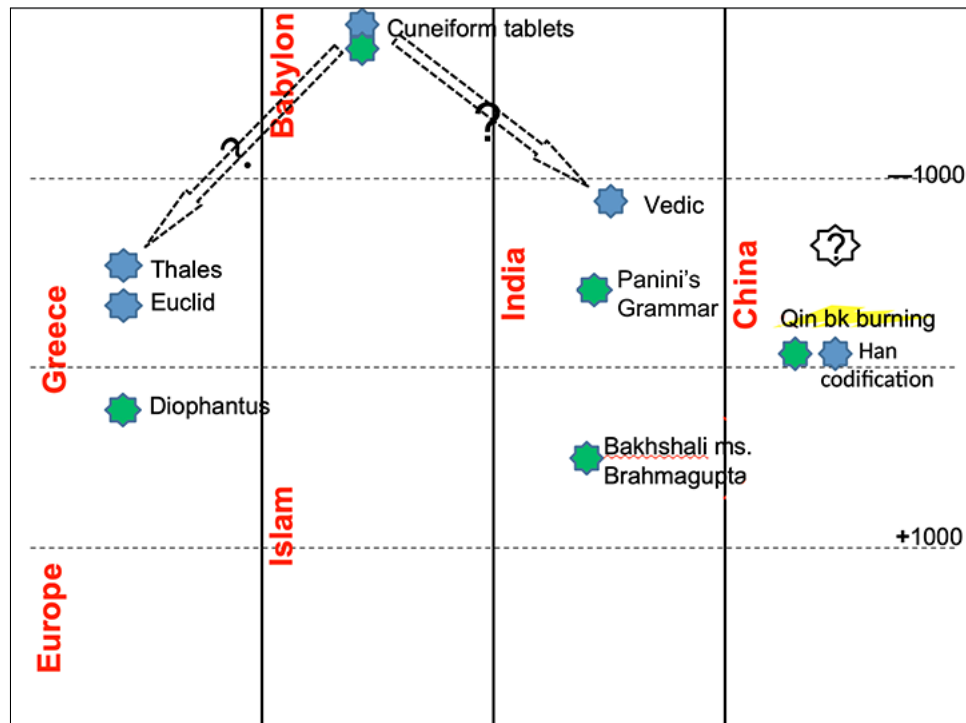
Next we look at the history of Algebra. This is an instance of clearly independent inventions all over the world. Again, it begins in Mesopotamia in problems posed for scribal students. But then it springs up in quite distinctive ways, seemingly both independently and idiosyncratically, in India, China and Greece. I argue that its real beginning in India was in Panini's famous grammar of Sanskrit. As discussed in the last chapter, his grammar uses symbolic references and organizes sets much like contemporary computer science and this is continued in Pingala's combinatorics arising from his analysis of Sanskrit prosody. Learning is cultivated by Brahmins, especially for its use in math and astronomy and in the large Buddhist "universities" in Nalanda and Taxila. In some mysterious way, a full-blown system of formulas and equations emerges in India by mid first millennium CE, found in both the Bakhshali manuscript and Brahmagupta's deep mathematical work. It is passed on verbally, by memorizing cryptic verses, from teacher to student, called the *guru-shisha* system. Meanwhile, China recovers some early algebra from the Qin ruins but mainly for commercial use in the marketplace, codified in the Han dynasty book "The Nine Chapters." But they *never adopt symbols for unknown numbers*, using only counting boards on which tokens are arranged much like the math in computers today. And finally Diophantus, outside of any clear Greek tradition, concocts his own formulas for rational number problems, aka "Diophantine" equations. It is interesting to compare the rudimentary formulas in Diophantus with those in the Bakhshali manuscript – they are not so different. But what follows a few centuries later is an extraordinary synthesis.
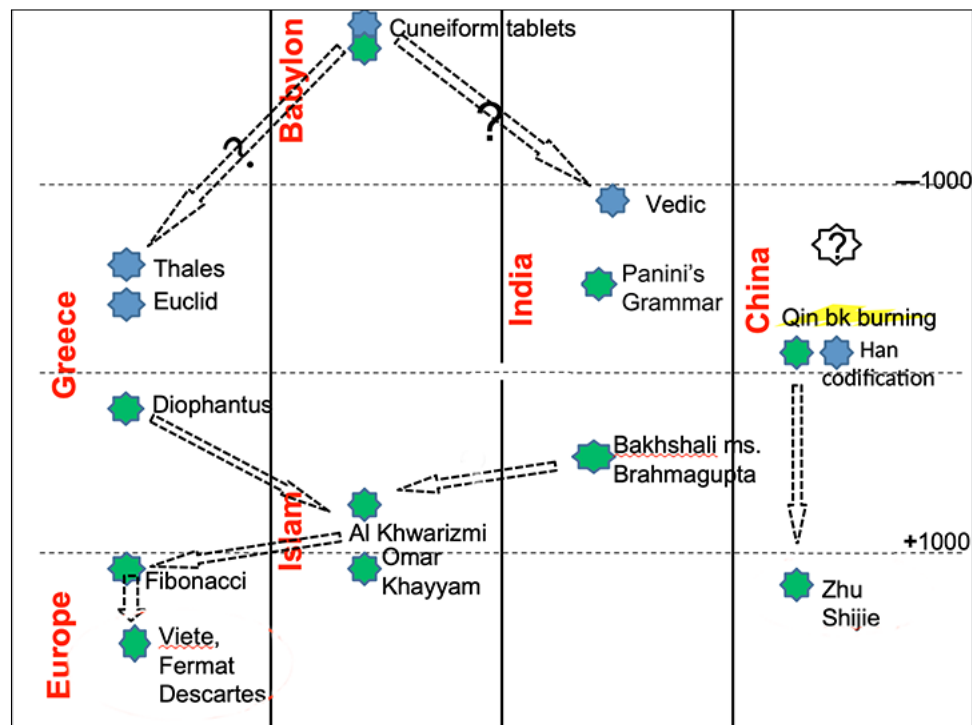
The synthesis was in the House of Wisdom, in Baghdad under Caliph Al-Mamun. Here works in Greek and Latin from Constantinople and some works in Sanskrit from India were collected. Al-Khwarizmi, a Persian from central Asia, wrote a text promulgating, first of all, the decimal system (a huge improvement on the sexagesimal system, not to mention Egyptian unit fractions where all fractions are described by sums of unit fractions $1/n$), but also some of the basics of algebra. Although not deep mathematics, this was a hugely important step creating a truly useful and learnable arithmetic. Then this was passed to the medieval Europe by Fibonacci whose book, *Liber Abaci*, also plays at great length with difficult algebra problems, perhaps not many relevant to his fellow Italian international traders. Meanwhile, another apparently solitary genius appears in Song dynasty China, Zhu Shijie. Still without using any symbols for an unknown, he invents the algebra of polynomials and devises the basic ideas of elimination theory (finding a polynomial $f(x)$ in the ideal $(g(x, y), h(x, y))$. His work is later taken up in Korea and Japan (e.g. by Takabe, c.1700) but not in China. My diagram ends with the Renaissance explosion of math in Europe: Viete, Fermat and Descartes whose algebra now gets close to ours today. They, however, still had a problem accepting negative numbers, clearly because in Euclid numbers were always positive. Truly modern algebra waited until first John Wallis and then Isaac Newton fully legitimized negative numbers (see [**E-2010c**]).
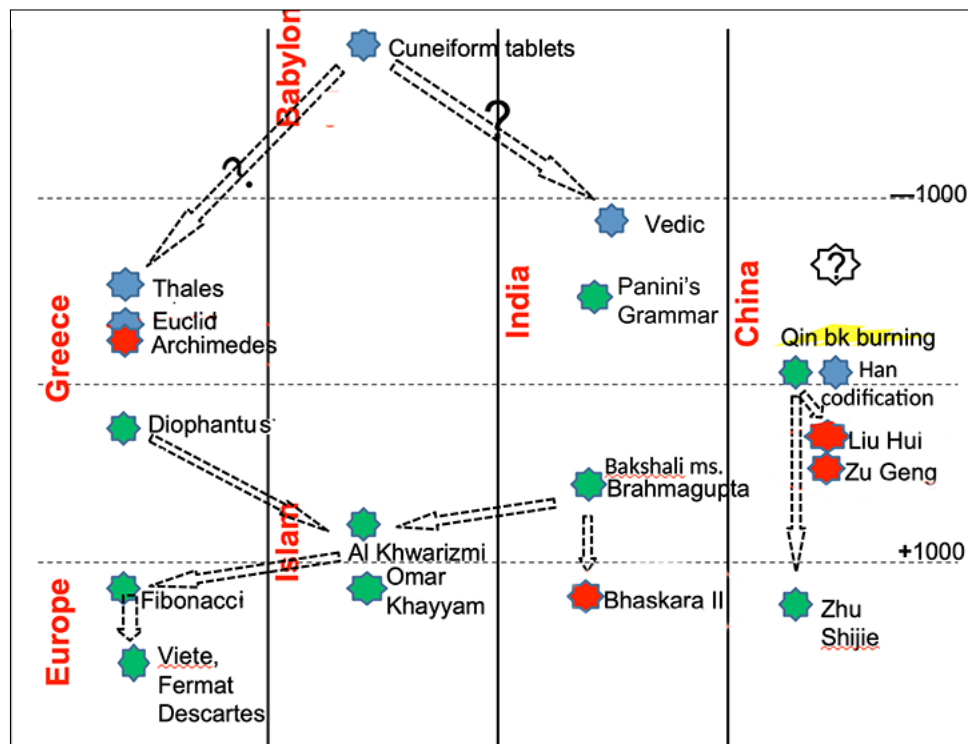
We now look at the beginnings of calculus in its use to work out the area and volume of a sphere. This seems to be an instance of quite independent discoveries that do indeed show strong parallels. Especially, what we now call *Cavalieri's principle* was hit upon in Greece, India and China apparently without any contact. Archimedes was the first in his famous palimpsest (i.e. a manuscript written on twice, once horizontally, once vertically) "The Method of Mechanical Theorems." In a nutshell, his method was to slice and dice objects and hang their pieces from a balance at different distances from the fulcrum but so that they balanced. Much of integral calculus including the volume of the sphere falls out. A similar method but with a totally different decomposition of the sphere was used by Liu Hui and Zu Geng. They start by showing that the volume of the "double umbrella," $x^2 + z^2 \leqslant 1, y^2 + z^2 \leqslant 1$ is $4/\pi$ times the volume of the sphere $x^2 + y^2 + z^2 \leqslant 1$ by comparing z-slices and then breaking up the double umbrella in an ingenious way. Zu found the correct result in the 5th century CE (although oddly using the silly approximation $\pi = 3$). Both Archimedes and the Indian mathematician Bhaskara II (or Bhaskaracharya, in the 12th century) worked out the *area* of the sphere quite independently by breaking it up into small slices via longitude (in spherical coordinates, $\theta \in [k\epsilon/2\pi, (k+1)\epsilon/2\pi]$). Essentially, they both evaluated $\int_0^\pi \sin(\phi)d\phi$.
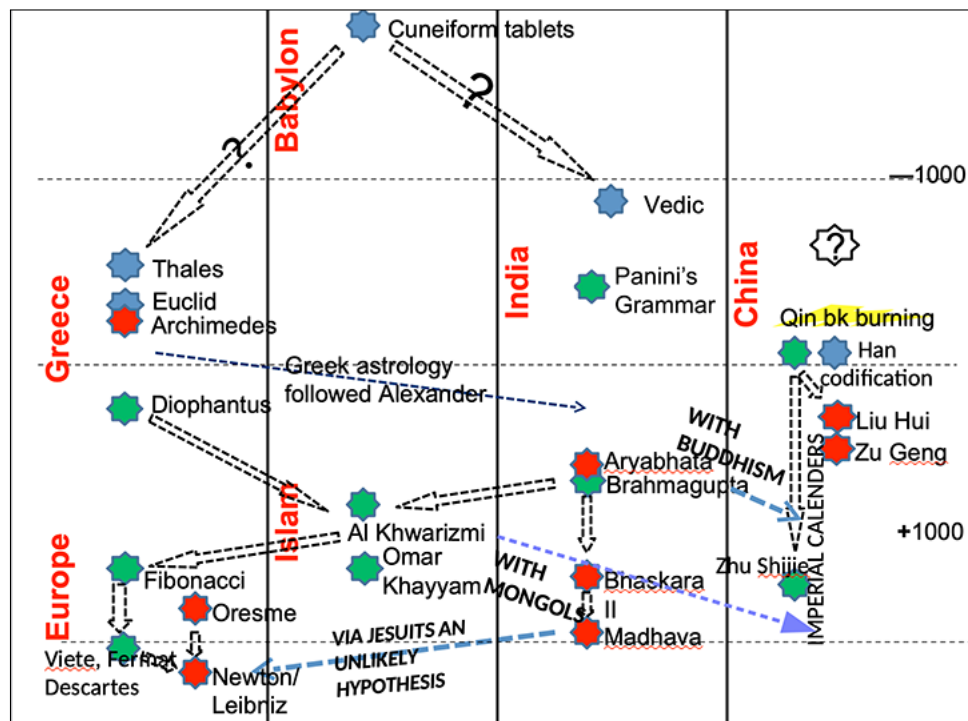
Now, let's look at calculus proper. Again, there is a remarkable parallel evolution but with some possible influences. In India, the differential calculus was discovered first, by Aryabhata, c.500 CE, a century before Brahmagupta. Astronomy, the need to calculate the positions of the sun, moon and planets, drove much of math in India, hence they were led early to sine and cosine and problems connected to circles and spheres. There had indeed been some transmission here though apparently not that of Archimedes' calculus. In the aftermath of Alexander's conquest, a Greek colony that eventually covered parts of Afghanistan, Pakistan and the Punjab (known as Gandhara or the Indo-Greek Kingdom[1]), Greek astrology was imported wholesale into India. How much astronomy was imported is debated but the use of epicycles and crude trig tables reflecting Greek ideas in the time of Hipparchus do seem to have made the jump. Where Aryabhata was really remarkable and totally original was in his discovery of the differential equation for the sine function, in a finite difference form (see [**E-2010a**],[Div18]). So the Indians started with differentiation, not integration! As mentioned, integration appears in Bhaskara II. But the dramatic flowering of calculus was in Kerala, along the southwest coast of India, in the period 1400-1600 CE, based on the brilliant work of a true genius, Madhava of Sangamagrama. They introduced infinite series and, for example, developed the power series expansions of sine,

---

[1]I bought a lovely silver coin from the time of Menander's rule c.300 BCE that is marked both in Greek and in Pali, using "Kharosthi" letters, a vivid demonstration of the interaction of the two cultures.

cosine and arctan. Thus they found (what was later called Gregory's expansion):

$$\tfrac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \cdots .$$

Two big questions about this work have been argued about: a) did they actually *prove* these results and b) did their math find its way to Europe, reaching Leibniz or Newton? As for (a), in my opinion, yes, their arguments are readily made fully rigorous by contemporary standards. But they didn't use ϵ, δ style estimates the way Archimedes had done in his "On the Sphere and Cylinder." There is no doubt that the ancient Greek mathematicians were the first to formulate totally rigorous proofs, even for calculus, arguably something not done again until Cauchy sat down to make calculus rigorous. Their math instead focussed on finding a *yukti*, a Sanskrit word cognate to the English word "yoke." Its original meaning was indeed "yoke" but, metaphorically, it can mean a device, an idea, a skill. Used for math, I understand it as meaning that their results had to be yoked together in a convincing way that clarified the whole and bound it together. As for (b), the question is whether, when the Jesuits came to Kerala, they sent manuscripts back to the Vatican and, from there, word got out to the 16$^{\text{th}}$-century intelligensia of Europe. However, no traces of such manuscripts, or even letters about them, have been found in the Vatican and the consensus is that this transmission is quite unlikely.

Another figure appears in my slide: Nicole Oresme (1325-1382), a French polymath bishop, who arguably restarts math in medieval Europe. In a sense, he is the first analyst (as opposed to geometer) inventing the idea of graphing and pointing out the importance of the area under the curve as the total quantity of something. He graphs mundane things like heat along a bar or velocity of an object varying in time, but also exotic things like a person's level of pain or their state of grace as functions of time. He uses the fundamental theorem of calculus in asserting that the area under the graph of velocity is the distance travelled and even considers improper integrals when the graphed value grows infinitely. Obviously, he paves the way for Newton and Leibniz.

A final influence and potential transmission is shown in my slide: the Chinese contacts with Indian astronomers in the Tang Dynasty and with Islamic astronomers in the Mongol Dynasty of China under Kublai Khan. Transmission of ideas from India or from the West have all come about either using the overland "silk road" or by sea. But the silk road is hardly a road: it traverses deserts and mountains and many intervening potentially warlike peoples. It is described in some detail in Claudius Ptolemy's *Geography*, written c. 150 CE, see the English translation [Pto11]. He states that he learned the geographic details of the road from merchants and used their reports to estimate the distance to China. A beautiful description of people using it in the second half of first millennium CE is in Whitfield's book [Whi15]. It was a major route in the Tang Dynasty (618-907 CE), a period when Buddhism was flourishing in China and manuscripts from India were much sought after. The monk Xuanzang travelled the road and spent 16 years in India, bringing back many such Buddhist writings.

What does this have to do with math? We need to know that one of the jobs of the Emperor was to issue, from time to time, a "calendar" (Chinese word *li*). This was not just a list of dates. It was also a whole ephemeris, describing celestial events and, in particular, eclipses. If the calendar failed to predict an eclipse, the Emperor might be thought to have lost the "mandate of heaven" and this was not a good thing! Better get it right. So the Bureau of Astronomy was an important office and astronomy, on the whole, was considered more significant than mundane mathematics, useful mainly for merchants. The remarkable point was that, along with Buddhism and Buddhists, actual astronomers from India came to live in the Chinese capital Chang'an (present-day Xi'an). Indian astronomy was quite advanced, at roughly the Ptolemaic level, since Aryabhata's masterpiece, the *Aryabhatiya*. They knew the size of the earth quite accurately and had a geocentric model of planetary motion with epicycles. Chinese astronomy in its whole history, up to the arrival of Matteo Ricci in 1582, had *no substantive geometric model of either a spherical earth*[2] *nor of planetary motion.* In other words, there appears to have been no transmission of the true geometry of the earth and planets from India to China.

I find this amazing and have studied it quite a bit ([**E-2012a, E-2016**]. If you dig deeper, you find an Indian astronomer named Gautama Siddha who wrote a calendar called the *Jiuzhi Li* in 718 CE, as well as compiling a huge Treatise on Astrology and Astronomy. But this was not officially adopted and it did not contain any of the Indian geometric model. Instead the calendar of the Buddhist monk Yi Xing (I Hsing in older transliteration) called the *Dayan Li* was issued in 728 CE. Yi was a very remarkable brilliant man, skilled in engineering as well as astronomy. At one point he travelled north and south measuring the altitude of the north star and the sun at solstices from near Lake Baikal in the north to Vietnam in the south, establishing how the angle of the tilt of an armillary sphere (a rotating model of the celestial globe) varies linearly on a meridian, see [Cul82]. It is impossible for me not to believe that Yi realized that the earth is round and that he calculated its circumference, comparing his data with the estimates the Indian astronomers had brought to China. Unfortunately, the model of a flat square earth, with China at its center, covered by a round celestial globe was entrenched in Chinese culture. Flat earth maps ruled by NS and EW lines go back to the *Yugong* in the Confucian canon with its 3×3 grid of provinces and continue through the 1136 CE Song Dynasty map, the *Yujitu*, carved in stone, with orthogonal rulings asserted to be equally spaced (see my analysis in [**E-2016**]). Moreover, Pei Xiu in the 3rd century CE wrote a treatise on how proper maps based on a flat square earth should be made. But realistically, if you're going to cover all of China, you must allow for the convergence of meridians or else seriously distort the geography. Apparently it was too radical to say so. Transmission occurs not merely when the first party wants to share an idea but when the second party wants to learn it.

The failure of transmission repeats itself in the Yuan Dynasty, under Kublai Khan.

---

[2]There was a philosophical speculation that the earth was like a yoke in the center of a cosmic egg, called the Han Tian theory, but without numbers, these remained abstract dreams.

Once again, China comes into contact with sophisticated astronomers, now Muslims, and once again the emperor has two calendars written, one by Muslim astronomers and one by Chinese. The latter, the *Shoushi Li* is the one that is propagated and, once again, it contains no geometric models. In fact, it contains a mysterious procedure for incorporating lunar parallax into eclipse prediction, a problem that screams for a geometric model. My own conjecture is that anything official had to be approved higher up in the bureaucracy and they could not allow anything that questioned the Confucian canon to be published. Nonetheless, I believe the Bureau of Astronomy maintained an understanding of the true picture as secret esoteric knowledge. My only argument for this is that the Bureau of Astronomy did issue calendars that made a decent stab at estimating lunar parallax and I can't see how they came up with this without a geometric model. It wasn't until the arrival of Matteo Ricci in China in 1582 CE that Western mathematics had any impact and how did he manage to do that? He translated the first five books of Euclid into Chinese!

I want to throw in one final comment. There are some Greek mathematical results that were neither transmitted to other cultures nor independently discovered. Two prominent examples are a) the list of the 5 Platonic solids and b) the idea of prime numbers including the unique factorization theorem and the fact that there are infinitely many of them. Both (a) and (b) are in Euclid's *Elements* but remarkably neither India nor China happened upon them.

Chapter 7

"Modern" Art/"Modern" Math and the Zeitgeist

My hypothesis in this chapter is that there has been an uncanny linkage between the underlying intellectual currents in Art and in Mathematics in the last two centuries. I first began to believe that this occurred when I noticed an amazing coincidence that occurred just after WWII.

i.   Beauty and power through randomness

The discovery that randomness can be harnessed to create both math and art seems to have taken place in the short period 1945-1950. It was expressed very explicitly in art by Jackson Pollock.

When the German emigré artist and intellectual, Hans Hoffmann, suggested to Jackson Pollock that he "observe nature" or his painting would become repetitious, Pollock – born in Cody, Wyoming – famously responded "f**k you, I am nature." Janson, in his History of Art (p.846),[Jan63], describes his paintings like this

> *Strict control is what Pollock gave up when he began to dribble and spatter ... The actual shapes were largely determined by the dynamics of the material and his process: the viscosity of the paint, the speed and direction of its impact on the canvas ... The result is so alive, so sensuously rich, that all earlier American painting looks pale by comparison.*

At almost exactly the same time, Nick Metropolis, Stan Ulam, and Johnny von Neumann at Los Alamos were proposing the same approach to modeling partial differential equations: *the Monte Carlo Method.* As von Neumann wrote to General Richtmeyer in 1947:
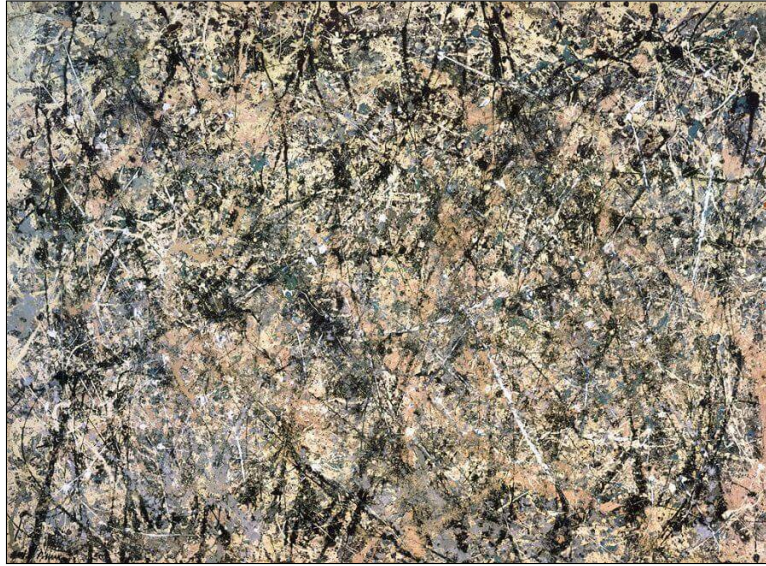
Figure 7.1: Jackson Pollock, *Lavender Mist*, 1950. From Wikimedia Commons, public domain.

> *I have been thinking a good deal about the possibility of using statistical methods to solve (nuclear devices) in accordance with the principle suggested by Stan Ulam. The more I think about it, the more I become convinced the idea has great merit.*

What is the Monte Carlo Method? An actual bomb has some

$$100000000000000000000000000$$

neutrons flying around inside it. Traditionally, one would try to model

$$d(x, y, z, u, v, w, t)dxdydzdudvdw$$

how many neutrons were at each point with each velocity at each time. Von Neumann, Ulam and Metropolis said – let's follow a small pollster's sample of them – say 100! – using the ENIAC. Instead of keeping track of all the uranium nuclei, let's just find the odds of each neutron hitting a nucleus at any given point, the odds of it splitting the nucleus, the odds of how many neutrons will come out and at what speeds and directions. We need to flip a lot of coins, so we get 100 pretend histories. Also we must keep track of how the uranium heats up, how it explodes (photons), etc. etc. It's a mini-simulation with dice. And this is actually how the H-bomb was designed!

Randomness is cool. Pollock found that spatter painting made a wonderfully energetic image, full of life. The school of abstract expressionism made this one of their favorite

THE BEGINNING *of the* MONTE CARLO METHOD

by N. Metropolis

... why mathematics is a delight to study, such a challenge to practise and such a puzzle to define.

—George Temple
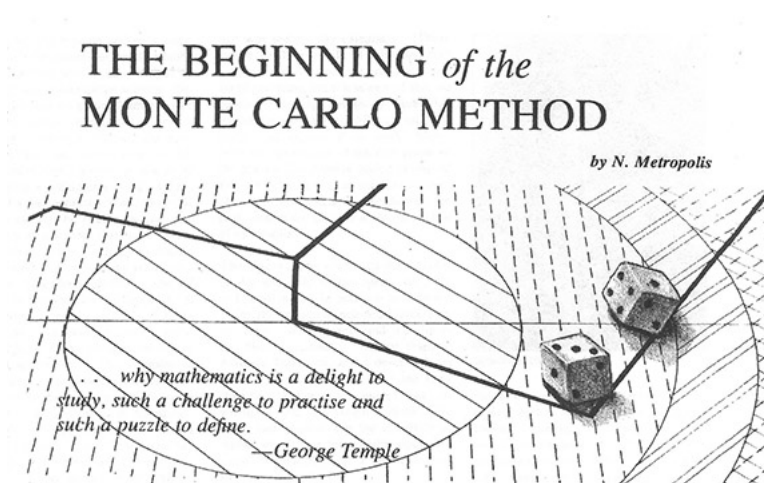
Figure 7.2: Neutron paths and dice on a sketch of a reactor, from N. Metropolis's article in *Los Alamos Science* [Met87], by permission Triad National Security LLC.

tools though few dared to be as wild as Pollock. Metropolis-Ulam-von Neumann found that throwing dice created realistic pseudo-worlds by which one can compute stuff in the real world. The Monte Carlo method is huge today in many types of calculations (like finding very large primes for banking encryption). Is it a coincidence that they happened nearly simultaneously in the late-40's!?

ii. When did abstract, non-figurative art & math start?

Surely, you say, all math is abstract and non-figurative! NO: what is abstract depends on the perceiver. Dealing with numbers as in Diophantus, geometry as in Euclid and processes in the world as in Newton are the concrete "representational" sides of math. Abstraction is a relative term: there are always "higher" levels of abstraction. The first stage of the movement towards abstraction was in the first half of the 19$^{\text{th}}$ century, focussing on one aspect of a concrete situation and throwing out irrelevant details to get to the essence. We see this clearly in the late work of Turner see here.

Meanwhile, what happened in Math was similar. Breaking the ties with the concrete, we get Galois (1811-1832); and Abel (1802-1829), two romantics whose ideas were rejected by the Academy, which could not understand what they were doing – it was too abstract. Galois died in a duel, Abel died of TB. Both were jobless and penniless, though their ideas were among the deepest of the 19th century. What did Galois do: he focused on one key aspect of the formulas, throwing out all details, like Turner painting light and air alone. Galois considered any possible formula for the solutions; if the degree is n there are n solutions. He proposed that you to rewrite all parts of the solution formula as expressions

Figure 7.3: J. M. W. Turner, *Steamer in a snowstorm*, 1842. Objects dissolve and he paints pure light, water and air, mixed in mist, spray and clouds. From Wikimedia Commons, public domain.

in these n solutions and then ask what rearrangements of the solutions don't change these expressions. Here's a simple example for the cubic polynomial. Start with the equation $x^3 + bx^2 + cx + d = 0$ and say x_1, x_2, x_3 are its roots. del Ferro's algorithm gives the solutions as:

$$x = \sqrt[3]{-\frac{b^3}{27} + \frac{bc}{6} - \frac{d}{2} + \frac{1}{6}\sqrt{\frac{4c^3 - b^2c^2 + 4b^3d - 18bcd + 27d^2}{3}}}$$
$$+ \sqrt[3]{-\frac{b^3}{27} + \frac{bc}{6} - \frac{d}{2} - \frac{1}{6}\sqrt{\frac{4c^3 - b^2c^2 + 4b^3d - 18bcd + 27d^2}{3}}}$$

Then the part inside the square root comes out in terms of the roots like this:

$$\frac{4c^3 - b^2c^2 + 4b^3d - 18bcd + 27d^2}{3} = (x_1 - x_2).(x_2 - x_3).(x_3 - x_1)$$

The expression on the right is preserved by cyclic permutations of the roots but not by interchanging two of them. He understood a solution formula as a way of step by step decreasing the number of these rearrangements until there are none – and you have one solution by itself.

iii.   Brave new worlds

The next step is the systematic creation of alternate and more vivid realities by counterfactual experiments with each part of our artistic/math tool-kit. In each experiment , the real world is modeled differently, one or another element is omitted or changed. These enlarge the aesthetic, while nature, in its richness, offers beauty/depth in new things classically deemed ugly. Here and on the next page are some examples, first showing the break with classical beauty and then three versions of brave new art worlds in which various classical conventions are discarded.

Figure 7.4: Left: Ingres, *The Valpinçon Bather*, 1808, the classical ideal of beauty. Right: Renoir, *Dance at the Moulin de la Galette*, 1876, dappled shade forms a fractal pattern over faces, classical ideals of beauty are disregarded. From Wikimedia Commons, public domain.

One instance of such experimentation in math is Karl Weierstrass's 1872 creation of a nowhere differentiable continuous function, one with no derivative at any point in its domain. This looks ugly but it tells you something essential about the universe of functions. The beautiful classical functions were fun – but to describe the non-smooth messy world, 'ugly' functions are needed too.

A clear math example of such experimentation is Hilbert's *Grundlagen der Geometrie*, 1899, where he made the ultimate analysis of Euclid's geometry (in 3D, using planes as well as lines), taking each axiom in turn and making an alternative geometry in which everything but that one axiom held. It is an elegant mind-game, seeking partially real alternate models of many kinds.

(A) 8 axioms of connection (e.g. given 2 distinct pts, there is a unique line containing both)
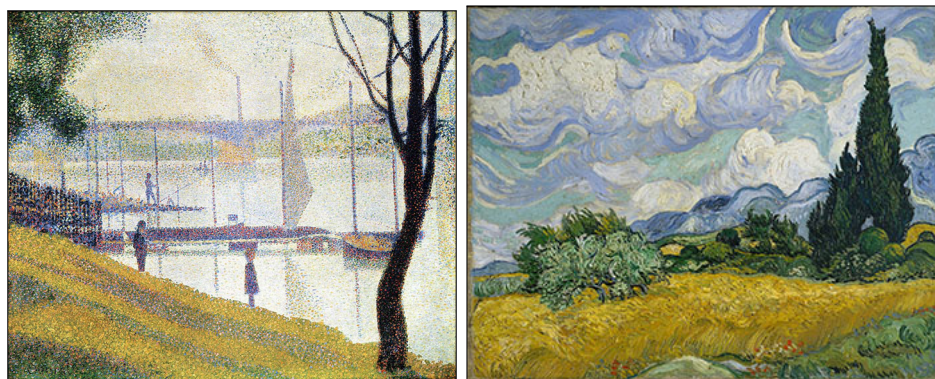
Figure 7.5: Left: Seurat, *The Bridge at Courbevoie*, 1886, form and texture are created out of dots, 'discrete geometry'. Sorry, you need to look closely to see the dots. Right: Van Gogh, *Cornfield with Cypresses*, 1889, form and texture are created out of fluid swirls, 'vector fields', From Wikimedia Commons, public domain

(B) 4 axioms of betweenness – based on work of Pasch, none in Euclid! (e.g. given 3 distinct pts on a line, exactly one is between the others)

(C) 5 axioms of congruence (e.g. 2 triangles with 2 sides and the included angle equal are congruent)

(D) 1 parallel axiom – in a plane, let ℓ be a line and P a pt off the line, then there is a unique line through the pt not meeting the line

(E) 2 axioms of continuity: Archimedes axiom: successive equal intervals cover the whole line and the 'Eudoxian' axiom: a sequence of nested intervals has a pt in the middle.

Of course, non-Euclidean geometries were those that dropped axiom D but no one had looked at the rest of his zoo.

iv.   Full blown abstraction

The final stage is to throw away all connection to conventional reality, the "reality" of the painting/math theory is not something it refers to, but something constructed by the art/math itself. See Figure 6 with works by Mondrian and Malevich.

In Piet Mondrian's own words, [Mon45]:

> *Art makes us realize that there are fixed laws which govern and point to the use of the constructive elements of the composition and the inherent inter-relationships between them. …. Non-figurative art is created by establishing*
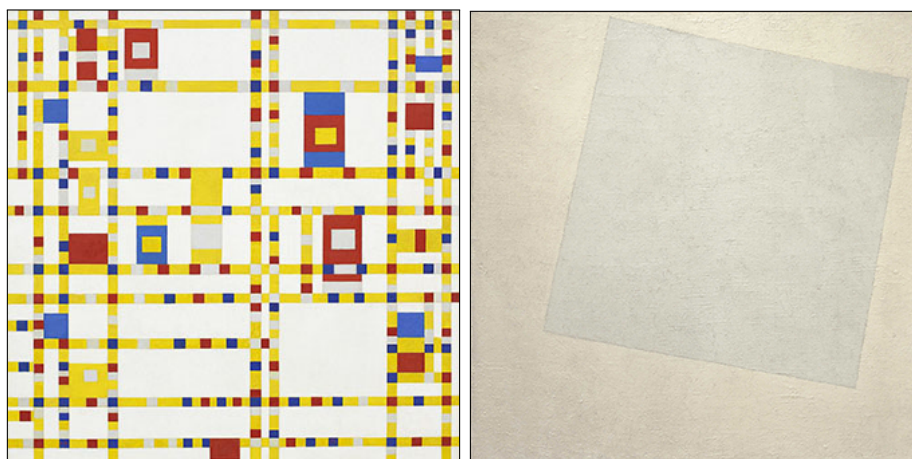
Figure 7.6: Left: Mondrian, *Broadway Boogie Woogie*, 1942, a dance of color and lines, a metaphor for Broadway. From Wikimedia Commons, public domain. Right: Malevich, *White on white*, 1918, truly minimal color and shape. From Wikimedia Commons, public domain.

> *a dynamic rhythm of determinate mutual relations, which excludes the formation of any particular form.*

Albers also wrote about what he was doing in his gallery notes (1965), explaining his minimalism:

> *(The) choice of the colors used, as well as their order, is aimed at an interaction – influencing and changing each other forth and back. …. Though the underlying and quasi-concentric order of squares remains the same in all paintings – in proportion and placement – these same squares group or single themselves, connect and separate in many different ways.*

Minimalism is an extreme example of abstraction, the reduction to the simplest conceivable structures. It goes back to Malevich but is seen today in Kelly, LeWitt, Serra and Noland.

Math (or better Pure Math) – at the same time – decided that every mathematical object should be built up from sets and their mutual relationships, members, subsets, sets of pairs. This is the ultimate reductionism and, in my mind, a clear equivalent to what Mondrian and Albers were doing.

1. Everything is a set, e.g. 5 is the set of the five smaller numbers 5={0,1,2,3,4}

2. The natural numbers is the infinite set $\mathbb{N} = \{0,1,2,3,....\}$

3. Addition is the subset of $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ of all triples $\{a, b, a + b\}$

4. A positive fraction is a maximal subset of $\mathbb{N} \times \mathbb{N}$, with any 2 pairs $\{a, b\}$ and $\{c, d\}$ in it satisfying $a.d = b.c$.

5. The plane P is the set of coordinate prs $\{x, y\}$, i.e. $\mathbb{R} \times \mathbb{R}$.

6. A line L is a subset of P of pts in the plane satisfying $ax + by + c = 0$.

7. etc., etc.

In Math, the best example of minimalism is the theory and classification of finite, simple groups (1955-1983). These are the most basic finite sets of elements, in which any 2 elements can be "composed" to get a 3rd. Think of this as like the smile of the Cheshire cat, they are the essence of symmetry when the symmetric object is taken away.

Today, we are in "post-modern" times. Both art and math have become more eclectic without a single focus. I think of this as like Mao's "let a hundred flowers bloom" campaign. Indeed, art is going in many directions (though its glamour and prices go in only one). I believe math flourishes with wild stuff like "infinity categories" on the pure side and the many wildly diverse branches of applied math on the other.

Interlude: Intelligent Design in Orion?

Looking at the sky from my hot tub in Tenants Harbor, as night falls earlier and earlier in the fall, I wait for the first sighting of Orion.[1] One evening, there it is, a warrior resplendent against the southeastern sky. Its seven principal stars all carry names - Rigel, Betelgeuse, Bellatrix, Saiph, Mintaka, Alnitak and Alnilam - and are among the 67 brightest stars in the whole sky[2]. The constellation is unmistakable not only as a cluster of so many very bright stars but also by its strikingly humanoid shape: Betelgeuse and Bellatrix form the shoulders, Saiph and Rigel the knees and Alnitak, Alnilam and Mintaka the belt. In addition, below the belt are the three stars, one the great nebula of Orion, forming Orion's sword. Every culture has recognized this striking cluster of stars: it was the god Osiris in Egypt, the Vedic creator of the universe, Prajapati, in India, one of the mansions of the White Tiger in China and the great father Hunhunahpo in Mayan Mexico. It is even conjectured to be the carving in a tusk dating from 32,500 BCE ([Rap03]).

This year the thought crossed my mind: is it not very improbable, if 67 stars were scattered at random in the celestial sphere, that such a pattern would be present? Aha, surely this is evidence of God's intervention, of the intelligent design proposed as an alternative to Darwinian evolution.[3] Having worked in computer vision, it is conceivable that the statistical models used in object recognition could quantify this. However, full human body models are not really ready for 'prime time'. But at least we can ask whether it is probable or not that 7 out of the 67 brightest stars should wind up so close to each other?

Moreover, the key component in what is sometimes called 'early vision' - that is the first steps in the analysis of the patterns of an image - is the identification of straight lines and extended curves in images. Psychophysics, esp. the experiments of the gestalt school,

---

[1]I trust the reader will enjoy a brief frivolous digression. This piece originally appeared in the *Svenska matematikersamfundet Medlemsutskicket*, (the Swedish Math Society Member Mailing), Feb. 2009, thanks to its then editor, my former student Ulf Persson.

[2]Because of variable and binary stars, there is some ambiguity in ordering stars by brightness, but using the listing in `http://www.astro.uiuc.edu/~kaler/sow/bright.html` the seven principal stars in Orion have ranks 7,11,26,29,30,52 and 67

[3]There's a great passage in "The Adventures of Huckleberry Finn" where Huck and Jim discuss how the stars came to be and Huck says there are too many for God to have made them all, so they just "happened."
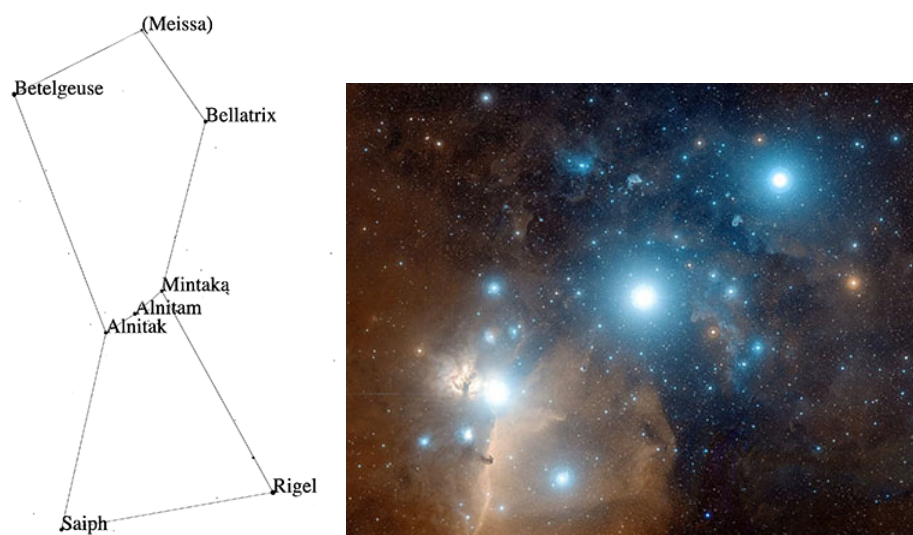
Figure 7.1: Left: The constellation Orion and its 7 major stars. Right: The three "belt" stars and the flame and horsehead nebulas next to Alnitak. Digitized Sky Survey, SOHO (ESA/NASA), public domain.

have confirmed that human perception recognizes these patterns in the midst of clutter with amazing sensitivity. Such curves can be contours of objects or parts of objects (such as limbs of trees). The three stars in the belt of Orion are striking not only because they are very close but because they are almost exactly regularly spaced in a line. Now the occurrence of such a linear pattern is easy to quantify.

Firstly, in the table below, we give the key facts about the seven main stars of Orion.

Star	Magnitude	Right Ascension	Declination	Distance
Alnitak	1.74	05 40 45.5	-01 56 34	815 ly
Alnilam	1.70	05 36 12.8	-01 12 07	1340 ly
Mintaka	2.23	05 32 00.4	-00 17 57	915 ly
Betelgeuse	0.70	05 55 10.3	+07 24 25	425 ly
Bellatrix	1.64	05 25 07.9	+06 20 59	245 ly
Rigel	0.12	05 14 32.3	-08 12 06	775 ly
Saiph	2.06	05 47 45.4	-09 40 11	720 ly

The data is from the Yale Bright Star Catalog (available via `ftp://cdsarc.u-strasbg.fr/cats/V/50/catalog.gz`), with recent distances from the Hipparcos satellite data, (found in `http://www.astro.uiuc.edu/~kaler/sow/bright.html`).

One checks that all seven stars are within $9.82°$ of Alnilam, the central belt star. Within the belt, Alnitak and Alnilam are $1.356°$ apart, Alnilam and Mintaka $1.386°$ apart,

a difference of only 2.2%. And that the exterior angle in the polygon joining Alnitak, Alnilam and Mintaka is only 7.5°.

To quantify the improbability of this, we turn to hypothesis testing. Hypothesis testing is the gold standard, for instance, of medical tests. Does some treatment improve a patient's chances of getting better? Well, suppose you know from past history that p_U is the probability of recovery in untreated patients. Now you take e.g. 1,000 patients and give them the treatment. Suppose p_T is the proportion of the treated patients who get better. Of course p_T had better be bigger than p_U or you can stop there. Then you imagine a game in which p_U is the chance of winning and you calculate the probability p of winning this game $1000 * p_T$ or more times if you play it 1000 times. In other words, we consider the null hypothesis that the treatment had no effect and then ask, if we assume the null hypothesis, what is the chance of seeing a proportion p_T or larger of patients being cured in a population of 1000. If $p < .01$, it is customary to give the treatment a seal of approval. In other words, when your health is at stake, if there is 1% or less chance of the medical test results coming out the way they did under the assumption that the treatment is worthless, you declare to the world at large that the treatment is worth taking. We want to apply hypothesis testing to Orion. We use the null hypothesis that the stars are scattered at random in the sky and we ask: what is the probability that the circle of radius 9.82° around one of them should contain 6 others. This is trivial to compute:

$$\text{Prob} \leqslant 67 \times \binom{66}{6} \times \left(\frac{\text{area(spherical disk),r=9.82}^o}{4\pi} \right)^6 \sim .001$$

BUT we are now committing the cardinal sin of hypothesis testing: we are choosing our test after we have the data, not before. This is the standard problem with people noticing "coincidences." Some striking thing occurs (Barlow used to talk of seeing five yellow VW bugs on the street one morning) and you say - "the probability of this happening by accident is tiny, so there must be some reason." What you don't do is try to imagine how many million other odd things might have happened but didn't. You picked the one test for which your reality had a low probability. What you need to do is apply the Bonferonni correction: if there are N possible remarkable events of which one actually occurred, you should take the p-value of that event, its probability under the assumption that everything is normal, and multiply it by N and ask if this probability is small, e.g. less than .05.[4] In the case of Orion, we chose to test for a tight cluster of 7 stars from the brightest 67. But there are many other possibilities, e.g. the Pleiades, a much tighter cluster but not all as bright. This was considered by John Mitchell in 1767 as we shall discuss later. If we put ourselves in the shoes of a person who has not seen the stars and ask what tests they might make to see if there are remarkable clusters, one approach, for example, would be to use the classification of stars by magnitude. Visible stars range in magnitude from -1 (the

---

[4]If all the tests are made at the same level p of significance, then the probability of one occurring under the null hypothesis is $1 - (1 - p)^N$ which is about Np

brightest) to 5 (maybe 6 but this requires very clear dry air which is in short supply these days). The seven major stars of Orion are all of magnitudes 0, 1 or 2. The six brightest stars of Pleiades are of magnitudes 3 and 4. There are, by one count, 2, 6, 14, 69, 192, 610 and 1929 stars of magnitudes respectively -1, 0, 1, 2, 3, 4 and 5. We might assign the significance level p and form a test for each magnitude n and cluster size m. If there are $N(n)$ stars s_k of magnitude at most n, we take as our test statistic:

$$f_{m,n}(s_1, \ldots, s_{N(n)}) = \min_{i(1),\ldots,i(m) \in (1,\ldots,N(n))} \left(\max_{2 \leqslant k \leqslant m} (\text{dist}(s_{i(k)} - s_{i(1)})) \right)$$

We need to find the value $t(n,m)$ such that:

$$\text{Prob}\left(f_{m,n}(s_1, \ldots, s_{N(n)}) \leqslant t(m,n) \Big| \text{position stars random} \right) = p$$

Then we check the values of this test statistic on the actual stars. The seven major stars of Orion are of 2nd magnitude at most and there are 91 of these on Kaler's web site referred to above (counting double stars as one). Then

$$\text{Prob}(f_{7,2} \leqslant 9.82°) \leqslant 91 \times \binom{90}{6} \times \left(\frac{\text{area(spherical disk)}, r = 9.82°}{4\pi} \right)^6 \sim .009$$

Aha: this means that if we chose p equal to the standard level 0.01 of statistical significance, we would find Orion causes us to reject the null hypothesis and conclude that the stars were not randomly distributed. But we have still committed the sin of fitting our statistic to the data by choosing the numbers $n = 2$ and $m = 7$. We can apply the same criterion to the belt, where 3 stars are within 1.386° of the center star:

$$\text{Prob}(f_{3,2} \leqslant 1.386°) \leqslant 91 \times \binom{90}{2} \times \left(\frac{\text{area(spherical disk)}, r = 1.386°}{4\pi} \right)^2 \sim .008$$

This is similarly 'statistically significant' - but not with a truly tiny p-value. I have not systematically examined for which m and n such significant clusters exist. This would be necessary if we went on to ask whether, if the stars were random, this collection of clusters was unlikely. Instead, I want to turn to a more unlikely situation which appears to be present in Orion.

Let's examine the belt more closely. Its amazingly symmetric configuration - three almost equally spaced stars very nearly on a line - is highly unusual. Such a configuration is called a 'linelet' in computer vision. If you consider clusters of three stars, there are only two striking special geometric configurations: equally spaced on a line or the vertices of an equilateral triangle. The Gestalt school of psychophysicists[5] investigated at great

---

[5]See, for example, Gaetano Kaniza's book [Kan80]

length what patterns in an image caused its points, lines and other parts to be grouped, to be seen as part of one object. The belief is that, for ecological reasons, what humans see is determined by what 2D patterns are most helpful in working out the 3D world around us. *Proximity* and *alignment* turn out to be the two strongest factors leading to visual grouping. An equilateral triangle is not a configuration found by the Gestalt school to be highly salient to the human visual system. This is presumably because equilateral triangles are not common in our visual experience whereas straight lines, whole or partially occluded, repetitive texture patterns and linear motion are very common.

We need to develop a specific statistic to measure the linearity of the belt. The most natural is the discrete second derivative, the angular distance from the middle star to the midpoint of the first and third star:

$$c = \text{dist}(s_2, \text{midpt}(s_1, s_3))$$

If $b = \text{dist}(s_1, s_3)$ is the overall size of the 'linelet', then we are associating to every triple of stars the simple, elementary and natural pair (b,c) which measures how closely it is a small 'linelet' (to use the terminology of computer vision). To develop a test, we need to combine b and c. It is easy to see that for three random stars, they are independent and have a distribution with density $\sin(b)\sin(c)/4\,dbdc$. Since they are independent, we take as our test statistic $T = b.c$. But this being small is not surprising unless both b and c are reasonably small, e.g. b should be less than the expected diameter of the smallest triple among the 91 randomly placed stars. A Monte Carlo simulation shows this to be about $6.7°$ or 0.117 radians. For the triple to look remotely like a linelet, we ask $c \leqslant b/8$ which means the spacing at worst 3:5 and the exterior angle at the middle star is less than $29°$. Then if we observe $T = T_0$, the p-value of this event among all stars of magnitude at most two is:

$$91 \times (91 \times 89/2) \times \iint\limits_R \frac{\sin(b)\sin(c)}{4}\,dbdc,$$

$$\text{where } R = \{b, c | bc \leqslant T_0, \ c \leqslant b/4, \ b \leqslant 0.117\}$$

In the case of Orion's belt, b $\sim$ 0.048 radians, c is merely 5.5 arc minutes or about 0.0016 radians, thus $T0 \sim 0.000076$. To evaluate the integral, we approximate $\sin(b)$ by b and $\sin(c)$ by c and find easily that $p \sim 0.00034$.

Now this is *much more* significant from a statistical viewpoint. But we still ought to allow for alternate tests for events that might have occurred but did not. While looking for unusual alignments, perhaps our cutoff at 2nd magnitude is arbitrary and perhaps 4 aligned stars should be considered too. This part of the argument really cannot be made precise. A common procedure is to allow some factor for this: I suggest 3, making the conservative p-value for the alignment of Orion's belt 0.001.
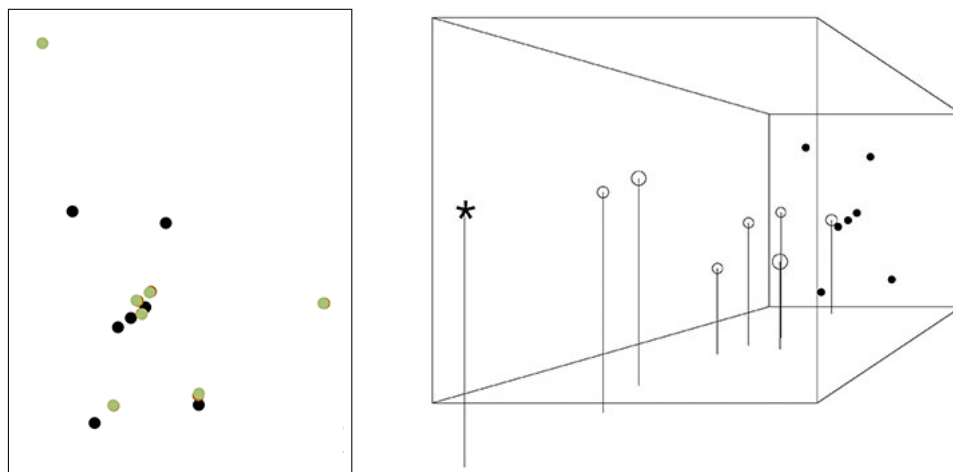
Figure 7.2: Left: Black = Orion today, Orange = Orion 2 million years ago; right = a 3D view of the whole constellation, with earth shown as an asterisk and the back wall showing an earthling's view, created by and used by permission of Prof. Ulf Persson. Note that the three belt stars are not at all aligned in 3D.

Now if the null hypothesis is rejected, what can be the cause of this alignment? In Gestalt psychology, alignment of some points in an image leads the perceiver to assume the world points projecting to these points on the image are aligned in three dimensions, unless there is strong evidence to the contrary. Aligned points in the world will be seen as aligned on the retina no matter what the viewpoint. Likewise, a cluster of salient points in an image is assumed to be caused by a cluster of points in the world. As we mentioned, John Michell in 1767, [Mit67], applied statistics to the Pleiades. Using the null hypothesis that the stars are scattered at random over the full celestial sphere and neglecting the caveats we have discussed, he asked how likely was it to find six stars as close together as they are in the Pleiades, among all the stars at least as bright. He found $p = .000002$ for the Pleiades occurring by random chance. He deduced from this that the null hypothesis was wrong and proposed that the Pleiades must be clustered in 3-space so that their positions in the sky were correlated, not independent. He actually went a bit farther and for this he was greatly criticized: he proposed assigning *prior* probabilities to the possibilities that these stars were close in 3-space vs. being distant in 3-space and merely close from the earth's vantage point. He could then apply Bayes's rule to deduce that .000002 was also the probability that the Pleiades were not a cluster in space. In fact, his conclusion was right: the Pleiades is indeed a cluster designated M45 in Messier's catalog.

How are the 3 stars of Orion's belt aligned in space? Fortunately, the Hipparcos satellite has provided excellent data on stellar distances. The result for Alnitak, Alnilam and Mintaka is shown in figure 2 right. It is clear that if the sun were positioned a little bit

above or below the plane of the belt, the three stars would fall out of alignment immediately, and the central star, Alnilam would move away from the other two. So Mitchell's alternate hypothesis does not explain Orion's belt.

As the sun and the seven stars of Orion move around our galaxy, the shape and the very existence of the grouping we call Orion will not remain. Betelgeuse is moving the fastest relative to the rest of Orion, flying left and up (north). Bellatrix is moving to the right and down nearly as fast. The rates are roughly a degree every 200,000 years or so. Alnitak is leaving the other 2 belt stars by a degree every 1-2 million years: enough to break its symmetry. According to Rappenglueck [Rap03], the shape of Orion has altered enough since Neolithic times that this can be detected in the prehistoric carving he analyzed. Figure 2 left shows a reconstruction of how Orion looked 2 million years ago with Betelgeuse off to the left, Bellatrix at the top.

What alternate hypotheses are we left with? Some might indeed infer from this evidence for intelligent design: that the creator has caused these 7 stars to assemble themselves as a great warrior just as homo sapiens emerged on earth. My son Steve suggested, tongue in cheek, that this could be described as God "micro-mangaging" the world. Frequentist statistics is a wonderful tool. Bayesians, on the other hand, put priors on alternate hypotheses such as intelligent design and, depending on your personal religious prior, this can radically alter your conclusions.

Part III

AI, Neuroscience and Consciousness

I began studying AI, computer vision and neuroscience in 1982. Although David Marr had just written an inspiring book [Mar82] on these subjects, neither AI nor computer vision at that time could boast any great successes. But talking with my colleagues at Brown, Ulf Grenander, Stu Geman and Elie Bienenstock, I became convinced of several key ideas. Firstly, that reasoning was statistical, not rule-based, and specifically used Bayesian models, implemented cortically by feedback [**B-1991, B-1994, V-1994b**]. Secondly, that an essential component of every form of thinking was grammar. This was often described as *compositionality*, the idea that higher order concepts were constructed by composing a cluster of components that fit into a learned higher order structure. We met in an inspiring conference in the Abbaye de Royaumont in 1991. Traditional grammars of language were assumed to be the tip of the iceberg as compositionality was ubiquitous. For example, it appears throughout the grouping principles of the Gestalt school of psychology. Chapter 8 describes this point of view and I later wrote a book, *Pattern Theory*, [**V-2010**] with Agnes Desolneux in which both Bayesian reasoning and grammars are key topics.

There was, however, all along an alternative point of view: the proposal that a very simple architecture, called *neural nets*, implemented stochastic reasoning and could automatically learn complex tasks without being helped along by being told about any fixed grammars. For almost three decades, I dismissed this view as wishful thinking until its manifest successes in the 2010's became undeniable. Chapter 9 describes my change of heart and attempts some synthesis. In fact, a key paper by Chris Manning and John Hewitt showed how grammars can be hiding in the neural nets. I describe this as well as describing some speculations on the cortical instantiation of this architecture updating the role of feedback loops.

Finally Chapter 10 concerns ideas about *consciousness*. A small revolution has taken place in the scientific community. Previously, the word 'consciousness' was forbidden in any scientific journal. Now, quite suddenly, it is all the rage. I have my own thoughts here, especially involving animal consciousness and whether or how physics connects with consciousness. Whether robots will be conscious in any sense is a huge and very significant question that the next few generations will undoubtably have to face.

Chapter 8

Parse Trees are ubiquitous in Thinking

i. Language

The field of linguistics has been split for most of my working life between those who followed the grammatical framework laid out by Noam Chomsky and those who resisted, claiming that language was richer and more idiosyncratic. This split was explained to me succinctly by the linguist Jean Gleason through a simple English sentence: "This dress zips up the back." Of course, 'zips' should be passive, not active, (the dress "is zipped") but idioms allow you to do most anything, to violate every rule. Grammar is a flexible task master and, in my opinion, seeking to codify every twist and turn is a fool's errand. People love to play with the language they speak. More recently, this has broken out into a feud between Chomsky and Daniel Everett over whether recursion and other grammatical structures must be present in all languages. Chomsky famously holds that some mutation endowed early man with a "language organ" that forces all languages to share some form of its built-in "universal grammar." Everett, on the other hand, was the first to thoroughly learn the vastly simplified language spoken by the Amazonian Piraha (pronounced *peedahan*) that possesses very little of Chomsky's grammar and, in particular, appears to lack any recursive constructions (aka embedded clauses), [Eve09]. What I want to claim is that both are wrong and that grammar in language is merely a recent extension of much older grammars that are built into every part of the brains of all intelligent animals to analyze sensory input, to structure their actions and even formulate their thoughts. All of these abilities, beyond the simplest level, are structured in hierarchical patterns built up from interchangeable units but obeying constraints, just as speech is.[1]

I first encountered this idea in reading my colleague Phil Lieberman's excellent 1984 book "The Biology and Evolution of Language." [Lie84] Most of this book is devoted to

---

[1]This chapter is based on the post "Grammar isn't merely part of language" dated Oct.12, 2014.

the still controversial idea that *Homo Sapiens* carries a mutation lacking in *Homo Neanderthalensis* by which its airway above the larynx was lengthened and straightened allowing the posterior side of the tongue to form the vowel sounds "ee," "ah," "oo" (i,a,u in standard IPA notation) and thus increase hugely the potential bit-rate of speech. If true, this suggests a clear story for the origin of language, consistent with evidence from the development of the rest of our culture. However, the part of his book that concerns the origin of syntax – and in particular Chomsky's language organ hypothesis – is in the beginning, esp. chapter 3. His thesis here is:

> *The hypothesis I shall develop is that the neural mechanisms that evolved to facilitate the automatization of motor control were preadapted for rule-governed behavior, in particular for the syntax of human language.*

He proceeds to give what he calls "Grammars for Motor Activity," making clear how parse trees almost identical to those of language arise when decomposing actions into smaller and smaller parts. It is curious that these ideas are nowhere referenced in the review paper of Hauser, Chomsky et al [HYB+14].

My research connected to the nature of syntax came from studying vision and taking admittedly somewhat controversial positions on the algorithms needed, especially those used for visual object recognition, both in computers and in animals. In particular, I believe grammars are needed in parsing images into the patches where different objects are visible and that moreover, just as faces are made up of eyes, nose and mouth, almost all objects are made up of a structured group of component smaller objects. The set of all objects identified in an image then forms a parse tree similar to those of language grammars. Likewise almost any completed action is made up of smaller actions, compatibly sequenced and grouped into sub-actions. The idea in all cases is that the complete utterance, complete image, complete action respectively carries many parts, some parts being part of other parts. Taking inclusion as a basic relation, we get a *tree* of parts with the whole thing at the *root* of the tree and the smallest constituents at its *leaves* (computer scientists prefer to visualize their "trees" upside-down with the root at the top, leaves at the bottom, as is usual also for "parse trees"). But at the same time, each part can be a constituent of other trees making a different whole and any part can be replaced by other compatible parts making a possible new whole – i.e. parts are interchangeable and re-usable within limits set by compatibility constraints. In other words, different parts can fill the same slot and the same part can appear in multiple slots in multiple trees. There is a very large set of potential parts and each whole utterance (resp. image, resp. action) is built up like legos of small parts put together respecting various rules into larger ones and continuing up to the whole. Summarizing, all these data structures are *hierarchical* and made up of *interchangeable, re-usable* parts and subject to *constraints* of varying complexity. I believe that any structure of this type should be called a grammar.

Let me start with examples from languages. Remember from your school lessons that an English sentence is made up of a subject, verb and object and that there are modifying

adjectives, adverbs, clauses, etc. Figure 1 is the parse of an utterance of a very verbal toddler [L.H29]. It contains two classical parse trees of the words plus a question mark

Figure 8.1: Grammar in the parsed speech of Helen, an especially verbal $2\frac{1}{2}$ year old.

for the implied but not spoken subject of the second sentence plus two links between non-adjacent words that are also syntactically connected. The idea of interchangeability is illustrated by the words "for Margaret," a part that can be put in infinitely many other sentences, a part of type "prepositional phrase." The top dotted line is there because the word "cake" must agree in number with the word "it." For instance, if Margaret had said she wanted to make cookies, she would need to say "them" in the second sentence (although such grammatical precision may not have been available to Margaret at that age). A classic example of distant agreement, here between words in one sentence with three embedded clauses is "Which problem/problems did you say your professor said she thought was/were unsolvable?" Plural nouns require the plural declension for verbs. This has been used to argue for the transformational grammars by Chomsky. It certainly shows that parsing sentences with simple *trees* and context-free grammars is not adequate for representing the full complexity of natural speech. Chomsky's adoption of transformational grammars is not unreasonable but we will argue that identical issues occur in vision, so neural skills for obeying these constraints must be more primitive and cortically widespread. In the next chapter, we will discuss a possible answer from deep learning experiments, the idea of low-dimensional projections of high-dimensional representations encoding sentences.

In other languages, the parts that are grouped almost never need to be adjacent and agreement is typically between distant parts, e.g. in Virgil we find the latin sentence

Ultima Cumaei venit iam carminis aetas.

which translates word-for-word as "last of-Cumaea has-arrived now of-song age" or, re-arranging the order as dictated by the disambiguating suffixes: "The last age of the Cumaean song has now arrived." Thus the noun phrase "last age" is made up of the first and last words and the genitive clause "of the Cumaean song" is the second and fifth

words, while the verb phrase "now arrived" is in the very middle. The subject is made up of the four words with orders 1,2,5 and 6. So word order is not essential for the tree structure if the relations of the underlying set of parts is determined by case and gender. Another example is Russian: *Tanja ubila Mašu*, 'Tanya killed Masha' can be said grammatically in all six orders! In other languages, e.g. Sanskrit, words themselves are typically compound groups, made by fusing simpler words with elaborate rules that systematically change phonemes, as detailed in Panini's famous c.400 BCE grammar. An example is in Chapter 5, section iv. Relative to Sanskrit, my good friend Prof. Shiva Shankar drew my attention Frits Staal's study *Ritual and Mantras, Rules without Meaning*, [Sta96]. Vedic rituals integrate speech and actions and embody a very precise abstract grammar. Finally, Comrie [Com81] gives an example of from Siberian Yupik , a sentence made up of a single word with a pile of suffixes: *Angya-ghlla-ng-yug-tuq* "Boat-[AUGMENTATIVE]-[ACQUISITIVE]-[DESIDERATIVE]-[3SING]," meaning "'He wants to acquire a big boat." A simple utterance but nonetheless with recursion, one sentence inside another unpacking to "He wants x; x = 'he acquires a big boat' ." Thus the parse tree leaves can be syllables of the compound words but there is still an implied tree of the familiar sort.

ii.   Vision

In the early twentieth century, the Gestalt school of psychology [Kan80, Ell99] was the first to develop grammatical grouping principles in the analysis of images. It was a real eye-opener to me when it became evident that images, just like sentences, are naturally described by parse trees. For a full development of this theory, see my paper [**V-2207**] with Song-Chun Zhu. Song-Chun, here and elsewhere, likes to describe grammars as "and/or graphs" (or AOGs), writing all possible expansions of a node as an "OR" node, and the required components of each expansion as "AND" nodes. The biggest difference with language grammars is that in images there is no linear order between parts. And even, when one object partly occludes another, two non-adjacent patches of an image may be parts of one object connected by an inferred hidden patch.

Figure 2, due to Zhu, shows the sort of parse tree that a simple image leads to. The football match image is at the top, the root. Below this, it is broken into three main objects – the foreground person, the field and the stadium. These in turn are made up of parts and this would go on to smaller pieces except that the tree has been truncated. The ultimate leaves, the visual analogs of phonemes, are the tiny patches (e.g. 3 by 3 or somewhat bigger sets of pixels) which, it turns out, are overwhelmingly either uniform, show edges, show bars or show "blobs." These visual "phonemes" emerge both from the statistical analysis of image databases and from the neurophysiology of primary visual cortex (V1) going back to Hubel and Wiesel's work, see my papers [**V-2003a**] and [**V-2006d**], §1.2.4.

Grammatical constraints are present whenever objects break up into parts whose relative position and size are constrained so as to follow a "template." The archetypal example
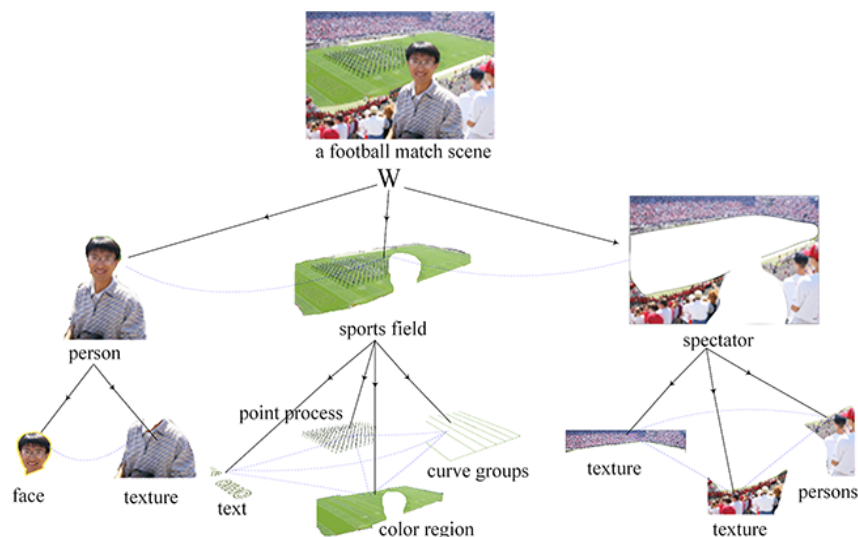
Figure 8.2: Parsing a scene at a football game, by permission of Prof. Song-Chun Zhu

is the face with 2 eyes, a nose and a mouth in their universal configuration. The gestalt psychologists worked out more complex rules of the grammar of images (although not, of course, using this terminology). They showed the way, for example, that symmetry and consistent orientation of lines and curves creates intermediate scale groupings between smaller phoneme-like patches as well as larger entire objects, even of non-adjacent patches. Thus two visible partial curves that *can* be smoothly joined into a single curve *are* so linked in our minds. They concocted multiple figures that demonstrated how powerfully parts of an object hidden by occlusion are inferred by people automatically. Another 2D and 3D parsing operation uses the graph formed by axes of the parts of an object. The archetypal example here is the representation of people by stick figures.

Figure 3 illustrates how occluded parts can be added to a parse tree. The blue lines indicate adjacency, solid black arrows are inclusion of one part in another and dotted arrows point to a hidden part. Thus H1 and H2 are the head, separated into the part occluding sky and the part occluding the field, and joined into the larger part H. S is the sky while VS is the visible part of the sky and H1 conceals an invisible part, Similarly for the field F and VF. The man M is made up of the head H and torso T. The top blue triangle MSF should be thought of as the largest groupings under the root.

Parsing images that require the full set of gestalt grouping principles has been pursued by Zhu's team, see for example [HZ09]. Figure 4 left is an example analogous to the above sentence concerning the professor's unsolvable problem, in which a chain of partially occluded objects acts similarly to the chain of occluded layers and creates constraints of illumination as well as texture and color. We also show on the right an example of a deeply shaded face. It is stunning how the mind can not only disregard edges formed by shadows
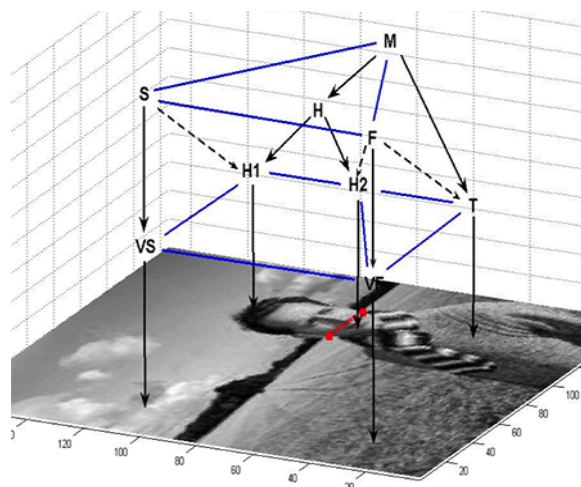
Figure 8.3: Occlusion complicates the parse tree. The forest and sky both continue behind the man's head. See text for abbreviations.

but mentally reconstruct missing edges where the face ends in the bright white glare or the deep black shadow.

Finally Figure 5 illustrates the power of a template for a compound object, so that these numbers are instantly recognizable in spite of added texture, thickening and thinning, outlines and shadows.

iii.   Actions and plans

Returning to motor actions and formation of plans of action, it is evident that actions and plans are hierarchical. Just take the elementary school exercise – write down the steps required to make a peanut butter sandwich. No matter what the child writes, you can subdivide the action further, e.g. not "walk to the refrigerator (for the peanut butter)" but first locate the refrigerator, then estimate its distance, then take a set of steps checking for obstacles to be avoided, then reach for handle etc. The student can't win because there is so much detail that we take for granted! Clearly actions are made up of interchangeable parts and clearly they must be assembled so as to satisfy many constraints, some simple like the next action beginning where the previous left off and some subtler.

The grammars of actions are complicated, however, by two extra factors: *causality* and *multiple agents*. Some actions cause other things to happen, a twist not present in the parse trees of speech and images. Judea Pearl has written extensively [Pea09] on the mathematics of the relation of causality and correlation and on a different sort of graph, his *Bayesian networks* and *causal trees*. Moreover, many actions involve or require more than one person. A key example for human evolution is that of hunting. It is quite remarkable

Figure 8.4: Left: An urban image of two of my grandsons in strong sunlight: note that the direction of the illumination is clear from their faces and *must be consistent* with their shadows and the illumination of the background. This is the exact effect present in the consistency of the number (singular/plural) of the unsolvable problems in the sentence given above: agreement of characteristics carried from one parse level to another. Right: Missing contours can also be caused by shadows and lighting but, as in this deeply shadowed face, the mind reconstructs them. Images by author and author's lab.

that Everett describes how the Piraha use a very reduced form of their language based on whistling when hunting. From the standpoint of the mental representation of the grammar of actions, a third complication is the use of these grammars in making plans for future actions. An example where some of the many expansions of one plan are shown in Figure 6 where the horizontal arrows represent causality (which may be in the past, present or future) and the vertical bracket is the expansion into component parts. Pursuing planning further, one encounters the need to model the knowledge and goals of multiple agents. In the human case, we also create and think about fictional worlds. Clearly new nodes have to added to specify the various contexts (in whose head, at what time, in what novel, etc.) in which an event or belief or desire takes place. My favorite sentence whose full parsing involves a lot of this complexity is "James turned out to be not as tall as he thought he was."

iv.   The big picture

I want to step back and look a bit more broadly at the parse trees and graphs we are proposing. Looking at *thought itself* as some kind of really big graph with links between related nodes has a long history. I believe one can trace it back to when Peter Mark Roget (1779-1869) sat down and decided to write a catalog of words he called a "Thesaurus," not

Figure 8.5: Visual "slang": graphic artists play games with our skill recognizing numbers in endless variations. We find the same underlying parse no matter what embellishment is present. Design by GeorgeTscherny Inc., School of Visual Arts.
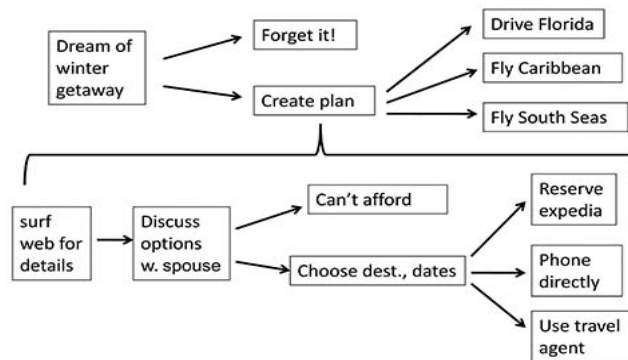
Figure 8.6: The grammar of planning is more complex as time sequences actions in ways that may or may not be causal.

a dictionary, but a huge graph. Its words are *classified* by their different meanings within the universe of thought and linked to each other by their similar meanings. On the highest level, he had 6 primary classes: a) Words expressing Abstract Relations, b) Words relating to Space, c) Words relating to Matter, d) Words relating to the Intellectual Faculties, e) Words relating to the Voluntary Powers, and f) Words relating to the Sentiment and Moral Powers. Each was divided and subdivided, until a fairly precise idea emerges described by one of 1000 key words. Then he gives a list of all words related to that key. For example, the table below shows how the particular key word "grammar" comes out through the successive subdivisions of the thesaurus.

I find it staggering that anyone should undertake such a project! Recent editions, still using the same name, drop the higher level categories and are a lot more mundane and sloppy. A remarkable analysis of the Thesaurus was carried out by Ron Hardin at Bell Labs after they digitized the whole thing (and before copyright pests stopped them). He found that using related words, the "distance" between an adjective and its opposite was never very far. For example, consider the chain 'generous' ↔ 'lofty'↔ 'superior' ↔ 'exclusive'

Table 8.1: How Roget's Thesaurus zeros in on the key word "grammar."

CLASS IV:	Words relating to the Intellectual Faculties
DIVISION II:	Communication of Ideas
Section III:	Means of Communication (vs. Modes, Natures of Ideas)
subsection 2:	Convential Means (vs. Natural Means)
subsubsction 1:	Language generally (vs. Spoken, Written Language)
Key word # 567:	*Grammar*, plus a list of related words, including
	"syntax," "parts of speech," "declension," "parse," ...

↔ 'selfish' ↔ 'ungenerous'. Each adjacent pair are words that refer to each other in the Thesaurus. (Here 'superior' is the linchpin as it may it can be ascribed for different reasons to both generous and ungenerous people.) The Thesaurus, though fascinating, has links that are more general than those in grammars, using symmetric links that may result from quite different relationships, and not considering the idea of groupings of parts.

Ulf Grenander has worked extensively on his version of a graph to explain mathematically every kind of cognitive task, even for consciousness itself. He calls this *Pattern Theory* [Gre81, GM07, Gre12] and, in my mind, his theory is a natural development of Roget, especially in the third of the cited books. It was, from its first formulation, based on graphs each node of which comes with what he calls "bonds," using which they are assembled into the graph and carrying attributes that constrain the bonds joining them.

In the grammars of this chapter, some of the links are shown vertically, thus are oriented and are used specifically to mean that a specific word/image piece/action is *part* of a larger group of words/image-pieces/actions. Others are horizontal and connect nodes that share some attributes (e.g. adjacency in sensory data or some agreement). But, as mentioned, the consistency can also be long range. Barbara Grosz once described to me taking transcripts of an expert helping a novice assemble a complex object *over the phone*. While assembling part A, there was a digression on parts B and C. Then the expert says "Now pick it up ..." and it was clear to the novice that part A was meant though it had not been discussed for the last few minutes. Context is all important and few utterances are context free.

There are many aspects of thought besides language, static vision and action/planing that have grammatical groupings. Here are a few:

- Videos: a video is a spatio-temporal signal, a function of both space (1D, 2D or 3D) and time. As such, understanding it depends on segmenting it. Tracking an object through time creates a tube-like subset of space-time. Or, another extreme, something like a door can be open or closed so you have a binary signal with jumps when an event occurs. The parsing is then similar to that with static images but with one new feature: some interactions of a human object on another object can be called *causal*. Zhu's team has pursued this extensively [PSY+13] parsing videos from

his lab as well as surveillance videos of a parking lot.

- Social groupings: there are a vast number of groupings of people relevant to their social behavior: families, clans, corporations, militias, nations, associations of every kind. Each such grouping has a parse-like structure and roles for its nodes, quite similar to what I have described above. Creating and reasoning with such parse structures is central to human life. Grenander proposed analyzing historical events with his type of graph.

- Categories: I'm thinking of the "is-a" graphs, introduced in early AI attempts to codify common sense knowledge, as in "a robin is a bird." Here the nodes are static categories of objects or actions, etc. thought of as sets, one including another. This is a natural more abstract extension of the grammars we have discussed.

For every link in the parse, there are constraints that must be satisfied between attributes of smaller and larger group. These attributes may come from the top, e.g. a face has slots for two eyes, one on the left, one on the right or it may be a consistency between the two, e.g. the ratio of the size of the face and of the eyes must be in a certain range or it may come from the bottom, e.g. certain items of clothing are usually worn by men, some by women.

Finally, all this should come with likelihoods. Constraints are seldom black and white. One encounters closely and widely spaced eyes, cross-dressing, confusion over the reference of a pronoun so all grammars should be statistical, not proscriptive.

To summarize, I believe that all animals with senses will also develop grammatical representations of the world around them from the signals they convey to the animal. Moreover, they typically carry out complex actions involving multiple steps by developing cortical mechanisms using further grammars. These grammars involve a mental representation of trees-like structures sometimes with extra long range linkages, built from interchangeable parts and satisfying large numbers of constraints. Language and sophisticated planning may well be unique to humans but grammar is a much more widely shared skill. How this is realized e.g. in mammalian cortex, is a major question, one of the most fundamental in the still early unraveling of how our brains work.

Chapter 9

Linking Deep Learning and Cortical Functions

One of the earliest ideas for programming Artificial Intelligence was to imitate neurons and their connectivity with *neural nets*. In the turbulent boom and bust evolution of AI, this remained a theme with strong adherents, but it fell out of the mainstream until around 2010 when these ideas were implemented with really huge datasets and really fast computers. The field of AI has now had a decade of tremendous progress in which neural nets, along with some major improvements, have been the central character. The purpose of this Chapter is to describe the further parallels between the software implementation of AI and the instantiation of cognitive intelligence in mammalian brains. I conjecture that, for better or for worse, all future instances of artificial intelligence will be driven to use these algorithms even though they are opaque and resist simple explanations of why they do what they do.[1]

i.   Neural Nets

Rectifying neural nets (ReLU nets), mathematically speaking, are just the class of piecewise linear functions $\phi : \mathbb{R}^k \to \mathbb{R}^l$ but defined in *a very specific way, as a composition of simple functions*, given by formulas of the following type:

$$\phi(\vec{x})_i = \max\left[0, M_{i,1}x_1 + M_{i,2}x_2 + \cdots + M_{i,k}x_k + b_i\right], \quad 1 \leqslant i \leqslant l \text{ or}$$

$$\phi(\vec{x}) = \max[0, (M.x + \vec{b})], M \text{ a given } k \times l \text{ matrix}, \vec{b} \text{ a given } l - \text{vector},$$

$$\max \text{operating componentwise.}$$

---

[1] This commentary started as a blog post "The Astonishing Convergence of AI and the Human Brain" that was put on the arXiv:2010.09101 as "The Convergence of AI code and Cortical Functioning."

Such a composition is always diagrammed as a set of layers, with a variables $x^{(n)} \in \mathbb{R}^{k_n}$ at layer n and with functions

$$\phi^{(n)} : \mathbb{R}^{k_n} \rightarrow \mathbb{R}^{k_{n+1}}$$

computing the next higher layer from the layer below and the running value after each composition being called the "activity" $\vec{x}^{(n)}$ in layer n. The components of these activities are called "units", as these are supposed to be analogs of neurons in the biological interpretation. The whole net depends on the weight matrices $M^{(n)}$ and the bias vectors $\vec{b}^{(n)}$ for each level, all of which need to be learned by fitting data via gradient descent, called "back propagation" from the shape of the formula for the gradient. In typical statistical settings, you assume you have access to a potentially infinite set of inputs and know, for each, what the output should be. In the simplest case, you have a binary output $\{0, 1\}$ and are just separating inputs into two classes. How does the learning work? We assume you are given a set of inputs at the lowest level and, for each input, the desired output at the top level, and then you measure how well your net works by the sum of the squared differences between what actually comes of the neural net and the desired output. It is easy to compute the partial derivatives of this measure with respect to the weights and biases, hence get a gradient – the direction in which this decreases as fast as possible. This is gradient descent in a situation known as *supervised learning*, i.e. you assume that for each set of training data, the desired output is given. Of course, the "proof of the pudding" is to test the neural net on new input data and, inevitably, the net doesn't work so well on this testing data because it has "overfit" the training data, making use of various unnoticed quirks.

It is easy see that ϕ is a piecewise linear continuous function: if the vector space of inputs is divided into polyhedral cells, each defined by the set of units whose activity is zero, then the output is a linear function on each of these cells. This whole apparatus is just an example of regressing data with a particular class of functions. A miracle (for which to my knowledge nobody yet has a good explanation) is how well gradient descent works to train the neural net: tested on new data, its performance is usually not that much worse than it was on the training data. Somehow, it rarely overfits the training data even if it has a truly huge number of weights. Except for some small bells and whistles, *this is the whole thing*. Calling it "deep learning" was pure PR.

The motivation for this algorithm was an extremely simplified model of what animal neurons do. Neurons in all animals, from a jelly fish up, form a directed graph: the vertices are the neurons, every neuron has a single axon (its output) and multiple dendrites (its inputs), its axon branches multiple times and each branch contacts the dendrites of some other neuron at synapses which then form the edges of the graph. Electrical signals do indeed propagate from neuron to neuron, out along the axon, across the synapse and into a new neuron via a dendrite. The signals, however, (with a few exceptions) come in short (1 or 2 milliseconds) identical pulses, its spikes, so the message sent from one neuron to another is called a spike train. Simplification #1: take the rate of firing, spikes per second,

as a real number signal emitted by each neuron. Thus, in neural nets, each unit is taken to correspond to one neuron, its real value $x_i^{(n)}$ being the associated firing rate. And when does the receiving neuron emit a spike in its turn? Simplification #2: assume that all neurons add up their active dendritic inputs linearly with weights indicating the strength of each synaptic connection (positive if it is an excitatory stimulus to the receiving neuron, negative if it is inhibitory) and that their firing rate is this sum after some kind of rectifying function is applied because the firing rate must be positive. Bingo: this is exactly what the function ϕ does, so we now have a neural net that is a rough caricature of the biological reality. Well, there is also Simplification #3: assume that neural synapses do not form loops so we can put the neurons in layers, each speaking only to neurons above them. *Unfortunately, none of these simplifications are true!* The precise timing of neural spikes is believed to be carry much information, the output of a neuron is known to be much more complicated function of its synaptic input and there are many loops in the graph formed by synaptic connections between neurons. I discuss some of this below.

Some modifications that make the neural nets a bit more realistic have been known for some time to also make them work better. First, there is no rigid layer structure in the cortex and neural nets often work better when there are layer skipping links, i.e. layer n can have some inputs from layer $n-2$ or lower. A special case are "residual" networks where the variable $\vec{x}^{(n-2)}$ is added to the variable $\vec{x}^{(n)}$, forcing the intermediate layers to seek not a totally new signal but an additive correction to $\vec{x}^{(n-2)}$. Another modification, known as "dropout," trains the network to work even with a certain percent of variables $x_i^{(n)}$ set to zero. This forces the neural net to be redundant just as our thinking seems to be resilient to some neurons malfunctioning. A third improvement is called "block normalization". This introduces an extra variable at each unit that, together with its bias, moderates the mean and variance of each unit's response to a random batch of data, something like regulating chemicals in the neuron.

ii. Tokens vs. distributed data

The neural recordings of David Hubel and Torsten Wiesel in the 1960's found a remarkable thing. They recorded from V1, the primary visual cortex, in cats, and discovered that each neuron seemed to have a definite preferred stimulus, like an bar with a certain orientation or an edge between a light and a dark region also with some orientation or even an isolated blob in a specific location on the cat's retina. It was this stimulus that caused the neuron to fire. Note that this is not just one neuron for each image pixel, it is more like one neuron for each of the simplest elements that make up images, this element being what that neuron is attending to. This led to Simplification #4, that *all* neurons were waiting for some event, some situation, stimulus or planned movement and that they fired in its presence. This was humorously called the *grandmother hypothesis*, e.g. why shouldn't there be a cell somewhere in the brain that fires if and only if you are looking at your grandmother? More

to the point, for each word we hear or speak, there should be a cell which fires when that word is heard or pronounced. If the grandmother hypothesis were true, all we needed to do was figure out the dictionary, neuron to stimulating situation, and an exhaustive recording of neural activity would tell us what the animal is "thinking." Although there were a few successes in this direction, it hit a brick wall when recordings were made in the higher visual area V4 and in the visual inferior temporal cortex (IT). It was quickly discovered that these cells were indeed paying attention to visual input and seemed to be looking at more complex features of the retinal signal: shapes, textures, perhaps the identity of objects in the scene. But no one could pin this down because there seemed an explosive number of combinations of features that stimulated each cell to varying degrees. In other words, the *simultaneous firing pattern* of large populations of cells seemed to carry the information, instead of each cell separately telling us one thing about the stimulus. Thus the stimulus seems to be encoded as a high dimensional vector that captures what was going on, perhaps thousand dimensional or more. The information is distributed over an area in the cortex and no simple meaning can be attached to the firing of single cell. Here's a new confirmation of neural net architecture: the idea that it is the simultaneous real values of all units in a layer that carries the data while the values of single units have no easy interpretation. It now seems as though Hubel and Wiesel's result, though true, was quite misleading when applied to rest of cortex.

Meanwhile, the AI people were trying to solve problems not only with understanding images but especially understanding language. Raw images are represented already by a big vector of real numbers, the values of their pixels. Typical problems are face recognition and general object recognition. Words, on the other hand, are just a list in a dictionary. Typical problems are sentence parsing, machine translation and internet question answering. How should neural nets be applied to such language tasks? A breakthrough arrived when a team of researchers at Google published an algorithm in 2013 called *word2vec*, [MCCD13]. The idea was to represent each word as a real valued vector in a high dimensional vector space, an instance of what has been called *vector symbolic architecture*. The constraint was that words which often occur near each other in speech or written text should correspond to nearby vectors, their distance reflecting how often they co-occur. One way to think of this is that a word has many aspects to it such as its syntactic role, its semantic classification in many senses, as well as other reasons why it co-occurs with other words, and high dimensional vectors have enough freedom to be able to capture much of this. For this to work, the high-dimensional vector must somehow encode a great deal about both the language and the world. If we describe the vector attached to a word by putting it in square brackets, then the most famous example of how it works is that the closest word vector [x] to the vector [king] + [female] – [male] turns out to be [queen].

What is remarkable is that this represents a major convergence of AI programs with actual neural activity. Needless to say, no neurons have ever been found the human brain

that respond to a single word and only that word[2].

iii.   Transformers and context

Remarkably, the Google language team went further in the 2017 paper entitled *Attention is all you need*, [VSP$^+$17]. It introduces a completely new architecture that enhances neural nets in a powerful way, the *transformer*. The authors are looking at linguistic tasks involving whole sentences which build word representations that encode the meaning of its words *in the context of a whole sentence*. The linguistic tasks they sought to solve all involve outputting a new sentence, e.g. translating the input sentence into a second language, answering the question posed by the first sentence (similar to the quiz show "Jeopardy") or more simply finding the word that was purposely omitted from the first sentence. Thus the algorithm has two parts: an encoder creating vector representations of all words occurring in sentences in the database and a decoder reversing the process producing a new sentence.

What does the transformer do? Transformers are made up from adding *attention heads* to conventional neural nets. Each head is a linear projection, from the full data representation of a word at some level key in the encoder, to a significantly lower dimensional vectors that, after being isolated, can play an essential role at later stages of the computation (for instance projecting from a 512 dimensional vector to a 64 dimensional one). Looking at the encoder net, for a given small set of layers of a conventional neural net, they add an extra layer each with a small number of these attention heads. (In the referenced paper, they happened to use a set of 6 layers, each made up of 512 units and added to each of them 8 attention heads.) For each of these heads, one trains three linear maps from the 512 dimensional layer data to shorter 64 dimensional vectors, the maps being called *queries*, *key* and *values* respectively. Remarkably, this introduces $6 \times 8 \times 3 \times 64 \times 512$ (or roughly 5 million) more coefficients that need to be trained, that is, learned by gradient descent from the dataset of sentences! Before the advent of contemporary super-fast computers with so-called GPUs, such an algorithm would have been impossible to implement. Assume you are processing the words in a specific sentence in the database. The idea is first to find, for each *query* applied to the current word (by matrix multiplication with the vector for this word at this level), the *key* is applied to all the other words in the sentence resulting in weights measuring from different perspectives the relevance of the context word to the current word: more precisely, scale a dot product measure of the distance between this query and this key to [0,1]. Finally use this to weight and then add up the head's *value* vectors applied to the context word (see formula below). Concatenate these over the 8 heads, bringing the dimension back up to 512 and train a final 512×512 matrix to jumble it all up like a fully connected layer of a neural net and add this to original layer vector. OK, this sounds complicated but, expressing it with a formula, this comes out fairly simply

---

[2]Recordings from the exposed brain of awake patients are employed in some operations for severe epilepsy

and unambiguously:

$$\text{Cat}_h \sum_\alpha \text{softmax}_\alpha \left(C.(X.W_h^Q).(Y_\alpha.W_h^K)^t \right) Y_\alpha W_h^V$$

where "Cat" stands for concatenate, h indexes the heads, "softmax" means exponentiating a set of numbers and normalizing to make their sum equal to 1, C is a constant, X is the input vector, Y_α are context vectors, the W^Q's are the query matrices, W^K the keys and W^V the values.

One of the most convincing demonstrations of what transformers do comes from the 2019 paper *A Structural Probe for Finding Syntax in Word Representations* by Chris Manning and John Hewitt, [MH19]. They took the public domain Google program "BERT-large" that, when given a database of sentences, produces vector representations of all its words in context. The program comes with fixed queries, keys and values from its training on two tasks: i) inputting normal English sentences from which one word has been excised, it is asked to output the full sentence and ii) the task of determining, for a pair of sentences, whether the second was a logical continuation of the first or has nothing to do with it. The point here is that it has *not* been trained on any tasks explicitly involving syntax. They then took sentences with known parse trees from a different database and looked for low dimensional projections of BERT's word representation at various levels such that, for any two words in the sentence, the squared distance between their projected word vectors approximated how many links in the parse tree connected the two words. Amazingly, they found that the best projections to say 20 dimensions allowed them to reconstruct the true parse tree with 80% accuracy. In other words, BERT's transformers were implicitly finding the underlying syntax of the sentence, but hiding it in the 512-dimensional vector representation, but then presumably using it in order to solve the missing word or the sentence continuation problem. This goes a long way, I think, to clarifying why these programs are so good at language translation.

The really significant conclusion of this demonstration is that, yes – the neural net is learning syntax, but no – it doesn't make *explicit* use of the syntax to solve problems. In the previous chapter, we have argued that grammars and their graphs are one of the main components of thought. It appears, however, that these graphs need not be an *explicit* part of cognitive algorithms, that they may merely be implicit. I will return to this in a discussion of the Whorfian hypothesis below.

iv.   Context in the brain

The take away from the success of transformers would seem to be that calculations that incorporate context require more than the simple weighted summation of vanilla neural nets. And, indeed, it has also long been clear that *neurons do a great deal more than add up their synaptic inputs.* To explain this, we need to review more of the basic biology.

<u>iv.a: Pyramidal cells</u>

The cortex is the structure common to all mammals which clearly is responsible for their cognitive intelligence (as opposed to muscular skills, instinctive emotional responses and routine behaviors). It is composed of six layers, each with its distinctive neurons and connections. Something like 2/3rds of its neurons are *pyramidal cells*, large excitatory neurons oriented perpendicular to the cortical surface, with up to 30,000 synapses in humans. They are the workhorses of cognition. They occur in most layers, as shown in the figure below.

Figure 9.1: Sample dendritic arbors of excitatory cells from mouse cortex with cortical layers shown on the left. All are pyramidal except for the 3rd (a *stellate* L4 cell) and last (a *multipolar* L6B cell). The cell body, called the *soma*, is the dark blob near the bottom of each cell. All the lines are dendrites, those at the top called "apical", those at the bottom "basal". From [RF18], figure 1, licensed by Creative Commons, Radnikow and Feldmeyer.

Modelers have long known that a pyramidal cell does something more complex than simply add up these 30,000 inputs. For one thing, their dendrites are not merely passive conductors but they have voltage gated channels that, like the axon, allow them to create moving spikes, [SSH16]. These can propagate either from synapses to the soma or, retroactively, from soma to synapses. In addition, they have special receptors on their basal dendrites, the *NMDA receptors*, that detect coincidence between arrival of new excitation and prior depolarization of the same part of the dendrite. These can depolarize part of the dendrite for periods of 100 milliseconds or more, known as NMDA plateaus or spikes, [AZM+10].

One hypothesis, the "Two Layer Model", is that the various branches of its dendritic tree are each doing some first stage of a computation and then, in a second stage, the cell as a whole combines these in some fashion, see Bartlett Mel's paper [Mel16]. But there is no consensus model for this yet, only suggestive bits and pieces. Another hypothesis is that, at any given time, some branches of the tree may be activated in such a way that its depolarization creates spikes in the dendrite that carry their responses to the soma, while

other branches are silenced. This amounts to a set of gates on the branches, allowing the cell to compute quite different things depending on which branches are activated. Finally, when the cell fires, emitting a spike on its axon, it can also generate a back propagating spike in the dendrites, altering their subsequent activity perhaps in some context specific way.

It is tempting to seek a transformer-like algorithm that uses all this machinery. However we need to face one way in which the mechanisms of computers and brains will never converge: signals in the brain are trains of spikes, not real numbers. It is true that the membrane potential of a neuron is a real number (in fact, a real-valued function along each dendrite and axon) but the cell's output is a stereotyped spike, always identical. What varies between spike trains is the *timing* of the individual spikes. Brains have no central clock and many modelers have speculated that precise spike timings, especially synchronous spikes, are integral parts of the ongoing cortical computation. This could allow spike trains to carry much more information than merely its spike count.

The essential idea of a silicon transformer is to seek ways in which the signal $\vec{x}$ being analyzed has certain definite connections to some part $\vec{y}$ of the context (e.g. some other sensory data or some memory, etc.). Transformers do this by computing products $x \cdot M \cdot y$ for learned low rank matrices M. It's quite conceivable that interlaced synapses, some with NMDA receptors, along basal dendrites of pyramidal cells, could do something similar if they carry synapses for the both the x and the y signals. For example, see Bartlett Mel's paper cited above. The interaction of NMDA receptors versus the conventional (AMPA) receptors may well implement a nonlinear version of $\sum_i x_i y_i$. This might be the basis of a transformer-like mechanism linking *local* neurons, e.g. linking the signals from two words in a heard sentence or from two objects in a scene being viewed.

iv.b: Feedback

However, there is another challenge about which I made speculations 30 years ago [**B-1991, B-1997a**]. It's well established in neuroanatomy that the cortex can be divided into high level and low level areas with processing streams going both "forward", e.g. from the primary sensory areas to association areas as well as "backwards", usually called feedback pathways. These connections are set up by long distance pyramidal axons in specific cortical layers and these have been meticulously worked out. A current diagram of these pathways from the paper [MVC$^+$14] is reproduced below.

My proposal some decades ago was that feedback was connected computationally to Bayes's rule. Naively, the rule by itself could be implemented, for example, if the feedback path carried a vector of prior probabilities of possible high level states that was combined with the locally computed conditional probabilities of the data by a dot product. My proposal was more complicated but whatever high level data is sent to a lower area, this is a natural place for biological versions of transformers. More specifically, I sought an architecture for connecting long term memories like knowledge of the sounds of words or of
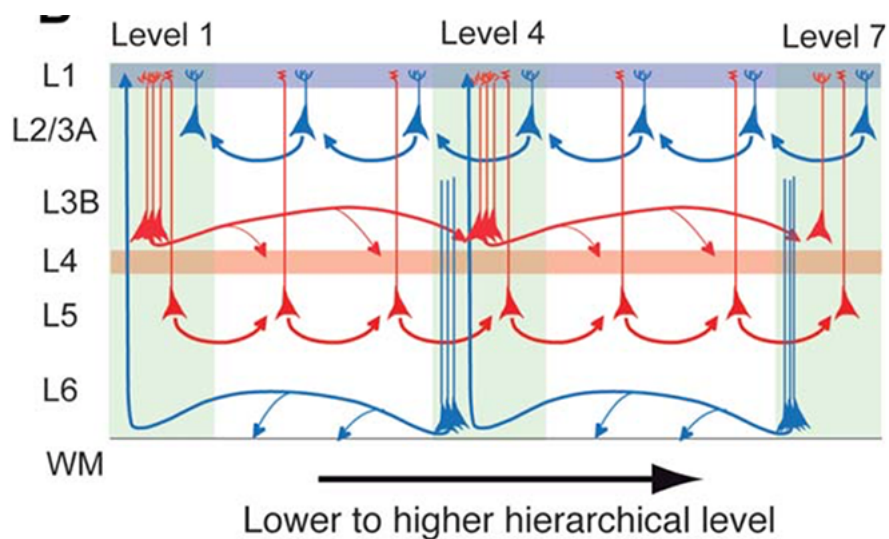
Figure 9.2: The red arrows are feedforward processing involving layers 3B, 4 and 5, while the blue arrows are feedback pathways involving layers 1, 2, 3A and 6. WM is white matter. The triangles are pyramidal cell bodies with the vertical lines indicating their dendrites. From [MVC⁺14], figure 12B, licensed by Creative Commons, J. Comparative Neurology.

the shape of objects, etc. to current sensory data but registering their differences. For all such tasks, we need to relate information stored in a higher cortical area with the current incoming signal in a lower area.

The diagram suggests strongly that the neurons in layers 2/3A and layer 6 are places where transformer-like algorithms can be implemented. Although many pyramidal cells in middle layers have long apical dendrites connecting the soma to layer 1 synapses at the end of feedback pathways, it is hard to see how the sparse signaling along the apical dendrite can allow very much integration of top-down and bottom-up data. But layer 2 pyramidal cells as well as multipolar layer 6 neurons have much more compact dendritic arbors and might do this. Layer 6 feedback is perhaps the strongest candidate as this is less focused, more diffuse than layer 1 feedback (see [MVC⁺14]). I strongly believe that some such mechanism must be used in mammalian cortex and that this is an exciting area for future research.

iv.c: Scaling

If there is one thing human society and human economy teaches you, it is that scaling up any enterprise, any organization by a large factor requires many many adjustments, even radical re-organization. Most things don't just scale up easily. So how is it that the cerebral cortex of mammalian brains scales from a mouse with about 13 million cortical neurons

to a human with about 16 billion cortical neurons, more than 3 orders of magnitude, with hardly any changes at all? The architecture of mammalian brains is totally different from those of birds and reptiles, so different that there are no universally agreed homologies between them. The mammalian neocortex just appeared from nowhere, apparently in its full blown form, having in all species the same pyramidal cells, the same 6 layers, the same basic areas and the same links to thalamus, etc. But once formed, almost all mammalian cerebral cortices seem essentially identical *except for size.* OK, the human brain has a uniquely large prefrontal lobe but this requires no major rewiring as well as a small set of peculiar "von Economo cells" and whale brains are an exception, simplifying their layer organization. But whatever algorithm makes mice smart seems to be the same thing that works for us humans too.

A very simple observation, but one that I think is fundamental, is that present day AI, in both its functioning and its training, seems to have the same remarkable resilience to scaling. I like to demonstrate in my lectures the way neural nets work with a "Mickey Mouse" example of a neural net with only 12 weights that learns nearly perfectly in front of the live audience to discriminate points in the plane inside a circle from those outside, using simple gradient descent. OpenAI's most recent language program GPT-3 is based on the same ideas as BERT but has 175 *billion* weights and is trained by the same old gradient descent. Who would have expected that such a scaling was possible? The fact that simple minded gradient descent continues to work is astonishing. Yes, there are a few tricks like dropout, pre-training, etc. and OpenAI and Google have the best programmers tuning it up but it is basically still just gradient descent on very similar architectures.

v.   What is missing?

Although on some problems with some measures, the so-called "leader-board" shows AI programs approaching or even surpassing human skills, there are many ways in which they still fall far short of human skills. For example, GPT-3 when asked how many eyes your foot has, said your foot has two eyes. I guess they didn't train it on the classic folk song "Dem dry bones" or it might have had a little better anatomical knowledge under its "belt".

v.a: Vision problems

More significantly, the idea of transformers are only beginning to make any significant headway in computer vision where the central problem is segmenting images into objects and then identifying the objects in possibly cluttered images. Several teams have attacked vision with transformers, calling this self-attention, [BZV+19, LLC+22]. Their approach is to start with a convolutional pyramid-style neural net (called a CNN) that uses translation-invariant weights (to deal with the very large number of pixels) and gradually reducing the image size by using units representing whole windows in the image by a vector of

values. This is followed by the same transformer architecture as the linguistic programs, but replacing words by the representation of the windows calculated by the CNN, sentences by the whole image. Self-attention seeks useful attention links from a vector representing one window to the similar vector at some other window. This means transformers may need to link any pair of windows, a huge challenge even for GPUs. Indeed a fundamental issue with all vision computations is handling the large size of data of a single image: any recognizable image needs a lot of pixels. I have argued that because of the size of image data, animal cortices can only afford to have one set of neurons that keep full resolution. In other words, V1 must be the only high resolution buffer in which to do things that need accuracy (like comparing the proportions of two faces). Be this as it may, one should note that both the programs using transformers and conventional neural nets are already doing tremendously better than pre-2010 algorithms using hand designed filters instead using neural net learned filters. For example, the papers [LDG+17, LMW+22], built on a *Convolutional Pyramid Network* (without transformers) outperforms all hand engineered programs on the so-called COCO benchmark and does combine all pyramid levels in one master representation.

My sense is that understanding static 2D images, without the stereo 2 eyes give us or any motion data (for us, pre-computed in the retina), is a really tough skill to master. Dogs only rarely recognize the content of a photo, e.g. most dogs don't recognize photos of their masters, but they recognize dogs on TV and can crash through the cluttered woods at top speed. It's important to realize that human babies as well as dogs learn vision in a moving world (and also making use of the tectum, the reptilian brain stem visual structure). When either you or the perceived object moves, objects at different distances shift relative to each other and this makes it easy to separate figure and background. Further motion reveals their 3D shape. I suggest that transformers will solve vision problems better when trained on movies or from robots moving around, equipped with cameras. More data makes the task easier. Actually, babies start off in the cradle learning hand-eye coordination, using both external and internally generated motion. And they have stereo vision as well which amounts to seeing everything from two places, separated by a small movement. Hand-eye coordination is very similar to the challenge of driving autonomous vehicles: vision and motor control must be integrated. Both should be learnable by transformers and I'm sure this is being implemented somewhere even now.

v.b: General AI

To analyze the next steps towards "general AI", let's consider the following model for the child's acquiring basic knowledge of the world around it. Starting with raw sensory input, the infant sees/hears/feels many confusing patterns, "one great blooming, buzzing confusion" as William James famously put it. But it soon recognizes simple recurring patterns. And then it sees patterns among the patterns, co-occurences, and learns to recognize larger more complex patterns. This leads to a tree in which various bits that

have been linked are *reified* into a higher level concept. As it goes on, the resulting tree is very much like the parse trees in conventional grammar. Each new step results in learning what my colleague Stuart Geman calls "reusable parts." It frequently happens that the pattern found in one context also occurs in a second quite different context. It is well established that in language acquisition, there are definite steps when the child acquires a new rule or concept and suddenly is able to apply it to new situations. This can be syntactical like seeing that most English verbs have a past tense formed by adding "ed" (love/loved) in contrast to a few very common exceptions ("see/saw"). A new word may be learned after only hearing it spoken *once*. Or it may be discovering a semantic class along with the word for the class, e.g. "car." This process of growing your cognitive framework, often in discrete steps, continues your whole life. Human brains do not even get fully connected until adolescence when the long distance axons that connect the most distant cortical areas are fully activated (*myelinated* is the technical term).

Much of this learning is already being done with neural nets. Really complex neural nets are being trained in stages. They may start with a net trained to answer simple low level questions about the data. Then layers are added that use the representations formed in the first net and are trained with more complicated questions. But suppose things computed in the higher layers suggest a modification to activity in the lower layers? In animal cortex, there is always feedback from the higher areas to which the lower areas project. This suggests that a new kind of transformer is needed for this, something with queries and values in the original net and keys in the new higher layers. This creates circular computations and raises an issue of timing. However, this is a mechanism known to occur in the brain.

Another example is a robot learning hand-eye coordination. In humans, the infant connects efferent muscle signal patterns with afferent retinal stimuli, but this is a complex relationship and needs to be learned in order to coordinate activity in the corresponding visual and motor parts of the cortex. The robot may have a pretrained visual program and a pretrained motion program but now it needs to join them together with transformers that pick out aspects of each representation that the other needs to use. It needs to learn what muscle commands lead to what visual stimulus, and more, to merge the representations of space both nets have formed.

In general, *distinct neural nets need some way to merge*, to train a larger net containing them both. As in the hand-eye situation, there may well be that some concepts implicit in the distributed representations of both neural nets, but how would the nets "know" that they have hit on the same reusable idea? Connecting two neural nets should certainly not need starting from scratch and relearning each set of weights. One needs instead to add new layers and transformers to create a larger net on top of the two others. I think this is an ideal task for a second generation of transformers, layered on top of the two pre-trained nets. The queries are in net #1, the keys in net #2 (or vice versa) and the training involves tasks where both nets need to work together. In terms of graphs, a parse tree or AND/OR graph should be present implicitly in the representations of the two nets and these new

transformers should find the common nodes leading to the creation of a larger, but still implicit, graph for the merged net.

The issue of feedback is central in the task of comparing memories with current stimulus. At every instant, you are usually experiencing a new configuration of events related to various old events and you merge memory traces as much as possible with the new sensory input and new situation, a process that appears to be mediated by feedback between neurons as we discussed above. For example, everyone has an inventory of known faces, e.g. the faces of your family, friends and co-workers. They are likely stored in the fusiform face area (FFA) or adjacent areas of inferior temporal cortex. When you see them again, you must perform some sort of matching before you can say you recognize them. As shapes, sizes, relative positions are involved, my own belief is that V1 must play a role via feedback, all the way from FFA or IT. But in all cases, the memories will not match the new stimulus exactly: there will always be changes, they will not be exact repeats. You must notice the changes in order to understand best what's happening now. This role of feedback – noticing the differences – was central in my papers referred to above. Is this needed and, if so, will transformers be needed for this?

All of this suggests that to reach general AI, neural nets will need to have something like memory in higher levels and feedback to lower levels, to be more modular and to have structures specific for both feedforward and feedback data exchange. So far, neural nets have been only a little modular. BERT has two pieces, the encoder and the decoder, and recent segmentation algorithms have more, some even looking a bit like analogs of the distinct mammalian visual areas V1, V2, V4. How many would be needed if general AI is achieved?, a big question. Finally, cortical architecture has a very specific architecture with the hippocampus acting like the highest layer, storing current memories for variable periods but eventually downloading some of them into appropriate cortical areas, forgetting many others. Should AI's imitate this if they seek human level skills?

Finally, I want to add that I await, not without some trepidation, the day when a robot is trained to see, hear, move and is turned loose, probably in a lab, *indefinitely*. It would have a charging station and its computations would be done externally in a major computer, and its task would be interacting with humans, understanding what motivates them and learning to help them.

v.c: The Whorfian hypothesis

I want to go back to *word2vec* where the idea was that high dimensional vectors are better carriers of linguistic data than the discrete tokens called words. There is one school of thought that asserts the opposite: that words are what has enabled humans to think so well and that, as a result, the way you conceptualize something mimics how your language expresses it. This is the so-called *Whorfian Hypothesis*, named after the linguist and engineer Benjamin Whorf who developed this idea together with Edmund Sapir. To some extent, this feels right, that it expresses well the content of consciousness. And yet, often

words just come out of your mouth unconsciously, without any reflection, any sense of your having had a choice. Your consciousness then looks like a supervisor watching what emerges from the unseen machines grinding away below. This is the model Stanislav Dehaene propounds in his book *Consciousness and the Brain*[Deh14]. In other words, we can understand thought either as manipulating word tokens à la Whorf, or, alternatively, think of the words as a gloss your consciousness puts on the output of a vast set of firing neurons, a sort of executive summary, à la Dehaene. From a computational perspective, this is simply the choice between discrete token-based representations and distributed real vector representations.

My own belief is that distributed representations are here to stay. I see no reason why we need single neurons or single neural net units that learn to respond to unique features of an ongoing thought process. Yes, we need outputs made of discrete signals. In brains, it seems that Broca's area processes a distributed representation of a thought into a grammatical utterance made from a sequence of words; and Google's BERT has a decoder half that outputs a sentence, retrieving the word tokens, so-to-speak, at the last minute. This is all about as far from Chomsky's Universal Grammar and from the Whorf-Sapir theory as it could be. It asserts that we know the grammar of our mother tongue not by rules but by endless experiences of its usage and nuances and by playing the game of sometimes using precise rules, sometimes ignoring them. It has emerged in the last few decades how much cortical activity is unconscious, how little makes its way into consciousness. Maybe what we are conscious of is the output of a decoder, like that in BERT, and is more token-like while the unconscious stuff are all embodied in distributed representations.

What is very disconcerting is that, if thinking must use distributed representations, then all future AI machines will be hard to impossible to understand, to know why "it", the machine, has concluded something. Indeed, we are truly living in a "Brave new world" with wonders aplenty.

ADDED IN PROOFS: I am lucky to have just discovered the course given in 2023 by Stanislas Dehaene at the Collège de France, [Deh23]. Part of his course concerns new discoveries about "Face Cells" in Inferior Temporal Cortex of macaque monkeys (lecture on Jan.13). They were first discovered by Charlie Gross in the early 70s, [MD19], a friend of mine as well as a remarkable neuroscientist who wasn't afraid to buck the prevailing tides. Recent work now shows that many neurons in this area have very specific focus on qualities of the face, e.g. gender, hair length, skin color, age, smiling vs. angry, angle viewed, [FTL09, HCL$^+$21]. In other words, some neurons do focus on very significant specific qualities of the stimulus. What this suggests is that the content of "thoughts" may not be entirely hidden in high dimensional representations but may be, to some extent, accessible in single cell recordings! Moreover, by the use of another neural net architecture, known as β-variational auto-encoding, in which the data is forced to pass through a low dimensional "bottleneck" layer, AI's have been designed that replicate the actual neuronal recordings. Another breakthrough described in Dehaene's lectures is a model for "one-shot

learning", e.g. the ability of children to learn a new word from one encounter.

Chapter 10

Does/Can Human Consciousness exist in Animals and Robots?

Human consciousness is the thing that starts up in each of us when, as in Iris DeMent's song "Let the Mystery Be," we "come from" a place that no one knows and leaves us when "the whole thing's done." Many people have sought theories of consciousness. I recently read an op-ed piece in the New York Times that I especially liked. Instead of the antiseptic word "consciousness," the author, Sean Kelley, calls his piece "Waking up to the Gift of Aliveness." The article is a commentary on the sentence "The goal of life, for Pascal, is not happiness, peace, or fulfillment, but aliveness" that he traces in some form to his teacher Hubert Dreyfus. He confesses that he knows no definition of aliveness but gives us two examples: looking at your lover's face when you have fallen in love; and lecturing to a class (he is a Professor) when your students are truly engaged and the classroom is buzzing. I take it that aliveness should be thought of as the most fully realized states of consciousness. While consciousness is the substrate of everything we do when we are alive in the mundane sense, the aliveness he is talking about is found in its most real moments, when all of life feels like it makes sense. He says aliveness should have the passion of Casanova without his inconstancy and the routine of Kant without his monotony. I'd like to think this is also the state of an enlightened Buddhist during meditation. And for me, I think this was how I have felt sailing, when the physical, the mental and the emotional strands of life all wove together.

This chapter has 5 sections. The first reviews what neuroscientists are saying. The second discusses the evidence for consciousness in animals, from bacteria to primates. The third is a digression on *emotions* which seem to me central to consciousness and the hardest to incorporate in robots. The fourth looks at what physics says about consciousness. Finally, I try to pull things together.[1]

---

[1]This chapter is based on three blog posts and one published paper: "Let the mystery be," April 13, 2018; "Can an artificial intelligence machine be conscious?," April 11, 2019; "Can an artificial intelligence

i.   What do neuroscientists say about consciousness?

Now science has had a real problem here: for a long time, even the word consciousness was taboo to practicing scientists. When I was a student, psychology had been overtaken by behaviorists and biology was being reduced to biochemistry. In this atmosphere, the mind/body problem had been left to philosophers (and a few quantum physicists – see below). The first time I encountered the taboo breaking was when I read the neuroscientist John Eccles' 1977 book, joint with the philosopher Karl Popper, entitled *The Self and its Brain*, [EP77]. Both Popper and Eccles are believers in a *Three World* view of reality: (I) the objective physical world, (II) the inner world of conscious beings and (III) the world of ideas, that is objects of thought. Concerning the first two, they sought a detailed model of how in particular the physical brain interacts with conscious experience. Eccles developed their ideas further in his 1990 paper [Ecc90]. His hypothesis is first that the cerebral cortex can be broken up into about 40 million columnar clusters, each made up of about 100 pyramidal cells which stretch from near the inner to the outer cortical surface, clusters that he calls *dendrons*. Secondly, each dendron "interfaces" with a corresponding unit of conscious thought that he calls a *psychon* via an interaction allowed on the physical side by quantum uncertainty. A figure from his Royal Society paper is reproduced in Figure 1. This is a breathtakingly bold and precise answer to the mind/body problem but one that has not drawn many adherents,
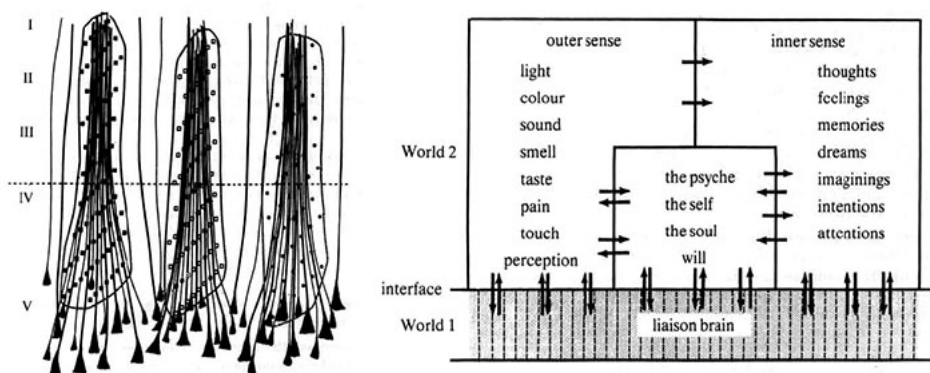
Figure 10.1: Eccles theory of the mind/body problem: left, his *dendrons*, right, how cortex interfaces with conscious thoughts, both from [Ecc90], figures 13 and 1 respectively, by permission of the Royal Society London.

More recently, scientists realized they could study *access consciousness*, that is the stuff that people report they are thinking, as opposed to consciousness as the ineffable,

machine be conscious, part II?," July 12, 2019; and "Thoughts on Consciousness," Journal of Cognitive Psychology, 2019 **E-2019**.

subjective sense of being alive, and then consciousness became something on which they could do experiments. Of course, they now exclude things like the reportedly heightened consciousness of Buddhists deep in meditation when all distracting thoughts of that involve the rest of the world are put aside. The goal of this research is to elucidate the *neural correlates of consciousness* with the aid of tools like fMRI (functional magnetic resonance imaging), i.e. can one identify the large-scale neural states in which a person will report being conscious of something. It turns out that this is not as simple as one might hope: there are many sensations that cause measurable activity in primary sensory and other cortical areas that people are consciously unaware of and there are neat experimental ways of producing them such as binocular rivalry, masking, and attentional distraction. Even the order in which two sensations occur can be experienced consciously as the opposite of what actually happened. Strikingly, the conscious decision to do an action seems to occur *after* there is brain activity initiating the action. Moreover a certain amount even of reasoning can also be accomplished quite unconsciously by the brain. Freud would have told them that even strong emotions and actions resulting from these emotions often do not reach consciousness – but his work was another taboo to scientists. The limitations of the self-awareness that consciousness provides were clearly summarized in Alex Rosenberg's NY Times piece *Why you don't know your own mind*, [Ros16]. His conclusion is "Our access to our own thoughts is just as indirect and fallible as our access to the thoughts of other people. We have no privileged access to our own minds."

What does make things conscious, according to many neuroscientists, is that the activity expressing some thought should spread over large parts of the brain, an idea known as the *global workplace theory* of consciousness. Almost the exact opposite of Eccles' theory, this proposes that activity over large parts of the cortex, often synchronized via 40 Hertz brainwaves (so called gamma oscillations) and where many parts of the cortex contribute to the full thought, is necessary and sufficient for the thought to be conscious. By every area, we mean the primary sensory areas can be involved but also perhaps the pre-frontal cortex and the so called association areas of parietal cortex have to be involved. I recommend Stanislav Dehaene's book *Consciousness and the Brain* [Deh14] for a detailed description of this theory. Dehaene writes:

> *Consciousness is like the spokesperson in a large institution. Vast organizations such as the FBI, with their thousands of employees, always possess considerably more knowledge than any single individual can ever grasp. ... As a large-scale institution with a staff of a hundred billion neurons, the brain must rely on a similar briefing mechanism. The function of consciousness may be to simplify perception by drafting a summary of the current environment before voicing it out loud, in a coherent manner, to all other areas involved in memory, decision, and action.(op.cit. p.99-100)*

Access consciousness cannot be used to interrogate animals without speech and, in any case, it hardly captures the full experience of aliveness. Moreover, speech is not all that
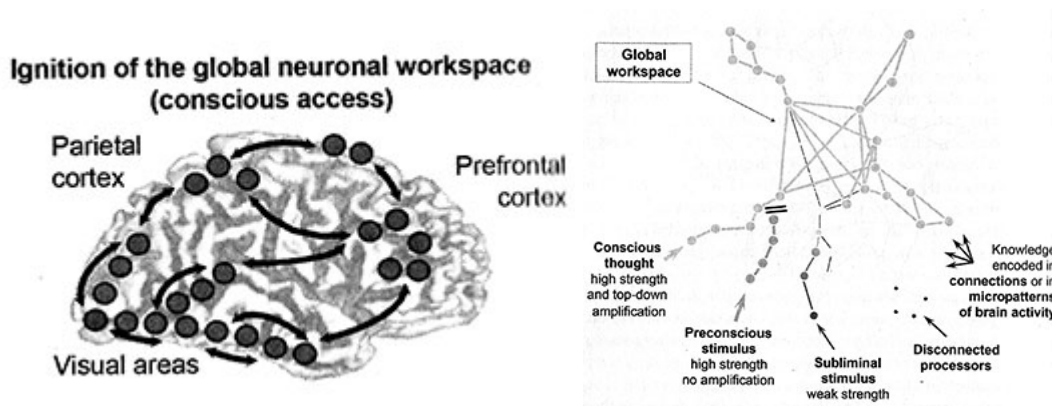
Figure 10.2: Diagrams of the global workplace theory: left: "Ignition of the global neuronal workspace," right: a diagrammatic version, both from [Deh14], figures 27,28, by permission Stanislav Dehaene.

reliable an indicator that we are in touch with another sentient being anyway: computers have occasionally been able to pass the Turing test and fool observers into thinking a real person is talking to them over a phone. So speech alone is an unreliable token of real consciousness. What can a scientist use for assessing consciousness in mute creatures? The main experimental tool in testing monkeys has been to train them to respond to a stimulus in different ways, e.g. by pressing various buttons, assuming that producing such a response means that the stimulus has activated something we can call their consciousness. In this way, a whole body of research has confirmed that consciousness in monkeys follows patterns similar to that in humans. For example, some stimuli do not reach consciousness and when they do, large parts of the monkey's neocortex show activity, often synchronized by gamma waves.

But a third theory of the cortical locus of human consciousness goes back to Wilder Penfield's operations on patients with intractable epilepsy. To locate the exact cortical area whose excision would cure the epilepsy, he operated with local anesthesia and interrogated his awake patients while stimulating their exposed cortices on the operating table. In this way, he almost always found some area where the trigger for the epilepsy was located. But one form of epilepsy, *absence epilepsy* in which the patient briefly looses consciousness without any other symptoms, did not correspond to any unusual cortical electrical activity. This led him to propose that consciousness is related not to the neocortex but instead to activity in the *midbrain*. This theory has been extended by Bjorn Merker (see his paper *Consciousness without a cerebral cortex*, [Mer07], who filmed and worked notably with hydranencephalic children, children born with no neocortex (though the paleocortex and thalamus are usually preserved). His claim is that when given full loving custodial care, and within the limits imposed by their many weaknesses, they exhibit behavior much like

normal children. See Figure 3 for diagrams of Penfield's excisions and for the location of the midbrain. For me, however, a really stunning piece of evidence for this theory was a sort of "Turing test" of this issue was that carried out by Jaak Panksepp (see Panksepp's commentary to Merker, in [Mer07], pp.102-103). He surgically removed the neocortex in 16 baby rats, paired them with normal rats and asked 16 of his students to each watch one such pair play and guess which rat was intact, and which lacked their neocortex. Only 25% of normals were correctly identified, while the decorticates were judged to be the normals 75% of the time! It seems the major role of the neocortex was to make rats that possessed one more cautious, leaving the decorticates more playful. The bottom line is that, if you subscribe to the midbrain location hypothesis, one ought to ascribe some form of consciousness to all vertebrates.
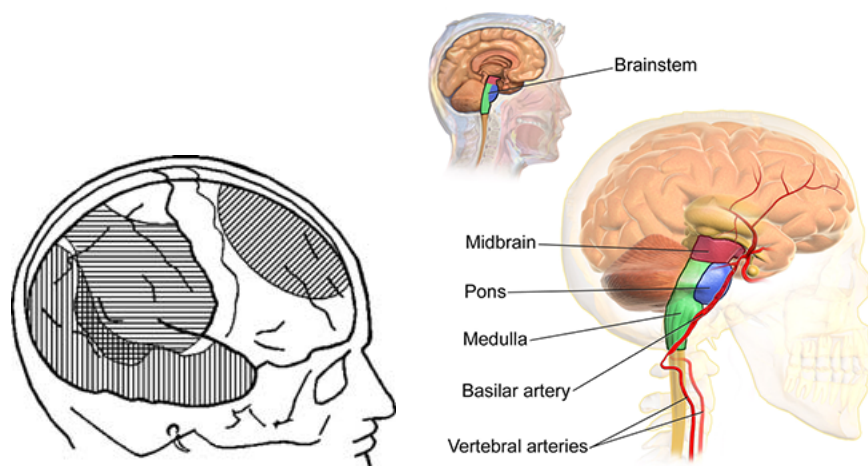
Figure 10.3: Left: Large cortical excisions performed by Penfield for the control of intractable epilepsy in three patients. In no case was the removal of cortical tissue accompanied by a loss of consciousness, even as it took place. From [Mer07],p.65, by permission of Cambridge University Press. Right: a diagram showing the location of the midbrain or mesencephalon [Bla14], Creative Commons, Wikijournal of Medicine.

ii. Consciousness in animals

If we seek a scientific theory of consciousness, we must first face squarely the question of whether and/or what animals have consciousness. Let me start by saying to my reader: *I believe that you, my friend, have consciousness.* Except for screwy solipsists, we all accept that "inside" every fellow human's head, consciousness resides that is not unlike one's own consciousness. But in truth, we have no hard evidence for this besides our empathy. So should we use empathy and extend the belief of consciousness to animals? Arguably, people

with pets like dogs and cats will definitely insist that their pet has consciousness. Why? For one thing, they see behavior that is immediately understood as resulting from similar emotions to ones that they themselves have. They find it ridiculous when ethologists would rather say an animal is displaying "predator avoidance" than say it "feels fear." They don't find it anthropomorphic to say their pet "feels fear," they find it common sense and believe that their pet not only has feelings, but also consciousness. Our language in talking about these issues is not very helpful. Consider the string of words: <u>emotion</u>, <u>feeling</u>, <u>awareness</u>, <u>consciousness</u>. Note the phrases: we "feel emotions," we are "aware of our feelings," we say we possess "conscious awareness," phrases that link each consecutive pair of words in this string. In other words, standard English phrases link all these concepts and make sloppy thinking all too easy. But to clarify, for me an emotion is a kind of feeling and every feeling is part of our consciousness and awareness is a synonym for consciousness. One also needs to cautious: in our digital age, some lonely elderly people are being given quite primitive robots or screen avatars as companions and such patients find it easy to mistakenly ascribe true feelings to these digital artifacts. So it's tempting to say we simply don't know whether non-human animals feel anything or whether they are conscious. Or we might hedge our bets and admit that they have feelings but draw the line at their having consciousness. But either way, this is a stance that one neuroscientist, Jaak Panksepp, derides as *terminal agnosticism*, closing off discussion on a question that ought to have an answer.

All mammals have virtually identical brains, differing only in the size of its constituent parts. Thus human brains are distinguished by having a greatly enlarged pre-frontal cortex that appears to endow us with greatly increased planning activity and skills. Given the extensive organ-by-organ homology of all mammalian brains, I see no reason to doubt that all mammals experience the same basic emotions that we do, although perhaps not so great a range of secondary emotions. And if we all share similar emotions, then there is just as much reason to ascribe consciousness to them as there is to ascribe consciousness to our fellow humans. This is a perfect instance of "Occam's Razor": it is by far the simplest hypothesis that explains the data.

Going beyond mammals, it is useful to review the various stages of life, both living today and reconstructed from fossils, with a view to their potential for consciousness. I am inspired in doing this by the book *Other Minds: the Octopus, the Sea and the Deep Origins of Consciousness* by the philosopher and diver, Peter Godfrey-Smith, [GS16]. At the base of the tree of life, we have two superficially similar kingdoms, the Bacteria and the Archaea. Both are prokaryotes, that is, are simple cells without nuclei, mitochondria, ribosomes or other organelles. On the other hand, both already possess proteins from the majority of protein families, as well as the universal genetic code (implemented by the same set of tRNA molecules) and, very significantly, they use the same complex electro-chemical mechanism as all higher life to synthesize ATP, their energy storage molecule. This mechanism uses ion pumps that make the cell membrane into a capacitor, the same mechanism that is used in higher animals as *the key to information transmission in nervous systems* (vividly described in Nick Lane's book, *The Vital Question*, [Lan15]). These simplest forms of life

also sense their environment chemically via channels in their membranes and most can move in various directions using their flagella, thus reacting and seeking better environments. This is the beginning, a primitive form of *sentience* that started up c. 3.5 bya (billion years ago). Although I personally prefer to be agnostic, it is perfectly possible that a mite of consciousness resides in these cells.

The next step was the formation of much much bigger, more complex single celled organisms, the *eukaryotes* c. 2 bya. It is hypothesized that they started from an archaeon swallowing a bacterium, the bacterium becoming the mitochondrion in this new organism and, by folding its membrane again and again, hugely expanded the cell's ATP factory, hence its available energy. Its skills sensing and moving got significantly better but I'm not aware of any change that might have brought it closer to consciousness. But after that, around 0.65 bya (or 650 mya), multi-cellular animals formed. These were larger and obviously needed significantly better coordination, better senses and better locomotion. It is believed that the first nervous systems arose almost immediately to coordinate the now complex organisms. These creatures were soft and left no fossils but modern day jellyfish and sponges may be similar to organisms of that time. Sponges do not have nervous systems but jellyfish (and comb jellies) do and are the simplest organisms with nervous systems today. The environment is described as a mat of microbial muck covering the bottom of a shallow sea over which jellyfish like creatures grazed. Anyone for consciousness in this world?

The world becomes much more recognizable with the advent of predation, bigger animals eating smaller ones and all growing shells for protection, all this in the Cambrian age 540-485 mya. Now we find the earliest vertebrates with a spinal cord. But we also find the first arthropods with external skeletons and the first cephalopods, predators in the phylum mollusca who grew a ring of tentacles and who, at that time, had long conical shells (Figure 4 has an image of a reconstruction of the cephalopod *Orthoceras* from the following Ordovician age).

Figure 10.4: A reconstruction of the cephalopod Orthoceras that lived in the Ordovician era, c.370 mya, from Wikimedia Commons, Nobu Tamura.

In all three groups, there are serious arguments for consciousness. One approach is

based on asking what animals *feel pain* and that feeling pain implies consciousness. There are experiments in which injured fish have been shown to be drawn to locations where there is a pain killer in the water, even if this location was previously avoided for other reasons. And one can test when animals seek to protect or groom injured parts of their bodies: some crabs indeed do this whereas insects don't. (See Godfrey-Smith's book, pp. 93-95 and references in his notes). Unfortunately, this raises issues with boiling lobsters alive, an activity common to all New Englanders like myself. Damn. Another approach is the mirror test – does the animal touch its own body in a place where its mirror image shows something unusual. Amazingly, some ants have been reported to pass the mirror test, scratching themselves to remove a blue dot that they saw on their bodies in a mirror, see Cammaerts & Cammaerts' remarkable paper [CC15], and Figure 5. As this paper notes, firstly ants are very social animals and secondly, their initial reaction to seeing themselves in a mirror seems to be puzzlement, even touching their reflection with their mouth parts. Yet, somehow, eventually, they do try to clean off the blue spot seen only in the mirror!

Figure 10.5: Left, an ant sees itself in a mirror with an unexpected blue dot on its "clypeus" (located where a nose would be). Right, an ant attempts to clean off the blue dot with its right antenna after seeing itself in the mirror, both from a lecture by and by permission of M. C. Cammaerts.

With octopuses, we find animals with brain size and behavior similar to that of dogs. Godfrey-Smith quotes the second century Roman naturalist Claudius Aelianus as saying *"Mischief and craft are plainly seen to be characteristic of (the octopus."* Indeed, they are highly intelligent and enjoy interacting and playing games with people and toys. I knew the famous neuroscientist Jerry Lettvin who worked with octopuses in Naples and (personal communication) was convinced that they were conscious beings and loved playing practical jokes on him. This has been confirmed by many observers. It seems they enjoy immensely playing with human toys. A beautiful book, *The Soul of an Octopus: A Surprising Exploration into the Wonder of Consciousness* by Sy Montgomery, [Mon15], develops this thesis drawing on extensive personal interactions (or should I say 'relationships') with octopuses.

See also this wonderful lecture by Montgomery on her experiences in the same bibliography entry. They know and recognize individual humans by their actions, even in identical wetsuits. As for neurology, their brains have roughly the same number of neurons as a dog, though, instead of a cerebellum to coordinate complex actions, they have large parts of their brains in each tentacle. This is not unlike how humans use their cerebral cortex in a supervisory role, letting the cerebellum and basal ganglia take over the control of detailed movements and reactions. If you can read both these octopus-related books and not conclude that an octopus has just as much internal life, as much awareness and consciousness as a dog, I'd be surprised. The most important point here is that there is nothing special about vertebrate anatomy, that consciousness seems to have arisen in totally distinct phyla with no common ancestor after the Cambrian age.

Finally, looking at vertebrates, a key point is that all non-mammalian vertebrates have brains which are fairly similar to each other but only similar to the mammalian brain *if you remove the neocortex*. The neocortex has a unique 6-layered structure not found in non-mammals although some recent thinking suggests that its parts are present, just not assembled and wired with pyramidal cells (as in Eccles' theory) as they are in mammals. These parts, called the pallium in birds and just the cerebrum in all classes, the 3-layered paleocortex, especially the hippocampus, as well as the thalamus (sometimes considered as a seventh layer of neocortex) are found in other vertebrates. The class of birds shows that this brain structure can produce great intelligence. Many people are convinced that birds, especially parrots and crows, are conscious beings, every bit as intelligent and responsive as, e.g. dogs and cats. A wonderful review is the book by Jennifer Ackerman, *The Genius of Birds*, [Ack16]. In the video from which Figure 6 top is taken, the parrot uses both its foot and beak together to insert the rod into the hole in the box, then lines it up with the food pellet and rotates the stick to push the pellet off its support! The frame has been modified to make the pellet more visible but whole video is well worth watching. Personally, I find it quite convincing that indeed birds and octopuses as well as mammals have consciousness. But note that while the range 200-2000 million neurons includes octopuses, rats, cats, dogs, crows and owls, humans have 100 *billion* neurons, though only some 20 billion in neocortex.

My personal view, after studying all this, is that the evidence suggests that consciousness is <u>not</u> a simple binary affair where you have it or you don't have it. Rather, it is a matter of degree. This jibes with human experience of levels of sleep and of the effects of many drugs on our subjective state. For example, *versed* is an anesthetic that creates a half conscious/half unconscious state. As our brains get bigger, we certainly acquire more capacity for memories but some degree of memory has been found for example in fruit flies. When the frontal lobe expands, we begin making more and more plans, anticipating and trying to control the future. But even an earthworm anticipates the future a tiny bit: it "knows" that when it pushes ahead, it will feel the pressure of the earth on its head more strongly and that this not because the earth is pushing it backwards, i.e. they anticipated the push back ([GS16], p.83). My personal belief again is that some degree of consciousness is present in all animals with a nervous system. On the other hand, Tolkien and his Ents

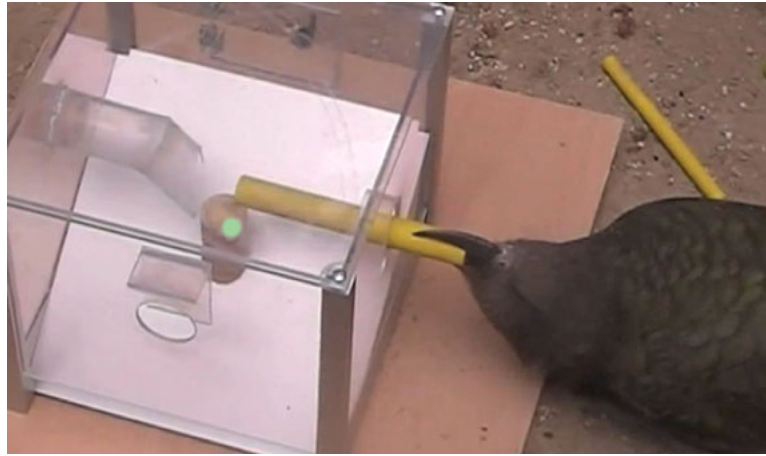Figure 10.6: Top: an intelligent Kea (a New Zealand parrot) uses a tool in a frame from a terrific video in [AvBG$^+$11] on Wikimedia Commons, (modified, see text). The Kea has just pushed the food pellet off its support (see text). Below: an octopus unscrews the top from a jar with food in it (the jar is upside down and one can see the lid clearly). From Creative Commons, thanks for the great shot Matthias Kabel.

notwithstanding, I find it hard to imagine consciousness in a tree. I have read that their roots grow close enough to recognize the biochemical state in their neighbors (e.g. whether the neighbor tree is being attacked by some disease) but it feels overly romantic to call this a conversation between conscious trees.

But returning to the hypotheses of neuroscientists in §1, if there are traces of consciousness in lower animals, then it is likely that consciousness in humans has dual neural location – partly neocortical, partly midbrain. The next section makes a stronger case for this.

iii.   We need Emotions #$@*&!

Intelligence is one important guide to the presence of consciousness. But what is intelligence actually? An essential ingredient of human intelligence is missing in IQ tests: <u>emotions</u>. (For those who did not grow up reading American comic books, the bizarre string of symbols in the heading of this section stands for a sequence of strong swear words, i.e. you'd better not forget emotions damn it.) In many ways, emotions seem more closely connected to consciousness than purely intellectual behavior. Without this, a person, an animal or a robot will never really connect to the humans around them/it. I find it strange that, to my knowledge, almost no computer scientists are endeavoring to model emotions for use by robots. Even the scientific study of the full range of human emotions seems stunted, largely neglected by many disciplines. For example, Frans de Waal, in his recent book *Mama's Last Hug*, [dW19], about animal emotions, says, with regard to both human and animal emotions:

> *We name a couple of emotions, describe their expression and document the circumstances under which they arise but we lack a framework to define them and explore what good they do.*

(Is this possibly the result of the fact that so many of those who go into science and math are on the autistic spectrum?) One psychologist clearly pinpointed the role emotions play in human intelligence. Howard Gardner's classic book *Frames of Mind: The Theory of Multiple Intelligences*, [Gar83], introduces, among a variety of skills, "interpersonal intelligence" (chiefly understanding others' emotions) and "intrapersonal intelligence" (understanding your own). This is now called "emotional intelligence" (EI) by psychologists but, as de Waal said, its study has been marred by the lack of precise definitions. A recent "definition" in Wikipedia's article on the EI is:

> *Emotional intelligence can be defined as the ability to monitor one's own and other people's emotions, to discriminate between different emotions and label them appropriately, and to use emotional information ... to enhance thought and understanding of interpersonal dynamics.*

OK, we can't define it but surely it is clear that possessing high EI is likely the best predictor of a successful career.

The oldest approach to classifying emotional states is due to Hippocrates: *the four humors*, bodily fluids that correlated to four distinct personality types and their characteristic emotions. These were: sanguine (active, social, easy-going), choleric (strong willed, dominant, prone to anger), phlegmatic (passive, avoiding conflict, calm), melancholic (brooding, thoughtful, can be anxious). They are separated along two axes. The first axis is extravert vs. introvert, classically called warm vs. cold with sanguine/choleric being extraverted, phlegmatic/melancholic being introverted. The second axis is relaxed vs. striving, classically called wet vs. dry, sanguine/phlegmatic being relaxed, choleric/melancholic always seeking more.
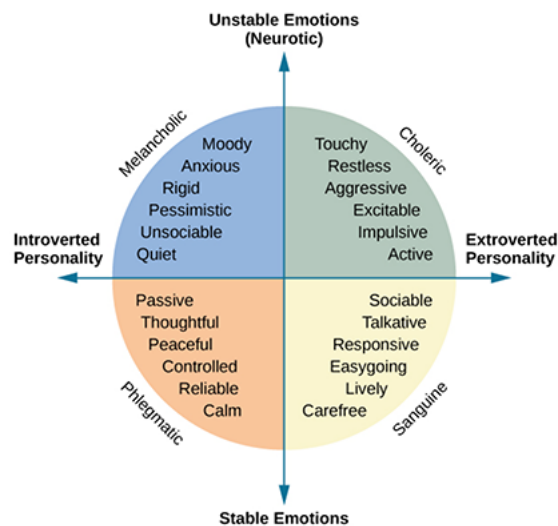
Figure 10.7: Hans Eysenck's colorful version of the 4 humors, licensed under Creative Commons.

The modern study of emotions goes back to Darwin's book *The Expression of the Emotions in Man and Animals*, [Dar72], where he used the facial expressions that accompany emotions in order to make his classification. His theories were extended and made more precise by Paul Ekman and led to the theory that there are six primary emotions each with its distinctive facial expression, Anger, Fear, Happiness, Sadness, Surprise and Disgust and many secondary emotions that are combinations of primary ones, with different degrees of strength.

There really is an open ended list of secondary emotions, e.g. shame, guilt, gratitude, forgiveness, revenge, pride, envy, trust, hope, regret, loneliness, frustration, excitement, embarrassment, disappointment, indignation, admiration, jealousy, empathy, etc., etc. which don't seem to be just blends but rather grafts of emotions onto social situations with mul-
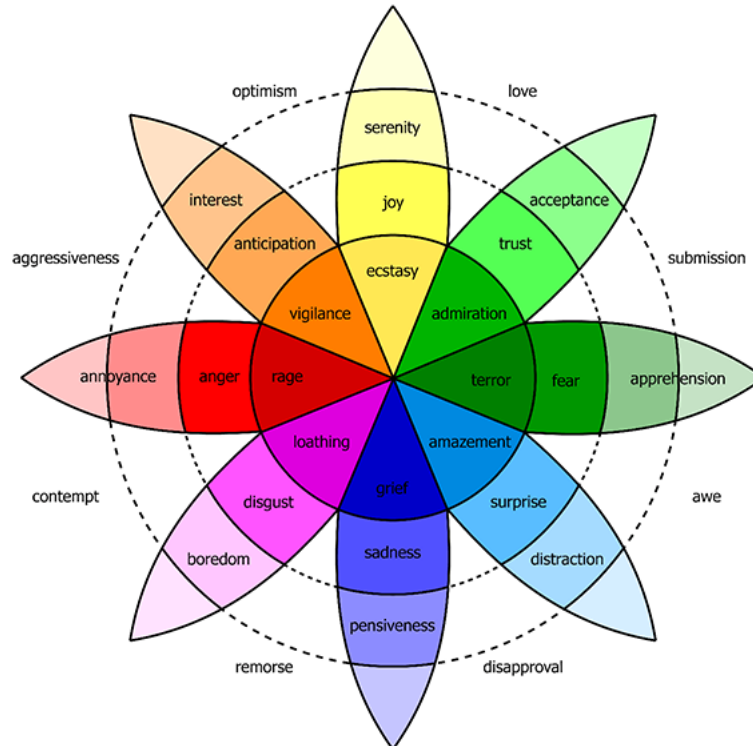
Figure 10.8: Robert Plutchik has extended Ekman's list to eight primary emotions and named weaker and stronger variants and some combinations, resulting in this startling and colorful diagram, from Wikimedia Commons, credit CaptainCyboorg.

tiple agents and factors intertwined. In the last few decades animal emotions have been studied in amazing detail through endless hours of patient observation as well as testing. Both Frans de Waal's book referred to above and Jaak Panksepp's books, [Pan04, PB04], the latter with Lucy Biven, detail an incredible variety of emotional behavior, in species ranging from chimpanzees to rats and including not just primary emotions but some of the above secondary emotions (for instance, shame and pride in chimps and dogs). Panksepp and collaborators have shown that young rats are ticklish and show the same reactions as human babies when their bellies are tickled (see [PB04], p.367). For me, these books and many others and, of course, my own meagre experiences with owning dogs, chickens and pigs, and with watching zoo animals makes a totally convincing case for animal emotions. Frans de Waal book (p.85) defines emotions by:

> *An emotion is a temporary state brought about by external stimuli relevant to the organism, It is marked by specific changes in body and mind – brain, hormones, muscles, viscera, heart, alertness etc. Which emotion is being trig-*

> *gered can be inferred by the situation in which the organism finds itself as well as from its behavioral changes and expressions.*

A quite different approach has been developed by Panksepp and Biven in [PB04].. Instead of starting from facial expressions, his approach is closer to the Greek humors. Panksepp for a long time has been seeking patterns of brain activity, especially sub-cortical midbrain activity and the different neuro-transmitters sent to higher areas, that lead to distinct on-going affective states and their corresponding activity patterns. Their list is quite different from Darwin's though partially overlapping. They identify 7 *primary affective states*:

1. seeking/exploring

2. angry

3. fearful/anxious

4. caring/loving

5. sad/distressed

6. playing/joyful

7. lusting

An aside: I am not clear why he does not add an 8th affective state: *pain.* Although not usually termed an emotion, it is certainly an affective state of mind with sub-cortical roots, a uniquely nasty feeling and something triggering specific behaviors as well as causing specific facial expressions and bodily reactions. They go further in Chapter 11 to propose that one specific midbrain area, the *periaqueductal gray* (PAG) (possibly together with its neighbors, the ventral tegmental area and the mesencephalic locomotor region) coordinates all the above affective states and gives rise to what they call *core self* or consciousness.

Yet another very influential classification is the work of Jonathan Haidt [Hai12] on *moral emotions.* Starting from the observation that moral judgements are arguably more emotional than the result of rational thought, he has gone on to separate 5 axes of moral vs. immoral behavior whose relative power varies strongly from individual to individual. These are:

1. Caring vs. Harming

2. Fairness vs. Cheating

3. Loyalty vs. Betrayal

4. Authority vs. Subversion

5. Sanctity vs. Degradation

Violation of any of these precepts causes outrage in many individuals. But incorporating these into robots is a key issue in what is called "alignment", that is ensuring that the robot's aims are aligned to human aims. The possible terrible consequences of misalignment are vividly illustrated by Goethe's 1797 poem "The Sorcerer's Apprentice" ("Der Zauberlehrling") so chillingly portrayed with Stravinsky's music in the Disney film "Fantasia". Computer scientists are well advised to heed Haidt's analysis. Here we see how tightly patterns of social behavior and their emotional drivers are tied together.

What I think is completely clear after all this research is that all mammals share the same basic repertoire of emotions and that this is a key component of both their intelligence and their consciousness. But how about robots? An excellent way to probe how human emotions may be mimicked by robots is to see what novelists have to say! In Ian McEwan's latest book *Machines Like Me*, [McK19], a small group of seemingly conscious robotic men and women are manufactured and sold around the world. His novel makes the concept of a conscious robots seem both plausible and frightening. The two human protagonists Charlie and Miranda have no doubt that their robot Adam (as he is named) is conscious nor does his character Turing (a version of Turing in the book who lives a long and amazing scientific life). But it does not end well!

McEwan plunges right in with their robot Adam falling in love and sleeping with Miranda, Charlie's girlfriend. Although needing to be regularly recharged by a plug in his navel, he has been loaded with basic human emotions, partly by Charlie and Miranda clicking a set of online choices. Next he breaks Charlie's wrist when Charlie inadvisedly reaches for the off button on his neck that turns him off. But they soldier on when Adam apologizes to Charlie, only to find in the denouement that his idea of moral behavior is totally out of sync with humankind's waffling moral compromises, with actions that send Miranda to jail. Charlie, out of his love for Miranda, smashes in Adam's skull and Turing brands him a murderer.

McEwan certainly makes hay from my precise point: that human emotions are extremely complex and convoluted and thus one has to question whether a robot can ever truly "understand" them. Yet I would argue that an essential part of being conscious is precisely "feeling" emotions. I put this in quotes as feeling and understanding are words that touch on what consciousness is. It seems to me that McEwan is making too fine a point by allowing Adam many intense emotions yet failing to give him any deeper understanding of how emotions work.

Adam's failure highlights the human behavior pattern expressed by the word "loyal." This word refers to a mix of emotions and of patterns of actions, both past and future and is typical of the complex interweaving of emotions and social activities in human beings. For instance, the central principles of Scottish ethics might well be thrift, honesty and loyalty, all three being emotionally freighted activities. Adam is thrifty and honest but fails on the demands of loyalty. On the other hand, my cousin Ruth Silcock wrote a series of children's books (see e.g. [Sil80]) about a cat named Albert John. In her first book, she wrote "Albert John was a loyal cat," *assuming that this concept was perfectly clear to her*

young readers. But not so for Adam. Thus, by and large, McEwan is agreeing with my belief that modeling human emotions and their resultant activities in a robot is a huge hurdle, even though his characters do see their robot as emotional enough to be deemed conscious.

No wonder de Waal said that as yet there is no definitive framework for emotional states. Perhaps what is needed to make a proper theory, usable in artificial intelligence code, is to start with *massive data*, the key that with neural networks now unlocks so much structure in speech and vision. The aim is to define three way correlations of (i) brain activity (especially the amygdala and other subcortical areas but also the insula and the anterior cingulate area of cortex), (ii) bodily response including hormones, heart beat (emphasized by William James as the core signature of emotions) and facial expression and (iii) social context including immediate past and future activity. An emotional state should be defined by a cluster of such triples – a stereotyped neural and bodily response in a stereotypical social situation. To start we might collect a massive dataset from volunteers hooked up to IVs and MRIs, listening to novels through headphones. I am reminded of a psychology colleague whose grad students had to spend countless hours in the MRI tube in the wee hours of the night when time on the machine was available. Like all clustering algorithms, this need not lead to one definitive set of distinct emotions but more likely a flexible classification with many variants. All humans in all cultures seem to recognize nearly the same primary and secondary emotions when they occur in friends, although the words used giving boundaries between related emotions often shift. Conscious artificial intelligences will need to be able to do this too although AI's not shooting for full consciousness will have no need for such a skill. Without this analysis of emotions, computer scientists will flounder in programming their robots to mimic and respond to emotions in their interactions with humans, in other words to possess the crucially important skill that we should call *artificial empathy*. I would go further and submit that if we wish an AI to actually possess consciousness, I believe it must, in some way, have emotions itself.

iv.   What do physicists say about consciousness?

Quantum mechanics has also grappled with the concept of consciousness. To explain this, we need a few technical ideas. To model the subatomic world, quantum mechanics uses wave functions. In the simplest case (without fields), these are small sets of complex-valued functions of the spatial coordinates of all the particles present $\psi_\alpha(\vec{x}_1, \vec{x}_2, \cdots)$ called Schrödinger wave functions. The details don't matter. What does matter is how the wave functions relate to the world: are they ontological, describing objective material reality or are they epistemic, describing an observer's knowledge of the world? The problem is that they are both! They are ontological in the sense that it has been shown, convincingly to essentially all physicists, that there can be "no hidden variables," meaning any more detailed descriptions of the state of the world than the wave functions ψ. They are "all

that there is" when looking at an atom. Yet they are epistemic in that if, in a lab, some aspect of a subatomic event has been amplified and then is observed by a human, then that human knows something new and he/she must reset the wave function if they wish to best predict future events. This quandary has disturbed Heisenberg, von Neumann, Bohr, Einstein, Feynman – all the great physicists. We will discuss this at length in Chapter 14.

The meaning of ψ is part of the broader question of reconciling an inherently indeterminate description of subatomic events with the determinate classical description of macroscopic events. Physicists almost religiously resist the conclusion that human consciousness might enter into the reconciliation of these two descriptions. But Wigner and many others[2] are resigned to this conclusion. Wigner writes "The preceding argument for the difference in the roles of inanimate observation tools and observers with a consciousness ... is entirely cogent so long as one accepts the tenets of orthodox quantum mechanics in all their consequences." [Wig62], Chapter 13. A key point for him is what happens if there are two physicists A and B, A making measurement A′ and B, later in another room, making a second measurement B′. The probabilities of the outcome B′ will be altered by the outcome of A′, so the resulting ψ must reflect both measurements. This makes the epistemic viewpoint reflect the joint knowledge of both physicists. Pushing these ideas further, one is led to believe that our whole civilization is creating a bubble in space-time in which ψ is forced to reflect all the measurements all of us have done, the deterministic realities of our lives. As this is all a consequence of what physicists call the "Bohr" or "Copenhagen" interpretation, I call this our *Bohr bubble* in which, weirdly enough, our consciousnesses do alter the objective world. Another way out is the *multiverse* theory which proposes a gargantuan proliferation of simultaneously existing worlds. This, for me, is even screwier.

Relativity theory connects to the nature of consciousness in an equally fundamental way, shaking our ideas about time. Firstly, Newton, in his *Principia* states:

> *Absolute, true, and mathematical time, of itself, and from its own nature flows equally without regard to anything external.*

OK, this is indeed a good description of what time with its present moment feels like to us mortals. We are floating down a river – with no oars – and the water bears us along in a way that cannot be changed or modified. Central to this view is the division of time into the past, present and future. The whole universe, right NOW, has a fixed past leading up to the present state while an unknown future lies ahead. This NOW, however, is always moving, changing future events into past ones according to the laws of physics. But Einstein totally changed this world view by introducing a unified space-time whose points are events with a specific location and specific time. He asserted that there is no physically natural way of separating space and time, no god-given way to say two events are simultaneous when they occur in different places or that two events took place in the same location but at different

---

[2]Rather than explicitly naming consciousness as a factor, the "information-theoretic" school of thought formalizes the information possessed by observers.
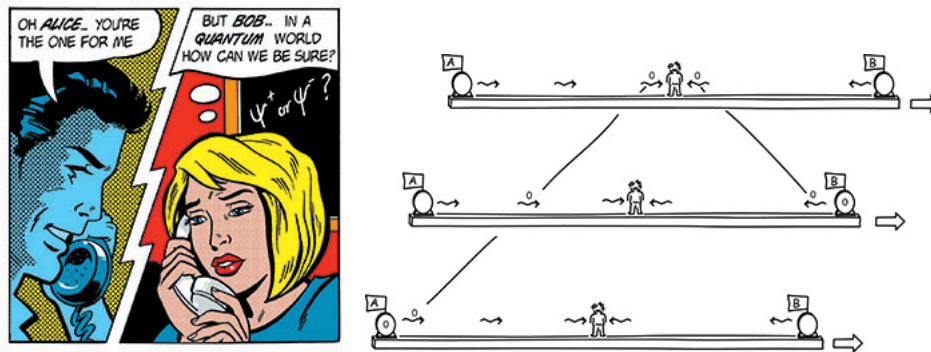
Figure 10.9: On the left: love in a quantum world, image by John Richardson, by permission of IOP publishing; on the right, from lecture notes of Prof. John Norton with his permission, a platform with clocks at each end is moving steadily to the right, an observer in the center watches two clocks A and B. The observer ascribes simultaneity to the times when he receives A and B's signals, times that are not simultaneous to the stationary observer.

times. This can only be done approximately and by using conventional coordinates, usually by setting up clocks in many locations and by exchanging signals. People's lives form a path in space-time and there is a natural length to this path, the thing we call our subjective time or body clock. But there is nothing in physics that corresponds to Newton's time, especially nothing corresponding to a physical NOW, the present. Not just that but science, essentially by definition, only studies correlations between events that can be reproduced exactly enough to show something is repeating itself, that here is a law of nature valid at least throughout some region of space-time. Sure, physics studies unique events such as the explosion of the Crab nebula seen on Earth in 1054 CE, but this is a fact of history, not a scientific law. The science of astrophysics explains this explosion by equations which are then applicable to infinitely many stars and in this way removes the historical uniqueness of that supernova. Thus it refuses to deal with any special instant that someone might call the present. The word "now" only enters our vocabulary through our conscious experience. For us, an experience is *never* reproducible (though we often try to make it so). As the saying has it, "you only go round once."

Hold on though. In quantum mechanics, experiments lead to "collapsing the waveform," resetting to the state vector to its projection onto an eigenspace for the observation, maintaining the classical macroscopic world we know and love. Einstein was fully aware of this issue and wrote about the seemingly paradoxical consequences when quantum theory and relativity are combined (in his famous paper with Nathan Rosen and Boris Podolsky [EPR35]. He seems to have wondered if the notion of the present, could have a place in physics. Though he never wrote about this, late in his life he had a conversation with

Rudolf Carnap in which he made this point. (My thanks to Steven Weinstein for telling me about this conversation.) Here is how Carnap described it:

> *Einstein said that the problem of the Now worried him seriously. He explained that the experience of the Now means something special for man, something essentially different from the past and the future, but that this important difference does not and cannot occur within physics. That this experience cannot be grasped by science seemed to him a matter of painful but inevitable resignation. He suspected that there is something essential about the Now which is just outside of the realm of science.*

Yes, yes that's what I'm talking about! How wonderful to hear it from Einstein.

It's interesting to recall a famous debate between Einstein and the philosopher Henri Bergson (thanks to my son Peter for telling me about this). They met in 1922 arguing about the nature of time. For Bergson, the important notion of time was not that of clocks but that of people's immediate conscious experience (*Les Données Immédiates de la Conscience* was the title of his dissertation). When you focus on the subjective time of a person, it is indeed not bound up with his spatial location. I find his ideas hard to follow but I think the key one is that time is heterogeneous, not homogeneous. Each instant for a conscious being is a thing in itself and their totality cannot be counted, like a flock of sheep. Time, he says is a temporal heterogeneity, in which "several conscious states are organized into a whole, permeate one another, [and] gradually gain a richer content" (Stanford Encyclopedia of Philosophy). In contrast, Einstein would say that a person's lifetime is a curve in space-time, time-like meaning the person moves more slowly than light, bounded by the space-time points representing his birth and his death, and along which integrating the Lorentz metric computes each person's subjective time. You can see these guys are not going to reach a consensus. Apparently Bergson's denial of Einstein's theory of time was the reason his Nobel Prize was awarded instead for his work on the photo-electric effect.

I find philosophical writings like Bergson's awfully hard to follow. But one thing seems totally clear to me and this is the central point of this post. I want to argue that it is this experience of a present instant, the NOW that is always changing yet is always our one and only unique present, the one that each of us owns, that that is the real core of what we call consciousness. You see this explains the Buddhist meditator: his mind may be empty of worldly distractions and his cortex may have no sensory, motor or memory activity but he still lives fully his present moment.

Although sentience, that is sensing the world and acting in response to these sensations, together with the corresponding brain activity, is often considered an essential feature of consciousness, I don't believe that. I think all scientists are missing the essential nature of consciousness. Sure we are conscious of what our eyes see and our ears hear, sure we are conscious of moving our body and making plans to do stuff and sure we can even fill our consciousness with the imaginary world of a novel or the proof of a theorem.
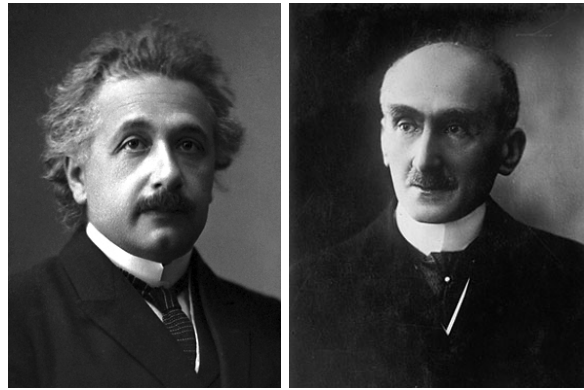
Figure 10.10: Einstein and Bergson about the time of their debate. Wikimedia Commons.

But I think all this misses what makes consciousness absolutely different from anything material: consciousness creates for us a present moment and it does this continuously moment after moment. I propose instead that the experience of the flow of time is the true core of consciousness, somewhat in the vein of Eckhart Tolle's "The Power of Now" [Tol97]. It rests on the idea that experiencing the continual ever changing fleeting present is something we experience but that no physics or biology explains. It is an experience that is fundamentally different from and more basic than sentience and is what makes us conscious beings. I believe that an experienced Buddhist meditator can put his or her self in a state where they wipe their mind clean of thoughts and then experience pure consciousness all by itself, free of the chatter and clutter that fills our minds at all other awake times. Accepting this, consciousness must be something subtler than the set of particular thoughts that we can verbalize.

v.   The Philosopher and the Sage

Philosophers and sages are not deterred by the failure of science. I want to start with the ideas of the German philosopher Thomas Metzinger, as presented in his book *The Ego Tunnel*, [Met09]. This is an exhaustive examination of what consciousness is from biological, psychological, information-theoretic and philosophical perspectives. It presents very relevant data from Out-of-Body Experiences, lucid dreaming and much else. After an analysis of what is going on in human brains, he writes a section entitled "How to build an artificial conscious subject and why we shouldn't do it" outlining how it might indeed be done.

Metzinger's book is easily the most readable dissection of the nature of consciousness by a philosopher that I have read. His basic thesis is that our brains construct for us a *phenomenal self-model*, by which he means "the conscious model of the organism as a whole that is activated by the brain" and that he also calls the Ego (p.4). He elaborates

this as follows (p.7):

> *First our brains generate a world simulation, so perfect we don't recognize it as an image in our minds. Then they generate an inner image of ourselves as a whole. This image includes not only our body and our psychological states but also our relationship to the past and the future as well as to other human beings. The internal image of the person-as-a-whole is the phenomenal Ego, the "I" or "self" as it appears in conscious experience.*

He says we feel we are consciously *having* the experiences that our bodies encounter in the world because this integrated inner image of ourselves is firmly anchored in our feelings and bodily sensations and because we are unable to recognize our self-models *to be just models*, because they are *transparent* like a glass window through which we see the world. Thus he is led to describe the life we lead as an *Ego Tunnel*. Our minds are filled by a model that we take for reality, hence we are in a tunnel through which we move as time goes on. Although he does not mention Schopenhauer, much of this theory seems similar to Schopenhauer's ideas: *Die Welt ist meine Vorstellung* (The world is my representation) is the assertion with which he opens his magnum opus *Die Welt als Wille und Vorstellung*.

Metzinger makes a great deal of the so-called rubber hand illusion. Here, the subject sits at a table with his left hand behind a barrier, but a rubber left hand is placed on the table in front of him. Then the rubber hand is tickled by a feather while, invisibly, his real left hand is also tickled. After a certain amount of time, the subject begins to feel the rubber hand is his own, that an invisible arm connects it to his body and tickling it alone causes him to feel his real hand is tickled. Metzinger interprets this as tricking the mind into altering its self-model into an unreal representation that still feels totally real. Similarly, he discusses at length phenomena like out-of-body experiences and lucid dreams (where you are aware you are dreaming but still feeling you are living a vivid convincing dream world). Oddly, he doesn't describe some of the other virtual reality experiments like the one where, wearing goggles that show you walking over a virtual cliff, you fall down with genuine fear (though actually onto a carpet in an empty room). I was a subject and experienced this at Brown. Nor does he discuss the vast virtual world in the movie "The Matrix" and the present vogue for virtual reality goggles and immersive entertainment. But surely these only reinforce his argument that we live in a self-model and can all too easily be tricked into taking an alternate world as reality.

Let us next look at an ancient Indian sage. My favorite story from the rich legacy of Hindu Mythology is the story of the sage Narada and his quest to understand Vishnu's *Maya*. It illustrates that Metzinger's phenomenal self-model has antecedants that go back at least to the first millenium BCE. It starts with Narada performing so many austerities that he acquires the spiritual power to ask Vishnu for a boon. He asks for an understanding of Maya (an ancient Sanskrit word for "illusion"). The story goes on, in the telling of Heinrich Zimmer in his wonderful book *Myths and Symbols in Indian Art and Civilization*, [Zim46], pp.32-34:

"Show me the magic power of your Maya ," Narada had prayed, and the God replied, "I will. Come with me;" with an ambiguous smile on his beautiful curved lips. From the pleasant shadow of the sheltering hermit grove, Vishnu conducted Narada across a bare stretch of land which blazed like metal under the merciless glow of a scorching sun. The two were soon very thirsty. At some distance, in the glaring light, they perceived the thatched roofs of a tiny hamlet. Vishnu asked: "Will you go over there and fetch me some water?" "Certainly, O Lord," the saint replied, and he made off to the distant group of huts. The god relaxed under the shade of a cliff, to await his return.

When Narada reached the hamlet, he knocked at the first door. A beautiful maiden opened to him and the holy man experienced something of which he had never up to that time dreamed: the enchantment of her eyes. They resembled those of his divine Lord and friend. He stood and gazed. He simply forgot what he had come for. The girl, gentle and candid, bade him welcome. Her voice was a golden noose about his neck. As moving in a vision, he entered the door. The occupants of the house were full of respect for him, yet not the least bit shy. He was honorably received, as a holy man, yet somehow not as a stranger; rather, as an old and venerable acquaintance who had been a long time away. Narada remained with them impressed by the cheerful and noble bearing, and feeling entirely at home. Nobody asked him what he had come for; he seemed to have belonged to the family from time immemorial. And after a certain period, he asked the father for permission to marry the girl, which was no more than everyone in the house had been expecting. He became a member of the family and shared with them the age-old burdens and simple delights of a peasant household.

Twelve years passed; he had three children. When his father-in-law died he became head of the household, inheriting the estate and managing it, tending the cattle and cultivating the fields. The twelfth year, the rainy season was extraordinarily violent; the streams swelled, torrents poured down the hills, and the little village was inundated by a sudden flood. In the night, the straw huts and cattle were carried away and everybody fled. With one hand supporting his wife, with the other leading two of his children, and bearing the smallest on his shoulder, Narada set forth hastily. Forging ahead through the pitch darkness and lashed by the rain, he waded through slippery mud, staggered through whirling waters. The burden was more than he could manage with the current heavily dragging at his legs. Once, when he stumbled, the child slipped from his shoulder and disappeared in the roaring night. With a desperate cry, Narada let go the older children to catch at the smallest, but was too late. Meanwhile the flood swiftly carried off the other two, and even before he could realize the disaster, ripped from his side his wife, swept his own feet from under him and flung him headlong in the torrent like a log. Unconscious, Narada was stranded eventually

> *on a little cliff. When he returned to consciousness, he opened his eyes upon a vast sheet of muddy water. He could only weep.*
>
> *"Child!" He heard a familiar voice, which nearly stopped his heart. "Where is the water you went to fetch for me? I have been waiting more than half an hour." Narada turned around. Instead of water he beheld the brilliant desert in the midday sun. He found the god standing at his shoulder. The cruel curves of the fascinating mouth, still smiling, part with the gentle question: "Do you comprehend now the secret of my Maya?"*

In Metzinger's language, I would interpret this story as follows: Vishnu put a fork in Narada's Ego Tunnel and led him down the new fork by his request for water. The new fork was long and ultimately led to Narada experiencing his own drowning. But then Vishnu made the new fork rejoin the old with a touch of a cruel smile on his face. Thus Maya can be seen as a description of one's phenomenal self-image, a convincing reality but only a small window into what is out there, constructed by our limited consciousness.

Very similar ideas about the nature of consciousness have been proposed by Manuel and Lenore Blum in their theory of a "Conscious Turing Machine" (CTM), [MB21]. Their goal is formulate as precisely as possible an architecture that could underlie a conscious robot. Their CTM has a large number of interconnected processors, working in parallel and carrying long term memory, that model the unconscious activity of our brain. Chunks of data from these processors compete and one such chunk at a time gets to the small short term memory whose activity is the stream of consciousness. Like all computers, the CTM has a clock that defines its internal time and thus the sequence of conscious chunks. Among the many processors, there is a key one called the "Model-of-the-World" processor that, together with "Inner-Speech," "Inner-Vision," "Inner-Sensation" processors create the "feeling of conscious awareness," they "give the CTM its sense of self." This processor handles multiple worlds in which there are both "self" and "not-self" objects.

Concerning the Now, Metzinger states "My idea is that this simultaneity is precisely why we need the conscious Now" (in the section "The Now Problem: A Lived Moment Emerges," p.34-36). It is well known that the mind plays fast and loose with simultaneity, so that two signals may be perceived consciously as occurring in the opposite order to their occurrence in the physical world. Temporal order seems to be, to some extent, a construction the mind makes as best it can. But now Metzinger reverses the logic. From the implication that experiencing a Now implies experiencing simultaneity, he wants to say that experiencing simultaneity creates the experience of the Now. He argues that creating a common temporal frame of reference for all the mechanisms in the brain leads to the inner model of the world around such a Now (p.36). I cannot follow this and the Blums don't consider this a problem: all computers have a clock and organize their computations and communications accordingly and most programs have no pretense of carrying consciousness.

As mentioned above, Metzinger spells out the application of these ideas to the construction of conscious robots in the later section "How to build an artificial conscious subject

and why we shouldn't do it" (p.190). On p.192, he describes the construction in four steps. The first is to endow the machine with a continuously updated integrated inner image of the world. The second is to organize its internal information flow temporally, resulting in a psychological moment, an experiential Now. The third is to be sure that these internal structures cannot be recognized by the artificial Conscious system as internally generated images, so they are transparent. The fourth step is to integrate an equally transparent internal image of itself into the phenomenal reality. None of these seem to Metzinger or me be considered to be impossibly difficult. But it's the second step where I feel he is assuming too much happens as a result of the silicon activity. The Now seems to me the truly magical step, the step that creates what Popper calls world II and others call spiritual.

The fact that our lives are lived as a trip down the river of time, that we are always conscious of being at a specific place surrounded by a local bit of the 3 dimensional world which changes as "time goes on," all this seems obvious and commonsensical, not magical at all. But this is because this experience of time is the core of everyone's consciousness, everyone's daily lives, not because in physics or in any other science is there anything like the flow of time with a present instant, lit up like a lighthouse. In physics, time is static, simply one way to put coordinates on the 4-dimensional panoply of all events, past, present and future and in any place whatsoever. We can artificially construct a mathematical "flow," a one-dimensional group of homeomorphisms of space-time, but no such flow is given by physics and there is no distinguished set of points called the present moment. To live in a world of time seems to me a wonderful gift and I have no clue how, like God and Adam on the ceiling of the Sistine Chapel, one might give this gift to a robot.

I want to summarize what I believe are the essential features of consciousness that emerge from all this discussion.

1. Consciousness is a reality that comes to many living creatures sometime around birth and leaves them when they die, creating a feeling of "moving" from past to future along a path in space-time as well as feeling sensations, emotions and their body movements.

2. Consciousness has degrees, varying from utterly vivid (e.g. positive feelings like love and negative feelings like pain) to marginal awareness. The brain has, moreover, a huge unconscious part whose activities and thoughts do not reach consciousness.

3. Consciousness occurs in many creatures including, for instance, octopuses, birds and all mammals. It arises from multiple neural structures, and is always connected in some way to an internal model that includes self and non-self objects to some degree.

4. Consciousness has many ingredients, including, in increasing order of how essential they are, a) *cognitive skills*, b) *subjective feelings* like pain and emotions and c) *the experience of the flow of time.* The first is the basis of what psychologists now study via human reports of their own thoughts. But it's easy to imagine your own state

without any cognitive activity, e.g. after a stroke or after trauma leaves you in a disoriented daze. Emotions are the spice of life but they come and go and their departure doesn't interrupt your sense of time passing. This leaves the last as the true core of consciousness.

5. Consciousness is not describable by science, as it is a reality on a different plane.

Finally I'd like to add some comments involving the word "soul". This word has not received the scientific respectability that the word "consciousness" has. But, from my perspective, "having a soul" and "having consciousness" are synonyms. I think this is pretty accurate in terms of their historical usage. When I first got involved in AI, there were acronyms for every variety of computer and I proposed that the human brain ought to be referred to as a "SOUL" machine, that is a "**S**ingle **O**pportunity for **U**se **L**earning" machine. Not everyone will be comfortable with souls, but, returning to the Iris de Ment song with which I began this Chapter, clearly the soul, whatever it is, is captivated by all the emotions its embodiment affords. So we had better take seriously endowing a robot with emotions if we want it to be conscious. I'll return to emotions and the soul in Chapter 18.

Part IV

And Now, Some Bits of Real Math

Being a mathematician, I can't stop loving actual math and, from time to time, felt that I really wanted to share something I found exciting. Chapter 11 was inspired both by a lecture by Barry Mazur about the Riemann Zeta function and by the comment made by Freeman Dyson that the imaginary parts of the zeroes of the Zeta function and the sequence of logs of the primes are, more or less, Fourier transforms of each other. The connection can be made by von Mangoldt's formula as explained in this chapter. But as an applied mathematician who was accustomed to numerical experiments with Fourier transforms, I wondered whether the smallest zero is produced, at least approximately, by small primes. Lo and behold, it shows itself immediately in the log's of the first two primes, $p_1 = 2, p_2 = 3$. If the primes show an oscillation, it should have one period between $\log(3)$ and $\log(2)$, i.e. the frequency should be around $2\pi/(\log(3) - \log(2))$, that is about 14.50. And the imaginary part of the smallest zero is about 14.13. Pretty close! This chapter describes how I ran with this for fun, even though I have never gotten deeply into analytic number theory.

Chapter 12 started with an email from Al Osborne, an expert on the mathematical theory of waves. As a sailor, I had always been fascinated by rogue waves and had also worked on some of the non-linear PDE's that Al used. Moreover, I had spent a decade working with Peter Michor on the PDE's of geodesics on various infinite dimensional Riemannian manifolds arising from *shapes*, e.g. spaces of plane curves, submanifolds or diffeomorphisms. There is a beautiful collection of diverse properties in this area of geometric analysis and I wanted to write a sketch of this, with the hope of enticing young mathematicians to look at it. The chapter begins with setting up the PDE for water waves, then goes on to outline a few of high points of my work with Peter and ends by linking the two, following V. E. Zakharov's work where you find the water wave PDE is a flow that is geodesic with a startling Riemannian structure plus the gravitational potential.

Chapter 13 is much longer and deals with my long term fascination with the foundations of math. Foundations have ceased to be a central topic in pure math, having been relegated mostly to the specific area of *set theory*. I think this is a mistake. I have found wonderful work being done by Harvey Friedman and Steven Simpson in what is called *reverse mathematics* (work out what minimal foundational formalism is needed to prove various theorems in mainstream math). And I have felt that the perspective of applied mathematics has not played the role it should in the foundations. Much of this chapter is expository, describing what I feel are the key points that one needs to know to talk seriously about foundations: the limits of Peano arithmetic, Ramsey theory, second order arithmetic, coding Borel sets, constructible sets, higher infinities and Brouwer's "free choice sequences." It ends with some ideas I hope will bear fruit.

Chapter 11

Finding the Rhythms of the Primes

The question addressed in this Chapter[1] arose when I listened to Barry Mazur's excellent lecture on Riemann's zeta function to the "Friends of the Harvard Math Department'" sometime in the early 2010's. Barry went on, with William Stein, to write a book on this function entitled *What is Riemann's Hypothesis?* addressed to people with minimal mathematical background. The book leads up to Riemann's 'explicit formula' which, in von Mangoldt's form, is the formula for a discrete distribution supported at the prime powers:

$$\sum_{\text{primes } p} \sum_{n \geqslant 1} \log(p)\delta_{p^n}(x) = 1 - \sum_k x^{(\rho_k - 1)} - \frac{1}{x(x^2 - 1)}$$

where $x > 1$, ρ_k ranges over the zeros of the zeta function in the critical strip $0 < \text{Im}(\rho) < 1$ and the sum over k converges weakly as a distribution. This relates primes to the zeta zeros. But, having been doing applied math at the time and thinking like an engineer, I asked: *Can we find approximately the smallest and maybe more small zeros hidden in the very smallest primes without resorting to analytic continuation?*. Although thousands of pages have been written about ζ, this, to the best of my knowledge, seems to be a new way of analyzing the periodicity of the primes.

For readers not familiar with the zeta function, let me orient them with a few words. Riemann in 1859 defined:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \text{ where we assume } s > 1.$$

$\zeta(s)$ goes to ∞ when s approaches 1, but he showed that if you allow s to be complex, it has an analytic continuation to the whole complex plane except for a simple pole at $s = 1$.

---

[1]This Chapter is a slightly edited version of my post "The lowest zeros of Riemann's zeta are in front of your eyes," dated October 30, 2014

With the help of various manipulations of contour integrals, he finds that ζ has zeros at $-2, -4, \cdots$ and at infinitely many points $a_k + i\rho_k, 0 \leqslant a_k \leqslant 1$. He conjectures that all a_k equal $1/2$ – this is the famous Riemann Hypothesis – and he then gives essentially the above formula.

Riemann called the terms in ρ_k the oscillating terms because if $\rho_k = 0.5 + i.\omega_k$, as he hypothesized, and we pair symmetric roots $\pm\omega$, then

$$\sum_k x^{\rho_k - 1} = 2 \sum_k \cos(\log(x).\omega_k)/\sqrt{x}.$$

Thus Riemann showed that *the logs of the primes* show periodic behavior. Let's start from scratch and ask if we find periodic behavior in the logs of the smallest primes or, as they get larger, clusters of primes.

The ratios of the lowest primes 2, 3, 5, 7, 11 are roughly 1.5, 1.67,1.4,1.57 which all cluster around 1.55. But then $13/11$ is only about 1.18. To fix this, after 10 we shift from single primes to prime pairs, replacing the pair by the even number in the middle, getting the new sequence:

2, 3, 5, 7, 12 for (11,13), 18 for (17,19), 23?, 30 for (29,31), 37?, 42 for (41,43), 47?.

Skipping the isolated primes 23 and 37, the ratios are now 1.5,1.67,1.4,1.71, 1.5,1.67,1.4. If you make a linear fit to the logs, you find a sequence that approximates the primes:

$$1.27 \cdot (1.557)^n \approx 1.98, 3.08, 4.80, 7.47, 11.64, 18.12, 28.22, 43.94. \cdots$$

Hmm: not bad. Also note that we ignored prime powers, which explains why the prime 5, dragged down by 4 became 4.8 and the prime 7, dragged up by 8 and 9, became 7.47. Even more startling, this power law would come from a periodic term in log-prime density of form $\cos(2\pi \log(x)/1.557)$ and $2\pi/\log(1.557) = 14.185...$, which is very close to the true first zero of Riemann's zeta, namely 14.1347...! In other words, the basic idea behind Riemann's periodic terms is indeed apparent in these small primes. This is especially startling because the convergence of the explicit formula is very slow: there are very many rapidly oscillating terms beyond the first one so there is no compelling reason why the lowest ω_k should nail these primes this well. This suggests there might be other formulas relating the primes with the zeros clarifying this correspondence.

Let's go back to the explicit formula and change coordinates to $y = \log(x)$. Again writing the zeros as $(0.5 + i.\omega_k)$ where ω_k is real under the Riemann hypothesis, being careful with the deltas and summing only over k with $\omega_k > 0$, you get:

$$\sum_{p,n} \frac{\log(p)}{p^{n/2}} \delta_{\log(p^n)}(y) = e^{y/2} - 2 \sum_k \cos(y\omega_k) - \frac{1}{e^{y/2}(e^{2y}-1)}$$

Note that instead of thinning out logarithmically as the primes do, the logs of primes now get dense at an exponential rate. After weighting the prime powers as shown, they still have density $e^{y/2}$, the first term on the right. But after that we get oscillations. Curiously

an immense amount of work has been done on very large primes and very large zeta zeros while this formula for small values of y doesn't seem to have been looked at. A graph of the small log-prime-powers weighted as in this formula and smoothed out with a Gaussian is shown in Figure 1. The oscillation given by the lowest zeta zero is now really clear.
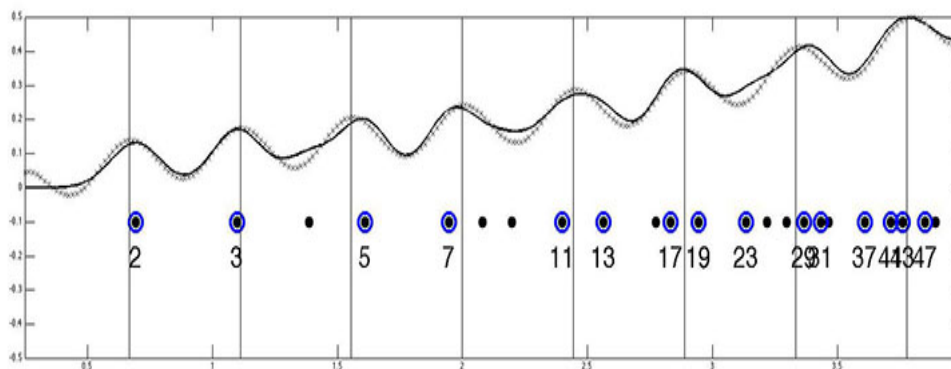
Figure 11.1: Prime powers up to 50 and its period. The horizontal axis is log scale, the filled circles are the logs of the primes up to 50, the dots the prime powers. The solid line is the convolution of the weighted sum of deltas as above with a Gaussian with standard deviation 0.1. The line of hatch marks is its approximation with the above explicit formula but using only *ONE* zero of zeta and the vertical lines are its peaks where the cosine equals -1. Note that 23 and 37 are being ignored and will require the next zero of zeta as will separating 5 and 7 from adjacent prime powers.

How many of the zeta zeros are hidden in the primes up to 53? Let's sample the interval [0,4] in the log-prime line discretely so that the sum of weighted deltas becomes a function on a discrete space and take its discrete cosine transform. We find chaos in the high frequencies but terms $cos(\pi \log(p)(k-1)/4)$ for $1 \leqslant k \leqslant 50$ seem to be coherent and give us oscillating terms whose discrete frequencies correspond to

$$\omega = 14.1, 20.0, 25.0, 30.4, 32.9, 37.6 \pm 0.4$$

Remarkably, these are quite close to the true zeros 14.1, 21.2, 25.1, 30.6, 33.0 and 36.9. Figure 2 shows the low frequency part of the DCT.

Can we find the oscillations in larger primes directly from tables of primes (not using von Mangoldt's formula)? The simple answer is that they get drowned in the exponentially increasing density of log-primes. Extending the above plot to higher primes, one finds that the slope of the large exponential function $e^{y/2}$ erases the local minima apparent for the small primes. There are several ways to find them however. One can simply subtract the mean density $e^{y/2}$ or one can convolve the weighted sum of deltas with a suitable filter that kills the average. An engineer knows how to form filters that not only do this but also pick
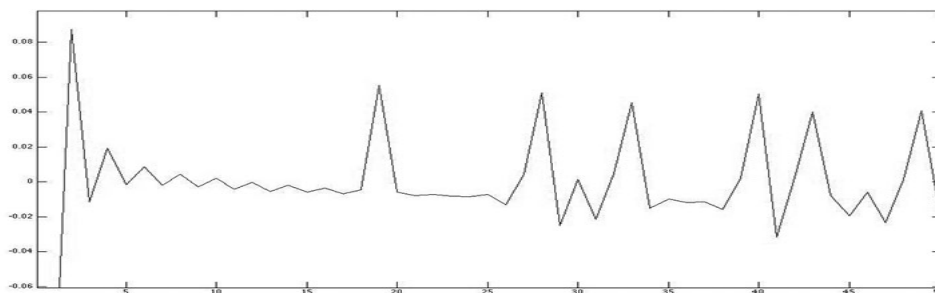
Figure 11.2: Powers of frequencies 0 to 50 of the discrete cosine transform of the weighted sum of delta functions at prime powers 2 through 53. The peaks approximate the first six zeta zeros.

out some range of frequencies. This can be used to find the oscillations caused by all the zeros of zeta.

Let's stick to the simplest case. If we want to both kill a constant term and suppress higher frequencies, a simple way is to convolve with the second derivative of the Gaussian. But we want to kill $e^{y/2}$, so we need to first premultiply by $e^{-y/2}$, then convolve with the second derivative and finally multiply back by $e^{y/2}$. In one step, this amounts to convolving with:

$$(y^2 - \sigma^2) \cdot e^{-\frac{1}{2\sigma^2}(y - \frac{\sigma^2}{2})^2}.$$

For $\sigma = 0.2$, the value we will use, the result is shown in Figure 3.
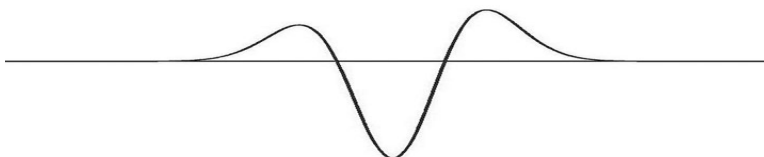
Figure 11.3: The modification of the second derivative of the Gaussian kernel that kills the exponential of y/2.

If we use this filter and convolve the weighted sum of deltas at the logs of all prime powers up to 3 million, we finally get the curve in Figure 4 where now the negative peaks show high density of primes, positive peaks low density.

The large negative peak on the left is almost exactly at log(2), the next at log(3), etc. The 10th peak is about 106 which lies the middle of the streak (101,103,107,109) of 4 primes (because 105=3.5.7). Looking to the right hand side of the plot, there is another negative peak around 1.9 million (log about 14.4) and another around 2.9 million (log about 14.9). I don't know if anyone has noticed this extra density of primes around these values. Note
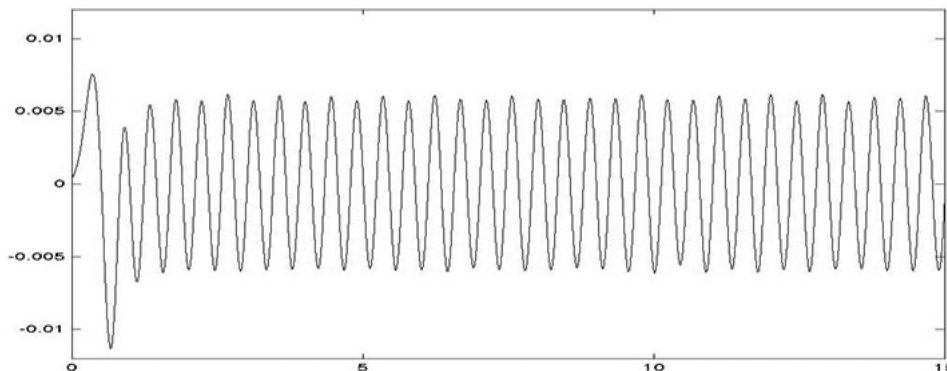
Figure 11.4: The result of convolving the weighted sum of deltas at log-prime-powers with the previous filter.

that we are not looking at one precise value but at a range, e.g. 1.75 million to 2 million and comparing the number of primes in that range with dips in density before and after. One wonders whether Gauss noticed this during his numerical exploration of $\pi(n)$, the function counting primes.

As Barry asked me, the fact that the lowest zeros of zeta show themselves in the very smallest primes seems to extend to Dirichlet L-series too. The simplest case is the mod 4 series, giving the sign +1 to primes congruent to 1 mod 4, and -1 when congruent to 3 mood 4. In fact, just as the lowest zero of Riemann's zeta is close to 2pi divided by the log of the ratio of the two lowest primes 3 and 2, the lowest zero of the L series (6.02) is close to π divided by the log of the ratio of 5 and 3 (6.15). This is because 5 and 3 are the two lowest odd primes and they have opposite residues mod 4, hence should differ by π, not 2π, in the oscillation caused by this zero. A plot, convolving the signed and weighted sum of deltas with a Gaussian of standard deviation 0.2 is on Figure 5. Note how we have negative peaks at 3, 7 and the pair [19 23], all congruent to 3 mod 4, and a positive peaks at 5, the pair [13 17] and the pair [37 41], all congruent to 1 mod 4. The vertical lines are half periods of the lowest frequency L-function oscillating term.
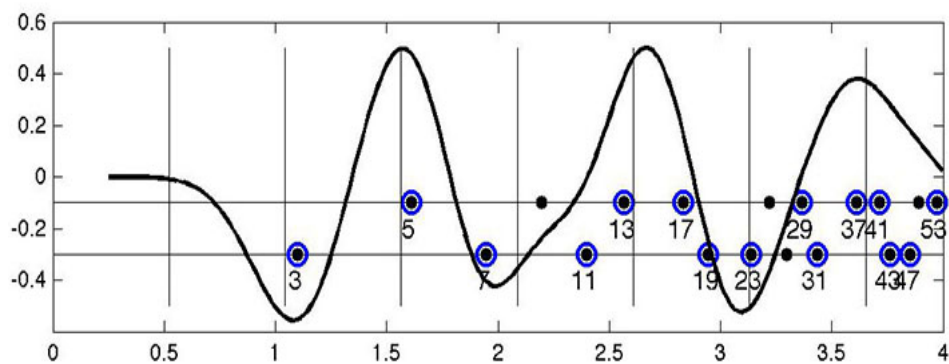
Figure 11.5: Odd primes and powers up to 53 and and the periodic behavior after convolution. The horizontal axis is log scale, the filled circles are the logs of the primes up to 53, the dots the prime powers with numbers congruent to 1 mod 4 above, 3 mod 4 below. The solid line is the convolution of the signed and weighted sum of deltas as above with a Gaussian with standard deviation 0.2.

Chapter 12

Spaces of Shapes and Rogue Waves

Back in 2020, I got an unexpected email from Al Osborne, a physics Professor at the University of Torino and researcher at the Office of Naval Research in the US. I discovered that he is one the preeminent world experts on *rogue waves*, the 50-100 foot monsters that can arise even in moderate sea conditions and sink ships. There's a fabulous BBC documentary on these waves on youtube at `https://www.youtube.com/watch?v=mC8bHx gdHH4`. As a life-long sailor who has made ocean passages, I was immediately drawn to this phenomenon. Al turned out to be a fan of theta functions on which I worked decades back, as they produce soliton-type solutions of the non-linear Schrödinger equation which are a possible model for such waves. I was doubly fascinated because this was also something that my student Emma Previato had worked out for her thesis (cf. [Pre85]). And after struggling with the literature, it dawned on me that this also fits in with my work with Peter Michor on the infinite dimensional manifold of simple closed plane curves and the idea of *shape spaces*. I'll start with the waves and then insert a digression on shape spaces and finally put them together.

i.   Nonlinear gravity waves

Like almost all physics, one begins by simplifying the problem! Water is incompressible, ok, so their velocity vector field has no divergence. But their theory gets truly messy and complicated by their *vorticity*, the curl of that vector field. Well, don't forget that vorticity is preserved along streamlines in the absence of any external force. And when water truly settles down, as it does from time to time, even in mid-ocean (I have seen this and swam in deep ocean water as flat as a pancake), then its velocity vector field is zero! So mostly ocean water can be modeled by curl free divergence free vector fields. Sure, the wind is an external force creating, among many things, what is called *Langmuir Circulation*, long cylindrical-shaped structures in the surface layer counter-rotating from one roll to the next. And shelving bottoms create external forces near shores causing

further vorticity. However, in deep water and ignoring the topmost layers being blown around, it is irresistible to assume the curl is zero too. Aha, harmonic functions now make their appearance and lots of standard math can be used. I want to thank Darryl Holm for explaining some of the complexity of actual waves!

Let's do the math, making this vorticity-free simplification. First a domain: assume z is the vertical dimension and we wish equations for the time varying surface of an ocean $\Omega^{(t)}$ of infinite depth. Denote the ocean's surface by $\Gamma^{(t)}$ and its equation by $z = \eta(x, y, t)$, (excluding breaking waves whose tops outrun the troughs). Let $\vec{v}(x, y, z, t)$ be the velocity vector of the water. Letting $\vec{N}_{\Gamma^{(t)}}$ be the unit normal to the surface, the motion of the surface is given by the normal component of $\vec{v}$:

$$\frac{\partial \Gamma^{(t)}}{\partial t}(P, t) = \vec{v}(P, t) \cdot \vec{N}_{\Gamma^{(t)}}(P), \text{ or}$$
$$\frac{\partial \eta}{\partial t}(P, t) = \vec{v}(P, t) \cdot \left(-\frac{\partial \eta}{\partial x}, -\frac{\partial \eta}{\partial y}, 1 \right)$$

Next, there potential $\phi(x, y, z, t)$ on $\bigcup_t \Omega^{(t)}$, *harmonic* in (x, y, z) such that $\vec{v} = \nabla \phi$. Euler's equation becomes now the *definition* of the pressure:

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} \|\nabla \phi\|^2 = -p - gz$$

where we take the density of water to be 1, and g to be the force of gravity on earth's surface. However, on the surface, p must equal the atmospheric pressure, which we can absorb into the normalization of z, hence set p at the surface to zero. We assume that, at the bottom of the ocean, ϕ and $\nabla \phi \to 0, z \to -\infty, p \to +\infty$. Finally, for every simply connected domain, one has the Poisson kernel $\mathcal{P}_\Omega$ that computes every harmonic function on the domain from its boundary values. For flat seas, for instance, the domain is the lower half space and the kernel is $-z/2\pi(x^2 + y^2 + z^2)^{3/2}$. Thus we complete the set of equations for the evolution of gravity waves using:

$$\frac{\partial \phi}{\partial t} = -\mathcal{P}_{\Omega^{(t)}} * \left\{ \left(gz + \tfrac{1}{2} \|\nabla \phi\|^2 \right) \big|_{\Gamma^{(t)}} \right\}$$

The majority of work on gravity waves deals with "wave trains," waves which are independent of one of the horizontal coordinates, e.g. y, leaving (x, z). In this case, $\Omega^{(t)}$ can be taken as a plane domain and harmonic functions are the real parts of complex analytic functions of $x + iz$. Their real and imaginary parts are conjugate harmonic functions that determine each other by an integral transform generalizing the Hilbert transform. But very few people use these equations. Instead, they start with the *ansatz*:

$$\eta(x, t) = \text{Re} \left(A(x, z, t).e^{i(kx - \omega t)} \right)$$

where A is a *slowly varying* "complex wave envelope." Then, by discarding judiciously terms thought to be small, one derives the result that A satisfies the non-linear Schrödinger

equation with coefficients expressed in terms of k, ω. The beauty of this is that one has explicit solutions of the non-linear Schrödinger equation arising from theta functions on Jacobians of algebraic curves that appear to produce "rogue waves," (cf. Osborne's book [Osb10]). But wouldn't it be more fun to avoid the ansatz?

ii.   Shape Spaces

Starting from completely different questions and motivations, Peter Michor and I had been studying, since the early 2000s, the infinite dimensional manifolds formed by the totality of a large variety of geometric structures. For example, if you fix an ambient manifold and look at all its submanifolds with some given invariants, then the totality of such submanifolds is itself a manifold, albeit a pretty big, infinite dimensional one. Following algebro-geometric traditions, we called these the *differentiable Chow manifolds* (cf. [**S-2013a**]). Riemann himself had noted the existence of such manifolds in his famous Habilitation lecture, manifolds where the coordinates of a point are given by an infinite sequence or by a function. There are many other examples but to fix ideas, the prime example, the one that has given rise to the most work, is this: take the ambient space to be simply the plane and consider in it all simple closed plane curves, making this a manifold in its own right. What continues to amaze me is the huge diversity of the geometric properties of this one space in the many natural metrics that it carries. A caution: I have on purpose *not* said how smooth or how jagged the curves are that define points in this space. Because of this, we don't have literally one space. It's exactly like the linear situation for function spaces: one has a core of smooth functions, but for each metric one forms its completion. These nest in each other in complex ways. OK, we have the same in the nonlinear realm: many instantiations of the space of "all" simple closed plane curves, all being completions in different metrics of the core set of C^∞ curves. And there are also finite-dimensional "approximations" like the space of non-intersecting n-gons. I'll give three examples of Riemannian metrics on this space that illustrate well the diversity.

   Let's denote this core space by $\mathcal{S}$ and its members by Γ with interior Ω. Then, as above, for all Γ, let T_Γ and N_Γ be their tangent and normal bundles in the plane. A section of the normal bundle $a : s \mapsto a(s).\vec{N}_\Gamma(s)$ represents a tangent vector to $\mathcal{S}$ at the point representing Γ. A Riemannian metric on $\mathcal{S}$ is then defined by a quadratic norm on every such section. The simplest possible one is just the L^2 metric $\|a\|^2 = \int_\Gamma a(s)^2 ds$. where s is arc-length. The resulting Riemannian manifold is a strange bird indeed: although it has geodesics, a) they can develop infinite curvature and end in finite time and b) the infimum of path lengths between any two points of $\mathcal{S}$ is zero. Geometrically, what's happening is that the sectional curvatures are all non-negative and, at any point, unbounded so that conjugate points are dense on geodesics. Visually, the intermediate curves can grow rapid wiggles that shorten the above distance along any path as much as you want. Two references: [**S-2005, S-2006a**]. Figure 1 illustrates these properties of this metric.
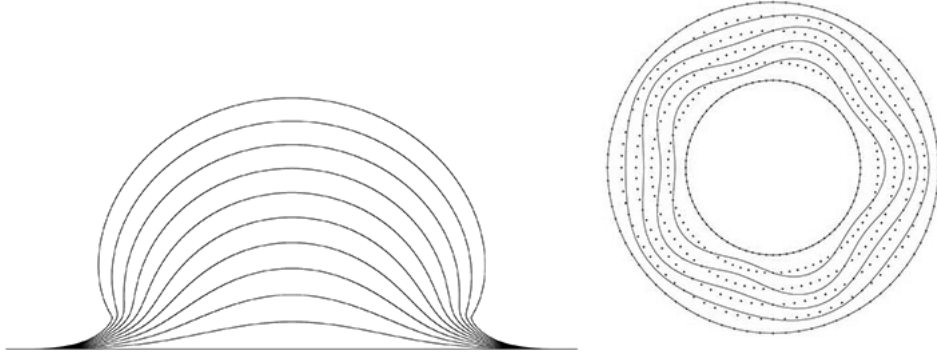
Figure 12.1: The L^2 metric: left, a geodesic starting from a straight line and moving in the direction of a small 'blip' that develops a corner with infinite curvature in finite time; right, two geodesics between two concentric circles, one simply intermediate circles, one with wiggles that grow and shrink, illustrating conjugate points in this metric.

To get metrics that behave more normally, the standard way is to use Sobolev-type metrics. The best way to do this is by viewing $\mathcal{S}$ as a quotient of the group of smooth diffeomorphisms of the plane, $\mathrm{Diff}(\mathbb{R}^2)$ by the subgroup of diffeomorphisms that map the unit circle to itself. The Lie algebra of this group is the vector space of smooth vector fields on the plane and one can put Sobolev norms on them component-wise:

$$\|\vec{v}\|^2_{Sob-n} = \int_{R^2} ((I - \Delta)^n \vec{v} \cdot \vec{v}) dx dy.$$

If one extends this norm to be one-sided invariant and takes cosets on the same side, you get a quotient metric on $\mathcal{S}$ for which the map from Diff to $\mathcal{S}$ is a submersion: the tangent bundle "upstairs" splits into a vertical part tangent to the cosets and a horizontal part that is the pull back of the tangent bundle "downstairs." This is an isometry between the *quotient* metric on $\mathcal{S}$ and the horizontal part of the one-sided invariant metric on Diff. All geodesics on $\mathcal{S}$ for this metric lift to horizontal geodesics on Diff. A simple way to understand this definition is:

$$\|a\|^2_{Sob-n} = \inf \left\{ \|\vec{v}\|^2_{Sob-n} \middle| \vec{v} \text{ on } R^2, \vec{v} \cdot \vec{N}_\Gamma(s) = a(s) \right\}$$

In the land of pseudo-differential operators, there is such an L_n for which $\|a\|^2_{Sob-n} = \int_\Gamma (L_n(a) a . ds$. Here n need not be an integer but, in all cases, L_n has degree $2n - 1$. So long as $n > 1$, these manifolds behave well, having geodesics and curvature etc., just like finite dimensional manifolds. Michael Miller's group at Johns Hopkins has used the 3D version of these metrics extensively to analyze medical scans (cf. [MTY02]). An example is shown in Figure 2.
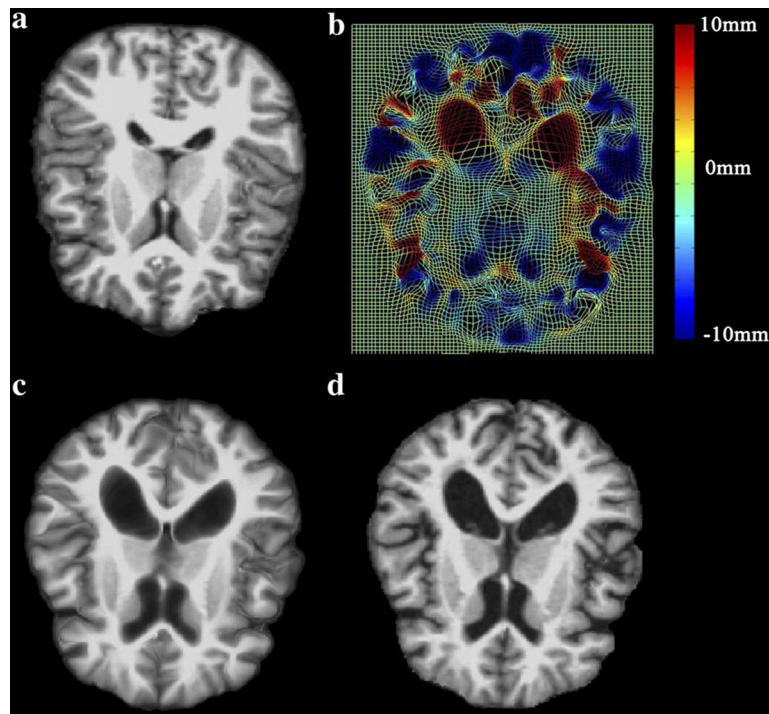
Figure 12.2: A paper of Du, Younes and Qiu [DYQ11] uses geodesic warping to match pairs of combined MRI brain scan, shape of cortical boundary and extracted sulcal/gyral curves (named "6D-LDDMM"). Here (a) is a normal scan, (d) a scan of a person with dementia, having major white matter loss and ventricle enlargement, and (c) the endpoint of a geodesic close (allowing for noise) to (d). (b) color codes the warping for subsequent analysis. By permission Elsevier Press.
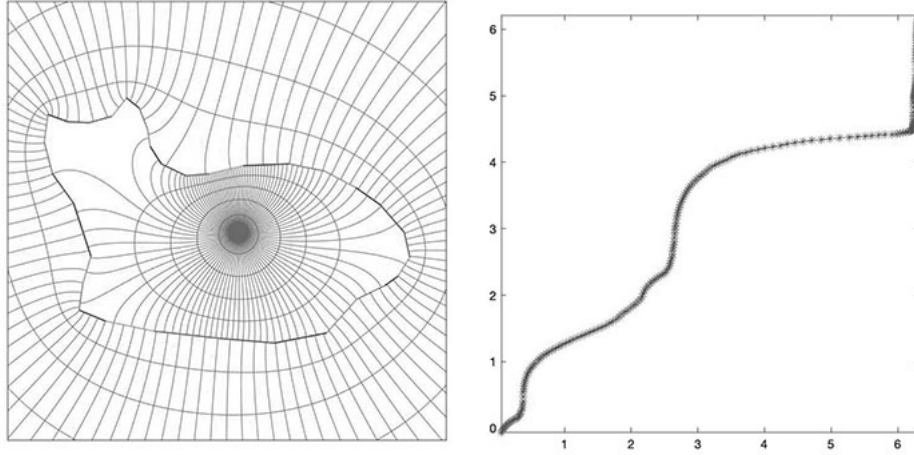
Figure 12.3: Left: the interior and exterior Riemann mapping of a Cat silhouette; Right: the welding map, the horizontal axis is the interior angle, the vertical the exterior. The extremes of large and small derivatives.

What makes these Sobolev metrics really great is that, because they arise from a one-sided invariant metric on a group, the lifted geodesics conserve their "momentum." It is transported by the diffeomorphisms in the lifted geodesic, leading to very simple geodesic equations. The cotangent space to $\mathcal{S}$ at Γ can be thought of as the space of 1-forms ω to $\mathbb{R}^2$, but given only along Γ and that kill the tangent space of Γ. The inverse L^{-1}_{Sob-n} defines a norm here which has degree $1 - 2n$, i.e. it's given by an integral kernel. Upstairs, the kernel is just convolution with a modified Bessel function, namely $\|\vec{x}\|^{n-1} K_{n-1}(\|\vec{x}\|)$ times a constant. As Darryl Holm pointed out to me, if $n > 1$, this is a continuous function at 0 so the completion of the cotangent bundle contains δ functions. This means we can set the momentum to a sum of delta functions on Γ and get ODEs for the resulting geodesics which may be thought of as a kind of soliton. Note that, in these cases, the metric on the cotangent bundle is always weaker than that on the tangent bundle.

The final example is given by the Weil-Petersson metric in a suitable model of the universal Teichmüller space. One starts with the Riemann mapping from a) the inside of the unit disk to the inside of Γ, call this ϕ_{int}, and b) from the outside to the outside, called ϕ_{ext}. The latter can be normalized by asking that infinity is mapped to infinity and that the derivative there is positive real. Then $\psi = \phi^{-1}_{\text{ext}} \circ \phi_{\text{int}}\big|_{S^1}$ is a diffeomorphism of the circle called the *welding* map and is unique up to composition on the right by an a conformal self-map of the unit disk, i.e. a Möbius map. This is illustrated in Figure 3.

It can be shown that this map $\mathcal{S} \mapsto \text{Diff}(S^1)$ creates an isomorphism between $\mathcal{S}$ mod translations and scaling and the group of smooth diffeomorphisms of S_1 modulo right multiplication by the three-dimensional Möbius subgroup of Diff. Once again we have a
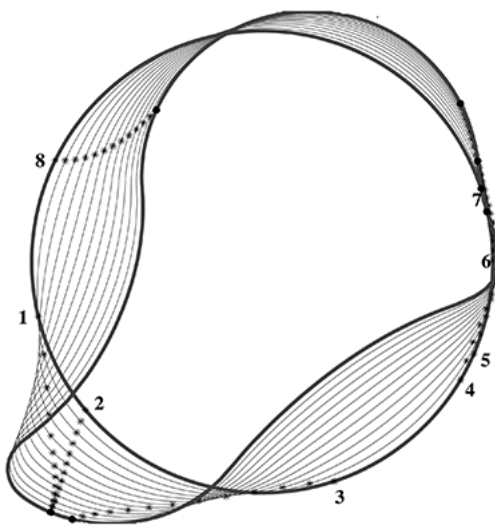
Figure 12.4: A Weil-Petersen geodesic ("Teichon") with momentum at 8 points, taking the unit circle to the outline of Donald Duck, by permission of Prof. Sergey Kushnarev.

one-sided invariant metric on $\text{Diff}(S_1)$ given on the Lie algebra by the formula:

$$\|a(\theta).(\partial/\partial\theta)\|^2 = \int_{S^1} H(a''' - a').a.d\theta = \sum_{n>0}(n^3 - n)|\hat{a}(n)|^2$$

where prime is the θ derivative and H is the Hilbert transform for periodic functions. This defines a homogeneous norm of Sobolev degree $3/2$ on $\mathcal{S}$. The dual metric is given by a simple explicit continuous kernel, hence we have what Holm called "Teichons," geodesics with discrete momenta at a finite set of points. A droll example is given by the Donald Duck head in Figure 4, [Kus09].

This is the famous Weil-Petersson metric. It turns out to be Kähler-Einstein metric with all negative sectional curvatures. The Einstein property says that its sectional curvatures must be small enough to make the Ricci trace finite, so in some sense, I think it is nearly flat. I think it's a gem of a space. The completion of the set of smooth curves in this metric has been shown recently by Chris Bishop to be the set of rectifiable curves that, in their arc length parametrization are Sobolev $3/2$ [Bis20].

Essentially all the material in this section is available on my website, especially the notes from some Pisa lectures [**S-2012b**].

iii.   Zakharov's Hamiltonian

Returning to the notation of the first part, the clue to linking these ideas on shape spaces to gravity waves is to consider the kinetic energy $\int_\Omega \frac{1}{2}\|\nabla\phi\|^2$ as a metric on $\mathcal{S}$. OK, not exactly $\mathcal{S}$ but now curves $z = \eta(x)$ which are suitably tame at infinity (i.e. near the real axis) bounding a 2D slice of an oceanic domain with infinite depth below them. Call this $\mathcal{S}_Z$. We assume the domain has fixed volume, meaning the mean of η is zero. A tangent vector to $\mathcal{S}_Z$ at Γ is a normal vector field $a(s)\vec{N}_\Gamma(s)$ to Γ such that $\int_\Gamma a(s)ds = 0$. The Neumann boundary problem then defines a unique harmonic function in the interior with a as its normal derivative along the Γ and that also goes to 0 at -∞;. If K_{Neu} is the corresponding Neumann kernel for the domain ω, the metric is:

$$\|a\|_Z^2 = \iint_\Omega \tfrac{1}{2}\|\nabla\phi\|^2 dxdz, \quad \phi = K_{\text{Neu}} * a.$$

Note that because ϕ is harmonic, the integral can be rewritten:

$$\iint_\Omega \|\nabla\phi\|^2 = \iint_\Omega \operatorname{div}(\phi.\nabla\phi) = \int_\Gamma \phi.\frac{\partial\phi}{\partial n} = \int_\Gamma \phi.a$$

Thus we can interpret $\phi/2$ as being the dual 1-form of the tangent vector a. In the simplest case $\eta \equiv 0$, $K_{\text{Neu}}(s, x+iz) = \frac{1}{\pi}\log|s - (x+iz)|$, hence $\|a\|_Z^2 = \frac{1}{2\pi}\iint a(s).a(t)\log|s-t|dsdt$. This is exactly the Sobolev $H^{1/2}$ norm because its Fourier transform is $\int|\hat{a}(\xi)|^2 d\xi/\xi$. So we are doing the opposite to what we did strengthening the L^2 norm via derivatives. Here we have a weaker norm on the tangent bundle whose dual is stronger than it is!

On the other hand, to regularize the situation, we have potential energy as well as kinetic energy. This means the gravity wave equation is not a simple geodesic flow but a Hamiltonian flow where the potential is added to the norm squared. This is V. E. Zakharov's beautiful discovery in his paper [Zak68]. The idea is to identify the cotangent space $T^*\mathcal{S}$ with pairs (Γ, ϕ), ϕ harmonic on ω and going to zero at $-\infty$, taking Γ and ϕ as canonical dual variables. The Hamiltonian now is $H(\Gamma, \phi) = \iint_\Omega \left(\frac{1}{2}\|\nabla\phi\|^2 + g.z\right) dxdz$ where the z term, after subtracting an infinite constant, should be interpreted as $\int_\Gamma (g\eta(x)^2/2)dx$. One then checks that, if we write $\delta\Gamma = a$, then

$$\delta H = \iint_\Omega \langle\nabla\phi, \nabla\delta\phi\rangle + \int_\Gamma \left(\tfrac{1}{2}\|\nabla\phi\|^2 + g\eta\right) \delta\Gamma.ds$$

Rewriting the first term the way we did above for the metric, we find $\iint_\Omega \langle\nabla\phi, \nabla\delta\phi\rangle = \int_\Gamma a.\delta\phi$, and we see that the Hamiltonian equations are the same as the equations for gravity waves: $\frac{\delta H}{\delta\phi} = a = \frac{\partial\Gamma}{\partial t}$ and $-\frac{\delta H}{\delta\Gamma} = -(\frac{1}{2}\|\nabla\phi\|^2 + g\eta)|_\Gamma = \frac{\partial\phi}{\partial t}|_\Gamma$.

Can we compute with such a system of equations? A key point is that the Hamiltonian is conformally invariant, hence one can shift everything to the unit disk using the time
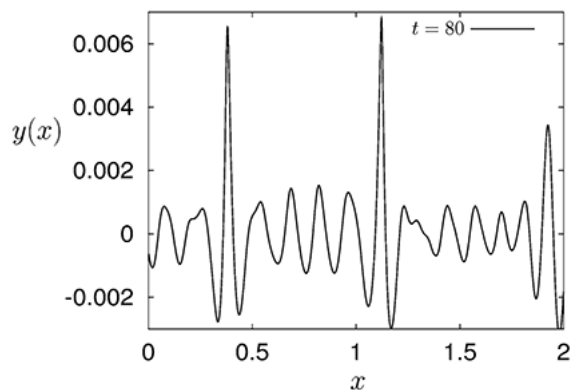
Figure 12.5: Results of a numerical simulation from [ZDE02] showing freak waves developing. By permission Elsevier Press.

varying conformal map from the unit disk to ω. This has been worked out by Dyachenko et al: [DKSZ96, ZDE02].

Perhaps an easy way to do numerical experiments is to replace the infinitely deep 2D ocean with the interior of a simple closed curve, close to the unit disk, as in the shape section above, while making gravity into a central force field based at the origin. Then Fourier series can be used and simulations without changing coordinates might be possible. Finding the rogue wave solutions by this route is a fascinating challenge and might even be of use to the study of genuine ocean rogue waves.

Chapter 13

An Applied Mathematician's Foundations of Math

As a student, I read about the controversies on the Foundations of Mathematics, about the three schools of thought: logicists like Russell and Whitehead, formalists like Hilbert, and intuitionists like Brouwer. However I soon learned that the naive contradictions in set theory (e.g. the barber who shaves all the people in town who don't shave themselves; who shaves the barber?) had been seemingly been put to rest with the acceptance of Zermelo-Fraenkel set theory as the basis of math and that math itself was proceeding just fine. So I fell in line with the Bourbaki program: logic $\rightarrow$ set theory $\rightarrow$ (axiomatic structures, a.k.a. categories) $\rightarrow$ (groups, rings, topological spaces, Lebesgue integration, etc.). The foundations of math had ceased to be an area to which most mathematicians paid attention. The universe of sets is now accepted as a comfortable place to work while set theory itself has become an exotic field, not in the mainstream although recognized as important and legitimate math.

But I had worked two summers for Westinghouse simulating submarine nuclear reactors with primitive computers and learned about the attractions of applied math. Now, having switched in mid-career from pure math back to applied math, I saw that something is missing in the discussion of "Foundations," namely the perspective of applied math. Well before Euclid, math had been invented all around the world as a way to *model* the world's quantitative aspects. With the exception of the contentious Greeks, practitioners had never had much need for abstraction. So is there a fourth way to build the Foundations of Math, to build it all on tangible models, not on thin air? This chapter is a small step arguing for such a radical realignment.

i.   A Warm-up: Arithmetic

Without a doubt, the most basic level of math is the activity of *counting*. In virtually every human society, counting things is practiced and children enjoy doing this from a very early age. This was essential for accounting in the earliest Mesopotamian city-states. Accounting also necessitates the use of addition, subtraction, multiplication and division. This fundamental first stage of math has been very neatly formalized in what's called *Peano Arithmetic* or PA, a set of rules expressed in simple logical terms. These date from Guiseppe Peano's 1889 book *Arithmetices principia* [Pea89] (following earlier work by Pierce and Dedekind). It is now always based on assuming the variables stand for "natural numbers" 1,2,3,... on which there are two operations, plus and times. We will use the standard mathematical symbol $\mathbb{N}$ for the set of all natural numbers. The axioms are the usual rules of arithmetic plus the all-important axiom of induction:

$$\text{For all predicate calculus propositions } P(x),$$
$$P(0) \wedge \forall x\big(P(x) \Rightarrow P(x+1)\big) \;\Rightarrow\; \forall x P(x).$$

This is all a great success, especially because it turns out to be easy to code finite sequences of natural numbers by a single number. Such a coding allows you to formalize arithmetically anything you want involving finite structures like graphs and trees. But there is one little worm eating at its heart: Gödel proved that one can construct statements in PA that, in effect, assert their own unprovability. Obviously, if such a statement could be proven, it would create a contradiction in PA and hence it must be true and unprovable! It's a formal version of the barber paradox. More precisely, he first creates a method of coding Propositions P in PA by numbers $\ulcorner P \urcorner$ and similar codes $\ulcorner Q \urcorner$ for proofs Q (expressed as a sequence of symbols), which allows him to construct an arithmetic proposition $\mathrm{pf}(x,y)$ such that $\mathrm{pf}(\ulcorner P \urcorner, \ulcorner Q \urcorner)$ is true if and only if Q is a PA proof of the Proposition P. He also constructs an arithmetic expression $\mathrm{subst}(n,m)$ such that, starting from any Proposition $Q(x)$ with one free variable x with code n, it gives the code for the Proposition $Q(m)$ obtained by substituting m for x. Then set $\mathrm{B}(x) = \neg(\exists m)\mathrm{pf}(\mathrm{subst}(x,x),m)$. The Proposition $\mathrm{B}(\ulcorner \mathrm{B} \urcorner)$), when you work it out, says that it is not provable! Hence it cannot be proven without making a contradiction, hence it must also be true.

Of course, we can't have a contradiction in PA because PA is model of stuff like accounting in the real world, hence everything would fall apart if Peano arithmetic were not consistent. As is well known, he proved the same awkward result for any formal system of axioms at least as strong as Peano arithmetic. And he went on to show that, in particular, the formal statement of consistency: $\neg(\exists m)\mathrm{pf}(\ulcorner 0 = 1 \urcorner), m)$, is also not provable. But for arithmetic this awkward fact has been explained in a beautiful really illuminating way: it was found that the basic issue is all about defining sequences of numbers that grow at truly HUGE rates.

Kids are always asking "what is the biggest number?," a gazillion perhaps now that trillions have become commonplace in economics. The Rig Veda defines some real biggies

and Archimedes went further with his famous Cattle of the Gods problem. But this is child's play now. Some big numbers are "one off" but the fun ones come from rules that giving an increasing infinite sequence like $10^{\left(10^{\left(10^{\cdots}\right)}\right)}$ where the tower of exponents has n 10's, or, expressed concisely, $x_{n+1} = 10^{x_n}$. It is easy to create a provable proposition of the form $\forall n \exists m P(n, m)$ in Peano arithmetic so that the m whose existence it asserts is exactly x_n. That's a pretty rapidly growing sequence. It raises the question: working only in PA, how rapidly can we make m grow as a function of n, using more sophisticated P's?

The way to get rapidly growing sequences is by composing functions. Composing $x \to x + 1$ with itself n times starting from m yields the addition $m + n$. Composing $x \to x + n$ with itself m times starting from 0 yields the multiplication $n.m$. Composing $x \to n.x$ with itself m times starting from 1 yields the exponentiation n^m. Then things really take off: composing $x \to n^x$ with itself m times starting from 1 yields $n^{n^{n^{\cdots}}}$ with a tower of m nested exponents. This is the basis of the killer construction of what is called the *Ackermann function* after Hilbert's pupil who invented it. f_n is the sequence of functions from natural numbers to natural numbers given by:

$$f_{n+1}(m) = f_n(f_n(\cdots(f_n(1))\cdots)), m \text{ repetitions of } f_n.$$

Even faster growing is Ackermann's function, the sequence $\text{Ack}(n) = f_n(n)$. But there is no reason to stop here! There is a hierarchy of fastness associated to the countable ordinals that we bring in in the next section. But is most interesting is that there is a limit to the growth rate of any function $m = f(n)$ definable by a Peano arithmetic provable formula of the form $\forall n \exists m P(n, m)$.

To get such functions, we can use Ramsey theory. The simplest example of this theory is this: consider a party with N people, some of whom know each other and others are strangers (we exclude any in-between "maybe I met you ..." cases). You ask about "homogeneous" groups at the party in the sense that either everyone in the group knows everyone else or, conversely, no-one knows anyone else. The result is this: for any number n, if the party is big enough, there will always be a homogeneous group of one of these types with n people. For example, if you want a group of size 4 of all friends or all strangers, the party must have at least 18 people there. Let's generalize: take the set $S = \{1, 2, ..., N\}$ and define a (k, r)-coloring to be a rule assigning one r colors to every subset of k numbers. For our party example, $k = r = 2$, whether a pair knows or doesn't know each other is the 'color' of the pair. Then it's a nice theorem that for all n, there is an N such that, for *any* (k, r)-coloring of 1,2,...,N, there will be some subset S of n objects that is mono-colored, i.e. all subsets of size k in S have the same color.

This was Ramsey's lovely original theorem. How big N must be as a function of (n, k, r) turns out to be really hard to work out exactly, although lots of upper and lower bounds are known and N does grow exponentially fast. But the real kicker comes when you ask a bit more of your mono-color subset S: require that the *size* of S be bigger than both k and of the minimum of the numbers of its members. Paris and Harrington called such sets *relatively*

large. With this added requirement, the required N as a function of $p = \min(n, k, r)$ grows really fast. In particular, Paris and Harrington proved the beautiful fact that $N(p)$ grows faster than any function definable by a PA provable formula $\forall n \exists m P(n, m)$, [PH77].

By the way, the proof of this souped-up Ramsey theorem is not especially difficult but it leaves PA by relying on the *infinite* case of theorem: "Given any (k, r) coloring of the entire set $S = \mathbb{N}$ of all natural numbers, there are mono-colored infinite subsets." The proof of this is very neat. Take the case $k = r = 2$ for simplicity (the general case is almost identical) and call the two 'colors' red and blue. Start with any $i_0 \in S$ and divide the rest of S into ones forming red and blue sets when joined to $i_0 \in S$. One of these sets must be infinite, say the red one and call it S_1. Choose any $i_1 \in S_1$. Divide the rest of S_1 into forming red and blue sets when joined to i_1. One of these is again infinite, say the blue one now. Continue in this way defining an infinite sequence $\{i_0, i_1, i_2, ...\}$. Either red or blue must have come up infinitely often! Take the corresponding $\{i_j\}$'s and one checks this is mono-colored. The finite case is reduced to the infinite one by contradiction: if the finite case is false, one considers all 'bad' (k, r) colorings of $\{1, 2, ..., N\}$ and makes a tree out of them by asking when one example extends another. If the extended Ramsey were false, this tree would be infinite and thus have an infinitely long branch and this would be a contradiction to the infinite Ramsey theorem. None of this is especially complex but it does involve infinite sets that take it outside Peano arithmetic.

Let's summarize: Gödel showed Peano arithmetic could not prove its own consistency. But now we have a clear explanation of this: no theorem of the form $(\forall n)(\exists m)P(n, m)$ can be proven in this weak system if the required m grows too fast. Moreover, we have theorems that are readily proven using standard math tools in set theory that do define functions growing at least that fast. Clearly, Peano arithmetic is a great system but it is not a satisfactory foundation for mathematics. What really do we need for a full "Foundations of Mathematics"? I think there are three approaches: i) the minimal way, ii) "go for broke" set theory and iii) basing it firmly on what math seeks to model via type theory. Let me take these up one at a time.

ii. Being conservative with second order arithmetic

In 1975, Harvey Friedman started a major program to analyze the foundations of math that he called "reverse mathematics." Instead of seeking to derive mathematical theorems from axioms, one asked "what axioms are needed for each theorem." It tuned out that a remarkable fraction of present day math can be stated and proven using a weak system called second order arithmetic. This based on having two types of variables, one for positive integers (as in PA) and one for subsets of the first, connected by $\in$, "member of." With a whole series of axiom systems, ranging from weak systems to stronger ones, it can be seen as underpinning successively more and more math, especially analysis, as it is practiced today. Stephen Simpson has written a wonderful exposition of this approach in his book

[Sim10].

More precisely, second order arithmetic is a simple extension of PA, known as Z_2, in which you have variables: natural numbers $n \in \mathbb{N}$ and (ii) subsets $S \subset \mathbb{N}$, two operations $+$, $\times$ for natural numbers, two constant natural numbers 0 and 1 and two predicates $n < m$ and $n \in S$. These are subject to the usual Peano axioms with induction being expressed by set membership:

$$\left(0 \in S \wedge \forall n \left(n \in S \Rightarrow n + 1 \in S\right)\right) \Rightarrow \forall n \left(n \in S\right)$$

and with the all important set of comprehension axioms, one for each formula $\phi(n)$:

$$\exists S \, \forall n \left(n \in S \Leftrightarrow \phi(n)\right)$$

It may sound like basing mathematics on Z_2 is ridiculous – how are we going to construct things like measure theory and Lebesgue integration on such a discrete world of sequences of natural numbers? Amazingly, this is not so hard. It is based on a series of codings allowing quantifiers over more and more complex structures. First we embed $\mathbb{N} \times \mathbb{N} \subset \mathbb{N}$ by the invertible map

$$(n, m) \mapsto k = (n + m)^2 + m,$$
$$k \mapsto (n, m), m = (\max\{s \in \mathbb{N} | s^2 < k\}^2 - k), n = s - m$$

This gives us codings for pairs, triples, etc. Then secondly define a code for rationals by coding triples (n, m, s), that stand for $+n/m, 0$ and $-n/m$ and s is a code for sign and has three possible values, $\ulcorner + \urcorner, \ulcorner 0 \urcorner, \ulcorner - \urcorner$ (one can use any three codes for signs, e.g. 1,2,3). In the code for 0, require $n = 0, m = 1$ and for non-zero rationals, require n minimal among fractions representing this rational. All this gives unique codes for rational numbers and gives a predicate $Q(k)$ true for only such codes. Thirdly, we define real numbers á la Dedekind as subsets S of the rational number codes that are "cuts" as usual. After that, we easily define predicates $\mathbb{R}(S)$, $\mathtt{add}(S_1, S_2, T)$, $\mathtt{mult}(S_1, S_2, T)$ stating S is code for a real number, resp. T is the code for a real which is the sum or product of the reals coded by the S's, etc. Thus we have the full algebra of real numbers in Z_2.

Although Z_2 lacks powerful ways of dealing with infinity, it allows one big advance: it can define the hierarchy of countable ordinals. These begin by adding ω as an object bigger than all natural numbers and continuing to build bigger ordinals using a form of arithmetic:

$$\{1, 2, 3, \cdots, \omega, \omega + 1, \omega + 2, \cdots, 2\omega, 2\omega + 1, \cdots, 3\omega, \cdots, \omega^2, \omega^2 + 1, \cdots, \omega^2 + \omega, \cdots$$
$$, 2\omega^2, \cdots, \omega^3, \cdots, \omega^\omega, \cdots, \omega^{(\omega^\omega)}, \cdots, \omega^{\left(\omega^{(\omega^\omega)}\right)}, \cdots, \epsilon_0, \cdots\}$$

Just to recall: by definition, an ordinal is simply a linearly ordered set in which all descending chains are finite ("well ordered"). The ordinals themselves form a linearly

ordered set and, usually, each ordinal is identified as the set of all smaller ordinals, i.e. $4 = \{1, 2, 3\}, \omega = \{1, 2, 3, \cdots\}$ but this is not necessary. One can also think of all countable ordinals as sets of points on the real line with limit ordinals corresponding to limit points, all successor ordinals inside open sets by themselves. Ordinals have an algebra: adding two just means taking their disjoint union and putting one set after the other; multiplication means taking the product set and using "lexicographic" order, that is $(x, y) < (u, v)$ iff $y < v$ or $y = v, x < u^1$.

Returning to Z_2, you form ordinals by defining special codes for these objects that are manipulated by their own special operations. We formalize this as follows: abstractly, an ordinal is defined by a sequence of codes forming a subset $X \subset \mathbb{N}$ plus its own relation $\ulcorner < \urcorner \subset X \times X$ on $X \times X$, with a smallest element $\ulcorner 1 \urcorner$ making it linearly ordered and *well-ordered*, meaning it has no infinite decreasing subsequence:

$$\neg (\exists\, f : \mathbb{N} \to X)(\forall i)\left[\ulcorner < \urcorner (f(i+1), f(i))\right].$$

Such a pair $(X, <)$ is called a countable ordinal and it is easy to see that the "small" ones look like the above sequence written with ω's. Besides building a fun hierarchy, this allows us to extend PA induction to the more powerful "trans-finite induction." Given $S \subset \mathbb{N}$ and $(X, <)$, a countable ordinal. Then:

$$\left(\ulcorner 1 \urcorner\right) \in S \wedge (\forall j \in X)[(\forall i \in X)(i < j) \Rightarrow i \in S]) \Rightarrow X \subset S$$

Trans-finite induction allows us to now define codes for Borel subsets of reals, and from these to Lebesgue integration, Banach spaces, the whole machinery of analysis! What was quite remarkable to me, when I first read this, is that Borel sets can be described by countable rooted trees. The root stands for desired Borel set and the leaves of the tree carry codes for intervals with rational (or infinite) endpoints (open, closed or semi-open). The Borel set is built top down, all the leaves being a finite distance from the root and all branches starting at the root must lead to a leaf in a finite number of steps. Finally, we require all nodes to have countable (infinite or finite) sets of edges numbered 1,2,3, ..., and be labelled as additive or subtractive. The Borel set is built by working down the tree attaching a subset to each node, forming a union at additive nodes, an intersection at subtractive nodes. Countable unions and intersections of such sets are made by building a bigger tree with one more layer and complementation merely riffles up the tree flipping positive and negative, and complementing the ultimate leaf intervals.

The possibility of giving codes to Borel sets by means of subsets of $\mathbb{N}$ means that the cardinality of the set of all Borel sets is not greater than that of the reals themselves. Pretty much all of contemporary mathematics has no need for higher cardinality sets. Topological spaces satisfying the second axiom of countability can be defined, with points defined by

---

[1] The ordinal ϵ_0 in the sequence above is the limit of the sequence indicated on its left and is the smallest ordinal ϵ such that $\epsilon = \omega^\epsilon$.

equivalence classes of sequences of the basis open sets whose intersection is that point. Likewise, separable Banach spaces can be defined.

What can't be defined is set theory itself, and the analysis of non separable topological and function spaces, probably others. But mainstream math all seems to go through. For lots more detail, see Simpson's book [Sim10]. His book is mainly concerned with six sets of *axioms* with weakened forms of the all-important comprehension axiom. This is the rule that allows you to define new subsets $S \subset \mathbb{N}$ as the set of numbers satisfying some predicate. In the weakest system, RCA_0, which he calls *constructivist math* (following Errett Bishop), restricts comprehension to recursively definable predicates. An intermediate point between constructivism and full impredicative Z_2 is the system ACA_0 which allows arithmetical comprehension, that is comprehension only for formulas with number quantifiers but no set quantifiers. But, ignoring intuitionism, the most reasonable choice is the full comprehension axiom leading to the usual powerful math world.

iii.   The Standard Foundation: ZFC

As all mathematicians know, the world of math seems so much simpler if you have only one kind of variable, namely a set, and one predicate $\in$. The now universally accepted version is known as ZFC (for Zermelo and Fraenkel who developed it plus the axiom of choice). This is a first order theory in predicate calculus, i.e. there is only one type of variable, called a set, and two binary predicates $x \in y, x = y$. One can describe its axioms somewhat informally as follows: (i) We have an axiom of equality: two sets are equal if and only if they have the same members, (ii) an axiom of foundation which is equivalent to saying that there is no infinite sequence of members $x_{n+1} \in x_n, n = 1, 2, \cdots$ going "down" and "down"[2]. All the other axioms assert the existence of some new set:

- there exists an empty set $\varnothing$,
- for every set x, there is a singleton set $\{x\}$ whose only member is x
- for any two sets x, y, their union $x \cup y$ is a set, (hence we get unordered and ordered pairs via $\{x, y\} = \{\{x\}, \{y\}\}$ and $\langle x, y \rangle = \{\{x\}, \{x, y\}\}$,
- the infinite set ω exists (e.g. constructed via $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \cdots\}$, see below),
- for every set x, there is a set $\bigcup x$ whose members are the members of its members,
- (power) for every set x, there is a set of all its subsets $\mathcal{P}(x)$ (leading to products $X \times Y =$ set of ordered pairs of elements of $X, Y =$ an easily defined subset of of $\mathcal{P}(X \cup \mathcal{P}(Y))$,
- (choice) for every set x, there is a map $f : x \to \bigcup x$ such that for all non-empty members u of x, $f(u) \in u$,
- (replacement) for all formulae $\phi(x, y, A)$ such that for all $x \in A$, there is a unique y satisfying ϕ, then there is a set B formed from all these y's. (This axiom implies the

---

[2]In the presence of the other axioms, this turns out to be the same as saying that every x has a member y disjoint from it.

better known "bounded comprehension" axiom: ϕ definable subsets of any set are new sets.)

As is well known, we cannot ask for unlimited comprehension, that is making a set out of $\{x|\phi(x)\}$ for predicates with unbounded quantifiers without running into contradictions, e.g. with $y = \{x|x \notin x\}$. So one instead calls the objects $\{x|\phi(x)\}$ *classes*. For instance, one has the class formed by all sets whatsoever, called V. The key step is then to define an *ordinal* as a set κ whose members are well ordered by membership (or inclusion). This leads to the class of all ordinals "Ord" that is itself well ordered beginning with:

$$0 = \varnothing, 1 = \{\varnothing\} = \{0\}, 2 = \{\varnothing, \{\varnothing\}\} = \{0, 1\}, \cdots, n+1 = n \cup \{n\}, \cdots, \omega, \omega + 1, \cdots$$

We have successor ordinals whose members have a maximal element and limit ordinals without a maximal element. Cardinals are then those ordinals which cannot be mapped to an element of themselves bijectively. These allow the universe V to be structured into a tower of sets using transfinite induction:

- $V_0 = \varnothing$,
- $V_{\kappa+1} = \mathcal{P}(V_\kappa)$
- If κ is a limit ordinal, $V_\kappa = \bigcup_{\lambda < \kappa} V_\lambda$.

The *rank* of a set X is the smallest ordinal α such that $X \in V_\alpha$.

Right from the beginning, after Cantor's discovery that the cardinality of $\mathbb{R}$ was greater than that of $\mathbb{Z}$, or, more generally, for any set X, the cardinality of $\mathcal{P}(X)$, was bigger than that of X, it was clear that the power set construction created huge cardinal numbers. How huge can you get? Hausdorff, in 1908, introduced the concept of *inaccessible* cardinals. These are cardinals κ that, whenever they equal the least upper bound to a set S of smaller cardinals, the cardinality of the approximating set $|S|$ cannot be less than that of κ. A set of fewer than κ just can't get big enough to reach κ. The axioms don't prove such cardinals exist but then, *why not* add an axiom saying they exist and play with them? In fact, using this axiom, one can prove the consistency of ZFC because it implies that the set V_κ is an "inner" transitive model of ZFC. (Here transitive means that if any set is in V_κ, so are its members.)

It's awfully hard to believe that ZFC is not a consistent theory because everything it deals with is so simple and transparent. Moreover, adding one inaccessible cardinal κ doesn't seem very dangerous and it has a lot of advantages. A key one is that it produces an *inner transitive model* of ZFC, meaning the model is a set in V (hence "inner") such that members of its members are members (thus "transitive") and, using the restriction of $=$ and $\in$, forms a model of ZFC. It is easy to see that the set V_κ (with its inherited relations $+$ and $\in$) is such an inner transitive model of ZFC. A big question is how small can such a model be? The standard Löwenheim-Skolem theorem in predicate calculus shows that every consistent predicate calculus theory has a *countable* model. We would like an inner one for ZFC.

The key to getting small inner models is the second key tower, Gödel's constructible sets defined again by transfinite induction:

- $L_0 = \varnothing$,
- $L_{\kappa+1}$ = set of sets defined by predicates with quantifiers and constants from L_κ, called $\mathrm{Def}(\kappa)$,
- If κ is a limit ordinal, $L_\kappa = \bigcup_{\lambda < \kappa} L_\lambda$.
- $L = \bigcup_\lambda L_\lambda$.

Clearly $L_\alpha \subset V_\alpha$ for all α, i.e. constructible sets form a sub-tower of $\{V_\alpha\}$. If κ is an inaccessible cardinal, then L_κ is a *set* and, in fact, it is not hard to see that it is an inner transitive model of ZFC too. We can then ask for the smallest ordinal α for which L_α is an inner transitive model of ZFC. This is called the *minimal model* of ZFC and, by Löwenheim-Skolem and the "condensation" lemma of Gödel, it is countable! Wow: a nifty smallish structure satisfying ZFC. This sounds simple and natural but note that it requires the existence of an inaccessible cardinal.

This is where Paul Cohen took off, inventing *forcing* and constructing lots and lots of models of ZFC showing how lots of assertions about sets could be either true or false, i.e. if ZFC is consistent, then adding either the truth or its falsity of the new assertion as an extra axiom is consistent. In particular, he showed that it was consistent to assume the continuum hypothesis is false (Gödel had shown that CH is true in every L_alpha, hence consistent). Cohen's technique of forcing uses transfinite induction but now to define not special sets like the constructible ones, but the extra sets that have to exist if you add a new set, G, to the starting model M. He defines a tower of extra sets called "Names" which are potential sets in a bigger model $M[G]$ in which one new set G has been added, hence demanding a zillion more sets derived from G in order to model ZFC. One then uses trans-finite induction again to define a new $=$ and $\in$ relation between the names which collapses them into the desired model. A nice exposition is in Wikipedia. Robert Solovay extended forcing ideas, constructed what he called "random real numbers" x such that $M(G) = M(x)$.

The use of larger and larger cardinals has become the credo of modern set theory: find properties that create yet bigger cardinals so long as there is no obvious reason why they shouldn't "exist." This theory is quite beautiful and deep. I think it is worth spending some time with what are arguably two of the most significant of these gigantic cardinals. The first of these are the Ramsey cardinals κ, defined by possessing a strong "Ramsey theorem"-like property. For example, let $\mathcal{P}^{\mathrm{fin}}(X)$ be the set of finite subsets of a set X. Consider a "coloring" $f : \mathcal{P}^{\mathrm{fin}}(X) \to \{0, 1\}$. Then one such Ramsey cardinal κ is defined by requiring that, for all colorings, $\mathcal{P}^{\mathrm{fin}}(\kappa)$ has a subset S of cardinality κ all of whose finite subsets have the same color.

To explain why the Ramsey property is so important in set theory, I need to introduce the concept of *indiscernibles*. Given a set of propositions $\phi_\alpha(v_1, \cdots v_n)$, an ordered set X of arguments for the ϕ's is said to be indiscernible if and only if, for all α, $\phi_\alpha(x_1, \cdots, x_n) \equiv$

$\phi_\alpha(y_1, \cdots, y_n)$ for any two ordered sets of n elements $x_1 < x_2 < \cdots < x_n$, $y_1 < y_2 < \cdots < y_n \in X$, i.e. the ϕ's see no difference between increasing sequences in X. Ramsey comes in if you have an infinite set X and you seek an infinite indiscernible subset of X. When there is only a finite set of ϕ's, this is easy: you assign the color to every set of truth values for the propositions ϕ_α. Then the standard Ramsey theorem gives you a subset from which all sets of arguments make each ϕ_α either true or false. A strengthening of this is due to Ehrenfeucht and Mostowski [EM56]: if you start with a theory with possibly an infinite set of axioms (but with some infinite model), then you can add new constants c_x for all $x \in X$ and axioms stating that the c_x are indiscernible for all propositions in the theory and still have a consistent theory. The idea is that any inconsistency must result from using only a finite set of propositions of the original theory and we just saw that then indiscernability is consistent.

Assuming that Ramsey cardinals exist, Jack Silver and Robert Solovay [Sil71, Sol67] used the idea of indiscernibles in an astonishing way that explains in a remarkable way what the constructible universe L is all about and shows that it is not all that complex. For any model $\mathcal{M}$, one can ask for what propositions in this model are *true*. This called the *theory* of the model, $T(\mathcal{M})$, and, via Gödel numbering, it can be described as a subset of $\mathbb{N}$. What Silver proved is that, *assuming Ramsey cardinals exist*, there is a miraculous set of ordinals $I \subset Ord$, closed under limits, including all uncountable cardinals but starting with certain countable ordinals, such that for all $\alpha \in I$, $I \cap L_\alpha$ is a set of indiscernibles in L_α. Their key property (which seems miraculous to me) is that *all the theories $T(L_\alpha)$ for $\alpha \in I$ are equal!* (see [Kan03], Chapter 2, Theorem 9.14.) This theory is denoted $0^\sharp$, called "zero-sharp" and was shown by Solovay to be a so-called Δ_3^1 subset of $\mathbb{N}$.[3] This means it can be described in second order arithmetic by propositions of a natural number n of both the forms $\forall x \exists y \forall z \phi(n, x, y, z)$ and $\exists x \forall y \exists z \psi(n, x, y, z)$ (quantifiers here ranging over subsets of $\mathbb{N}$).[4] I is naturally called the class of Silver indiscernibles. It is awfully tempting to say $0^\sharp$ is the final set theory and, in principle, settles all of math, but this is a fever dream. Why on earth should all real numbers be constructible, that would be absurd. More on this in the last section.

What is astonishing here is that this skirts Gödel's incompleteness theorem. Theories including PA can never be complete yet here the complete definition of "truth" in *constructible* set theory is given by an explicit ZFC formula (in full set theory) and one that is not that complicated either. What makes this possible is that $0^\#$ itself is *not* constructible. The proof of this depends on the detailed analysis of models with infinite sequences of indiscernible cardinals. Having all these large indiscernible cardinals seems to mean that nothing new is going on in the higher layers of the ladder L_α and, in fact, the universe of

---

[3]This is actually a weakening of the theory that deals with propositions $\phi(x_0, \cdots, x_k)$ which are true in L_α) if you plug in any infinite increasing sequence $\alpha) < \cdots < \alpha_k < \cdots < \alpha$ of ordinals in I.

[4]Hugh Woodin pointed out to me that this is not unique: if $\mathcal{M}$ is any model such that $T(\mathcal{M}) \in \mathcal{M}$, then this model has a (class) forcing extension of $\mathcal{M}[G]$ in which this theory is defined by a similarly simple statement.

constructible sets is generated by "Skolemizing" the ordinals in I. This means converting propositions $\phi(x, y_1, \cdots, y_n)$ into functions $x = f(y_1, \cdots, y_n)$ where f picks out the "smallest" possible constructible set x satisfying ϕ with respect to the canonical well-ordering of L or just $\varnothing$ if none exists. You then plug in a sequence of the indiscernibles I for the y_i's. Details can be found in Jech's book [Jec97], Chapter 18 or Kanamori's book [Kan03], Chapter 2, §9.

Ramsey cardinals have also been central in another development. Harvey Friedman, following the concept of "Reverse Mathematics," has worked extensively on finding simple combinatorial assertions that are equivalent to the consistency of various strengthenings of ZFC. Much of his work deals with dense order relations, especially the ordering of $\mathbb{Q}$ and intervals within it. I give here a sketch of some of the ideas in his fully detailed 2011 preprint "Invariant Maximal Cliques and Incompleteness" [Fri11]. The main theorem concerns graphs whose vertices are $(\mathbb{Q}[0, n])^k$ and whose edges are "order invariant," meaning whether one vertex is connected to another only depends on which of the three $>, =, <$ order relationships hold in the $2n$-tuple formed the two vertices. For each k, n, this means there are only finitely many such graphs but he requires k, n to be really huge. He then asks for maximal cliques which are closed under the following curious equivalence relation: $(x_1, \cdots, x_n) \sim (y_1, \cdots, y_n)$ if and only if their order relations are the same and there is a $z \in \mathbb{Q}[0, n]$ such that $x_i = y_i$ whenever one of them is less than z and x_i, y_i are both positive integers whenever both are larger than z. He calls this "upper $\mathbb{Z}^+$-equivalence." His main theorem is that the existence of such maximal cliques is equivalent to the consistency of a set theory with a cardinal possessing the *stationary Ramsey property*. This is the usual Ramsey property on the cardinal κ but asking for a color homogenous set which is also stationary[5].

How, in heaven's name, can Harvey connect simple statements about finite sets of rational numbers with large cardinals? I was quite intrigued about how he managed this. In the direction, existence of large cardinals implies existence of invariant maximal cliques, the basic idea is to consider $\mathbb{Q}[0, 1) \times \kappa$ and put a linear order on it by: $(p, \lambda) < (q, \mu)$ if and only if either $p < q$ and $\lambda = \mu$ or $\lambda < \mu$. What happened is that he has filled the hole between every ordinal and its successor with a copy of a rational semi-open interval. This makes a dense linear ordered set. In here he uses Ramsey with fancy coloring and concocts the needed clique.

In the other direction, he defines a very intricate and curious sort of order-invariant graph that reminds me of Rube Goldberg cartoons. Using this and a maximal invariant clique, he defines an epsilon relation in the countable set $\mathbb{Q}[0, 1)^{14}$ and shows that it satisfies most of the axioms of set theory. What it most definitely lacks is the axiom of foundation. But it has a ladder of ordinals and he can define Gödel's constructible sets for this and –

---

[5]First we define "club" sets $C \subset \kappa$ by asking that their sup is κ and which are "closed": the sup of all subsets $C' \subset C$, bounded by a smaller ordinal, is in C. These are sort of really thick subsets of κ and two such always have a non-empty intersection. Then a subset is stationary if it has a non-empty intersection with all club subsets.

lo and behold – this is a model of ZFC+Con(SRP).

A much stronger hypothetical property, for cardinals, i.e. creating a seriously bigger universe of sets, are the cardinals κ called *measurable*. They are defined by the existence a strong type of *ultrafilter* on κ: a set $\mathcal{F}$ of subsets of κ (i) containing all supersets to any member, (ii) not containing any singelton, (iii) containing every subset or its complement and (iv) closed under the intersection of any λ members for all $\lambda < \kappa$ (the usual ultrafilter asks this only for finite intersections). The reason this is called measurable is that if you define a map $\mu : \mathcal{P}(\kappa) \to \{0, 1\}$ to be 1 on $\mathcal{F}$, 0 everywhere else, then this is a measure on κ that is λ-additive for all $\lambda < \kappa$.

Following work of Ulam, Solovay showed [Sol71] that if a measurable cardinal exists, there is a model of ZFC in which the real numbers already form a cardinal in which Lebesgue measure can be extended to a countably additive measure on all subsets of the reals! Although this sounds impressive, note that such an extension cannot be translation invariant because of the usual argument using a set $X \subset \mathbb{R}$ of coset representatives of $\mathbb{R}/\mathbb{Q}$, i.e. these are not very useful measures. But the vast zoo of subsets of $\mathbb{R}$ brings up issues of what is not merely equi-consistent with ZFC but also what is, in fact, *true* and this leads to another angle on foundations.

The theory of measurable cardinals and, especially, that of the large number of proposed even larger cardinals, is tied up with the construction of *classes* (not sets) $M \subset V$ and maps $j : V \to M$ that are "elementary embeddings". Now there is no definition of truth for V or other classes, no set $T(V)$, so what does elementary equivalence mean? You do the best you can: you ask for all sets X, that the restriction of j to X is an elementary embedding of X to $j(X)$. This is in the model theoretic sense for the structures $(X, \in)$ and $(j(X), \in)$. The mind-boggling idea, due to Dana Scott, was to consider the set U of all maps $f : \kappa \to V$ mod the equivalence relation

$$f \equiv g \text{ iff } \left[\{ x \in \kappa | f(x) = g(x) \} \in \mathcal{F} \right].$$

One defines the relation $\in$ for U in the same way. Then $(U, \in)$, by the simple Mostowski collapse, is $\in$-isomorphic to a unique class $M \subset V$ with its induced $\in$. j is defined by the constant maps $j(x)(a) = x$ for all $a \in \kappa$. κ is recovered from (M, j) by the fact that it is the smallest ordinal not mapped to itself by j.

To me, it feels as though taking the gigantic object V literally and playing with it like this, as though we might really know what it is, leaves the known world completely behind. I had always thought of V as a vague totality, a bit like the universe we live in whose totality seems unknowable. But this is truly the bread-and-butter of contemporary set theory. The next section describes my own favorite alternative.

iv. The Applied Perspective

My main aim in this chapter is argue for a third approach to the foundations of math, one growing out of science as a whole and not dealing with abstractions whose relevance

to the real world is doubtful. I really don't mean to insult anyone by saying this, as I believe ZFC set theory is as profound and as subtle as any branch of math that I know. But it does remind me of the 13th century scholastic philosophy of the Aquinas and others trying to merge Catholicism and Aristotle by combining relentless logic and very literal interpretations of mystical issues to which the words "true" and "false" don't apply in any transparent way. In fact, the two set theorists Aki Kanamori and Menachem Magidor wrote "The adaptation of strong axioms of infinity is thus a theological venture" , [KM78]. Moreover, in the first flowering of set theory in the early 20th century, the Russian set theorists Dmitri Egorov and Nikolai Luzin both linked their study of complex subsets of the reals to a mystical approach to God called "name worshipping", described in the excellent book [GK09]. They were strongly motivated to "name" as many such subsets of the real numbers as their theory revealed to them. Set theory is spinning off strange working hypotheses such as the existence of measurable cardinals (or the even wilder axiom of determinacy) that no one is sure are even consistent nor do many people feel they are true in any absolute sense. And what happened to the role of math as the embodiment of rock-solid certainty, of unimpeachable arguments, a role it played from the time of Euclid to the philosophy of Kant and beyond.

To an applied mathematician, the essence of mathematics is to find parts of the booming, buzzing world that can be described by numbers and finding the rules that the measured numbers obey. In its earliest stages, there were two things that led to mathematical models. One was counting, driven by the need to barter goods and keep accounts as well as keeping track of the cycles of time, like counting the days in a year. The other was geometry driven by construction and surveying. These led, of course, to integers and their operations on the one hand; and to triangles and ratios of distances on the other. Both are described beautifully in Euclid's *Elements*. But there we also find what I might call the "original sin" in Book V. The Greeks were deeply concerned with how discrete sequences of events combine with distances, such as in the paradox of Achilles and the Tortoise. The tortoise has a $100'$ lead and it takes Achilles 10 seconds to reach the tortoise's starting point (I'm not aiming for accuracy here). But by then the tortoise has moved $10'$ further. So Achilles needs 1 more second to reach the tortoise's new location. Now he has moved $1'$ further on. Then Achilles needs 0.1 seconds to reach this, etc., etc. In other words, Achilles must complete an infinite number of discrete actions to reach the tortoise. No problem you say, it takes him 11.11... seconds to reach the tortoise. Yes, we do have a way of *reducing* geometry to sets of whole numbers by using infinite decimals or, more generally, by approximating rationals. An amazingly abstract formulation of this reduction is in Book V, widely credited to Eudoxus. That part of the Elements is identical to the modern use of the Dedekind cut except that Euclid took both distances and whole numbers to be *given* and needing to be related, while Dedekind used the same set-theoretic technique to *construct* distances from whole numbers.

In the 21st century, we might say Dedekind used set theory while Euclid was using *Type Theory*, the approach to foundations in which the underlying variables belong to

more than one type. If we set out to build the foundations of math for science, it is more logical to use types than to use sets. Physicists, chemists, biologists need distances, time intervals, speeds, weights, densities, etc. and don't think of the right mathematical model for underlying number as sets of fractions forming a Dedekind cut! For them, every measurement comes with its "dimension," e.g. momentum is (mass)×(distance)/(time), or so-and-so-many gram·meter/seconds. On the other hand special, general relativity and quantum mechanics have found various "natural" units and, in the end, we may settle on using a real line on which we may fix both an origin, a positive direction and a unit depending on what the application needs in any given situation. But what is measured in the world is also an approximate real number, not an integer nor an exact real number. Going back to Dedekind cuts and Euclid Book V, it is certainly a theorem now that the ratio of two distances x and y is determined by the set of all pairs of positive integers n, m such that x repeated n times is greater than y repeated m times. This is the bridge between the discrete world $\mathbb{N}$ and the continuous world $\mathbb{R}$.

So we are led to work with a set theory with two constants, $\mathbb{N}, \mathbb{R}$, subject to axioms making their members integers and real numbers respectively, with the usual basic properties.[6] Of course we do need sets for virtually all abstractions describing the structures we find in the real world. So we seem to want two types, sets of each, sets of sets, functions between them, etc. But how much of ZFC do we accept as being *real things?* Finite sets are certainly unobjectionable and we need an axiom of infinity to cover the integers. Two of Zermelo-Fraenkel's axioms are problematic: the axiom of choice and the power set axiom. Both of these lead almost instantly to immensely complex sets. A radical approach to this was formulated by Saul Kripke and Richard Platek [Kri64]. Their theory, referred to simply as KP set theory, is the same as ZFC except that (i) it throws out both the choice and power axioms and (ii) restricts comprehension and replacement to predicates with bounded quantifiers. This is known as the "predicative" approach and radically handicaps mathematicians practicing it.

Taking choice and power in turn, what's the big deal? Well, introducing the axiom of choice gives us coset representatives of $\mathbb{R}/\mathbb{Q}$. This is a bizarre set. Projecting this to the circle $\mathbb{R}/\mathbb{Z}$, we get an unmeasurable subset because the whole circle is now decomposed into a disjoint union of the countably infinite set of translates of these coset representatives by $\mathbb{Q}/\mathbb{Z}$. No translation invariant measure can be assigned to the coset space because its measure would be the quotient of the measure of the circle by infinity. This was a precursor to the famous Banach-Tarski decomposition of the unit 3-ball into a finite set of pieces that can be rigidly reassembled into a ball of twice the size [BT24]. Personally, I don't believe this set of coset representatives is a "real" object, let alone the Banach-Tarski pieces. Neither are things that are met with in the scientific study of space. Turbulence creates a need for some awfully complicated subsets of space but nothing like the above.

---

[6] A more fundamental shift would be to have two categories, the category of rings and the category of topoi and add axioms stating the existence of the basic examples.

But the issue of modifying the axiom of choice has a nice solution: I believe that restricting it to *countable choice* suffices for the development of virtually all contemporary math. Paul Bernays realized that the right way to formalize countable choice is this:

AXIOM OF DEPENDENT CHOICE: *given a relation $R \subset X \times X$ such that for every $x \in X$, there is some $y \in X$ with xRy, and given $x_1 \in X$, there is a sequence $\{x_n\}$ with $x_n R x_{n+1}$ for all n.*

Using this axiom DC, all of core math is just fine although we need to jettison non-separable spaces. This is similar to what we found with Z_2. Solovay [Sol70] showed that ZF+DC+∃inaccessible is consistent with assuming all subsets of $\mathbb{R}$ are really nice, Lebesgue measurable and with both the perfect set property and the property of Baire. However, his model consists only in constructible sets so it's really small.

The issue of finding the "right" power set axiom is much subtler. The most visible problem is that once you introduce $\mathcal{P}(\mathbb{R})$, the set of arbitrary subsets of the real line, this leads immediately to the problematic issue of the continuum hypothesis: is there or is there not a subset $S \subset \mathbb{R}$, bijective to neither the integers nor the whole line? What I think is relevant here is to look back at one of the key ideas of L.E.J. Brouwer's *intuitionist foundations*, namely his concept of *free choice sequences* (FCS). Brouwer did not want to deal with infinite objects but he recognized the idea of a construction that can go on as long as you want. Some infinite sequences follow "laws," i.e. are generated by algorithms, but these are very special. A canonical example of a lawless sequence is the outcome of an infinite number of roles of a die.

Intuitionism is better known for rejecting the law of the excluded middle and rejecting objects that cannot be constructed. My canonical example is distinguishing between the value of the maximum of a continuous function and its argmax (the argument where the max is taken on). Let the function be real valued with domain $[0,1]$. If the function is given by some algorithm that delivers approximate values and ϵ, δ uniform continuity bounds, one can readily approximate the max. But if there are two competing maxima, it may take forever to settle which is larger or whether they are equal. Thus intuitionists reject the idea that there is always a point where any such f takes its maximum value. I spent many hours trying to understand this philosophy and seeking a middle ground with my good friend Gabriel Stolzenberg who devoted his career to constructivist ideas.. But, in the end, I side with conventional thinking except that I thoroughly support Brouwer's free choice sequences.

I believe that the proper mathematical formalism for Brouwer's free choice sequences are the concepts of a *random variable* and of *independent random variables*. Random variables are about as real as anything in mathematical models: they are everywhere in our everyday world. The clouds above us, the weather forecasts, the clusters and swerves of drivers on the road, the mosquitos that bite you – all these have not only probabilities but instantiations in our lives. This leads to a huge area of applied math, that of probability and statistics. Math needs to cover this area. But, paralleling the controversies about the foundations of

math, there has been a long debate on what are the proper foundations of both probability theory and its application in statistics. I am no expert in the subtleties here but have come face to face with some of this in my work in computer vision. The specific issue that affects set theory though is the reduction of probability theory to measure theory. This is usually described as dating from the Kolomogorov's book on the foundations of probability [Kol33]. Here a random variable is described as a function on a basic measure space (X, μ). The problem is that, in its application to the real world, this basic space is a fiction. It is as unreal as higher cardinals or as the set of coset representatives of $\mathbb{R}/\mathbb{Q}$. I have written elsewhere [**E-2000**] that random variables must be treated as a third type of variable along with real numbers and integers. To reduce them to real numbers via measure theory does not capture their full meaning, especially the concept of "independent" random variables.

My key example of the subtlety of random variables and their independence is Christopher Freiling's disproof of the continuum hypothesis [Fre86]. More specifically, in the presence of the axiom of choice, the continuum hypothesis is equivalent to the statement that the real interval $[0, 1]$ can be well-ordered such that, for any $x \in [0, 1]$, $\{y \in [0, 1] | y < x\}$ is countable. What Freiling disproves is that if you accept the existence of two independent real random variables, then there is no well ordering of the reals built as above from countable ordinals. Using darts as a colorful way to describe randomness, imagine that 2 people throw darts (well OK, replace $[0, 1]$ by the unit disk if you like darts). Obviously, given a countable subset of the dart board, a random dart is going to miss this subset. So if the two darts land at points x, y, is $x < y$ or $y < x$? Neither can be true so the well ordering cannot exist. The two throws deliver independent random points, i.e. you can treat either as being thrown first and then the second misses the countable set of lesser points. (Note that this is argument has a lot in common with quantum physicist's analysis of the collapse of the wave function when two observations are made at space-time points, neither in the future light-cone of the other.) Thus, free choice sequences contradict the continuum hypothesis pus choice not to mention the rather extreme reductionary hypothesis V=L. More precisely, it shows that ZFC + FCS implies ¬CH. It does, however, need a well-ordering of $\mathbb{R}$, and a well ordering of $\mathbb{R}$ is just as crazy a set as the coset space of $\mathbb{R}/\mathbb{Q}$. I don't believe either of them are "real" and like to think of the above proof as showing that the real line is truly a riotous garden of diversity. The central new idea here is not introducing one random real number, something that forcing arguably already did, but introducing countably many *independent* real numbers with the property that each can be treated as chosen after the other. Kolmogorov's approach to probability does not allow this because the graph of the assumed well ordering relation "<" in $[0, 1]^2$ is not measurable.

I think it is essential to try to express Freiling's approach in strict set theoretic terms. I come up with this[7]:

---

[7]This concept reminds me of generic points in algebraic geometry. Given a ground field k, and a universal domain Ω of infinite transcendence over it, an Ω-geometric point of a variety lying over the usual generic point over k is very like a random real, Σ playing the role of the ground field.

DART AXIOM: *Given a countable sequence of sets Σ, define $Z(\Sigma) \subset \mathbb{R}$ as the union of all sets A of reals of measure 0 definable with constants Σ, i.e. there exists a predicate $\Phi(x)$ in these constants such that $A = \{x | \Phi(x)\}$. ($Z(X)$ is then of measure 0 too.) Then for any "base" Σ, there exists a sequence $X = \{x_1, x_2, \cdots\}$ such that for all i*

$$x_i \notin Z(\Sigma, X - \{x_i\})$$

called independent random reals relative to Σ. In Freiling's case, we have simply $X = \{x_1, x_2\}$ with base Σ containing only the well ordering of $\mathbb{R}$.

A big question for applied math is what subsets of the real line make a good working hypothesis for its power set, adequate for applications of all kinds? The whole mathematical field of analysis from A to Z is based on the concept of Borel sets. And once you have these, it seems unreasonable not to accept the tower built by adding the projection operation to countable unions and complementation. The smallest set of subsets closed under these three operations is a good candidate. But perhaps a better candidate is the set of *hyperprojective* sets that are defined by trans-finite induction up to the least ordinal α for which $L_\alpha(\mathbb{R})$ models KP set theory [Mos09]. (Projective sets are the ones involving only a finite number of projections.) We now get an awful lot of unsolved problems such as "Are all hyperprojective sets measurable?." Set theorists have shown that more higher cardinal hypotheses (e.g. using "determinacy", axioms that certain infinite games have winning strategies) do imply that projective sets are measurable, but is this enough to convince conservative mathematicians that it is true? And Solovay's theorem that there is a model of ZF where all subsets of the the reals are measurable is based on a radically reduced, even countable model. His model is a quotient of a forcing model $M[G]$ where M might as well be countable. Given the FCS Axiom, this feels like an impoverished world.

But I'd like to ask instead: *what is true?* I think I'm in good company, that Gödel himself asked whether there are further axioms that will enable us to answer questions like this. My hope is that extending the use of random constructions, there may be more answers. Perhaps a suitable definition of random Borel sets, random projective sets and using arguments like Freiling's, may help. In any case, I think a workable "applied mathematicians" set theory should be a reduced version of ZFC:

1. Start with two types of variables, one forming the natural numbers and the other the real numbers forming two given sets $\mathbb{N}, \mathbb{R}$ with the usual operations and relations.

2. Replacing unrestricted choice with the above axiom of dependent choice,

3. Replacing the power set axiom by allowing constructions that lead e.g to hyperprojective sets of reals, i.e. assuming all other sets beyond $\mathbb{N}, \mathbb{R}$ are constructible from them.

4. The dart axiom for the existence of countably many independent random real numbers over any base Σ.

It is not clear, however, how to integrate this last axiom with the rest of set theory. I hope a theory along these lines may be developed.

Part V

Coming to Terms with the Quantum

I loved physics more than math in high school. I did the coolest experiment with a great physics teacher, Mr. Brinckerhoff at Phillips Exeter, mixing oil and water with higher and higher dilutions of the oil. At each dilution, we put some camphor flakes (if I remember well) on top to see if they spin. They won't spin on the oil, only on pure water. So at a high dilution, there isn't enough oil to make a film over the whole surface and the flakes spin. Bingo, you get the size of the oil molecules! Lord Rayleigh did this in 1899 and found about 10^{-9} meters. It worked for us too. Then the class tried to repeat the classic measurement of the charge of an electron by observing oil droplets in a potential field with the smallest charge. This didn't come out very well: I got 2 1/2 electrons! Experimental physics was not going to be my forte. But next I learned the math in special relativity and later, with von Neumann's book [vN55], the fascination of quantum mechanics. Then, in college, I made the mistake of auditing for a few weeks Schwinger's course on quantum field theory. This was impossible for me to follow and I realized I was more at home with the clean definitions of the math world than with the formulas of free wheeling physicists for whom the math was a window dressing, to help expressing the real stuff in the world.

I occasionally worried about Schrödinger's cat (see the next chapter) but left physics to physicists. But, more recently, I happened upon Gerald Folland's book "Quantum Field Theory: A Tourist Guide for Mathematicians" [Fol08]. The title was promising and, indeed, I found that fields, though complicated, could be understood a bit and I have read and re-read it in bits and pieces trying to come to terms with the physicist's wild blue yonder. As is well known, Fock spaces work great for free quantum fields without interactions but it turns out that the Hamiltonian expressing the interaction between electrons and photons still hasn't worked mathematically. For example, the deduction of the Coulomb law from exchange of photons has only been heuristically deduced. At present, it's still a case where bizarre unrigorous uses of math nonetheless lead to stunningly accurate numerical predictions.

But the much more basic problem of measurements in quantum mechanics always bugged me. It was hard for me to accept the "Copenhagen" approach, that classical physics operates in the human world while quantum physics operates in the atomic world and that "nature" collapsed the wave form at some stage during an experiment to keep quantum madness at bay. I worried that there are places in space-time, past or future, near or remote where there are no humans making observations and what would cause collapse there? And without collapse, would the macroscopic world come to look more and more like that of atoms? For instance, in the Paleozoic or Mesozoic eras did the world still include something like measurements that collapsed its wave form or did it run purely on Schrödinger's equation, creating species mixtures? I have struggled to say anything clearly about this question for some time. For what it's worth, I talk about both the clear ideas and the speculations that I came up with in the next chapter. To put one disturbing idea very bluntly, if a single ionization event in the wrong part of some DNA can cause cancer, who is to say who or what collapses the resulting wave function superposition (hence you do or don't get cancer).

I also got involved in physics when I gave a seminar on Peter Shor's work in quantum computing with my physicist friend John Myers. At their core, quantum computers consist in a finite-dimensional Hilbert space $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$ of so-called Q-bits and I wondered how Feynman's theory worked there. In fact, his "sum-over-histories" technique in this case is so simple and fun that it could be explained in an undergraduate linear algebra class. I describe this in Chapter 15.

Chapter 14

Quantum theory and the Mysterious Collapse

i.   Background: Measurements and 'Copenhagen'

Quantum mechanics (I'll abbreviate it to QM) is very strange scientific theory. As is well known, Einstein and Bohr argued for many years over what it meant and whether it was even a reasonable theory. Feynman often acknowledged that it was a truly weird theory and claimed that nobody really understood it. In this chapter, I want to add my two cents worth, posing a new way of looking at the classical/quantum puzzle and then asking whether DNA replication can cause macroscopic uncertainty. In this section, I will begin by describing what is so bizarre about QM and then review some of the interpretations of its meaning. For considerable help in this review, I want to thank Professor Jakob Yngvason.

I think the simplest way to present the strangeness of QM is this. Quantum theory proposes to describe the state of the world by a unit vector in a Hilbert space, $\phi \in \mathcal{H}$ (mod a phase change $\phi \mapsto e^{i\theta}\phi$). ϕ evolves in time by Schrödinger's partial differential equation or its fancier field-theoretic versions, but it also must be changed by *discrete jumps* when a measurement is made, the so-called "collapse of the wave function." The big question is simply this: does ϕ really represent something existing in the physical world or does ϕ measure what an observer knows about the world? If the former is correct, then what physical process could cause these discrete jumps? If the latter is correct, then human knowledge is inextricably tangled with what goes on in the microscopic physical world, and physics involves intangibles like consciousness. Or, more succinctly, is ϕ an *ontological thing* or an *epistemological thing*? It appears to be both. So long as it evolves via Schrödinger, it certainly looks ontological, an external reality; but when it jumps after a measurement, it surely is epistemological, representing a state of knowledge.

What is equally difficult to wrap your mind around is that vectors in a Hilbert space can be added, so two states ϕ and ψ can be combined into *superposition states* $(\alpha\phi +$

179

$\beta\psi)/\|\alpha\phi + \beta\psi\|, \alpha, \beta \in \mathbb{C}$. There is nothing analogous to this in classical physics. The most puzzling form of this enigma is given by the rather disgusting thought experiment proposed by Schrödinger: a cat is put in a sealed box along with some poisonous gas that will be released if and only if a sample of radioactive material emits an alpha particle in a certain time period. Atomically, a possible emission is not a black and white affair but puts an atom in a superposition state of having emitted and not having emitted an alpha particle. Thus if we apply Schrödinger's equation in the Hilbert space describing the cat, the box and the radioactive material and we find a superposition state of the whole ensemble in which the cat is simultaneously alive and dead. It would seem that until the box is opened and the cat is observed, hence indirectly an atom in such a superposition state is measured, does the wave form collapse and the fate of the cat get decided. But it is hard to imagine that the cat did not meet its fate before the box is opened. Any pet lover knows the cat has a consciousness too and thus is making its own "measurements" and physicists are reluctant to believe that such a gross superposition state is really possible. Nonetheless, this thought experiment has become the paradigm of macroscopic superpositions of two totally distinct situations, hence I will call all such states "cat-states." This thought-experiment shows that what constitutes a measurement and when and where collapse occurs is not a simple question.

But superposition raised yet another problem besides cat-states. In 1935, [EPR35], Einstein, Podolsky and Rosen suggested it should be possible to produce a pair of particles shooting off in opposite directions with indeterminate internal (spin or polarization) states yet the states of the two were entangled, that is the state of each one determines the state of the other. We can describe such a state as $\phi_{L\uparrow} \otimes \phi_{R\uparrow} + \phi_{L\downarrow} \otimes \phi_{R\downarrow}$ with L/R indicating the particles, $\uparrow, \downarrow$ two alternative internal states. Then if the internal state of one is observed, this observation determines the result of any later measurement of the state of the other particle. This is known as "spooky action at a distance" but it is turns out to be all too true. John Bell refined the test with an ingenious set of measurements to preclude the possibility that the internal states have somehow been fixed when the pair were generated. And when his refined tests were carried out, both the superposition and the presence of action at a distance were confirmed, see [Bel62, GZ15, SN15].

Werner Heisenberg adopted the full fledged epistemic interpretation of the wave function when he wrote:

> *We can no longer speak of the behavior of the particle independently of the process of observation. As a final consequence, the natural laws formulated mathematically in quantum theory no longer deal with the elementary particles themselves but with our knowledge of them. Nor is it any longer possible to ask whether or not these particles exist in space and time objectively … When we speak of the picture of nature in the exact science of our age, we do not mean a picture of nature so much as a picture of our relationships with nature.* from [Hei58], p.15, 28.

However, the epistemic interpretation was haunted by its dependence on human observation (or measurement) and Bell sarcastically asked, [Bel90],p.34:

> It would seem that the theory [QM] is exclusively concerned about "results of measurement," and has nothing to say about anything else. What exactly qualifies some physical systems to play the role of "measurer"? Was the wave-function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.?

Recently, the epistemic viewpoint has shed its apparent human dependence by being re-labelled as the "information theoretic" interpretation of QM or as the statistical variant "QBism" (Quantum Bayesianism) but it seems essentially the same to me.

But some, like Einstein, De Broglie and Bohm, refused to abandon the ontological interpretation. They worked extremely hard to add "hidden variables" to the wave function in terms of which a deterministic model of the microscopic world would be restored. De Broglie's key idea, for the case of non-relativistic quantum mechanics, was to allow the wave function ψ as usual to propagate as usual with Schrödinger's equation *with no collapse.* But he proposed that it also acts as a "pilot-wave" to guide bona fide particles that follow the gradient of the *phase* of ψ. The positions of the particles define the macroscopic world and they do make collapse-style choices on measurement outcomes and are what we experience consciously. A recent exposition is [Bri16]. There have been some attempts to extend the theory to Lorentz-invariant fields but it seems impossible to make it compatible with relativity except by either assuming ψ somehow implicitly defines a notion of simultaneity, hence restoring Newtonian geometry, or by requiring every space-time point to have a weird access to its past light-cone. Einstein was quite skeptical of this approach and, wrote, in a letter to Born, "That way seems too cheap to me." Note that in this theory ψ carries forever all the alternate outcomes of every measurement. David Deutsch, commenting on such a never collapsing ψ, wrote "Pilot-wave theories are parallel-universe theories in a state of chronic denial."[1]

Deutsch is here referring to Everett's wild interpretation of QM according to which, after every measurement of a superposition, the world itself splits into multiple worlds, one for each outcome of the measurement. This has become a kind of play what-if thing for science-fiction writers and science popularizers. But to me, this is just playing with words and has no empirical meaning whatsoever. We live in one world and measurements do have definite outcomes and imagining other worlds is pure fantasy.

In another direction, there is a school of thought that asserts that the problem is clarified by "decoherence." Instruments in a lab that amplify a microscopic signal inevitably involve large random molecular events, often a so-called "heat-bath" into which you can dump

---

[1]This appears in the article "Comment on Lockwood," *British Journal for the Philosophy of Science*, volume 47, pages 222–228.

entropy and allow the creation of macro-scale order. Thus the coherent superposition of two microscopic states can result in a superposition of two macroscopic states but only by linking the distinct read-out instrument states (often called "pointer" states) with distinct states of the heat-bath. One can model this by assuming the lab's Hilbert space is a tensor product $\mathcal{H}_{\text{instr}} \otimes \mathcal{H}_{\text{heat}}$. You replace the entangled state $\psi \in \mathcal{H}_{\text{instr}} \otimes \mathcal{H}_{\text{heat}}$ by the self-adjoint rank 1 matrix $\psi.\overline{\psi^t}$ and now take its *trace* with respect to the heat-bath factor. You get a small rank operator D on $\mathcal{H}_{\text{instr}}$, perhaps rank 2 if the measurement is binary, known as the *density matrix*. Using the randomness of the heat-bath, it becomes plausible that all off-diagonal terms of D nearly vanish so that the diagonal terms are now classical probabilities of the possible macroscopic outcomes that sum to one. But I fail to see why this solves anything: you still must observe the outcome and doing so collapses the wave-form, including its heat-bath component.

Finally we come to the standard way of dealing with this conundrum, known as the *Copenhagen approach* proposed by Niels Bohr. Here, one accepts that two theories are needed, a non-deterministic one for the microscopic world and a deterministic one for our macroscopic world and that certain simple microscopic measurements, such as the measuring both the position and velocity of a particle, cannot be made simultaneously. Essentially all physicists reject the idea that human consciousness can play any role in physics as this amounts to polluting their beloved physical theory with human involvement, biology or even philosophy. Instead they believe that there is some point called the 'Heisenberg cut' where *nature* makes the choice. Basically, this means "live with it," weird as it is. For example, as Jakob Yngvason pointed out to me, a standard QM text writes: "We emphasize that when speaking of 'performing a measurement' we refer to the interaction of an electron with a classical 'apparatus', which in no way presupposes the presence of an external observer." [LL65], p.2. A recent collection of the ideas of 17 physicists can be found in [Sch11].

But then where is this mysterious cut? Like De Broglie, some have sought non-linear stochastic modifications of Schrödinger's equation *that create cuts*. These are known as "collapse theories" and are based on the idea that, with very small probabilities, every particle sometimes decides that it should jump to some definite position allowed by the wave function. Most of the time, this has no large effect but, in cat-states, one particle deciding to be in the cat's live form versus to be in its dead form forces the entire cat to follow because the wave function is the sum of these two. And because there are so many particles in the cat, this happens more or less instantaneously. Nifty idea but, so far, no evidence that it might be true.

Another recent formalization of the Heisenberg cuts has been made by Jürg Fröhlich and is called the "Event-Tree-History" or ETH theory[2], [Frö19, Frö22]. He starts with an "isolated open local system" S, a part of the world essentially uninfluenced by the bigger universe but open in the sense that it can influence and even be entangled with events

---

[2]A pun on his institution in Zürich!

outside itself. He introduces the algebra of observables $\mathcal{A}_t$ of events inside the system at time t. $\mathcal{A}_t$ can shrink due e.g. to photons leaving this local area. The state of the system is given by a density operator ω_t which defines the expectation linear function on the observables in $\mathcal{A}_t$, $X \mapsto \text{tr}(\omega_t . X)$. In this setting, "events" are when the wave function collapses and these involve families $\mathcal{X} = \{\pi_\xi\} \subset \mathcal{A}_t$ of orthogonal projections forming a partition of unity: $\pi_\xi . \pi_\eta = \delta_{\xi,\eta}\pi_\xi, \sum_\xi \pi_\xi = I$. The key new point is that he gives a formal definition of when $\mathcal{X}$ ought to be collapsed, equivalent, I believe, to $\omega_t(\pi_\xi X \pi_\eta) = 0$ for all observables X if $\xi \neq \eta$. He then defines collapse via

$$\omega_{t+dt} = \pi_\xi \circ \omega_t \circ \pi_\xi / \|\pi_\xi \circ \omega_t \circ \pi_\xi\|$$

where ξ is selected by "nature" with the usual probabilities, see [Frö22], p.21 where this is described with exactly this word. This is his Heisenberg cut.

ii. AMU sets

After this quick review of the measurement problem, I want to postpone further discussion of it until the last section "Bohr bubbles." Instead, I will take for now an agnostic approach to the problem of measurements and collapse, because I believe there is another useful way to analyze the interaction of the atomic world and the world of classical physics. This is to imagine a world in which atomic events are predicted by Schrödinger's equation alone and *no atomic measurements are made* that force a collapse of the wave function. (This is what is proposed in the De Broglie-Bohm theory but without their added particles.) I want to ask: in the absence of physics labs where atomic events are intentionally magnified and measured, *would we know the difference?* If Schrödinger's equation goes its merry way forever, would quantum uncertainty somehow creep into our classical world? I recently came across the comment of Guido Bacciagaluppi "Nature has been producing macroscopic superpositions for millions of years, well before any quantum physicist cared to artificially engineer such a situation," [Sch11], p.143. I have wondered the same thing for quite a while and the focus of §ii-vi is trying to be more specific about this possibility.

To formulate this, I need to consider a QM model that includes a whole local human environment or even, for that matter, the whole earth. By my estimates, the earth contains roughly 5×10^{51} electrons, protons and neutrons but so what? If the Hilbert space $\mathcal{H}$ is large enough, why shouldn't it describe the whole earth, a pretty good "open local quantum system" as defined by Frölich.

There is a set of self-adjoint operators on $\mathcal{H}$ whose eigenvalues correspond to the observations we make by touching, seeing, listening and interacting with our environment as we go about our normal daily life. Outside physics labs where atomic experiments force dials to register superposition effects, this world definitely appears to be always in *near eigenvector states* (superpositions of eigenvectors with very similar eigenvalues) for all these human observations, i.e. deterministic. This means that your toothbrush always has a definite approximate location given by the eigenvalues of a suitable operator and if it is found
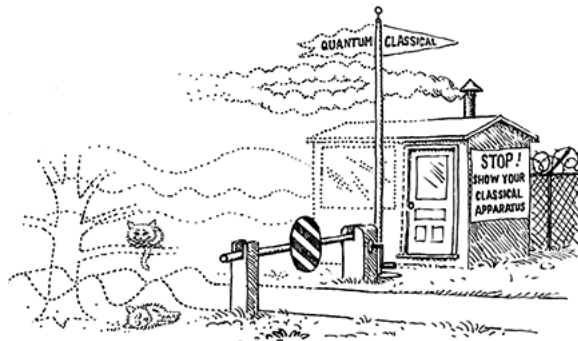
Figure 14.1: Zurek's excellent cartoon of the macro/micro issue from [Zur91]. Note the two "Cheshire cats," only one smiling. Reproduced by permission of IOP publishing.

in an unusual place, you are sure someone must have moved it, not that the toothbrush was in a cat-state. From a quantum point of view, such classical states are very special: clearly a superposition of two near eigenvectors is almost never another near eigenvector.

States that are near eigenvectors for some set of human observables thus define a class of fairly small open subsets $\texttt{AMU} \subset \mathcal{H}$, where $\texttt{AMU}$ stands for "Approximately Macroscopically Unique" states. In other words, these are the states that describe a world recognizable to us with no maybe-dead/maybe-alive cats. Once you make this definition, it raises the questions: what are the shapes of these subsets and to what extent do solutions of Schrödinger's equation stay there vs. how often is a collapse of the wave-form needed to stay there? Can the world of localized objects with definite shapes and behaviors survive without invoking collapse? These are big questions and this chapter will only scratch the surface of the issues this raises.

What are macroscopic variables? I am thinking of the position, motion, shape and mass of solid objects, the location, density and temperature of liquids and gases, the proportions and internal connections of constituent materials, the average strength of electric, magnetic and gravitational fields in small parts of space, etc., but I don't have an exhaustive list. Each comes with a dimension in terms of the primary quantities: meters, seconds and grams, and secondary derived dimensions: degrees centigrade, volts, amps etc. definable in terms of the primary ones using basic constants so that our senses and simple measuring instruments give us approximate values, numbers with explicit uncertainties that are also readily estimated. For example, lengths with millimeter accuracy are easily measurable with eyes alone and in microns with optical microscopes. Temperature is essentially the total internal kinetic energy per unit mass of some substance, up to a factor measuring the number of degrees of freedom and Boltzmann's constant. Simple devices measure the spatially smeared out electric and magnetic fields, usually filtered to particular frequencies. These variables are "observables" in quantum theory and hence define Hermitian operators

in $\mathcal{H}$. Our senses and instruments can only measure things within certain limits so these are, in fact, *bounded* Hermitian operators. In a given state $x \in \mathcal{H}_1 = \{x \in \mathcal{H} | \|x\| = 1\}$, the expected value when measuring an observable given by the operator A is simply $\exp_A(x) = \langle x, Ax \rangle$, a quadratic function on the unit sphere in $\mathcal{H}$ invariant under phase change $x \mapsto e^{i\phi}.x$.

What is central to this discussion are "near eigenvectors." No measurement can ever be a precise real number. There is always a limit to how exactly any quantity is recorded, hence, if there is a continuous spectrum, we can never land right on a mathematical eigenvector. In the simplest approach, near eigenvectors are defined by the *variance* or the *standard deviation* of the measurement made by a bounded self-adjoint operator A in a state x:

$$\text{var}_A(x) = \langle x, (A - \exp_A(x)I)^2 x \rangle = \langle x, A^2 x \rangle - \langle x, Ax \rangle^2,$$
$$\text{sd}_A(x) = \sqrt{\text{var}_A(x)} = \|(A - \exp_A(x)I)x\|.$$

The variance and standard deviation are always real and non-negative. It is defined even if A is unbounded, though it might then be infinite. If we measure A in an ensemble of preparations of the same state x, then this variance will approach the variance of the results. Note that, because of the square of the expectation, the variance is a fourth degree function of the state x (restricted to the unit sphere). To have an apparently deterministic world, we can define AMU by requiring that *the standard deviation of all macroscopic variables is less than the accuracy of your instruments.* Around 1700, it probably sufficed to have the variance of the position observables less than one millimeter. By 1850, perhaps it needed to be less than one micron. In any case, we now define a *family* of AMU sets by (i) listing the macroscopic variables A_n we are concerned with, (ii) assigning tolerances σ_n to each and setting:

$$\text{AMU}(\{A_n, \sigma_n\}) = \{x \in \mathcal{H}_1 | \forall n : \text{sd}_{A_n}(x) < \sigma_n\}$$

Because var is a fourth degree polynomial function on the Hilbert space, the sets AMU should be expected to have a complicated shape.

A lemma that will be useful below is:

Lemma. *If A_t is a family of commuting, self-adjoint operators and $B = \int w(t)A_t dt$ is a weighted average of them, then:*

$$\text{var}_B(x) \leqslant \int w(t)\text{var}_{A_t}(x).dt.$$

Proof. To simplify, first replace A_t by $A_t - \langle x, A_t x \rangle I$. Then, if you expand $\iint w(t)w(t')\|(A_t - A_{t'})x\|^2 dt dt'$, you get the difference of the two sides of the inequality. $\square$

It's not clear whether bounding variance is a strong enough definition to capture the certainty of the world we are all living in. It is tempting to believe that when a measurement is made, we know it is not exact but we are certain that it is not too far off. You lay a

tape on a doorway and measure so-and-so many inches and you know you might be off by an eighth but you're sure it's accurate to within a quarter inch. OK "measure twice, cut once", as the carpenter's advice goes, but the length of a rigid body just doesn't change in the classical world. This may be expressed by the projection valued measure defined by the operator A: let Q_λ^A be projection onto the subspace of A's eigenvalues $\leqslant \lambda$. Then this is expressed by requiring `AMU` states x to satisfy $(Q_{\lambda-\sigma} \circ (I - Q_{\lambda+\sigma}))\, x = 0$. This means that the measure on the spectrum of A corresponding to macroscopic knowledge has finite support of length at most 2σ. Unfortunately, this requirement is impossible to impose on both position and momentum because the distributions of their values are Fourier transforms of each other and, if one has compact support, the other is the restriction to the real axis of an entire function. In other words, QM demands that very rarely, *something inconsistent is inevitable in the macroscopic world.*

To flesh out this definition, let's look at the simplest quantum system of all: the motion of a single scalar particle on a line with coordinate x, treated non-relativistically. This is given by the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$ with two observables, position $X =$ multiplication by x, and momentum $P = -i\hbar.\partial/\partial x$. Then Heisenberg's inequality says $\mathrm{sd}_X(\phi).\mathrm{sd}_P(\phi) \geqslant \hbar/2$ for all $\phi \in \mathcal{H}_1$. Thus `AMU` is empty you choose $\sigma_X.\sigma_P < \hbar/2$. But Planck's constant is very small so this is not a problem for normal macroscopic accuracy. The states ϕ with the most precise position and momentum are given by the functions:

$$\phi_{x_0,\sigma,k}(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2 + i.k.x}$$

called *Gabor functions* by electrical engineers. These have $\mathrm{sd}_X(\phi) = \sigma, \mathrm{sd}_P(\phi) = \hbar/(2.\sigma)$. The `AMU` sets are natural open neighborhoods of this three dimensional locus of Gabors and clearly do not have a simple shape. If we further assume that the Hamiltonian has only kinetic energy and no potential, we can integrate Schrödinger's equation and see how these Gabor states evolve. Taking for simplicity $x_0 = 0$, an initial uncertainty σ_0 and m for the mass of the particle, we get what is usually called a Gaussian wave packet (see, e.g. the Wikipedia article on this):

$$\text{Let } S(t) = \sigma_0^2 + i\hbar t/m, \quad \sigma(t) = \sqrt{\sigma_0^2 + (\hbar t/m)^2 \sigma_0^{-2}}$$

$$\phi(x,t) = \sqrt{\frac{\sigma_0}{2\pi S}} \cdot e^{-\frac{x^2 + ik\sigma_0^2 k(2x - k\hbar t/m)}{2S(t)}}$$

$$|\phi(x,t)| = \frac{1}{\sqrt{2\pi\sigma(t)}} \cdot e^{-\frac{1}{2}\left(\frac{x-k\hbar t/m}{\sigma(t)}\right)^2}$$

Here $\mathrm{sd}_X(\phi(\cdot,t)) = \sigma(t) > t\frac{\hbar}{m}\sigma_0^{-1}$ We see that, as well as moving at speed $k\hbar/m$, also its spatial indeterminacy expands with time, growing until the particle looses its spatial localization and behaves more and more like a wave. Thus it inevitably leaves all `AMU`

sets. However, this smearing out only happens when t is comparable to a large multiple of $m.\sigma_0/\hbar$.

With single particles, this is not especially surprising but this analysis applies to larger objects too. Consider a meteorite in outer space of mass M whose vector position $\mathbf{x}$ and vector momentum $\mathbf{p}$ are being measured. We can model this in a non-relativistic, non-field theoretic way as is done in quantum chemistry. Let its constituent atoms be labelled by the subscript α and let the position of its atoms be given by $\mathbf{x}_\alpha^{(i)}, i = 1, 2, 3$. The Hilbert space is then $\mathcal{H} = L^2(\mathbb{R}^{3N})$ with position operators given by multiplication by the coordinates. The momentum operators are then $i.\hbar\partial/\partial x_\alpha^{(i)}$. The position of the rock is the *average* of the atoms positions, its momentum is the *sum* of the atoms momenta:

$$\mathbf{x} = \frac{1}{N} \sum_{\alpha=1}^{N} \mathbf{x}_\alpha, \qquad \mathbf{p} = \sum_{\alpha=1}^{N} \mathbf{p}_\alpha.$$

The position and momentum operators of distinct atoms commute so Heisenberg's commutation relation propagates to whole rock:

$$[\mathbf{x}^{(i)}, \mathbf{p}^{(i)}] = \frac{1}{N} \sum_{\alpha} [\mathbf{x}_\alpha^{(i)}, \mathbf{p}_\alpha^{(i)}] = i.\hbar.I$$

$$\text{hence} \quad \mathrm{sd}(\mathbf{x}^{(i)}).\mathrm{sd}(\mathbf{p}^{(i)}) \geqslant \hbar/2$$

Now the Hamiltonian is the sum of kinetic and potential energy and depends only on the relative position of the atoms, hence commutes with $\mathbf{p}$, which must therefore be constant. What this means is that the macroscopic observables $\mathbf{x}$ and $\mathbf{p}$ evolve exactly like those of single particles. In particular, if we measure the rock's position very accurately, after a while its macroscopic position will get more and more indeterminate and then we would truly be outside the `AMU` set.

But Planck's constant is awfully small, so e.g. if we took even a tiny space rock of size 1 mm, hence mass of about 0.001 grams and measure its position to within 1 micron, it will take some trillion years before it will have "spread out" by 1 mm. So this effect is not going to challenge macroscopic determinacy.

iii.   Constraints on macroscopic variables

It is evident that for `AMU` to be non-empty, the commutators of the macroscopic variables must be sufficiently small at states in `AMU`. In fact, if $x \in$ `AMU` and $\lambda_i = \exp_{A_i}(x)$, we have:

$$\begin{aligned}
|\langle [A_i, A_j]x, x \rangle| &= |\langle [A_i - \lambda_i.I, A_j - \lambda_j.I]x, x \rangle \\
&\leqslant 2|\langle (A_i - \lambda_i)x, (A_j - \lambda_j)x \rangle| \\
&\leqslant 2.\mathrm{sd}_{A_i}(x)\mathrm{sd}_{A_j}(x) \\
&\leqslant 2\sigma_i\sigma_j
\end{aligned}$$

Thus it is natural to assume all these commutators have norms with such bounds. Unfortunately, it seems unlikely that this is a strong enough. One really needs to consider the whole C^*-algebra $\mathcal{A}$ generated by the $\{A_i\}$s: all variables in this algebra are natural candidates for macroscopic measurements. But the C^* algebra will contain the amplified commutators and these are measurements we expect to be zero! Extending commutator bounds to the whole algebra brings up many problems and questions. For some time, I thought the following was might be true:

Query. *If A, B are bounded self-adjoint operators in a Hilbert space and f is a continuous function on the reals with Lipschitz constant C, then is:*

$$\|[A, f(B)]\| \leqslant C.\|[A, B]\|?$$

To my surprise, when I asked Alain Connes, he found a counterexample to this using $f(x) = |x|$ due to Alan McIntosh [McI71, Kat73]. But it does, however, hold if you put bounds on the the *third* derivative of f. Secondly, we not only want the set of global human friendly states `AMU` to be non-empty but also, we must have a procedure to "collapse" back to a state in `AMU` any state in $\mathcal{H}$ that might be produced by an experiment in a physics lab. This is just requiring that the ambiguity of Schrödinger's cat cannot be allowed to disrupt the human world. In other words, we need some kind of projection from "cat-states" where quantum ambiguity has penetrated the macroscopic world to states in `AMU`. The simplest way to achieve this would be to assume that we can construct *commuting* bounded self-adjoint operators A_i' such that the operator-norm differences $\|A_i' - A_i\|$ are all small. Then the macroscopic world can be sustained by projecting onto eigenstates of the $\{A_i'\}$.

But here we find a real obstacle. A result that goes back at least to Halmos (see [BH74], p.477, lemma 2) is the following: there exist pairs of self-adjoint operators A and B with arbitrarily small commutators and norm at most 1 such that, for all commuting pairs A', B', $\|A - A'\| + \|B - B'\| \geqslant 1$. Here's his result:

Halmos's Lemma. *If S is the right shift operator on $L^2(\mathbb{N})$ then $\|S - (N + C)\| \geqslant 1$ for every normal operator N and compact operator C.*

Proof. Assume N, C exist with $\|S - (N + C)\| < 1$. Note that S^* is left shift, hence $S^*.S = I$, hence $\|I - S^*.(N + C)\| < 1$, hence $S^*.(N + C)$ is invertible, hence $N + C$ is injective. Now apply the Fredholm alternative so that $N + C$ must be surjective too. This implies S^* is invertible, a contradiction. $\square$

To apply this, take C_n to the weighted right shift with entries decreasing slowly from 1 to 0. Let $A = S + S^* - C_n - C_n^*$, $B = i(S - S^* - C_n + C_n^*)$, self-adjoint with arbitrarily small commutator. It follows that for any commuting A', B', we always have $\|A - A'\| + \|B - B'\| \geqslant 1$.

This is in stark contrast to the theorem of Lin [Lin97] in finite dimensions, to the effect that for all ϵ, there is δ – independent of n – such that if A, B are self-adjoint $n \times n$ norm 1 matrices with $\|[A, B]\| < \delta$, then there are commuting self-adjoint A', B' with $\|A - A'\| + \|B - B'\| < \epsilon$.

The case of the (unbounded) position and momentum operators in $L^2(\mathbb{R})$ is an interesting one. One would like to construct an orthonormal basis of this Hilbert space of states in $\mathtt{AMU}(\{X, P\})$, of functions well localized in both space and frequency. From this, we could obtain operators diagonal for this basis that approximate both X and P. But the theorem of Balian and Low gives an obstacle: one cannot construct a function $g(x)$, well localized in both space and frequency[3] for which the doubly infinite set of functions $e^{2\pi inx/\delta}.g(x - m\delta)$ forms an orthonormal basis. However, a simple construction due to Daubechies and many others [DJJ91, CM91, AWW91] shows you can do this *if you allow a pair of opposite frequencies in the Fourier transform.* The first cited paper constructs an orthonormal basis where g has exponential decay, but replacing the periodic factor by sines and cosines, that alternate with the parity of $n + m$.

The upshot is that approximating self-adjoint operators by commuting ones is not a simple question. In fact, there are several different approaches for making a formal definition of a macroscopic system. At the least, one needs to assume that $\mathtt{AMU}$ is not empty. Better is to assume suitable commutator bounds. Strongest of all is to assume that the generating set $\{A_i\}$ is approximated by commuting $\{A_i'\}$. Finding the right definition looks like an important and interesting question.

iv. Molecules

Let's look at some actual quantum models an d their observables. In quantum chemistry for molecules, it's usual to approximate the full QED model by a non-relativistic, non field-theoretic model with pairwise potentials. Let's assume we have N particles with masses m_a and charges e_a with coordinates $x_a \in \mathbb{R}^3, 1 \leqslant a \leqslant N$. Further, assume their center of mass is the origin, i.e. $\sum_a m_a x_a = 0$. The state space is then $X = \mathbb{R}^{3n-3}$ and the Hilbert space is $\mathcal{H} = L^2(X)$. Let $p_a = -i\hbar \frac{\partial}{\partial x_a}$ (each is a 3-vector of operators). The Hamiltonian is:

$$H = -\sum_a \frac{p_a^2}{2m_a} + \sum_{1 \leqslant a < b \leqslant N} \frac{e_a e_b}{\|x_a - x_b\|}$$

For example, if $N = 2, m_1 \gg m_2, e_1.e_2 < 0$, we have a simplified spinless hydrogen atom. In this case, the Hilbert space breaks up into a direct sum $\mathcal{H} = \mathcal{H}_B \oplus \mathcal{H}_S$ where H has a discrete negative spectrum on $\mathcal{H}_B$, the bound states, and continuous non-negative spectrum on $\mathcal{H}_S$, where the atom is ionized, the electron and proton free. The discussion in §ii generalizes to the assertion that, for any state ψ with non-zero projection on the free subspace, the variance of position goes to infinity as time goes to $\pm\infty$.

---

[3]It suffices to assume $\|Xg\|$ and $\|Pg\|$ are finite.

The fundamental theorem of non-relativistic scattering under coulomb potentials generalizes this. Consider all partitions Π of the particles into 'clusters': $\{1, 2, \cdots, N\} = C_1 \cup C_2 \cup \cdots \cup C_M$ where C_i are disjoint non-empty subsets of the particles. Then $\mathcal{H}$ is an orthogonal direct sum

$$\mathcal{H} = \mathcal{H}_B \oplus \left(\bigoplus_{\Pi} \mathcal{H}_{\Pi}^+ \right)$$

where the states in $\mathcal{H}_B$ are the 'bound' states, sums of discrete eigenvectors of H and the states in $\mathcal{H}_{\Pi}^+$ are those in which, as $t \longrightarrow \infty$, the particles in each cluster C_i of Π remain bounded but the clusters scatter away from each other. This theorem is physically nearly obvious, i.e. if it did not hold, there would something amiss in the Hilbert space model, but is not easy to prove mathematically. A good survey is [HS00] where this result is theorem 7.2 but the estimates we need date back to [Ens83].

To state this formally, for any partition Π, we have (i) Hilbert spaces for each cluster $C \in \Pi$ and, in this, bound states $\mathcal{H}_B^C$; (ii) Hilbert spaces $\mathcal{H}_{\Pi}$ where each cluster $C \in \Pi$ is collapsed to a point x_C with momentum p_C. The theorem asserts there are isomorphisms $\mathcal{H}_{\Pi}^+ \cong \left(\bigotimes_C \mathcal{H}_B^C \right) \otimes \mathcal{H}_{\Pi}$ such that, if $\psi \in \mathcal{H}_{\Pi}^+$ and $\phi \in \mathcal{H}_B \otimes L^2(X_{\Pi})$ correspond to each other, then:

$$\lim_{t \to \infty} \left\| e^{-iHt} \psi - e^{-iH_{\Pi}t - iI_{\Pi} \log(t)} \phi \right\| \longrightarrow 0$$

$$\text{where } H_{\Pi} = - \sum_{C \in \Pi} \frac{p_C^2}{2m_C}, I_{\Pi} = \sum_{C,D \in \Pi} e_C.e_D \left\| \frac{p_C}{m_C} - \frac{p_D}{m_D} \right\|^{-1}$$

The meaning of the I_{Π} term is simply that for $t \gg 0$, even though the clusters will be separated by approximately $t \| \frac{p_C}{m_C} - \frac{p_D}{m_D} \|$, the long range Coulomb forces will still cause a slowly increasing cumulative displacement proportional to $\log(t)$.

A Corollary of this is that the only `AMU` states in $\mathcal{H}$ are those in $\mathcal{H}_B$. In fact, any state not in $\mathcal{H}_B$ must have a non-zero component in some $\mathcal{H}_{\Pi}^+$ and this must correspond, under the scattering isomorphism, to a state with a component of type $\phi_b \otimes \phi_s$. Then ϕ_s is evolving by the unitary operator $e^{-itH_{\Pi} - i\log(t)I_{\Pi}}$. These operators commute with the operators p_C, so the cluster's momenta are constant and thus the position operator for the cluster x_C evolves as:

$$x_C(t) = x_0(t) + t \frac{\hbar}{m_C} p_C + \log(t) \frac{\hbar e_C}{m_C^2} \sum_D \left\| \frac{p_C}{m_C} - \frac{p_D}{m_D} \right\|^{-3}.$$

Then, computing as above, the variance of x_C increases quadratically and the state cannot remain in any `AMU` subset.

v.   Fields

To do anything serious with QM, you must use fields. Full relativistic field theory with interactions remains to this day a mine field, even the theory of standard electrodynamics being based on heuristic perturbation expansions and limited theoretically by Haag's theorem (the problem of "vacuum polarization"), see [Haa96], pp.53-55. For this reason, I will discuss only free fields, those without interactions. Here we encounter an essential division between bosons and fermions. Photons, forming the electromagnetic field, are bosons that, by definition are made from operators that commute or commute up to Planck's constant. Part of the EM field can be measured macroscopically as our eyes demonstrate to those who are not blind[4] but every part of the spectrum is now grist for the scientist's mill. On the other hand electrons, protons and neutrons are composite fermions whose fields are made up from *anti-commuting* operators so their underlying field is fundamentally unobservable. What are observables for fermions are *quadratic* expressions in the components of the field operators that do commute up to Planck's constant.

The basic idea of boson field theory is to model all such particles of every type by a simple harmonic oscillator. This is the Hilbert space $\mathcal{H} = L^2(\mathbb{N})$ with orthonormal basis $e_0, e_1, \cdots$, in which the basic operators are a weighted left shift $\mathbf{a}(e_k) = \sqrt{k}.e_{k-1}$ called the "annihilation" operator, and its adjoint $\mathbf{a}^*(e_k) = \sqrt{k+1}.e_{k+1}$ called the "creation" operator. To be in the state e_k means there are k particles of this type present. The Hamiltonian is $H = \mathbf{a}.\mathbf{a}^* - \frac{1}{2}I = \mathbf{a}^*.\mathbf{a} + \frac{1}{2}I$ and it has eigenvectors e_k with eigenvalues $k + \frac{1}{2}$. We have a pair of conjugate self-adjoint operators $Q = (\mathbf{a} + \mathbf{a}^*)/\sqrt{2}$ and $P = i(\mathbf{a} - \mathbf{a}^*)/\sqrt{2}$. An important fact is that $Q^2/2 + P^2/2 = H$, hence the sum of the variances of Q and P at any state x is bounded by the energy of that state, hence is a potential macroscopic operator. This looks more familiar if we diagonalize Q, using an isometry of $L^2(\mathbb{N}$ with $L^2(\mathbb{R})$ so that Q becomes multiplication by x, the coordinate in $\mathbb{R}$. Here e_n goes over to $P_n(x)e^{-x^2/2}$, P_n being the Hermite polynomials, P becomes $i\partial/\partial x$ and H becomes the well-known $-\frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2}x^2$ (with units making Planck's constant equal to 1).

Back to photons. Each photon has a frequency, a direction of motion and a polarization. The first two are combined in a 3-covector, its momentum p, that defines its associated EM waves with components proportional to $e^{i(x.p - c|p|t)/\hbar}$ and whose energy is $c|p|$. To quantize the field, we require annihilation/creation operators for each momentum p and for two polarizations that can be given by choosing, for each p, an orthonormal basis $e_{p,1}, e_{p,2}, p/|p|$ in $\mathbb{R}^3$. I'll write these operators $\mathbf{a}(p, s)$, $s = 1, 2$. They are distribution-operator-valued functions. Avoiding details here, we can simply say that the three electric and three magnetic components F_k the quantized EM field are all given by operators of

---

[4]Interestingly, psychophysicists have found that dark adapted normal humans can detect light with only a handful of photons. The human/particle gap is small in this particular case.

the form:

$$F_k(x,t) = \int \sqrt{|p|} \left(e^{i(x.p-c|p|t)/\hbar} \mathbf{a}_k(p) + e^{-i(x.p-c|p|t)/\hbar} \mathbf{a}_k^*(p) \right) d^3p$$

$$\text{where } \mathbf{a}_k(p) = \sum_{s=1,2} c_k(p,s)\mathbf{a}(p,s), \{c_k(p,s)\} \text{ functions of } \{e_{p,s}\}$$

To make math out of this, we must specify the underlying Hilbert space and then give precise definitions of these operators. This is done by first defining a one photon space as $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$ and then taking the "Boson Fock space" over that, essentially the polynomial algebra over the former. We refer the reader to Folland's book, [Fol08], §5.2 and §5.4. There is one "little" problem: because of that pesky $\frac{1}{2}$ in the energy, $F_k.F_\ell$ makes no sense. $\int \mathbf{a}_k(p).\mathbf{a}_k^*(p)d^3p$ is infinite, even on the vacuum zero state. This is the first of the infinities that screw up quantum electrodynamics (QED). This one is usually solved by just insisting on the ad hoc requirement that $\mathbf{a}^*$'s should always come after $\mathbf{a}$'s, called Wick ordering. For our purposes though, the problem disappears when you convert distributions to functions by convolution.

So now integrate $F_k(x,t)$ against a test function in order to get actual self-adjoint operators. To pick out a location in 3-space and a set of similar momenta, we can use a Gabor function of x (we do not need to smooth over time) $g_1(x) = e^{-i(x-x_0).p_0}.e^{-\|x-x_0\|^2/2\sigma^2}$ with Fourier transform $g_2(p) = e^{ix_0.p}.\frac{\sigma^3}{(2\pi)^{3/2}}e^{-\sigma^2\|p-p_0\|^2/2}$. Convolving with g_1, we obtain:

$$F_k(g_1,t) \stackrel{\text{def}}{=} \int F_k(x,t).g_1(x)d^3x$$

$$= \int \sqrt{|p|}|g_2(p)| \left(\mathbf{a}_k(p).e^{ix_0.p-ic|p|t} + (\mathbf{a}_k(p).e^{ix_0.p-ic|p|t})^* \right) d^3p$$

Thus we can apply the lemma in section ii, and deduce that the variance of $F_k(g_1,t)$ at any state x is bounded by the photon energy of that state (here including the energy of the vacuum state). Summarizing, boson fields seem immune from seepage of quantum uncertainty into the macroscopic universe. This is highly relevant to astronomy today: photons are being observed that propagated for eons through outer space without interacting with other particles. The measurements that they afford astronomers have enabled them to extend humanity's knowledge (and what in the last section I will call their Bohr bubble) out in space and back in time billions of light-years/years respectively. The fact that EM fields stay in `AMU` means we can use Maxwell's equation to model them without worrying about their dissipation because of the uncertainty principle.

Fermions are a totally different picture. The basic idea of fermion field theory is to model all such particles of every type by a simple Qbit – every state is occupied or not – and two fermions of the same particle type must occupy different states. Because the field operators anti-commute, macroscopic measurements must relate to quadratic expressions in the field. Looking at spin 1/2 particles, Dirac's field operators are distribution-operator-valued 4-vectors, technically bi-spinors, $\psi_i(x,t), 1 \leqslant 0 \leqslant 3$. One can smooth the field by

convolution and form 16 products $(F * \psi_i^*).(F * \psi_j)$ which then commute up to Planck's constant, hence lead to approximate macroscopic measurements. These include positrons as well as electrons, both of which can have two values $\pm$ of spin. What can be measured are the particle density (positrons and electrons added) , the charge density (positrons and electrons subtracted), the current density as well as densities and currents related to spin, e.g. the difference of spin up and spin down in some orientation. These last have now been made measurable using MRI scans, requiring massive magnetic fields.

Our initial discussion of single non-relativistic particles or space rocks using Gaussian wave packets suggested that, in the absence of measurement related collapse, *the wave aspect of particles always eventually dominates the particle aspect* and thus leads to states eventually outside all AMUs. However, bound states can locally resist dissolution and apparently lattices with some randomness create long term stable states (see "Anderson localization" in Wikipedia or [ADJ$^+$16]). Whether anything like this happens for interacting fields depends, in even the simplest case of QED, on the full machinery of coupled fields and is beyond my expertise.

vi.   DNA

A quantum-lab measuring instrument must be a device that at one end is sensitive to atomic level events while at the other end delivers a macroscopic event that can be recorded. Pretty inevitably, the amplifying process involves contact with large scale random effects, contact with gases or plasma, hence it creates a state in which the microscopic event is entangled with a so-called heat bath, an object in a some kind thermodynamic equilibrium. These come in various guises. There was the original Wilson cloud chamber that depended on creating a volume of super-saturated moisture on the verge of condensation. Here the passage of a single charged particle creates a train of ionized water molecules that cause droplets to form along its path.

Then there are devices containing cascades like a photo-multiplier tube. The tube contains a sequence of cathodes held at higher and higher voltages. When an electron enters the tube, it is attracted to the first cathode where it triggers the emission of more electrons and, bouncing back and forth from cathode to cathode, an ever larger volley of electrons is created.

However, the major point of this chapter is to point out that our biology contains a truly amazing amplifying device: the DNA molecule. I was happy to find recently that some physicists have also noticed this: Bacciagaluppi, on the page quoted in §ii, went on to say "... genetic mutations induced by natural radioactivity can magnify quantum phenomena to the macroscopic level, quite analogously to the case of Schrödinger's cat." At human conception, two sets of 23 DNA molecules, the chromosomes, come together. A chain of events is set in motion that creates the adult life form, the macroscopic phenotype. Moreover, many microscopic events can cause atomic level mutations, altering a single base

pair of the genome. This can result from exposure to ionizing radiation or contact with mutagenic chemicals. A key point, however, is that there is an atomic event triggering every mutation and the outcomes of such events are not certain but always result in superpositions with varying probabilities. Thus at the particle level, the result is a superposition of a mutated and an un-mutated state. Then gestation forms a new phenotype, usually a change for the worse, but occasionally an improvement leading to evolution. At its core, the ability to reproduce and create macroscopic effects depends on the ability of a DNA strand to duplicate itself. If you think of a DNA strand as a sequence of units, each of one of 4 types G, A, T and C, it is not unlike a quantum computer. Of course, it is actually a double strand, each unit in one strand being paired with its complementary base on the other strand ($A \leftrightarrow T, C \leftrightarrow G$). In reproducing itself, the strands separate and each strand assembles a new partner, one nucleotide at a time. Figure 2 is an a cartoon of the process from :
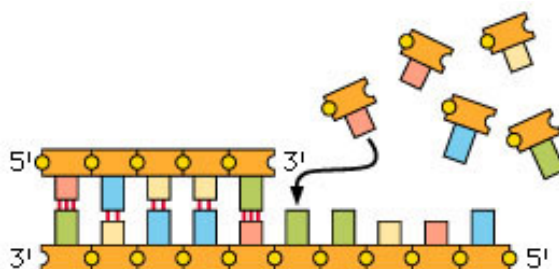
Figure 14.2: A partially assembled DNA leading strand extends itself through random interaction with nucleotides swimming in the cytoplasm. Here the small circles stand for the phosphate bonds that glue adjacent base pairs. In real life, many enzymes facilitate the process and the complementary lagging strand needs to replicate backwards in pieces as the strands have a natural orientation that complicates replication. Reproduced from *Essential Cell Biology* by permission W.W.Norton & Co.

Obviously a full model of this process would be very complex but we can imagine its salient features being modeled like this: we are given one strand of the helix and we imagine each location where a new nucleotide is to be placed being in a 5-dimensional quantum state with basis states consisting in the location being filled by G, A, T or C or being 'empty'. In the empty state, the external hydrogen atoms of this location in the given strand are not bonded and there is a triple phosphate attached to the last filled location ready to drive the bonding chemical reaction. Energetically, an empty slot is best filled by the complementary base and this almost always happens, one location at a time. The process involves a whole squad of attending complex enzymes (e.g. DNA polymerase, DNA primase, DNA ligase, etc.) that oversee the work and correct almost all mistakes that inevitably get made.

Let's make a model of this, simplified as much as possible. Imagine an infinite chain of qbits in interaction with a heat bath and replace the energetic bias supplied by the existing strand by an energetic bias towards the value of each pair of adjacent qbits having the same value. Then the process in the model is that, starting with the initial state $|\epsilon_1 \epsilon_2 \cdots \epsilon_n \cdots \rangle$, the chain iteratively replicates its first bit, changing at the nth step like this:

$$|\epsilon_1 \epsilon_1 \cdots \epsilon_1 \epsilon_n \epsilon_{n+1} \epsilon_{n+2} \cdots \rangle \rightarrow |\epsilon_1 \epsilon_1 \cdots \epsilon_1 \epsilon_1 \epsilon_{n+1} \epsilon_{n+2} \cdots \rangle$$

The process ends when the whole chain is in the state $|\epsilon_1 \cdots \epsilon_1 \cdots \rangle$.

Such a change is obviously impossible without the heat bath because the qbit sequence undergoes an irreversible change, throwing away the old value of $|\epsilon_n\rangle$ at the nth step. Fortunately this information can be stashed in the heat bath, so the process can go forward! One can mimic the action of the attendant enzymes by assuming that the Hamiltonian changes, one qbit at a time to favor change equal to the previous qbit. The full behavior of a qbit with all types of Hamiltonian in contact with a heat-bath was worked out with extensive calculations by Leggett et al [LCD$^+$87], section VII, based on the influence function technique of Feynman and Vernon [FV63]. Their result is that a single qbit, in contact with a heat bath in thermal equilibrium will converge to its preferred state provided that its bias towards this state is sufficiently big compared to the tunneling energy and heat bath coupling. The result works regardless of the spectrum of the heat bath and its temperature. This is a big simplification of the DNA biochemistry, but I see no reason why this same behavior would not occur for the more complex replication of DNA.

What I want to assert is that, in the absence of wave function collapse, mutations are going to create entities like Schrödinger's cat: a phenotype in a mixed state with positive probabilities of being two macroscopically different animals. Ionizing radiation for example interacts with DNA molecules via either the photo-electric effect (being absorbed and ejecting an electron from some atom) or the Compton effect (interacting with an electron, loosing some energy while also ejecting the electron). This can be described by an S-matrix and leads to to a superposition of un-ionized and ionized states. The resulting small mutation is often corrected by the squad of attending enzymes but not always. If not, through DNA replication, it will affect the fully developed organism in one way or another. The key thing to remember is that Schrödinger's equation is linear, so if the world is in a superposition at time t_0 and if there is no collapse, the result will still be a superposition at time t_1, now of the consequences of the original two states. In other words, the result is the superposition of an unmutated DNA strand and a mutated one and, from that, a superposition of an unmutated animal and a mutated one – a cat-state. Thus DNA replication is like an open spigot transferring atomic level indeterminacy to the macroscopic world.

vii. Bohr bubbles and speculations

I want to return to the issue raised in the first section about nature of measurements and whether the "collapse of the wave-form" is somehow done by nature when effects cross the border from microscopic to macroscopic or whether they are the result of human observation. Speaking for myself, I find the pilot-wave theories and the collapse theories unconvincing, especially because of their difficulty incorporating relativity theory, not to mention the absence of any experimental support. And many-worlds seems just silly. I am left with the various epistemic viewpoints. I prefer to call this the "anthropic" viewpoint, not concealing but emphasizing its dependence on humans. Personally I find this less weird than the idea that nature somehow takes care of it by some unknown mechanism. I believe Wigner also expressed this point of view in an essay entitled "Remarks on the Mind-Body Question," Chapter 13 of [Wig62]. Here he imagined a scientist conducting an experiment while he has a friend in the next room, that he first makes a measurement by himself and afterwords, goes and tells his friend the result. (Here we make the scientist and friend male only to avoid the awkward circumlocution "he/she.") From his friend's perspective, did the wave function collapse when the scientist did his measurement or when he, the friend, was told the result? This sounds nit-picking: clearly knowledge is shared by a whole community but it shows that there are issues even if you accept the epistemological interpretation of measurements. Each scientist has "local" knowledge of what's going on in his/her lab and also shares knowledge with the community, thus making it "global." This sharing means that the macroscopic world can continue in its classical nature, one and the same for all people in a community, without any microscopic indeterminacy ever affecting it. This connects with philosopher's concept of "common knowledge," whose subtleties have been discussed by economists and computer scientists as well, see [FHMV95] and the Wikipedia page on the "Two Generals' Problem."

I think the right way to think of this is to imagine that the macroscopic world our community lives in is part of a "Bohr bubble" in which information is shared and macroscopic observables always have unique values up observational error. Whether in physics labs where atomic events are being probed or giving birth to babies with mutations, we do not tolerate superpositions of grossly distinct states, cat-states, hence our community is collapsing its macroscopic world, keeping us in its `AMU` set. So long as we live in our Bohr bubble, this means we must be continually and actively maintaining its classical nature. Our analysis above showed that the free photon field does not disrupt our bubble. It is for this reason that astronomers have extended our Bohr bubble billions of light years out in our past light-cone without difficulty.

But our natural world, its flora and fauna, is another issue. If you accept my analysis of DNA replication, accepting or rejecting mutations of living things is a major effect of our observations. We are, for example, continually observing our own bodies and, by doing so, we may well be deciding whether an internal mutation has caused cancer or not. Curiously, many mind/body medical specialists have suggested that our mental

attitudes affect susceptibility and response to cancerous mutations. This then opens a huge can-of-worms: could it be possible for free will to affect wave-form collapse, skewing the probabilities dictated by the Born rule, when the choice between two alternatives carries strong emotions?

My own favorite conundrum of this sort is archeological. Maybe dinosaurs were not conscious enough to maintain their Bohr bubble and the mesozoic ended with quantum ambiguous animals. Then their fossils retained their ambiguous state until your diligent archeologist unearthed them. Then its wave form finally collapsed. This would mean that archeologists have extended our communal Bohr bubble back to the mesozoic and, in a sense, *created* the fossil history we now possess. We may have created that astonishing Jurassic park ecosystem by our exhaustive modern explorations. I know this sounds utterly crazy, but I find it hard to definitively reject such a possibility.

Being at every turn so astonishing and unintuitive, quantum mechanics lends itself to speculation. I already talked about the idea of many-worlds and I am certainly not the first person to suggest that free will might play a role in wave-form collapse. But I want to end this riff by proposing another wild idea. The Greeks were often occupied in puzzling over the infinite divisibility of space as in the "paradox" of Achilles and the Tortoise. But, for sure, Eudoxus and Archimedes both wrote with a very modern understanding of how to formalize the mathematics of the real line, Eudoxus with an equivalent of the Dedekind cut and Archimedes with ϵ, δ arguments. My reason for recalling this is that maybe quantum mechanics is telling us the time has come to abandon real numbers, abandon Cartesian coordinates. It sure feels as though space (and time) are *utterly different* on the atomic level, that it has a different texture. Electrons are only localized if we force them to be and the wave/particle duality suggests that localization of particles in space-time times energy-momentum is both flexible and limited. String theory has been one way to alter $\mathbb{R}^4$ but another would be to let points go completely. This could be done in the "net of algebras" approach (cf. [Haa96]), but it was also done by Grothendieck when he invented the theory of topos and used it for define étale cohomology. Not having plain numbers there "at the bottom" to describe position is scary and I don't see where this might go, but it feels plausible that a new theory might be lurking there.

Chapter 15

Path Integrals and Quantum Computing

In the previous chapter, we have discussed the basic incompatibility between quantum mechanics and the classical world that requires the process of "collapsing the wave form" in order for the latter to sustain itself in the presence of superpositions in the former. This incompatibility has long been seen as a kind of barrier between two worlds – see Figure 1 in the previous chapter. So long as this was only an issue in the half dozen physics research labs around the globe, it seemed a matter of concern for a small subset of the intelligentsia to argue about. But there is a definite possibility now that a more intimate connection will be forged between the atomic and classical worlds: namely *quantum computing*.[1] If only atomic events of some medium complexity can be tamed, isolated and then measured *before any collapsing or interaction with the great stew of external atoms*, computations of tremendous complexity can be carried out and this will change our lives. It is not at all clear whether this will be possible but an awful lot of money is being poured into labs where multiple approaches are being played with and a few small successes have encouraged their devotees. If and when this works, the apparent barrier between the atomic and classical worlds will look a lot less formidable and quantum mechanics will really become part of our lives.

What is a quantum computer? One starts by assuming that one is dealing a small number of particles in a situation where they are constrained so that their degrees of freedom are described by a finite dimensional Hilbert space $\mathcal{H}_{qc}$. Typically one assumes the system consists of a set of qbits which means $\mathcal{H} = \bigotimes^n(\mathbb{C}^2)$ but I just take any finite dimensional system here. Then one assumes there is a base Hamiltonian H that, in a non-relativistic way, gives an evolution via the unitary operators $e^{iHt} : \mathcal{H}_{qc} \to \mathcal{H}_{qc}$. One also assumes one can turn on and off various external events, e.g. EM fields, that add perturbations to H_{qc}, coupling it with the outside. One such might set the quantum

---

[1]This Chapter expands my blog post *An Easy Case of Feynman's Path Integrals*, dated Nov.1, 2014

computer to an initial vector, others might alter its "program" at intervals and a last one might allow a "read-out." The key idea is to take advantage of superposition in $\mathcal{H}_{qc}$ to, in effect, do exponentially many computations simultaneously, overcoming the "P vs. NP" obstacle. The most famous paper here is Peter Shor's demonstration that huge numbers can, in principle, be factored by such computers, [Sho94].

What I want to explain in this chapter, however, is not the details of programming quantum computers but how I came to understand Feynman's approach to quantum mechanics by asking what it said for quantum computers and also how it can be used to treat the effect of coupling the small computer space with the rest of the world. After I posted the blog on which this chapter is based, I discovered that there is considerable literature on the "sum-over-histories" approach to quantum computers, e.g. [DHH$^+$05, RG06, PKS17]. First, a little background. In the late 1940's, physics was abuzz with multiple ways to model fields. Feynman devised a scheme all of his own, computing the probability of measurement A being followed by B by *integrating over all possible paths of all particles leading from A to B including all possible interactions even with particles that appear and disappear.* Freeman Dyson describes it like this[2]:

> *Dick Feynman told me about his sum-over-histories version of quantum mechanics. "The electron does anything it likes," he said. "It just goes in any direction at any speed, forward or backward in time, however it likes, and then you add up the amplitudes and it gives you the wave-function." I said to him, "You're crazy." But he wasn't.*

Like many pure mathematicians, I have been intrigued over the meaning of Feynman's path integrals and put them in the category of weird ideas I wished I understood better. The idea that when asking to compute a quantum evolution given by a one-parameter group of unitary transformations U_t, you need to consider every possible way the underlying quantum system might go from one state to another is, in some sense, obvious. Namely, because $\left|\langle \phi_f, U_{(t_f-t_i)}(\phi_i)\rangle\right|^2$ is the probability of ϕ_i leading to the outcome ϕ_f and because

$$\langle \phi_f, U_{(t_f-t_i)}(\phi_i)\rangle = \langle \phi_f, U_{(t_f-s)} \circ U_{(s-t_i)}(\phi_i)\rangle$$
$$= \sum_k \langle U_{(t_f-s)}^{-1}(\phi_f), \phi_k\rangle \cdot \langle \phi_k, U_{(s-t_i)}(\phi_i)\rangle$$

for any intermediate time $t_i < s < t_f$ and any orthonormal basis $\{\phi_k\}$ of the Hilbert space, the group property shows immediately that you must sum something over all possible states at any intermediate time s. Matrix multiplication is even more clearly about paths: take any $n \times n$ matrix A, then its powers A^N are given by the usual formula:

$$(A^N)_{i_0,i_N} = \sum_{i_1,\cdots,i_{N-1}} A_{i_0,i_1} \cdots A_{i_{N-1},i_N}.$$

---

[2]Address of March 1979 at the Princeton Einstein Centennial published in [Woo80], p. 376

One can think of this in a new way: let $S = \{1, 2, \cdots, n\}$ be thought of as a discrete "state space." Then $\{i_0, i_1, \cdots, i_N\}$ is a discrete path from the set of "times" $[0, N]$ to space S, and the matrix coefficients of the power are sums of terms, one for each such path from some given column index to some given row index. This is so simple and obvious but it is the root of Feynman's remarkable idea.

Instead of powers of a matrix, we need to consider a 1-parameter group of unitary matrices obtained by exponentiating a fixed self-adjoint matrix H, namely $U_{qc,t} = e^{itH_{qc}}$. In our case, we fix an orthonormal basis of $\mathcal{H}_{qc}$ and S is the discrete set of basis vectors. A path in S will simply mean a sequence of constant intervals interspersed with jumps from one basis vector to another, like a frog jumping on lily pads. This is the finite version of what Feynman introduced in his path integral formalism for quantum mechanics. In a more challenging case, $\mathcal{H}$ could be $L^2(\mathbb{R})$, $\mathbb{R}$ is now the space S and U_t could be an integral operator given by convolution with a kernel $K(x, y, t)$. Then his goal was to write $K(x, y, t)$ as an integral over all paths $\gamma(s) \in \mathbb{R}, s \in [0, t]$ starting at x and ending at y of an expression involving the path and K. Feynman thought of these as paths of an underlying classical particle moving in $\mathbb{R}$. Of course, the set of paths is an infinite dimensional manifold and then to integrate over all paths one needs a measure on this set of these paths with respect to which one can integrate. Finding the appropriate measure is one problem and showing the integrand he needs is in some sense integrable turned out to be even harder. A crazy trick for "evaluating" highly oscillating Gaussian integrals with imaginary exponents is to add a tiny negative regularizing term to the exponent, evaluating as usual and then letting the nudge go to 0!

I want to work out his approach for finite dimensional U_t where everything is quite elementary and rigorous. The path integral formalism also turns out to be the convenient one to use when you treat the interaction of this elementary quantum computer with the external world from which it can never be totally insulated.

Start by fixing a large integer N . Then:

$$(U_{qc,t})_{a,b} = \left((U_{qc,t/N})^N \right)_{a,b}$$

$$= \sum_{a=k_0,k_1,\cdots,k_N=b} \prod_{\ell=1}^{\ell=N} (U_{t/N})_{k_{\ell-1},k_\ell}$$

Now if $N >> 0$, $U_{qc,t/N} = e^{itH/N}$ is approximately equal to $I + (it/N)H_{qc}$. Thus if at some ℓ, $k_{\ell-1} = k_\ell$, the term in the product is near 1 while otherwise it is a bounded number divided by N, hence very small. From this we see that the more jumps the sequence k_ℓ makes, the smaller the corresponding term in the product. So let J be the number of jumps and consider the sparser sequence of values $a = k_0, k_1, \cdots, k_J = b$ where now $k_{\ell-1} \neq k_\ell$ for all ℓ . The jumps take place at particular 'times' ℓ_j/N and we reformulate the above

expression as:

$$
\approx \sum_{J=0}^{\infty} \left(\frac{t}{N}\right)^J \sum_{\substack{a = k_0 \neq k_1 \neq \cdots \neq k_J = b \\ 0 = \ell_0 < \ell_1 < \cdots < \ell_J < \ell_{J+1} = tN}} \prod_{j=1}^{j=J} iH_{qc,k_{j-1}k_j} \cdot e^{i \sum_{j=0}^{j=J} \frac{\ell_{j+1}-\ell_j}{N} H_{qc,k_j,k_j}}
$$

It shouldn't be hard to quantify the approximation error here but let's skip this and pass quickly to the limit as $N \to \infty$ where the expression becomes exact again. This leaves the k sequence alone but now the ℓ_i/N's are replaced by intermediate times t_i in the interval $[0,t]$ where the jumps take place, the sum over ℓ's is replaced by an integral over the t's and you take into account the constant needed when the sum over the ℓ's is looked at as a Riemann sum for the integral over the t's. Note that the integrand is bounded by a constant to the power J and the integral is over a simplex with volume $t^J/J!$, hence we get convergence of the sum over J. What comes out is this:

$$
= \sum_{J=0}^{\infty} \sum_{\substack{a = k_0 \neq k_1 \neq \cdots \neq k_J = b \\ 0 = t_0 < t_1 < \cdots < t_J < t_{J+1} = t}} \int \prod_{j=0}^{j=J-1} iH_{qc,k_{j-1}k_j} \cdot e^{i \sum_{j=0}^{j=J}(t_{j+1}-t_j)H_{qc,k_jk_j}} dt_1 \cdots dt_J
$$

Going a step further, let X be the path space of piecewise constant functions $f : [0,t] \to \{1,2,\cdots,n\}$ with a finite number of jumps. X breaks up into pieces X_J according to the number of jumps and these into pieces depending the the sequence $\vec{k}$ of values of f and finally what remains are simplices in $\mathbb{R}^J$. We have the euclidean measure on these components, giving a finite measure μ_X on X. Let $X_{(a,b)}$ be the paths that begin at a and end at b. Then we get:

$$
(U_{qc,t})_{a,b} = \int_{k \in X_{ab}} e^{iS_{qc}(k)} d\mu(k)
$$

$$
S_{qc}(k) = \int_0^t \Big[H_{qc,k(t)k(t)} - i \sum_{\text{jumps } t_j} \delta(t_j) \log(iH_{qc,k(t_j^-)k(t_j^+)}) \Big] dt
$$

I guess the expression in the square brackets is what physicists would call the "Lagrangian" although I've never seen one like this. The term in brackets is real only if all the off-diagonal elements of H_{qc} have absolute value 1. Then we have the final theorem stated for any matrix H and dropping the i:

Theorem. *For any $n \times n$ matrix H, the matrix entries of e^{tH} are the integral over all piecewise constant paths $k \in X_{ab}$ of the exponential of $\int_0^t H_{k(t),k(t)} + \sum_{\text{jumps } t_j} \delta(t_j) \log(H_{k(t_j^-),k(t_j^+)})) dt$.*

I'm not sure one can convince college teachers of this but this result fits easily into the curriculum of undergrad linear algebra courses!

Feynman's more general theory describes the evolution U_t of very general quantum systems by integrating over a set of paths in an appropriate set of states:

$$\langle x, U_t(y) \rangle = \int_{\text{paths } \gamma(t)} e^{iS(\gamma)} d\mu(\gamma), \quad S(\gamma) = \int_0^t L(\dot{\gamma}(s), \gamma(s)) ds$$

where $\gamma(0) = y, \gamma(t) = x$, L is some kind of classical Lagrangian for the whole system, S is the action and μ is a measure on the set of paths (see [FV63], formula (2.2)). The states can be an orthonormal basis as in the above description of a quantum computer or the distributional eigenvectors $\delta_x(\cdot)$ of multiplication by the coordinate if the Hilbert space is $L^2(\mathbb{R})$. In the latter case, we realize the original idea that "the electron can go anywhere it wants," $\gamma(t)$ being its path.

A classic 1963 paper of Feynman and Vernon [FV63] extended his idea of computing the evolution of one system by summing over histories to describing the perturbation of one system, e.g. $\mathcal{H}_{qc}$ caused by coupling it with another system, e.g. the outside world. This is a beautiful application of his method of writing propagators U_t as integrals over all paths in state space.

As we said, quantum computers are made by isolating a tiny atomic setup from the whole buzzing, booming world so that its behavior is given by exponentiating a finite dimensional self-adjoint operator H_{qc}. Minimizing this interaction is the central challenge in manufacturing a real live quantum computer. But the rest of world always intrudes to some extent and this is often modeled by a tensor product $\mathcal{H}_{qc} \otimes \mathcal{H}_{\text{ht}}$. The second factor is the inevitable intrusion of the messy outside world into the system. It is another Hilbert space often referred to as a heat bath because it may be assumed to be at or near thermodynamic equilibrium. The evolution will then be described a joint Hamiltonian operator $H_{tot} = H_{qc} \otimes I_{ht} + I_{qc} \otimes H_{ht} + H_{int}$ where H_{int} is the interaction term. Then the system evolves according to $U_{tot,t} = e^{itH_{tot}}$.

Then if, in our coupled system, paths are made up of an independent pair, $k(t)$ in the quantum computer and $x_{ht}(t)$ in the heat bath, and if the action splits $S = S_{qc} + S_{int,ht}$, then, by separating the integration over the two sets of paths, one gets:

$$\langle (k(t), x_{ht}(t)), U_t(k(0), x_{ht}(0)) \rangle = \int_{\text{paths } k(s)} e^{iS_{qc}(k)} . \mathcal{F}(x_{ht}(t), x_{ht}(0); k) . d\mu(k)$$

$$\mathcal{F}(x_{ht}(t), x_{ht}(0); k) = \int_{\text{paths } x_{ht}(t)} e^{iS_{int,ht}(k(s), x_{ht}(s)) ds} . d\mu(\gamma_{\text{ht}})$$

where $\mathcal{F}$ is called the "influence function."

But we don't really know, nor are we interested in the exact state of the heat bath. We need to "trace out" the heat bath factor if we want to describe its effects on the quantum computer. This is done by retreating a bit from describing the system by a single state and accepting that we need to describe it as a *mixed* state. A mixed state is a probabilistic combination of many states described by a *density matrix*. If the mixture is made up of

a set of orthonormal vectors $\vec{x}^{(a)} \in \mathbb{C}^n$, each with a probability $p(a)$, so that $\sum_a p(a) = 1$, then one defines the density matrix describing this mixed state by the Hermitian matrix:

$$\rho_{i,j} = \sum_a p(a)\bar{x}_i^{(a)}x_j^{(a)}.$$

It's a sad fact of life that any system entangled with messy parts of the world needs to be described by these ρ's and is never "pure" anymore. A density matrix evolves by conjugating it with U_t. Using integration over paths, we now need *two paths* γ, γ' so that the density matrix, given here by its kernel $\rho(x, y)$, evolves like this[3]

$$\rho(x, y; t) = \iint e^{i \int_0^t S(\gamma'(s)) - S(\gamma(s))ds} \rho(u, v; 0) d\mu(\gamma) d\mu(\gamma') du dv$$

$$\gamma(0) = u, \gamma(t) = x, \gamma'(0) = v, \gamma'(t) = y.$$

Inserting an influence function factor $\mathcal{F}(\gamma, \gamma')$ in the above formula, we get a way to compute how one system is perturbed by coupling it with a heat bath.

I want to sketch how one explicit example of the beautiful idea of influence functions comes out: the simplest case of coupling a 1-qbit quantum computer with a heat bath. First of all, how do we model a heat bath? The idea is take a set of independent quantum harmonic oscillators, each tuned to its own frequency ω and coupled to the rest of the world by some linear function. Quantum harmonic oscillators were described in §v of the last chapter. We use a collection of simple harmonic oscillators with some spectral density $J(\omega)$, each oscillator starting at time 0 with the density operator of a thermodynamic equilibrium at a given temperature T (e.g. the probability of the n particle state proportional to $e^{-\beta n}, \beta = 1/k_B T$, T the temperature, k_B Boltzmann's constant). The simplest, and well studied, case is the two dimensional quantum computer, i.e. one Qbit, also known as a "spin boson" system. Conventionally, its two states are given the values $\{+1, -1\}$ so the paths just jump between the two values (aka "tunneling") at some sequence of times. We write its 2×2 Hermitian as H_{sb}. Because we're dealing with density matrices, we need to integrate over not one but *two* piecewise constant paths (k, k'). Then, putting everything together and evaluating some integrals, the final result comes out as:

$$(\rho_{\mathcal{H}_{sb}}(t))_{a,b} = \iint (\rho_{\mathcal{H}_{sb}}(0))_{k(0),k'(0)} \mathcal{F}(k, k') e^{i(S_{qc}(k) - S_{qc}(k'))} d\mu(k) d\mu(k')$$

$$\log(\mathcal{F}(k, k')) = \iint\limits_{0<r<s<t} iL_1(s - r)(k(s) - k'(s))(k(r) + k'(r))$$

$$- L_2((s - r)(k(s) - k'(s))(k(r) - k'(r))drds$$

---

[3]This is Feynman's original description in [FV63]. But for harmonic oscillators, an explicit formula known as Mehler's kernel solves Schrödinger's equation so Feynman's sum over histories and its semi-rigorous $d\mu$ are not needed – though see [Fol08] , Ch. 8 for how Feynman's integral is computed.

where L_1, L_2 are determined by the temperature, coupling, and frequency spectrum of heat bath. This is a complicated result but what needs to wrestled with if ever a useful quantum computer is built. We don't want to go through the proof here but, besides Feynman's original paper, a detailed description of the above and how this works out is in [LCD$^+$87], formula (4.5) and the recent book [Wei12], Ch.21, esp. formula (21.2).

Part VI

Nothing is Simple in the Real World

There is a curious similarity between mathematicians and politicians: both of them strive to achieve their goals by *simplifying a situation*. In the mathematician's case, they start with a bewildering set of questions, maybe some aspect of the real world, like modeling waves in water, or maybe some internally generated puzzle. But to make progress, they try to extract its essence in as simple a form as possible, eliminating everything but one hard problem inside the mess and then they work on that. In the politicians case, they decide to simply ignore 90% of the issues and harangue their voters on the one point which they believe will be heard and lead them to power. Make the voter's choice sound like good vs. bad and make it clear who wears the white hats, who the black hats.

In any case, this part of my book is focussed on my journey from a naive, privileged youngster doing math to events that startled and roused me to get involved in issues in the real world, and to study, sometimes even do some small thing, outside math. I slowly realized I needed to face up to the complexity of all real world problems. Without a doubt, one of the greatest privileges that all successful scientists enjoy is to travel the world and meet their colleagues in many countries. As a result, they see a bit about how other people live, what issues and religions in their country move their citizens. And then they may gradually become more sensitive to the complexities at home but also aware of its idiosyncratic nature. So this the "beyond" in the book's subtitle.

Many mathematicians have resisted getting involved in political issues, math being a wonderful place to escape. Recently when "woke" politics became unavoidable, more professors including mathematicians have found politics affecting their lives. In my early career, I just wanted to "do math" and I closed my eyes to the civil rights movement, the chaos of the late 60's, the hippy movement and the protests over the war in Vietnam. In fact, I didn't even join the AMS until Lipman Bers, as President, scolded me and even then, I only consented if he agreed to write me a letter saying "welcome but don't expect to be on any committee because you are unfit for such things." But increasingly, over the years, some events cracked my self-imposed shell and forced me to be a little less naive and I tried to understand some "big issues". This part of my book describes several such events. Some of them started with math friends who shared a bit of their struggles and their passions with me and I responded in a small way. Others are just the restlessness of a polymath trying to make some sense of human life.

Chapter 16 is about the crisis in math publishing and the advent of the internet in the 90's. I had reluctantly agreed to join the Executive Committee of the International Mathematical Union (the IMU) and through this, especially when I served a term as President, I was forced to get involved in publishing and the burgeoning internet. The chapter *Wake Up!* describes how my best intentions went nowhere. More specifically, the publishers Klaus and Alice Peters were close friends for many decades and I supported them when I could but this experience opened my eyes to the complexity of dealing with many parties with their own objectives and competing business models.

Chapter 17 starts with how I grew up and how I lived in a truly international world where cultures and religions mixed without rancor. But I first began to understand a

little about how complex the world was by the movement to rescind Igor Shafarevich's membership in the NAS on the basis of antisemitism. I knew Igor and our interaction is described here. And later I was caught up by anti-Dalit (the bottom "untouchable" strata of Indian society) actions of the Director of the Indian Institute of Technology-Madras relayed to me by my good friend Shiva Shankar. Again, I put my foot in it. More broadly, this chapter addresses the tension between the ideal of an international liberal democracy and the reality of strong national traditions and the communities these foster. I go on to describe a few other experiences in a variety of countries including the Middle East.

Chapter 18 discusses one of my few meaningful brushes with religion. It started with my deciding to read Bertrand Russell's "History of Philosophy" hoping to make some inroads into this formidable area. I soon learned that there was a big obstacle: all these early philosophers discuss "substances" and I had no idea what these were. Playing around like a dizzy first year grad student in philosophy, somehow I hit upon Spinoza and I was entranced. Amazingly, his magnum opus is written like math: all numbered propositions with proofs and cross references. It likely can be reduced to pure logic. And he spells out a uniquely attractive form of religion.

Chapter 19 started as a "Letter to my grandchildren" giving my best shot at intermediate term predictions. By this I mean, not predicting the short-term outcome of the chaotic political culture we have recently fallen into nor predicting anything about the world a few centuries hence, where no one has a clue. Rather it deals with the middle term future, say 50 year predictions. I am not a believer in the apocalyptic "singularity" theory of Kurzweil but I do discuss some of its components.

As these chapters involve many hot button issues, I feel I need to say where I stand quite simply. I believe all these political issues are complex and *do not have simple solutions*. The far left and far right in the US (and in many other countries) are both churning up anger, demonizing the other side and making rational discussion almost impossible. Whenever possible, I favor the middle path and believe that even in the most contentious issues, people with mutual respect can find such a path forward. Perhaps the Quakers have a better way of engaging people in conflict. The Quaker approach means working with both sides. Believing that violence begets violence, they advocate direct personal interaction in a neutral context with those who seem racist or intolerant. Ideally this leads to seeing the humanity in "the other side" and, one hopes, to actual friendships. I recently read the beautiful story "Apeirogon" , [McC20], based on the true friendship between an Israeli and a Palestinian both of whom lost a daughter to the conflict. Currently, abortion is one big issue looming large over the US. While Europe has found a middle course, varying slightly from country to country, the US is consumed with angry partisans who say "all this" or "all that" and neither side can hear the humanity in the other. This is crazy when polls show that the majority of people are OK with something like the European compromise. Indeed, Math is so much easier than politics!

Chapter 16

Wake up!

The world of professional publishing, of scholarly communication, has been in a state of profound transformation since the 90's when online publication became widespread. In some fields, for example physics and computer science, researchers have embraced this transformation and have forged new policies and better customs. In my experience, however, mathematicians are one of the most conservative research communities, clinging to old habits in spite of the opportunity to improve their working life. The impetus for this post on April 1, 2015 was the death of Klaus Peters, a publisher who, more than any other person that I have met, saw publishing in mathematics as a *service to the professional community* and strove tirelessly to find new ways to assist our community. The changes that have happened in the commercial publishing world deeply disturbed him. Some things have improved since then, some not, but I still want to suggest to my colleagues that they themselves really control the business model of research math publishing since it depends 100% on their writings and they should be open to radical changes. To paraphrase an old left wing slogan, you have nothing to lose but the chains that are binding you to exploitation by greedy for-profit publishers and you can gain a freer, simpler world to work in.

Book and journal publishing have been rocked by two major changes during my lifetime. The first was the takeover of smallish niche publishers by their Chief Financial Officers, subsequent mergers and the entry into this business of private equity firms. The second was the expansion of the internet to a state where it can provide instant availability of whole libraries everywhere at your fingertips.

i.   Springer and Klaus Peters

Let me start with what publishing used to be. In the 50's my first wife worked for Houghton-Mifflin, reading (and usually rejecting) submitted fiction. In those days, it was typical for an author to form a life-long relationship with a specific editor who would see him or

208

Chapter 16

Wake up!

The world of professional publishing, of scholarly communication, has been in a state of profound transformation since the 90's when online publication became widespread. In some fields, for example physics and computer science, researchers have embraced this transformation and have forged new policies and better customs. In my experience, however, mathematicians are one of the most conservative research communities, clinging to old habits in spite of the opportunity to improve their working life. The impetus for this post on April 1, 2015 was the death of Klaus Peters, a publisher who, more than any other person that I have met, saw publishing in mathematics as a *service to the professional community* and strove tirelessly to find new ways to assist our community. The changes that have happened in the commercial publishing world deeply disturbed him. Some things have improved since then, some not, but I still want to suggest to my colleagues that they themselves really control the business model of research math publishing since it depends 100% on their writings and they should be open to radical changes. To paraphrase an old left wing slogan, you have nothing to lose but the chains that are binding you to exploitation by greedy for-profit publishers and you can gain a freer, simpler world to work in.

Book and journal publishing have been rocked by two major changes during my lifetime. The first was the takeover of smallish niche publishers by their Chief Financial Officers, subsequent mergers and the entry into this business of private equity firms. The second was the expansion of the internet to a state where it can provide instant availability of whole libraries everywhere at your fingertips.

i.   Springer and Klaus Peters

Let me start with what publishing used to be. In the 50's my first wife worked for Houghton-Mifflin, reading (and usually rejecting) submitted fiction. In those days, it was typical for an author to form a life-long relationship with a specific editor who would see him or

Figure 16.1: Alice and Klaus Peters, photo courtesy of Stan Sherer.

her through the ups and downs of their creative muse and become an intimate friend. This sleepy world is nicely captured in J. L. Carr's satire *Harpole & Foxberrow, General publishers*, [Car92]. This is also the world in which the greatest mathematicians of the world (including Hilbert, Einstein, Courant, Caratheodory, Hecke, etc.) could write in 1923 a letter of appreciation to Ferdinand Springer for saving the then leading journals *Mathematische Annalen* and *Mathematische Zeitschrift* from bankruptcy. This letter, a copy of which was given to me by Klaus Peters, is displayed in Figure 2. There was at that time a partnership between authors and specialized publishing firms that understood their needs and tried to serve them while doing business. Klaus recalled this spirit when he met Ferdinand Springer sometime in the 1960's in these words:

> *One day my phone rang: "Springer here, please come to my office." Ferdinand Springer, the legendary publisher, did not usually deal with junior members of the staff nor had I been formally introduced to him. I went to his office unsure what this all meant. His personal secretary kindly advised that I should listen and quietly excuse myself when the 'audience' was over. On entering his office I was greeted warmly as the new mathematics editor. Mathematics was one of Springer's favorite programs. He then proceeded to explain the raison d'être of a publisher: to facilitate the work of the authors by taking away the burdensome aspects of editing, producing, and most importantly distributing their work widely. He made it very clear that these added values were the justification of a publisher's existence.*
>
> *His fierce loyalty to authors and editors is confirmed by another story. When Ferdinand Springer sought to leave the occupied city of Berlin after World War II to rescue his family, he was stopped at a military control post. The com-*

Figure 16.2: The 1923 letter from the leading German mathematicians of the day to Herr Ferdinand Springer expressing their appreciation for his "opferbereite unternehmungslust" –zest for action, ready even to make sacrifices (tr. Peter Michor).

manding Russian officer demanded an explanation. Springer identified himself as a publisher of scientific books and journals (in his mind that was explanation enough) whereupon the officer commanded, "Tell me the names of the editors of such and such journal!" Springer had retained the names of Russian scientists and editors on the masthead of the journals that they had served, despite the war. As he recited these names, the officer suddenly interrupted, "That's me, and I am honored to meet you." He provided Springer with free passage which allowed him to rejoin his family.

Klaus went on to nearly single-handedly rejuvenate Springer-Verlag's mathematical program, bringing it back to its pre-WWII status as the leading math publisher in the world. He introduced the Lecture Note series and got to know most of the leading mathematicians of his generation, often soliciting new books from the world's top experts. But things changed: in the late 70's, Springer's CFO was made the director with the final say. Klaus and Alice, his wife and partner in all his work, resigned in protest as they felt the editorial department should run the place. In Springer's own self-published history, Klaus's role was completely erased! At the same time, all the small math publishers were being swallowed up or their math series discontinued (van Nostrand, Wiley-Interscience, Benjamin, etc.). One saw journal prices for the leading journals go sky-high and prices of later editions of older books were raised to match those of the newest books. Circulation took second place to quarterly profits, often based only on library sales. Klaus and Alice continued to seek a position where the traditional values of publishing were respected, moving to the Swiss publisher Birkhauser until it was swallowed by Springer, then to Harcourt-Brace-Jovanivich until it was bought by General Cinema and finally striking out on their own as *AKPeters*.

Springer's turnover of control of its operation from editors to accountants did not take place out of the blue. The full story is told in a brilliant Guardian article written in 2017 by Stephen Buranyi and available online [Bur17]. The transformation of the scientific publishing business was driven by Robert Maxwell and his creation Pergamon Press (later bought by Elsevier). He was a larger than life character, tall, brash, Czech by birth but became British through intelligence work in WWII, eventually a multi-millionaire celebrity who drowned under mysterious circumstances. But his genius was to realize how scientific journals were cash cows, material produced and vetted *for free* by the scientific community with guaranteed librarian customers, no matter if the price was in the thousands. By minting journals in every subsubdiscipline and courting the scientific elite, Pergamon was amazingly successful and became the envy of all the other scientific publishers – so Springer was forced to abandon its lofty ideals and follow. But at the peak of its success, he sold Pergamon to Elsevier, which went on to invent an even more lucrative trick: they *bundled* their thousands of journals, so libraries were forced to pay for the whole package, including vast numbers of junk as well as their prestige journals

The buyout and merger mania in the pursuit of higher profits and the abandonment

August 22, 2002

Dear Dr. Mohn,

The world mathematical community is gathered in Beijing at its quadriennial International Congress. We, the presidents of the major mathematical societies in the world, write to you in connection with the anticipated sale of BertelsmannSpringer.

For more than one hundred years, there has been a close and mutually beneficial collaboration between Springer-Verlag and the international mathematical community. For example, in the 1920's Ferdinand Springer rescued the two pre-eminent mathematical journals of the time, Mathematische Annalen and Mathematische Zeitschrift, resulting in the letter of appreciation of leading mathematicians of the decade, that we attach. The close association of Springer-Verlag with the mathematicians of the world was reborn after World War II and has grown over the decades so that Springer remains not only the largest publisher of research level mathematics but also a valued partner in the international research enterprise.

The world of publishing is changing today in several ways. One is due to consolidation and loss of personal contacts between publishers and authors. Another is the introduction of electronic modes of publication, restructuring the business models but also possibly enabling or facilitating access to scientific information by people in developing or financially troubled countries. In this situation, a strong publisher like Springer with close ties to the mathematical community is needed more than ever.

Knowing your concern for the many communities with which you have been working for so long, we are taking the liberty of proposing to you a rather radical move. Envisioning an economically viable enterprise which will keep the academic spirit of Springer-Verlag, we suggest the reorganization of at least its mathematics portion as a not-for-profit foundation.

Rolf Jeltsch, the president of the European Mathematical Society, is the person to contact would you wish to discuss this issue further.

Respectfully yours,

International Mathematical Union
Jacob Palis, The President

International Council for Industrial and Applied Mathematics,
Olavi Nevanlinna, The President

European Mathematical Society
Rolf Jeltsch, The President

Deutsche Mathematiker-Vereinigung
Peter Gritzmann, Der Präsident

Société Mathématique de France
Michel Waldschmidt, Le Président

American Mathematical Society
Hyman Bass, The President

Sociète Mathèmatique de Canada
Christiane Rousseau, La Présidente

Figure 16.3: The 2002 letter from assembled Presidents of math societies from around the world at ICM2002 to Dr. Mohn asking him to consider formation of a not-for-profit to manage Springer's math program. Naive but hoping that Dr. Springer's lifelong dedication might still resonate. In my files as Past-President of the IMU at that time.

of "service" continued. A controlling interest in Springer itself was bought by the privately held publishing and mass-media conglomerate Bertelsmann in 1999. When they put Springer on the market in 2002, a group of us at the Beijing ICM tried a last ditch attempt to appeal to the Mohn family who owned Bertelsmann for an alternate solution. A letter signed by the Presidents of the IMU, ICIAM, EMS and the math societies of Germany, France, Canada and the US was sent to Dr. Mohn, recalling the partnership of Springer and the math community and asking him to consider the formation of a not-for-profit foundation to continue this partnership. The letter is reproduced in Figure 3. *We received no reply or response of any kind.*

Subsequently, Springer has been sold three times to private equity firms: in 2003, to the British investors Cinven and Candover who acquired and merged both Kluwer Academic Publishers and BertelsmannSpringer; next to the private equity firm EQT Partners and the Government of Singapore Investment Corp.; and again in 2013, to yet another private equity firm BC Partners. Only Mitt Romney seems to have missed the boat. And why does private equity scramble to own scientific publishing firms? The article [Bur17] cited above notes that in 2010, Elsevier's scientific publishing arm posted a profit margin of **34%**, a higher rate than Apple, Google and Amazon.

If any mathematician doesn't realize that a large part of his or her professional life is mortgaged to capitalists, perhaps they have spent too much time thinking only about theorems. Private equity buys a firm for one and only one reason: they believe they can squeeze more profits out of its operations, i.e. out of us mathematicians (and our societies and libraries). As Klaus put it in a piece entitled "A Vanishing Dream" on which he was working a few weeks before his death:

> *Alice and I feel that we have lived a dream to preserve and provide a service that was once considered worthwhile. I mean "publishing as a service." ... That this concept (with few exceptions of small individual publishers) is widely lost is no secret but what bothers me intellectually is the fact that publishing companies can be run financially successfully without an intellectual mission and without thought to optimize sales (by numbers of copies) or to produce well-edited and designed books. They compensate these shortcomings by optimizing the bottom line through skimping on editorial and production cost and offsetting revenue loss from smaller per-title sales (by number) by inflating prices.*

ii.   The Impact of the Internet

Let's talk about the cause of the second huge change in our professional life: the internet. It was not clear to me in the early 1990's how the internet would do anything to our working lives except speed up communication, replacing some types of letters by emails. My eyes were opened when Philippe Tondeur proposed that the math community could and should digitize *the entire corpus of mathematical books and journals* and make them available to

all and sundry: a *World Mathematical Library*. Wow!, was this really possible? Of course, its practicality is obvious now and Google has gone even further, seeking to digitize all written material. From this, it's only a small step to ask: why put math on paper at all? If something is on the web (and not password protected), anyone can get it and either read it on the screen or print it out if they prefer.

Full of enthusiasm for this brave new world, Peter Michor and I worked to involve the International Mathematical Union (the IMU). We set up its Committee on Electronic Information and Communication (CEIC) that, we hoped, would help mobilize the mathematical community in navigating this transition. Now I realize how naive this was, not because the early dreams were unrealizable, but because human nature is complicated and *fast action was needed to stay ahead of aggressive publishers*. A big meeting of all the groups doing digitization of math was organized in Washington DC where the various obstacles were discussed and it was proposed that the IMU could serve as an umbrella group coordinating the half dozen initiatives that had been started. But it was a case of "all Chiefs and no Indians" (as US children say when they can't form a team and don't tell me I'm not woke – I know that): none of the digitizers wanted to cooperate if this meant modifying their ongoing efforts in any way, shape or form. I had two chances to talk at length with John Ewing, then Executive Director of the AMS, but his conservatism made him very reluctant to consider any radical change in the math publishing business model. The AMS was at that time financially dependent on the traditional publishing model and John was building up its 100 million dollar nest egg. On the CEIC, John's deep knowledge of copyright complexities resulted in stymying all pro-active initiatives that we might have promoted at that point. It was not long before the commercial publishers asserted that their copyrights blocked wide electronic sharing of older articles and found a new source of revenue in these older articles that they had previously thought were worthless. Springer has locked up its back issues in "Springer Link." Note how different this is from the idea of a library where everything published is available for nothing. In yet another twist, "open access" journals with exorbitant per article charges (e.g. 3000 euros!) are now proliferating. More recently Springer realized that even books out of copyright could generate new revenue and offered authors the "benefit" of keeping their books in print indefinitely by voluntarily extending copyright to infinity. Actually, you can get nearly all math books free online at the rogue Russian "Genesis Library," with websites libgen.in and gen.lib.rec.ec (most of my books are there – help yourself). Which do *you* prefer: lunch money royalties once a year or wider free distribution of your books?

Let's speculate on what an internet-based professionally controlled working environment might be:

- All journals would be online and free, including all their back issues.

- A selection of libraries would maintain paper copies and mirror online content.

- Journals would all maintain their current refereeing policies so they continue to certify

the quality level they are known for, while unrefereed websites like the arXiv would offer immediate dissemination.

- All mathematical books would be available online, with the author(s) free to choose their business model, i.e. self publish (as in present day Springer Lecture Notes) or work with a publisher who provides editing, formatting, print versions and advertising by agreement.

Of course, I hear loud cries of "who pays?." Yes, many necessary services are not free. But moving to something like the above would free up large amounts of library money currently being spent for overpriced journals, e.g. Springer and Elsevier (maybe even shaming NYU into reducing its ridiculous price for *Communications in Pure and Applied Math*). The cost of running an online journal is certainly fairly small, though by no means zero. There are no printing, mailing and storage costs and no subscription record keeping. Refereeing is done for nothing, manuscripts are prepared by the author in latex with fixed formatting packages so they are ready to post, editing beyond a spell check is a luxury we can omit, esp. in our multi-lingual world where the niceties of grammar are increasingly forgotten or never learned by foreign speakers. (I can't resist describing the "law of conservation of **s** that I learned from my student Tai Sing Lee, namely – "several author**s** write; one author write**s**.") I don't feel that finding funds for such journals can be too big a problem, especially considering the above mentioned library funds. An ingenious combination approach called "Subscribe to Open" (S2O) is gaining traction, especially in the European Math Society. Here, the income from subscriptions to a journal is counted up year by year. Once, in any given year, it reaches a threshold high enough to support the publishers costs, the next year becomes fully Open Access to all. Clearly, the hope is that libraries will continue to subscribe and this should be enough to cover expenses, hence the journal becomes forever open access to all individuals, i.e. the library budgets will be funneled to creating Open Access to the community. If this takes hold, maybe we can break the strangle hold of commercial operations.

Mathematicians, by nature, want to concentrate on their work and resist worrying about the mechanics of communicating their results to their colleagues. But business models for publishing are changing rapidly in this digital age and whether the ultimate control rests in our hands, the hands of the professional community, or in the hands of financial concerns who shift money from sector to sector following the scent of profit, the choice is something we ought to be aware of. I hope that the new pro-active CEIC, the great interest shown at the Seoul ICM in three panels on the impact of the internet and mathematical publishing and the AMS's introduction of online journals all indicate that the whole community is moving towards this choice.

Most of the above was in my original post. Partly, I wanted to reprint that post because I'd like to keep alive the memory and legacy of Klaus and Alice Peters. And partly I wanted to illustrate how hard it is mount an international effort to address an international problem. My sense is that in the intervening 6 years, there has been gradual
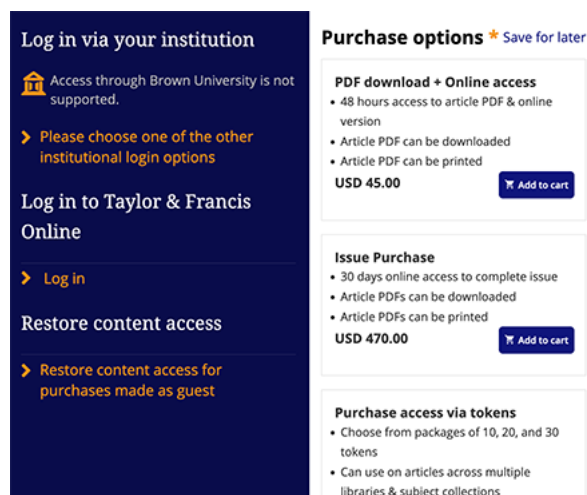
Figure 16.4: Journals are still milking the scientific community: a screenshot of my computer of a Taylor & Francis mailing asking $45 for a simple article not available even with my Brown connection and VPN.

improvement. For instance, NSF now has a policy that all NSF funded research has to be available via public access within one year of publication. This should start the ball rolling. Many Open Access journals have appeared. Mathematicians have begun to embrace the `arXiv` as a universal preprint server. On the negative side, there is an explosion of junk journals and a vast increase in the sheer number of publications that makes keeping up with any field ever more difficult. As a retiree playing with math in the Maine woods, I have come to realize how crucial VPN (virtual private networks) via a university connection is. Without VPN through Brown, it would be hopeless for me to do any remote math. Heaven help the amateur without a university connection whose library pays the hefty subscription fees of Springer and Elsevier. The fight with large corporations is not over, as this screen shot in Figure 4 from my computer today shows, taken while seeking to download an article from the journal *Cognitive Neuroscience and Molecular Genetics*.

Let me summarize my feelings about publishing in a simple assertion:

SCIENTIFIC RESEARCH AND COPYRIGHT ARE INCOMPATIBLE

They are "oil and water". Research demands unlimited sharing of ideas. Mathematical publications, except for a few textbooks for large undergraduate classes, are not done for money. I might add that getting copyright permissions for the figures in this book has been a major aggravation for more than 6 months.

Chapter 17

One World or Many?

i. My Own Experiences

I was raised in a very international multi-cultural setting. My father worked in the UN and had previously started a school in Tanzania whose goal was not to create British civil servants but instead was based on teaching the students basic technology and hygiene that they could bring back to their villages (think toilets, irrigation and fertilizer), [Mum30]. He had a PhD in Anthropology and sought to put these ideas into practice. My mother, though raised in privilege, rebelled against the business men in her family by strongly supporting Roosevelt and Wallace. We entertained an international group of visitors. It has always seemed an axiom to me that the world would gradually become one, each culture sharing its values with others and accepting the others' differences. How naive of me to expect anything so simple! Conflicts were far away from my sheltered neighborhood. The woes of the great depression were nowhere to be seen, the devastation of Hiroshima was a world away. Though nominally Christian, we had a number of Jewish friends (Jacob Epstein, who sculpted my mother, visited for a week) and we never went to church. The exception was that my father did like to don a top hat and flamboyantly appear in the Episcopal church on Easter.

The sciences, including math, are the most international professions. Freedom to travel and work with colleagues from *every* country in the world has been an essential ingredient in the explosion of scientific progress from the end of WWII to the present. And the ease with which – most of the time – foreigners could visit the US and also immigrate if they desired has made working in the US a paradise. The biggest exception was the tragic isolation of Soviet mathematicians. There was a curious exchange of letters when Grothendieck came to Harvard for a semester in 1958. In the McCarthy era, visitors had to sign a statement that they would not work to overthrow the government. Grothendieck said he could not do that and asked whether, if he was put in jail, he could get all the books and visitors he wished. The fact that he and Mireille were not married was another shocker in those

puritan days. But Oscar Zariski worked some magic and somehow he came and what an impact he made.

My graduate students and friends came from everywhere. While I was still a Harvard graduate student, Heisuke Hironaka came from Japan and, after I joined the faculty, Tadao Oda was my second graduate student. Then came students from all over – Birger Iverson from Denmark, Finn Knudsen from Norway, Bernard Saint-Donat from France, Ulf Persson from Sweden, Amnon Neeman from Israel and Australia, Emma Previato from Italy, etc. To verify the international nature of math, you only need look at lists of joint authors of papers and books, e.g. my favorite book on Teichmüller (an inspirational mathematician, though, sickeningly, a Nazi) theory is by Leon Takhtajan from Armenia and Lee-Peng Teo from Malaysia. I checked a recent preprint from Google and, without researching this deeply, find Indian, Russian, Israeli, Hispanic, and one Welsh name as its authors. The fully international world would seem to have arrived.

Unfortunately, today, nationalistic governments appear to have taken over a large part of the world. For people with a broader education than mine, the seeds of this nationalism might have been obvious. Let me describe some of my personal experiences that eventually gave me a greater understanding of nationalism.

In 1963 I visited Japan for two months and saw the still devastated Hiroshima with my own eyes. I was never called a "gaijin" (a pejorative term for a foreigner) in the largely closed society of Japan though I'm sure that is how I was seen. It took quite a while, talking to many Westerners who had spent more time in Japan, to realize the strength of native Japanese traditions and how difficult, maybe even impossible, it is for a foreigner ever to be fully absorbed there.

In 1967 I spent 2 weeks in Israel, mostly on a *Moshav*, obeying the Torah with regard to separate milk/meat meals. I saw the contrast between the brown earth on the Occupied Palestinian areas, AKA the "West Bank" and the green irrigated land in Israel but I did not see the absence of even the tiniest bit of cooperation between the Palestinian and Jewish peoples. In the period 1995-2009, I made multiple visits to Middle East and Turkey. It started with an invitation from my colleague and friend Professor Mina Teicher leading to two exciting weeks of science and touring in Israel with my wife Jenifer. This was a period when, despite persistent low grade violence, the Oslo accords had injected some hope. I remember a sign on the Israel/Jordan border where the words "Shalom" in Hebrew and "Salām" in Arabic were posted, one above the other. How similar, the words they spoke. It didn't last. I returned sometime later with my son Jeremy visiting Lebanon, Occupied Palestine and Israel. My guide to Occupied Palestine was the Palestinian mathematician Iyad Suwan, whose family home in Arab East Jerusalem is within feet of the wall. I described this trip in the Notices of the AMS (**E-2008c**). I made Turkish, Israeli, Palestinian and Lebanese friends and found universities much the same in every country, except for the occasional horror story I heard over lunch (e.g. imprisonment or explosions). Moreover, I was shocked to discover that the occupation is set up with "zones" that make it literally impossible for Israeli mathematicians to have a joint seminar with Palestinian

mathematicians. There was no way to ignore the intractable Palestinian/Israeli anger, each with their own nationalism. Reading the bible, one sees that this conflict literally has a three millennium history.

In 1967/68 I lived side by side with the highly visible poverty of third world Bombay. I saw people living in the streets and cleaning our apartment with rags but not that many of them bore the label of "Dalit" (traditionally called "untouchables"). Little did I know how strong Hindu culture is (though my wife, in love with Hindu myths, was enlightened by André Weil that she could not convert to Hinduism and the best she could hope for was to be born a dalit in her next life). Being inculcated with the open arms American way of life, I failed to fully appreciate in all three cases the passion with which Japan, Israel and India were all driven by their intact – and strongly exclusive – cultures.

A final exposure to Middle Eastern conflicts came with several further visits to Beirut. Michael Atiyah asked me to join the board of his fledgling math research center in the American University of Beirut, but, after my last visit, I replied to him that one needed a PhD in the chaos of Lebanese cultures to navigate any involvement there. I might add that I have had several mathematical invitations to visit Tehran where coincidentally my college roommate M.M. lives[1]. Another roommate explained to me, however, that M.M.'s actions are closely monitored and hosting the visit of an American might not be healthy for him. The last thing I wanted to do was cause any trouble for him so I never went to Iran.

As I see it now, there is a major conflict, not to be papered over, between the tolerant international liberal viewpoint and the passion with which each culture tries to maintain its traditions and pass them on generation after generation. I grew up completely committed to the former and my whole life working freely with colleagues from every part of the world reinforced this. But now I hear and read more and more voices that say "not so fast – our culture, our jobs, our very identities are vanishing." The rapidity with which technology is advancing and the immense growth of international wealth, private and corporate, all support only the "one per cent" and the educated with ties to multiple countries. Moreover, the ever expanding population of refugees relentlessly aggravates the conflict. Nationalistic governments have taken over China, India, Russia, Brazil, etc., etc. Every countries' unique identity is threatened by these forces and every country has plenty of right wing politicians riding the reaction to it.

I don't believe there is any simple right or wrong here. Much of the problem is due to the rapidity of change now. Everyone's lifetime is long enough for them to see whole livelihoods and communities disappear (see Fiona Hill's amazing book [Hil21]). It makes no sense to demonize either side. This was the core issue in the US election in 2016: Clinton represented the liberal "politically correct" internationalist standpoint and promised merely to fine tune the hurricane of change; Trump wildly asserted that he could restore a strong and prosperous America with mid-twentieth century values without giving a hint of how

---
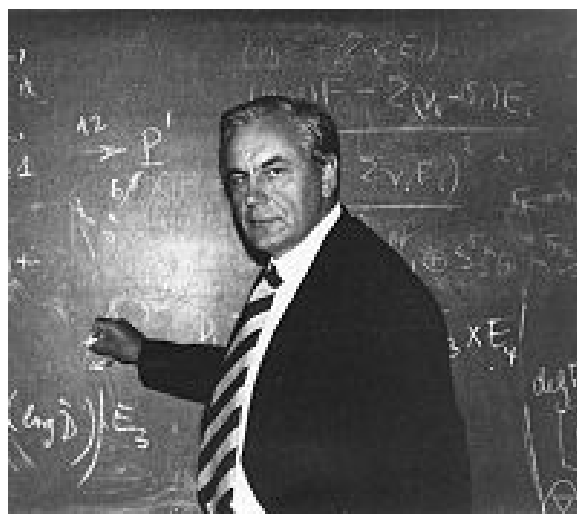
[1]Name omitted for his protection.

Figure 17.1: Igor Shafarevich lecturing. From Wikimedia Commons, credit Konrad Jacobs.

he intended to do this. I want next to make this conflict clearer by describing two vivid examples of nationalism that have involved my interaction with another mathematician.

ii.   Russia and Shafarevich

When I recently packed up my office files at Brown, how I could resist re-reading some of the old letters in my files?[2] In 1992, there was a major controversy at the US National Academy of Science over censuring their Foreign Member Igor Shafarevich for anti-semitic writings and actions. He was an old friend and we exchanged quite frank letters at that time (his letter to me is below). Of course, without a doubt, there was indeed a great deal of overt anti-semitism at that time in the USSR and, in particular, in the Moscow mathematical community. But personally I have not seen evidence that Shafarevich himself was anti-semitic, but rather that he was a fervent believer in his country, its people and its traditions – perhaps one should say its soul.

I met Shafarevich in 1962 at the Stockholm International Congress of Mathematics. I spent an evening getting to know Shafarevich and his young colleague Yuri Manin, enjoying their company and drinking a bit more vodka than was good for me. I met them next in 1979 in Moscow, neither having been allowed to travel to the West in the interim. (I recall Manin having a desk with a glass top under which he had kept all the many invitations he had been forced to decline.) But in the meantime, in spite of being so isolated, Shafarevich had built in Moscow one of the best groups of mathematicians working on the synergistic

---

[2]This section includes my post *Nationalism and the longing to belong, with best regards to Igor Shafarevich*, Sept. 15, 2016, with some small changes.

fusion of algebraic geometry with algebraic number theory. He had a strong personality, was a wonderful teacher and was also quite religious (Eastern Orthodox). In addition he had thought deeply about social science and how history molds the character of a country. Here is a quote from the last section of his essay "Russophobia," (from the English translation in [Sha90]), p.29) that provoked the 1992 controversy:

> *A thousand years of history have forged such national character traits as a belief that the destiny of the individual and the destinies of the people are inseparable in their deepest underlying layers and, at fateful moments of history, are merged; and such traits as a bond with the land—the land in the narrow sense of the word, which grows grain, and the Russian land. These traits have helped it endure terrible trials and to live and work under conditions that have at times been almost inhuman. All hope for our future lies in this ancient tradition. ...*
>
> *....... We most likely are dealing here with a phenomenon to which present-day science's standard methods of "understanding" are completely inapplicable. It is easier to point out why individual people need peoples. Belonging to his people makes a person a participant in History and privy to the mysteries of the past and future. He can feel himself to be more than a particle of the "living matter" that is for some reason turned out by the gigantic factory of Nature. He is capable of feeling (usually subconsciously) the significance and lofty meaningfulness of humanity's earthly existence and his own role in it. Analogous to the "biological environment," the people is a person's "social environment": a marvelous creation supported and created by our actions, but not by our designs. In many respects it surpasses the capacity of our understanding, but it is also often touchingly defenseless in the face of our thoughtless interference. One can look at History as a two-sided process of interaction between the individual and his "social environment"— the people. We have said what the people gives the individual. For his part, the individual creates the forces that bind the people together and ensure its existence: language, folklore, art, and the recognition of its historical destiny.*

I have to admit that I was deeply startled when I first read these lines. I had not heard such strong nationalistic sentiments before. But these words also seemed romantic and an expression of the core of conservative appeals to preserve a country's traditions and cohesiveness, an appeal that we now hear around the world. The bulk of *Russophobia* is an attack on writers who, he believes, have denigrated the Russian "people" and who claim that the Russian peoples' salvation lies in replacing native Russian values with Western liberal and internationally oriented ideas. Naturally enough, many of these writers are Jews hence his being called anti-semitic for writing this essay. This seems quite ironic to me as the whole rationale for the state of Israel has been the restoration of Jewish traditions,

language and religion, in a homeland free of outside coercion. Zionism and Shafarevich's notion of "peoples" seem to me to have a great deal in common.

Shafarevich had long been fascinated with history, his first love before he discovered mathematics. *Russophobia* was one expression of his mature views but another book was about his disgust with communism that he preferred to call "socialism." He was clearly not talking about its benign form in Scandinavian socialism, but rather with political movements that abolished private property, and might even abolish the custom of families and of religion. His book *The Socialist Phenomenon* [Sha80] describes such a state, obviously including the USSR, but also describing an extraordinary diversity of other cases, for example, a) the society the women create in Aristophanes' comedy *The Congresswomen*, b) in some extreme Protestant movements like the Anabaptists, c) in the Inca's empire in Peru and in many other places and writings. He writes:

> *Most socialist doctrines and movements are literally saturated with the mood of death, catastrophe, and destruction*
> and *One could regard the death of mankind as the final result to which the development of socialism leads.*

No wonder he was fired from Moscow University in 1975. He was clearly a man of passion with his own ideas of what humanity needed. His letter to me, reproduced here, responds directly to some of the criticisms that he received and reads as an important historical document for what was going on in Russia at this fateful time.

> *Nov.4, 1992*
> *Dear Mumford,*
> *Thank you for your friendly letter. Of course it is hopeless to explain "where I stand" in 1 or 2 pages but I will try to say what I can. Certainly the slogans of patriotism can lead to bad things, but I don't know what slogans can't. You know probably what were the consequences of the slogans "egalité, fraternité, liberté" during "la terreur" and how the idea of "God's own country" became a warrant for the genocide of North-American Indians. I do not see a danger of such tendencies in the movement of mild national flavor to which I belong. Of course, there is the famous "Pamyat" but it is (a) completely isolated, (b) extremely scanty, (c) without any influence at all in this country and (d) probably created exactly to draw a picture of "russian fascism" (but here I am not certain). I was interested to read about my participation in "political rallies where others have explicitly called for 'cleansing' the government of all Jews, the violent removal of Yeltsin and the re-conquest of the former Soviet Union." I never heard such appeals. Of course Yeltsin is a disaster but the common idea is to remove him by constitutional means which is quite possible and even probable if only he himself will not break the Constitution. Indeed it was exactly he who proclaimed the idea to "disperse the parliament." The idea to "re-conquest" the Soviet Union would*

be stupid if not insane. However, many people, including me, hope the country will re-unite in its principal parts – simply because the people will see what a tragedy its disruption brings. The lies that are written about me would be not very important. But it is really dangerous if your media are feeding you with information of the same quality on more important subjects. In our country this is exactly the case.

But I think one has to say truly that all fuss about me was provoked by what I wrote about Russian-Jewish relations. The subject is painful but it is never good to avoid difficult situations pretending they do not exist. I tried to write with greatest restraint. Some people say that what I have written may be correct but it can give rise to anger and violence. I do not believe this is probable. But what is the logic of my opponents? My paper is composed mainly of quotations. Why do they not address their appeals to to people who write or publish such things that even a quotation from them can provoke violence? But what I have read about myself in American newspapers is beyond any logic. The foreign secretary of the NAS accused me of interfering in the careers of young Jewish mathematicians and preventing them from publishing their papers. Probably such accusations are punishable by court! In reality I have taken many troubles to help my students of Jewish (or partly Jewish) origin – such as Golod or Manin – in their careers. Not, of course, because of their origin – I tried to do the same for all my students. The President of the NAS even makes me responsible for the policy of the Steklov Institute, while Arnold is in the same Institute and Fadeev is even its vice-director, both foreign members of the NAS. Novikov is head of a department there. Are all of them responsible? I also read how I advocated on television the views of "Pamyat" while I did not even mention the name. Formerly I believed that the novel of M. Twain about his attempt to be elected a governor was a parody and a vast exaggeration. Now I think it is a rather accurate description of American life. However I received many letters of support from the the US and this comforts me.

With best wishes,
Shafarevich

So was Shafarevich anti-semitic? Unfortunately Shafarevich's words "*individual people need peoples … (their) 'social environment': a marvelous creation supported and created by our actions.*" can be used to justify many reactions. Although we can empathize to some degree with Shafarevich's love for Mother Russia, it is hard to look at what has happened in Russia since he wrote *Russophobia* and even more since his death in 2017 and think he could possibly have approved of how Russia has evolved. One of the downsides of nationalism is the slippery slope towards dictatorship, the attraction of a strong decisive hand on the tiller. And a dictator is never satisfied with what he rules but insists his

country needs to control all its neighbors, to have more *lebensraum* (living space). Sadly this is exactly what has happened and continues to happen in Russia. Before returning to the dream of a fully international mathematical world, I want to describe another case in which I personally got involved in another country's wave of nationalism.

iii.   India and Castes

The society and culture of India and the US have remarkable parallels[3]. Both societies are a melange of peoples with very different traditions, mostly very religious but following many different rituals. Both India and the US are in the midst of strong nationalistic movements. And both have a large minority that has been and still is being denied opportunities: Blacks in the US (12%), Dalits (AKA untouchables) in India (25%). In addition, in the US, another 19% are Hispanic and suffer serious discrimination; in India, another 20% are Muslim and Christian and are also under great pressure from the BJP (or the Bharatiya Janata Party), the ruling nationalistic party led now by Narendra Modi.

I got first involved with India in 1963 when an unexpected letter with many exotic stamps arrived in my mailbox. It was from C. S. Seshadri (note: South Indians have no family names so Seshadri was his given name, C and S being the village where he was born and his father's name). By the enduring miracle of international math, he and M. S. Narasimhan had created the same moduli space as I had, but with totally distinct tools. Naturally, we decided to get together. He came to Harvard first and I went to Bombay (as it was then called) in 1967. In Bombay, I found the intellectual equivalent of Coleridge's *In Xanadu did Kublai Khan a stately pleasure dome decree.* The Tata Institute of Fundamental Research (TIFR) sits at the tip of the Bombay peninsula with a glorious lawn stretching out to the Arabian Sea. Air conditioned so that I kept a sweater in my office, it was a ferment of mathematical activity. Bombay was indeed a melange with dozens of districts where different communities lived together speaking dozens of languages. For example, the mother of an Indian roommate of mine from Harvard lived in the Gujarati speaking gold seller's district. And there was a district called Bhuleshwar with narrow winding streets, each packed with tiny stalls selling a different item, like a middle eastern souk. I also visited Seshadri's father in Conjeevaram, his town: he came back from the law courts in his wig, quoting Wordsworth and served us lunch on banana leaves.

My close relationship with India continued my whole career. Seshadri became one of my closest friends and I followed him to Chennai when he retired, to the Chennai Mathematical Institute (CMI) that, amazingly, he had founded. I adopted a daughter in India and one of my sons married an Indian. I studied the History of Indian math and some of this appears in Chapter 6. But, since it felt repugnant, embarrassing, one thing I never did was ask my colleagues in India about their caste. Then I met Professor Shiva Shankar at CMI. He is

---

[3]This section is an abbreviated version of my blog post *"All Men are Created Equal?"*, dated June 16, 2015.

a tireless force for the civil rights of the low castes, a true activist.

Here I really need to give some background about caste. The four categories, Brahman, Kshatriya, Vaishya, Sudra are not castes but *Varnas*, each divided into hundreds of actual castes or *Jatis*. Your caste is inherited, immutable and traditionally determined not only what your occupation was and whom you might marry but even with whom you might dine. Below these four Varnas are the outcastes or Dalits that are also subdivided. One of the lowest are the manual cleaners of latrines who are given no protective gear and must climb into pits to clean them. Not only that but you inherit your *karma*, a sort of bank account of your accumulated good and bad deeds in previous lifetimes. The caste system was codified in the last centuries BCE in the "Rules of Manu" (the *Manusmrti*). This is a long treatise that can be found online in English [Man]. These rules include hideous punishments for anyone who violates them. Shafarevich might have included this in his book on Socialism. Nietzsche loved it, Vivekananda struggled to make it seem less oppressive, not very convincingly.

The core of the problem here is that caste is built into Hinduism. The Rules of Manu are *smriti*, sacred writings, one level below *sruti*, revelations of truth. Ambedkar, the amazing brilliant Dalit, who wrote the Indian constitution at the time of independence, recognized that what was needed was not just extending civil rights to Dalits, but an overhaul of Hinduism itself, updating some of its ghastly practices. (*Sati*, forcing a widow to throw herself on her husband's funeral pyre, was another.) For a while, under the Congress Party, such a revolution might have seemed to have a small chance of coming to pass but what has happened instead is that the BJP has come to power and is riding the wave of strict constructionism, the literal interpretation of all ancient writings as well as demoting Muslims and Christians to second class citizens. Even more sinister is the RSS (the Rashtriya Swayamsevak Sangh), the paramilitary arm of the BJP that trains its members with weapons and whose rallying cry is *Hindutva*, the restoration of a mythical purely Hindu world after purifying the country. It was a former RSS member who shot Gandhi.

I got personally involved in 2015 after an official action at the Indian Institute of Technology in Madras (IIT-M) that "derecognized" its student run "Ambedkar Study Group"[4]. I wrote a letter to the Director decrying this move that Shiva, unbeknownst to me, forwarded to two national newspapers. Boy, did I get a lot of pushback, a deluge of email, a real education in Hindutva. What I found is that the right wing believes that Muslims, Missionaries and Communists are three groups of foreign enemies seeking to undermine true Hinduism. It really was not my business, but I had had good times staying there in the IIT-M guest house, giving lectures there and I saw no harm in politely expressing my opinion. America is no paragon here, but it is hard to abandon the memory of the tolerant state that Gandhi and Nehru tried to create. This exchange was the small thing that made what I was reading about present day India real for me. Once again, one

---

[4]I describe all this in greater detail in my blog post *"All Men are Created Equal?"*, dated June 16, 2015

has to acknowledge the power of nationalistic fervor and resign oneself that "one world" is a distant dream.

Let's step back a minute and look at the cases of Russian nationalism with Shafarevich and of Indian nationalism with the BJP in a broader context. Every country has minority and majority subsets of the set of its citizens and these subsets regularly clash and usually the majority dominates the others. In the US, we have the black and the Hispanic minorities; in Israel, the Arab/Palestinian minorities; in India, the Dalits as well as the Muslim and Christian minorities; and in many countries, the Jews are a prominent minority. Legally speaking, you have "standing" in expressing your opinions about your own country but not in any other country. But, as an intellectual, of course you inevitably form opinions of the actions of every government, of its moral principles and even its religion. In this Chapter, I know I have argued for some judgements I made that may offend some friends and colleagues. But I don't want to apologize or "take them back". Here is my bottom line: I, personally, want to feel I am not anti-India or anti-hindu if I criticize the BJP, not anti-American if I object to actions of a President or a Supreme Court, not anti-semitic if I criticize Israeli government policies. As a thinker, I want to be free to criticize any government, even though lacking in "standing". A major problem with the words like "anti-American", "anti-Russian", "anti-semitic/anti-Israel", and so on is the unfortunate confound between being against a government's actions and against its people. I believe you can love and respect many people you meet in a foreign country, yet be dismayed by what its government is doing.

Religion makes for particularly treacherous waters. In the case of Jews and Israel as well as the case of Hindus and India, the majority holds to a different religion from the minority and such differences frequently lead to irreconcilable moral outrage on both sides (as Haidt analyzes in his excellent book [Hai12] on the emotions behind political beliefs). The dispute over abortion in the US is another instance fueled by irreconcilable moral principles. I have purposely omitted from this Chapter a third personal entanglement arising from the very heated political disputes current in the US.

If we want to maintain international collaboration somehow in the small community of research mathematicians, we need to have some *modus operandi*, some ground rules with some tolerance. I think what is necessary is that one really must try hard to understand the views of the people espousing the point of view opposite to your own, to listen to and understand as much as possible those who support an action you oppose, hard though this always is. If you don't try to do this, you won't be able to maintain friendship, let alone collaborate in research, with those on the other side.

Chapter 18

Spinoza: Euclid, Ethics, Time

In our secular age, it is hard to bridge the gap between the long tradition of theistic philosophers and contemporary science-based speculation about the nature and fate of humankind.[1] My friends and family are all over the map – from avowed atheists to weekly church-goers. I have not been a regular churchgoer since graduating from Phillips Exeter where Sunday church attendance was compulsory. The word "God" was already an obstacle for me as the idea of "Him" as a super-powerful old man in white robes watching and judging every action of every human felt so absurd, there seemed no point in looking further. But in the back of my mind, I also knew that all those famous thinkers in the Judeo-Christian tradition were far from stupid. Struggling to find my own path, I stumbled last year upon Spinoza and, to my surprise, found a great deal that I could understand, though not without a struggle. I also found that Einstein, when asked about his religion, often replied that he believed in the God of Spinoza.

I believe this wall between science and religion is on the verge of crumbling. One big reason is that AI programs are beginning to act with remarkable intelligence and it now seems likely that humanity will be dealing with apparently conscious robots within the next 50 years. I discussed this at length in Chapter 10. Another reason is the pressure to acknowledge the messy relationship of human observers to quantum mechanics that I discussed in Chapter 14. In both of these discussions, the issue of time as a purely subjective experience (as necessitated by relativity theory) is unavoidable. But time via the span of human life is also central to all religions. So how did this truly remarkable thinker, Spinoza, deal with all this?

This chapter is about my efforts to understand Spinoza's writings and to understand their relation to other ideas in my philosophical thinking, e.g. Plato, Descartes, Buddhism and Physics. Another excuse for putting this in a scientific memoir is that Spinoza's major work, *Ethics*, [dS77], is written exactly in the style of Euclid's Geometry (and that of much modern math): *it is a numbered sequence of Definitions, Axioms, Propositions with cross*

---

[1]This Chapter is a slightly modified version of my 4/19/2020 blog post "Reading Spinoza".

referenced proofs and, here and there, a helpful Scholium. One wonders whether it might even be rewritten as a fully logical system in the modern tradition. On the other hand, though you might think this makes it easy reading for math people, this is not the case.
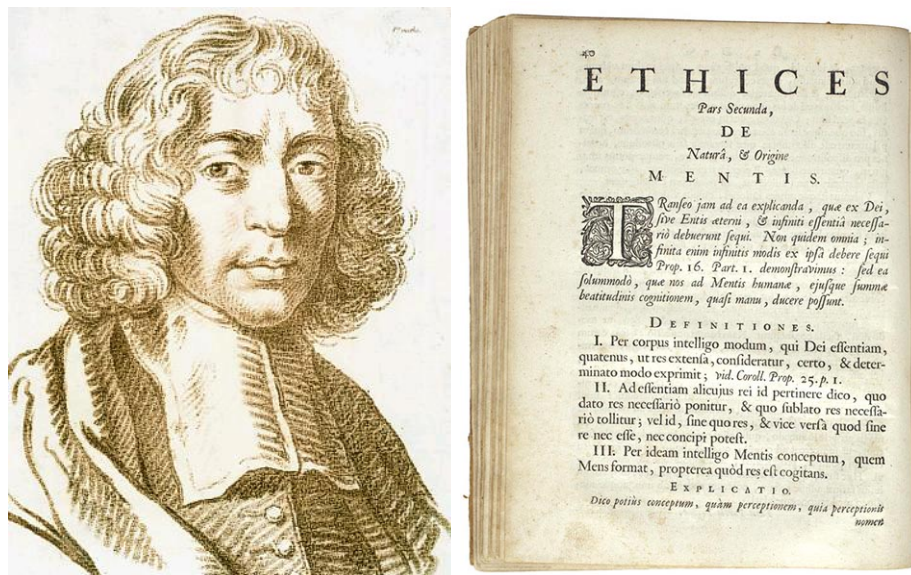
Figure 18.1: Left: portrait of Spinoza; right: opening page of Part II of his *Ethics*, both from Wikimedia Commons.

i. Spinoza and *substances*

Born in 1632 into a Jewish family that had immigrated to Holland to escape forced conversion to Christianity (or death) in Portugal, then excommunicated by the Jewish authorities for his views at age 23 and finally his books put on the Pope's forbidden list, Spinoza was still protected by this liberal state, an island in turbulent 17th century Europe. He died at age 44 in 1677, his lungs poisoned by his profession of lens grinding.

His books are not bed-time reading. Fortunately, there are also many good contemporary commentaries on his book *Ethics*. I would like to acknowledge the huge help I got from Prof. Beth Lord's book *Spinoza's Ethics*. But I had another problem: everyone from Aristotle through Leibniz describes their metaphysics using the key word *substance*. Leibniz's *monads*, for example, are sort of mini-substances. Everything depends on this word but, unlike the convention in math textbooks, no philosopher gave a list of simple examples of substances to help you get the feel. Spinoza helpfully gives a philosopher's definition in *Ethics*, Part I, namely:

> I-Definition 3: *By <u>substance</u> I understand what is in itself and is conceived*

> *through itself, that is whose concept does not require the concept of another thing, from which it may be formed.*

Clearly, he doesn't mean what we call substances, e.g. water, iron, wool. Going back to its origin, the word really seems to stem from a bad Latin translation of the Greek word $ουσια$ used by Aristotle. This word is simply the present participle of the verb "to be", nominalized, i.e. made into a noun, and, interestingly, in the feminine gender. So it means something like "beingness" or "an existing thing" so what you decide to call substances must be the core of your ontological beliefs. Aristotle distinguished primary and secondary substances (individual objects and classes of them) and it is the primary ones that Spinoza is talking about. In fact, after many preliminary logical arguments, Spinoza gets to this Proposition:

> I-Prop. 14: *Except God, no substance can be, or be conceived.*

In the proper scholastic tradition, he gives a *proof* of this! He spins a cat's cradle of extremely abstract concepts, e.g. essence, attribute, mode, that he ties together in a web of Propositions. Thus the above Prop.14 comes from Props.11 and 5, etc. Once you absorb his technical terms, I am willing to believe that his logic is sound. You are welcome to try to unwind his reasoning in the beginning of Book I. As far as I know Spinoza was the first to interpret substance with this laser-like restricted focus on the ultimate source of existence. He is saying that all being is part of God or that God is precisely the totality of being. He also uses the Latin phrase *Deus sive Natura*, God or Nature, to emphasize that he views God and Nature as synonyms, just two ways of thinking of the same thing. For this reason, he has been called a *Pantheist*, a short description that certainly captures part of his beliefs though by no means all as we shall see.

ii.   A short history of dualism and *substances*

But Spinoza knows well that a full description of substances is not so simple. From Plato to the present day, all philosophers have realized that the simple phrase *what is* is not at all simple and most of them have been forced to one or another form of *dualism*, a system of describing reality as having two parts or two aspects (or even three, e.g. in Popper). In order to put Spinoza's views in context, I need to first review some of notable high points in this history. Starting with Plato, his dualism is best understood through his metaphor that all humans are chained in a cave seeing only shadows of the true world, consisting of *ideal forms* outside the cave. For instance, I see my dog Gracie on the floor next to me, but I can only dimly understand the full essence of dogness, that is present in its ideal form outside the cave. Perhaps clearer is the example of the number five (the choice of 5 is arbitrary): I can see many collections of five objects, but the mathematical number five is an ideal form outside in the sunlight. In short, his dualism consists in the sensory world of our perceptions vs. an ideal world of true forms.

Aristotle developed much further the idea of *form* being the essence of everything in the world. In most of his writings, substances were compounds of *matter*, their material substratum, and *form*, compounds he called *hylomorphic*. He regards the "form" parts as the true primary substances but also asserts that his forms are *not* the same as Plato's ideal forms. His key examples are sculptures whose matter is just a hunk of bronze but whose form is its shape that makes it a representation of something. When talking about life, he states that the substance or form of all living things from plants to humans is their *soul*. Since this gives souls to both plants, animals and humans, it suggests that what he calls souls would be better translated as their "life-force". In the absence of any knowledge of biochemistry and DNA, the idea of matter self-organizing into living things was inconceivable at the time, and so endowing all living things with a special type of form called a soul was not an unreasonable idea.

His theory of souls is more or less Psychology 101. They have four parts: the nutritive/reproductive, the sensory, the intellectual/imaginative and the desire/motor parts. Human souls uniquely possess the intellectual part where the forms in the material world are mirrored. What we today call the *mind-body problem* is the question of how his intellect interacts with the material world. Does this raise any problems for Aristotle? In *De Anima*, he simply states *The thinking part of the soul must therefore be capable of receiving the form of an object* (Book III, part 4, my underlining) and *The instrument which desire employs to produce movement is no longer psychical but bodily: hence the examination of it falls within the province of the functions common to body and soul.* (Book III, part 10). Thus the issue of how the mind and body interact, that gave rise to so much discussion from Descartes to the present day and above in Chapter 10.i, is not at all an issue for Aristotle. He just states that they do interact. Although he does introduce God as the prime mover, his universe is strikingly materialistic and in many ways modern and common-sensical.

Saint Thomas Aquinas(1225-1274) attempted to integrate Aristotle's framework with Catholic doctrine, but it's an uneasy integration. On the one hand, he retains Aristotle's idea just described that a human soul is the *form* of a compound thing in which it is joined to its *matter*, namely a human body. But Catholic doctrine insists that human souls do not die and that there will ultimately be a resurrection in which they regain their bodies. So his synthesis required that our conscious souls can both shed their bodies and later get them back, just like doffing and donning a fancy suit of clothes. I confess that, for me, this feels plain weird.

But Christian metaphysics did make one major step, as I see it, through its belief that God was *outside* of time, that there was no special present for Him so that our past, present and future were equal parts of his vision. This idea is clear in Aquinas's writings but goes back to Saint Augustine, to his meditations in Book 11 of his *Confessions*. Here he pleads with God to let him understand the mystery of time and winds up saying that the passage of time, the never ending transformation of anticipated events into past memories, is unique to the experience of each human being. After rejecting as unreliable all objective methods of measuring the passage of time, he concludes that the passage of time is not part of either

the material world nor of God's understanding of his great creation, but is uniquely a part of our subjective experience. I have discussed this above in Chapter 10.iv where these ideas meet the 20th century theory of relativity. I will discuss this further below.

Skipping ahead, Descartes (1596-1650) lived only one generation before Spinoza and now shifts the dualism of substances from Aristotle's form vs. matter to intellect vs. matter, better known as the mind-body problem. Under the influence of incipient science being developed by Galileo (in turn, only one generation older than Descartes) and others, Descartes believed that the material world proceeded by strictly mechanical laws, or, as one says, by "clockwork". He extended this strict determinism from inanimate objects to bodies, human or otherwise, and to all material forms of life. In his theory, non-human animals lacked a mind, hence were automatons without consciousness or souls. I doubt he had a pet. He attempted to build a physics for all this using his concept of *vortices* but this was sadly a false start.

Humans, for Descartes, did have thoughts and souls and these thoughts were the bedrock of his metaphysics: the only indisputably existing thing. He expressed this, of course, in his famous words *Cogito ergo sum*. But what our senses tell us, he reasoned, might well be an illusion. Only by invoking a benign God did he feel one could dispel one's doubts about the genuine existence of the material world. All this works out in a neat way using the idea of substances. There is only one "true" substance, namely God but there are two sorts of substances in our daily lives: minds whose *mode* of existence is thinking and bodies (animate and inanimate objects) whose *mode* of existence is *extension*, that is, being extended in space, being 3-dimensional matter. With the discovery of the law of conservation of momentum, the problem arose of how the mind's decision to make a movement of any kind could alter the course of this mechanical universe. How could the soul have free will if the material world followed immutable mathematical laws. This has turned out to be the perennial problem of Cartesian dualists.

iii.   Spinoza's *Ethics*

At the risk of oversimplifying subtle things, I want to make this long section more readable by starting with a summary. I think Spinoza's thought has three pillars:

FIRST: God is in *everything*,

SECOND: God is *outside* time, his nature makes no distinction between the future and the past, and since he sees it all at once, the world is deterministic and praying for help is an error,

THIRD: For God, *all is good*; evil is a subjective notion caused by our limited perceptions that leave us in bondage to our emotions unless, through reason, we acquiesce to the love of God.

His definitive book *Ethics* is made up of five parts. I will take them up in turn.

Part I: Of <u>God</u>

Part II: Of Nature & the Origin of the <u>Mind</u>

Part III: The Origin and Nature of <u>Emotions</u>

Part IV: Of <u>Human Bondage</u>, or the Strength of the Emotions

Part V: Of the Power of the Intellect, or <u>Human Freedom</u>

I. God

We have already described some of the essential properties of God, as Spinoza conceives of it. God is the one and only genuine substance around. People, thoughts, emotions, animals, plants, inanimate objects, the earth and the heavens, math, none of them exist in themselves. All these "things" exist as part of God. Any sort of *being*, of *existing* must come about as an *attribute* of God. His God, as we shall see, is quite abstract, is not a warm loving spirit who listens to our prayers (more on this below). The word "God" itself is so fraught these days that I think Spinoza's God is better described by a compound "god-nature-beingness", a synonym for *everything that is.* This is essentially the same as the name Moses receives from God in Exodus, *Yahweh* or simply YHWH, the 3rd person singular of the Hebrew verb "to be".

Now Spinoza is fully aware of what Plato, Aristotle, Aquinas and Descartes wrote and how they all split things up and wrestled with ontology. Having lumped all substance into one, Spinoza's genius was to redefine the distinction between mind and body as the presence in God of two *attributes*. In fact, God, he believed, has infinitely many attributes but only two of them are manifest to our meagre human existences: *extension* and *thought*. One should think of these attributes as two of the very many faces of God's essence. The attribute of extension characterizes material objects that exist in space and time. In modern physics, we would certainly add "fields"; though non-material, they occupy space and time, so partake of extension. The attribute of thought characterizes all the contents, all the conceptions of our minds. Thus Descartes mind/body problem is solved by there being two attributes in God's substance.

The last key word in Spinoza's ontology is *mode*. Finite modes are the manifestations of the attributes that we are know directly but still owe their existence to the all-encompassing substance, i.e. God. Your body, the North star, a grain of sand are finite modes of the attribute of extension. Your loves, plans, understanding of the number 5 are finite modes of the attribute of thought. He states:

> I-Prop.16: *From the necessity of the divine nature there must follow infinitely many things in infinitely many modes.*

II. Mind

One might think Spinoza would get directly to human beings now. But instead, he must deduce everything from God's nature, the only substance. This may sound unduly abstract and indirect but it's all part of his precise logic, his answer to Descartes "Cogito ergo sum". Here are his conclusions:

> II-Prop.13, *Corollary: It follows that man consists of a mind and a body, and that the human body exists because we are aware of it.*
> II-Prop.21, *Scholium: The mind and the body are one and the same individual, which is conceived now under the attribute of thought, now under the attribute of extension.*

We are, in other words, each a mode of God's substance, an idea actively conceived by God, part of its infinite intellect. And Spinoza is denying that there is any separation between the mind and body, these are two faces of the same thing. And then he continues, giving all objects some sort of mind and making explicit his pantheistic conceptions in a Scholium:

> *The things we have shown so far are completely general and do not pertain more to man than to other individuals, all of which, though in different degrees, are nevertheless animate. For of each thing, there is necessarily an idea in God, of which God is the cause in the same way that he is of the idea of a human body. And so, whatever we have said of the idea of a human body must also be said of the idea of any thing.*

This surely sounds like something John Muir, another pantheist, might have said.

He continues with his epistemology of the mind. The most distinctive part of this is his concept of *adequate* vs. *inadequate* knowledge. As usual, his definition is rather opaque (Part II, Definition 4): *By adequate idea I understand an idea which, insofar as it is considered in itself, without relation to an object, has all the properties, or intrinsic denominations of a true idea.* Here, true ideas are ideas that are "in God" (II- Prop. 32, Demonstration). Since your mind's essence is part of God's infinite intellect, your finite mind can have some access to truth, hence adequate ideas. But most thoughts get confused with many other ideas and are inadequate as the Scholium to Prop.29 states:

> *I say expressly that the mind has, not an adequate, but only a confused knowledge of itself, of its own body, and of external bodies, so long as it perceives things from the common order of Nature, that is, so long as it is determined externally, from fortuitous encounters with things, to regard this or that, and not so long as it is determined internally, from the fact that it regards a number of things at once, to understand their agreements, differences and oppositions. For so often as it is disposed internally, in this or another way, then it regards things clearly and distinctly.*

In this quote, Spinoza is making the distinction between the mental processes he calls imagination (thoughts swayed by particular perceptions) and reasoning (assessing and integrating your experience). Should I comment that political discourse these days is a clear example of a cacophony of inadequate ideas?

To my understanding, this distinction of adequate vs. inadequate feels very close to Plato's ideal forms vs. perceptions of shadows in the cave, or to Karl Popper's distinction of what he calls "world 3 *knowledge*" vs. "world 2 *mind*". (His "World 1" are the things with extension.) By "mind" Popper refers to mental states, to the content of consciousness, perceptions, ideas and plans. Some mental states can simply be consciousness without thought, as in deep meditation. On the other hand, his "knowledge" consists in ideas whose meaning does not depend on any individual but has universal validity. Math is arguably the best example. Why is it possible for people speaking different languages to communicate? Popper would say it's a reflection of the universality of knowledge, the existence of adequate ideas.

Finally, near the end of this section, Spinoza launches his bomb shell: there is no free will.

> II-Prop.48: *In the mind, there is no absolute, or free, will, but the mind is determined to will this or that by a cause which is also determined by another, and this again by another, and so to infinity.*

As in all his assertions, he proves this, referring back to an earlier Proposition that "*God acts from the laws of his nature alone, and is compelled by no one*". He embodies these laws, so that's how it has to be! Praying for help from heaven is pointless, is misconceived. In our era of neurobiology and with the legacy of Freud's unconscious, it is hard to deny the logic in this. Arguably, quantum mechanics may give us some wriggle room. But, in this connection, I cannot resist quoting a humorous dialog Lars Gårding wrote between God and the then recently deceased mathematician von Neumann [Går05]. Von Neumann badgers him with questions and gets annoyed when God states the Riemann hypothesis is true but he can't give a proof, he just knows it. And next:

> Von Neumann (agitated): *Do you understand why there are so many problematic infinities in quantum mechanics?*
> God: *Understand and understand. When I invented quantum mechanics, I wasn't on my best form, but it hangs together all right.*
> Von Neumann: *Your answer is ridiculous. I find it more and more difficult to believe that you are God.*

III. Emotions

Spinoza's analysis of emotions is quite straightforward. There is one and only one basic desire:

> III-Prop.6: *Each thing, as far as it can by its own power, strives to persevere in its being.*
> III-Prop.11, *Scholium: Joy (is the) passion by which the mind passes to a greater perfection. Sadness (is the) passion. by which it passes to a lesser perfection*

This is not merely seeking survival but seeking to flourish in every sense. Joy results from success, sadness from failure. Love and hatred are simply joy and sadness in the presence of an human external cause. Passionate love is what arises when when our joy is reciprocated. He discusses fear, hope, pride, pity, shame, anger, etc. but also empathy (here and below I have replaced the word "affect" by "emotion", its synonym in Spinoza):

> III-Prop.27: *If we imagine a thing like us, toward which we have no emotion, to be affected with some emotion, we are thereby affected with a like emotion.*

IV.&V. Human Bondage and Freedom

These are arguably the most important sections of the book. Here is how it begins:

> *Man's lack of power to moderate and restrain the emotions I call bondage. For the man who is subject to emotions is under the control, not of himself, but of fortune, in whose power he so greatly is that often, though he sees the better for himself, he is still forced to follow the worse.*

But what he really wants to talk about the problem of good and evil. He has a radical solution to this huge question: From God's perspective, there is no evil; good and evil are always relative to an individual.

> Preface, part IV: *As far as good and evil are concerned, they also indicate nothing positive in things, considered in themselves, nor are they anything other than modes of thinking, or notions that we form because we compare things to one another. For one and the same thing can, at the same time, be good, and bad, and also indifferent.*
> IV-Def.1: *By good, I shall understand what we certainly know to be useful to us.*
> IV-Def.2: *By evil, however, I shall understand what we certainly know prevents us from being masters of some good.*

This sounds as though he is advocating purely selfish behavior. But he distinguishes people who are controlled by some emotion with those who are able to follow the dictates of reason. If they do so, they will recognize that what is good for others, is also the best thing for them. In a scholium, he expresses himself very eloquently:

> *So let the satirists laugh as much as they like at human affairs, let the*
> *theologians curse them, let melancholics praise as much as they can a life that*
> *is uncultivated and wild, let them disdain men and admire the lower animals.*
> *Men still find from experience that by helping one another they can provide*
> *themselves much more easily with the things they require, and that only by*
> *joining forces can they avoid the dangers that threaten on all sides.*

In short, one must use reason to restrain our emotions and then striving to flourish individually will lead us to work together. He goes on to preach the ethics of joy: "*to refresh and restore himself in moderation with pleasant food and drink, with scents, with the beauty of green plants, with decoration, music, sports, the theatre ... without injury to another.*" So what causes clearly evil actions like murder, etc.? It is a confusion or perversion of some action in our nature caused by some emotion, some inadequate thought. There is no force for evil, only inadequate knowledge.

I find myself diverging from Spinoza (and from Buddhism) here: moderation and the dictates of reason are too cool for me. I feel instead that it is part of our nature to revel in strong emotions from time to time and that this activity makes us truly alive.

In the last Part, he gives some council on how to restrain our emotions. Understand them as much as you can so you can stand back. Realize that hating someone is more harmful to you than to the other. Proposition 2 says that for both unhealthy desires as well as hates, reason will allow you to deal with them:

> V-Prop.2: *If we separate emotions from the thought of an external cause,*
> *and join them to other thoughts, then the love or hate towards the external cause*
> *is destroyed, as are the vacillations of the mind arising from these emotions.*

We must gain an adequate understanding of these emotions and then we can control them – today we call this basic psychotherapy.

He goes on to relate our understanding ourselves to our love of God:

> *He who understands himself and his emotions clearly and distinctly (i.e.*
> *adequately) loves God and does so the more, the more he understands himself*
> *and his emotions.*

More or less as an aside, he then adds that *God is without passions and is not affected by any emotion of joy or sadness*, that *No one can hate God* (essentially because if you thought you hated God, you wouldn't really know God) and that whoever '*loves God, cannot strive that God should love him in return.*

He ends the book with some quite deep and fascinating comments on eternity. I think the key to this discussion is that eternity, for him, does not mean an infinite duration from some unbounded past to an unbounded future. He says "*eternity can neither be defined by time nor have any relation to time*", i.e. it must be considered outside time altogether. He then says explicitly that when you talk about some aspect of the human mind being

eternal, you do not mean to attach to the mind any *duration* beyond the bounds of birth to death. But he asserts that some part of the mind is eternal in this "outside time" sense. You can read this yourself in the often misunderstood and mysterious passage in Part V-Prop.22-23, demonstration and Scholium. I find very attractive this idea that there is no simplistic "afterlife" but that, if the soul does exist free of a body, this existence is not in time at all. The "passage of time" and the notion of a "present" are experiences limited to our biological lifetimes.

iv.   Relations to various religions and to modern science

My partner, Alice Gorman, pointed out to me that there were really only two kinds of prayers. One is "help me, help me, help me" and the other is "thank you, thank you, thank you". Likewise, there are two aspects of religions, one where God (or gods) are tracking you and may intervene in your life and the other is where God is unknowable, ineffable and we are on our own. The first type of religion is by far the most universal and "help me" the most common prayer. The most straightforward but simplistic way to way to get a God to help you has always been to give it something dear to you. In the most brutal societies like the Mayan, this would be sacrificing a person's life. Or perhaps an animal would do the trick as in the Vedic tradition, e.g. the fire rituals of the Satapatha Brahmana. In violent Muslim groups but also in some very peaceful Jain sects, it is your own life you offer in sacrifice. In Hinduism, you may bring a simple coconut to a humble shrine and ring a bell loudly calling the God to be present in the idol and you can pray e.g. to Lakshmi for success in business. In China, I am told, even Buddhists pray for monetary success. Another kind of gift, practiced in Evangelical Christian sects, is to publicly vow to give *yourself* to God or Jesus. The variations are infinite.

Spinoza is an example of the iconoclast to both Judaic and Christian religions, who said it explicitly: you should love God but do not treat him like a father who will love you in return. His religion is uncompromisingly of the second sort, saying "thank you" but not asking for anything in return. Of course, many Christian saints epitomize a life that asks for nothing, e.g. St. Francis of Assisi, St. Theresa of Avila. The Sufi sect of Islam follows the same precepts. Buddha's life certainly exemplifies it. It is particularly interesting to compare the Book of Job with Spinoza's ideas. On the one hand, as in Spinoza's writings, Job's story says don't expect God to always reward you for your prayers and offerings and don't expect to be able to understand why all things happen the way they do in God's world. On the other hand, Job's God is very involved with his creation, tinkers with it and speaks directly to Job, things that Spinoza would find ridiculous. And the cruel irony is that Spinoza's own life had parallels with Job's: while never loosing faith, he was ostracized by his fellow Jews and afflicted with a dreadful lung disease that killed him at a young age.

In its original conception, though not always in practice, Buddhism consistently rejected asking some powerful force to help us mortals. "Om mani padme hum" is their "thank you,

thank you" prayer and it is up to you to work towards enlightenment through prolonged immersion in meditation. There are clear parallels between Spinoza and Buddhism with respect to the negative things in life. Spinoza says that the bad events in life are only seen as bad because of our limited understanding, our inadequate thoughts, and that mastering your emotions with reason will gradually let you pass to adequate knowledge in which the bad events loose their hold on you. Buddhism talks of *dukkha*, the suffering caused by poverty, illness and death, and that letting go of your emotions through meditation will let you pass to an enlightened state. They both give us tools to release our bondage to our passions.

This may be oversimplifying truly subtle things but I also see a parallel with regard to time. As we have seen, Spinoza talks of part of our essence being outside of time. Is it unreasonable to think Buddhism's seeking an enlightenment where the spirit breaks the endless cycle of rebirth as something similar? There is another possible link. I have not quoted the difficult Prop. 11 of part II that states:

> *The first thing that constitutes the actual being of a human Mind is nothing but the idea of a singular thing which actually exists.*

As I understand this, he is saying that your mind cannot exist before you have something to think about. I believe this a parallel to the Buddhist notion of the essential emptiness of all things. In this doctrine, we are like mirrors and only by reflecting other things do we exist. The world is like a web whose nodes are empty but it is held together by its links. Spinoza seems to be saying that about each human mind.

Relativity theory has gone a long way to illuminating the issue of time. First of all, it states that physics takes place, not in distinct 3-dimensional space and 1-dimensional time but in the merged 4-dimensional space-time. Thus the attribute of "extension" is clarified as "existing in space-time" and our lives are simply paths in space-time that have a beginning and an end and are "time-like", meaning we can never travel faster than light itself. Your subjective feeling of the passage of time is yours alone. To clarify the significance of this, an example is useful. As soon as we begin to travel to nearby stars, we will find that everyone's clocks record time differently. Specifically, you may return from a inter-stellar jaunt still a young person but find your children in their old age. This is *not* science fiction. It has been confirmed by experiments in analogous situations.

As described at some length in Chapter 14, quantum mechanics shakes things up even more. It starts with the assertion that atomic level events are not deterministic. Even more strangely, in its standard interpretation, quantum theory includes an interaction of human consciousness with this indeterminacy. Simply put, suppose an experiment in a lab records some atomic level event whose outcome is not determinate by quantum rules. Then the act of observing the recording creates a new atomic state overriding the indeterminacy. This is called "collapsing the waveform". If this sounds weird to you, you're in good company. But if, as Spinoza has it, thought and extension are merely two faces of one reality, perhaps it is not so strange after all. It's just one more way in which it is manifest that thoughts

and states of material in space-time are merely attributes of a single reality. Spinoza's metaphysics feels to me like it accommodates the complexities and counter-intuitive results with which modern physics has confronted us

I am deeply grateful to Spinoza for giving me his vision of how beautiful our world is. I find both his metaphysics and most of his ethics to be very persuasive, a very coherent way of making sense of the crazy world. But personally, I still find using the word "God" difficult because of all the associations that it brings with it. And his optimism that reason can overpower our inadequate thoughts sometimes seems hard to share. Being in the midst of a pandemic and a political breakdown, it is tempting to heed instead the beautiful words of *Amazing Grace*: "T'was grace that brought us safe thus far, And grace will lead us home", even though they do suggest a God who actively intervenes in our small lives. Well, who knows?

Chapter 19

Thoughts on the Future

There are endless discussions these days about whether the world is heading to some sort of catastrophe[1]. Everyone has their diagnosis of what caused this or that, what is wrong and how to fix it. Undeniably, the physical world we are living in and the culture by which we live in it are changing very fast, arguably faster than they ever did during the entire history of mankind. For millennia, almost all children lived the same way their parents did, hunting, farming, buying, selling or bartering like their parents. They formed families and perpetuated a seemingly constant way of life. Generation by generation, changes occurred but they were incremental. Wars, famines and epidemics periodically disrupted life but life recovered. I believe that none of this is true now. It is now *all disruption* and life is never going to go back to what it was even to what it was when I was born in 1937, let alone to a stable mythical past of contented people living in peace and plenty.

One obvious reason for all this rapid change is that mankind has been so incredibly successful in bending nature to its needs and wishes. Way back at the beginning of the neolithic, they found new sources of food through grains and domesticated animals. Soon after, metals were harnessed for axes and swords. Skipping ahead, it was a mere century ago that the magic of electricity and electromagnetic waves were harnessed for our use, leading to illuminating the world, then telephones, radio, TV, transistors, lasers, and eventually computers. Shortly after this advance started, antibiotics were discovered and medicine began to really work for the first time, leading to the unravelling of the biochemical secrets of our bodies. Socially speaking, the biggest change was caused by the invention of the birth-control pill which freed sex from propagation. Each advance was exciting but also had huge impacts on our culture. After transforming a hunting culture into a farming one, and then using metal plows and finally gas powered engines, now something like 3% of the US work force can produce all the food needed in the US and more. So now we can spend immense sums instead on entertainment and tourism. This is happening not merely

---

[1]This Chapter expands my two blog posts *Hard to be an Optimist*, dated Jan.1, 2017 and *Letter to my Grandchildren*, dated March 1, 2020

fast but at an accelerating pace. The power this is giving mankind is intoxicating and a large part of our culture cannot stop drinking from this faucet. One simple reason for the acceleration is simply that there are more people in the world making these discoveries and inventions hence they are more frequent. Let's look at that.

i. The Population Explosion

But I'd like to suggest that essentially *all the big problems* we face today can be traced to one basic cause: the explosive increase of the human population – Malthus's famous contention in *An Essay on the Principle of Population* [Mal98]. A staggering fact: World population has increased by a factor of 3.6 in my lifetime.[2] Recycle, buy solar panels – fine, but nothing any of us can do is going to control our vast and still growing numbers and all the problems this unprecedented multitude brings.

First of all, let's check some numbers. I used to like to say at cocktail parties "One out of every two people is alive today," meaning, if you consider all humans who have lived at any time since the origin of homo sapiens, then half of them are alive now. This turned out to be nearly but not literally true. But, using the classic estimates in the book *Atlas of World Population History* by Colin McEvedy and Richard Jones, [MJ78], I came up with a better, more plausible, summary statistic. After some thought, it seemed more practical to estimate *person-years*, not numbers of people. This is the *integral* of the population curve, the area under the curve made by graphing world population against time, and does not depend on conjectural longevity estimates. Moreover, it feels like the best way to measure total human existence. Then what I found is this: from the origin of homo sapiens through 1400 CE, about 650 billion people-years were lived (about half before the year 0, half after); from 1400 to the year 2000 CE, another 650 billion people-years were experienced; and if the mean lifetime of everyone alive today is 85 years (assuming medical advances prolong many lives while people in less advanced economies live fewer years), then the people alive today will experience a total of roughly 650 billion people-years. Ignoring some corrections (counting people alive at the year 2000 who may or may not have died by now), we can say that about 1/3 of all human existence is taking place NOW! This, to me, is mind boggling.

This increase is truly scary and I feel it viscerally, traveling, working, reading, like a phase change to another state. People often offer the following nostrum to sooth one's anxiety: "Population is leveling off due to the *population transition* caused by urbanization and the spread of a middle class life style in which it is no longer rational to have large families." First of all I don't think the evidence for this is compelling. As mentioned above, in my lifetime, the world population has increased in my lifetime from 2.2 billion to 8 billion (the latest figure). See `https://www.census.gov/population/international/data/worldpop/table\_history.php` for historical data and `http://www.worldometers.inf`

---

[2]The US population has increased only 2.6 times while India and Pakistan appear to have grown by a factor of roughly 4.5.

`o/world-population` for current data. These give an increase by a factor of 3.6 *in one lifetime.* Yes, the urbanization of over half the world has decreased birthrates but they are still very high in Africa (and might grow higher if Bill Gates eradicates some ghastly tropical diseases). Birthrates are strongly depressed in some countries like Russia and Japan due, it seems, to social malaise that could readily lift (it only took the gamekeeper to wake up Lady Chatterly's feelings and bump the English birthrate). And the Chinese birthrate, that was suppressed by draconian government measures, can now increase, esp. if they settle people in Tibet and Xinjiang, sparsely populated by non-Han Chinese peoples. The UN, with shaky extrapolations, allows for a range of 9.5 to 13 billion in 2100. Check out `https://esa.un.org/unpd/wpp/Publications/Files/Key\_Findings\_WPP\_2` `015.pdf` with many tables. An imminent population plateau is certainly possible, but given the fickleness of human society, I wouldn't bet on it. Urbanization has been driven by desperate landless people seeking some employment somewhere and has resulted in unplanned and ungovernable megacities, riddled with crime. To give some perspective, there were no cities with population over one million until the early 19th century (see `https://web.archive.org/web/20070929110844/www.etext.org/Politics/World.Sy` `stems/datasets/citypop/civilizations/citypops\_2000BC-1988AD`). The first were London and Beijing and when I was born, the largest city was New York with around 8 million. But as of 2015 there were 36 megacities with over 10 million inhabitants, some in every continent including Jakarta, Karachi, Mumbai, Manila, Mexico City and Lagos with over 20 million each. In fact, the UN estimates (see `https://esa.un.org/unpd/wu` `p/Publications/Files/WUP2014-Report.pdf`) that rural population has plateaued as workers seek employment in cities. But, as many movies have documented, the slums in these cities are not happy places and often are effectively controlled by criminals as social norms disintegrate.

ii.   The Consequences of this Explosion

I think the challenge of living with 8 billion fellow humans is best displayed in Figure 1. The diagram there feels to me almost like a mathematical theorem: each arrow a virtually inevitable consequence. Of course, one could pursue the chain of causative events further, asking why such a population explosion occurred now for the first time in human history. I would assume that a) the huge success of medicine, esp. with antibiotics, b) breeding grains that are twice as productive, and c) the whole industrial revolution are jointly responsible. Human dominance goes back to stone tools, harnessing fire, skinning animals for clothes, basically the fact that we have a bigger frontal cortex with which we plan, plan and plan some more. The discovery of electricity and microbes are just more recent events that have further enhanced our control of the world – *though not our wisdom.*

My biggest fear is not that the size of the present population couldn't be stabilized at some slightly higher level, but that managing a world of anything like this size requires
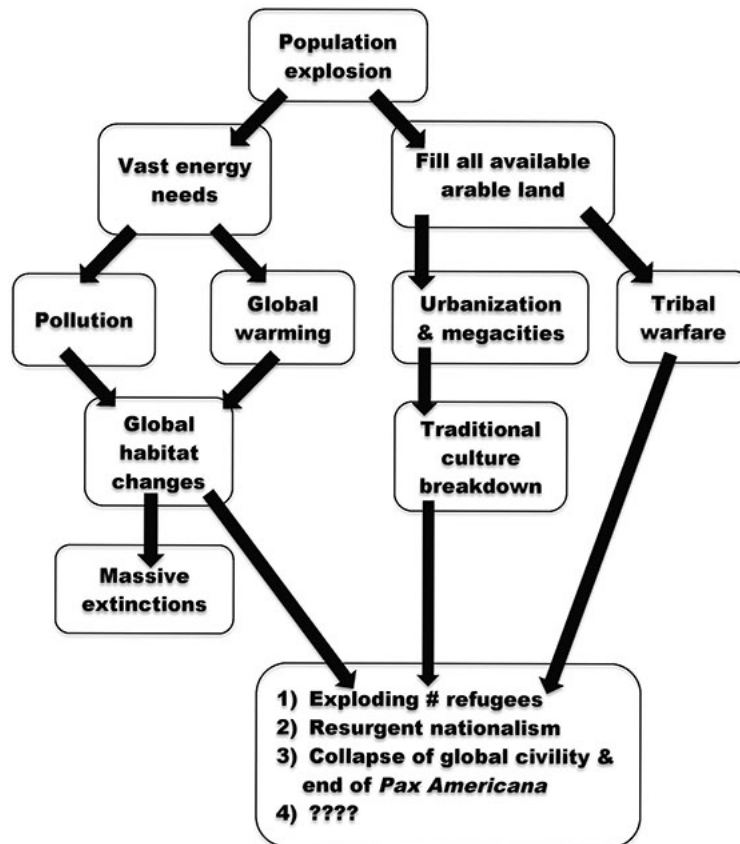
Figure 19.1: A chain of events and their consequences. Each arrow is a very convincing cause and effect.

reasonably rational and cooperative governments to deal with the huge number of problems it creates (e.g. managing megacities with vast slums, the need for new jobs in the face of automation, rising expectations for meat and consumer goods). And I don't see many countries with such reasonably rational governments, nor do I see any indication of actual financial cooperation between the nations. Perhaps the biggest problem a non-growing society will have to face is that capitalism is based on exploiting new opportunities and, without growth, where will these opportunities come from? Will there be work for everybody now that we are so efficient. Psychologically, adjusting to a non-growth society is a huge challenge.

Global warming is the box that everyone is focussed on right now. Things are not looking great for any international agreement. For example, the annual plans formulated at conferences to control global warming are ignored, never funded by the separate nations after the big meeting ends. Actually, I think the massive air pollution in India and China will prove to be more effective in forcing a phasing out of coal in these most populous countries than weak international agreements with no enforceability. But so much of the climate changes are irreversible, e.g. the melting of arctic ice starts a vicious cycle because open water absorbs more sunlight, hence accelerates the melting. If Greenland melts, sea level rise will be catastrophic. Even though the mathematical models are crude approximations and are based on inadequate data, they do all suggest that once a change starts, it has an inertia and is not easily reversed. I see no way to doubt that the changes in glaciers, coral reefs, the ranges of sea life all point to the same world-wide climate change, a change that is going to intensify. An object lesson is the state of the earth at the end of the Permian period (252 million years ago) when the climate became absolutely hellish and more than 90% of all species went extinct. Everyone should read Elizabeth Kolbert's meticulously documented book "The Sixth Extinction," [Kol14]. I personally saw a fully healthy coral reef for the first and last time in 1963 and it was unforgettable. But I must add, another figure that dismays me equally is the estimate that the *biomass* of domesticated animals exceeds that of all living wild animals by a factor of 15.

What's happening now in other boxes of the figure? The street gangs in many cities are out of control. The Tribal warfare in many areas, especially the Middle East, show no signs of abating (haven't Jewish people been fighting the other tribes in Palestine for three millennia – since the book of Exodus?). As for refugees, their number is exploding and no-one has any answer for what to do. Both Europe and the US have erected walls and are allowing in only a trickle of refugees in. The UN Refugee Agency (UNHCR) estimates that there are now about 90 million forcibly displaced people fleeing crime, war, drought etc. If indeed Bangladesh (population 150 million) becomes the victim of massive floods as many expect, where on earth would their refugees go? Climate change and sea level rise will give rise to literally *billions* of climate refugees. Most of the world is "full up". Arguably, the US is one the few places which could in principle still absorb hundreds of millions of

refugees and climate change might make Siberia a reasonable alternative[3]. Since they will have nowhere to go, I fear that wars are inevitable.

As a mathematician, all this reminds me of the Lotka-Volterra equation. For those who aren't mathematicians, this is a famous model of competing species taught in all introductory differential equation classes. It deals with foxes and rabbits and produces cyclical behavior in which the number of foxes explodes until they reduce the rabbit population to nearly zero, then the foxes starve until the rabbits reproduce and their population in turn explodes, etc. etc. In our case, humans are the foxes and all the rest of the earth – animal, vegetable and mineral – are the rabbits. We have gone through half the cycle: the ascendancy of the foxes/humans but not the second half, their collapse. Let us all pray the model fails to predict the future. Jared Diamond has outlined many of the ways previous cultures have blundered into terminal decline in his book "Collapse" [Dia05].

An extraordinary experiment on the effects of over-population was carried out by the ethologist John Calhoun in 1968-1972. Calhoun spent his whole career studying rats and mice confined to artificial environments that he built and documenting the effects of over crowding when the normal constraints of limited food supply, disease and predation were absent. His final experiment, "Universe 25" is documented in [Cal73]. 4 pairs of mice of reproductive age were introduced into a 8 foot square habitat with unlimited amounts of food and water and nesting boxes adequate for a population of nearly 4000 mice. The population grew exponentially for about a year, reaching 620. But then pathologies set in. In the wild, mice could leave their nest and strike out on their own but not here. Calhoun describes what happened instead (p.84, op. cit.):

> *In the experimental universe there was no opportunity for emigration. As the unusually large number of young gained adulthood they had to remain, and they did contest for roles in the filled system. Males who failed withdrew physically and psychologically; they became very inactive and aggregated in large pools near the centre of the floor of the universe.*

Fighting broke out both between the successful and withdrawn males and between the withdrawn males. The nursing females became aggressive, maternal behavior was disrupted resulting in increased fetal and pre-weening mortality. The population peaked about 8 months later at 2200 as things deteriorated further, leading to the final extinction of the colony in 4 1/2 years. OK, mice are not human but we share a great percentage of our emotional and motivational make up (see Chapter 8, §2 and 3). The details of Calhoun's paper are well worth reading and evoke irresistible comparisons with aspects the world today.

---

[3]As of 2020, the US has only about 36 people per square kilometer, whereas the Euro area has 128, China 150, India 470 and Russia merely 9. Check `https://data.worldbank.org/indicator/EN.POP.DNST`.

iii. A Safety Valve?

But there's actually one silver lining to our predicament: SPACE! There is a new frontier. Bit by bit, the prospect of actually settling in outer space, on the moon, on Mars, on a man-made planetoid as in the 2013 movie "Elysium," each of these is beginning to be taken seriously. The human urge to expand, to conquer new territory is not to be denied. We are captives to the mantra of perpetual growth, and a growing population that can spread beyond the earth, reshaping it for our purposes. This growth fuels our need for more power, both power in the physical sense but even more, personal power to control and dominate whatever comes within our grasp. Space may well become a new "wild west" with all the trappings Hollywood has depicted.

As the earth becomes jammed to its gills with its 8 billion people, expected to rise to 11 billion or more before leveling off, the idea of colonizing space is sure to become attractive. The other day I was watching a video of astronauts zipping around the space station in zero gravity, hanging out, playing games and laughing and it all seemed so natural. All that's lacking is a way to make money in space and then, like the investors who financed the pilgrims, people will pour money into space/asteroid/planetary settlements. Once again, perpetual growth will seem possible.

iv. Love those Robots

Besides rockets, there are two more technologies in particular that are on the cusp of changing our lives in a truly profound way, maybe even for the better (if you are an incurable optimist). The first is the construction of fully intelligent robots and the second is the mastery of biochemistry potentially allowing us to modify our babies in desirable ways. I know this sounds like science fiction but I think you are wearing blinders if you don't take a hard look at what is going on now. Let me start with robots.

Arguably, the most impressive specific skill that AI (artificial intelligence) has achieved to date is the mastery of language: the ability to accurately translate nearly anything from any one language to any other language, and the ability to compose coherent compositions on virtually any topic. The reason the first of these is remarkable is that, because languages differ so much in how information is presented, to translate a sentence accurately, you must "understand" it. And to understand it, you must know many pieces of basic information about the world. Every sentence has a context, including a situation in the world with a speaker and a listener. The recent algorithms learn how to translate from digesting massive amounts of bilingual text, just as a baby listens and sees the world continuously for a couple of years and then begins to speak and understand. The AI program learns from massive textual data, finding good values for over a billion numbers that express its knowledge. These numbers have no meaning to us, only the computer knows how to use them. That algorithms can learn in some way the meaning of a sentence is a huge step. In the early days of AI, people used to wring their hands over the seemingly impossible

difficulties of codifying "common sense" knowledge. These algorithms have now done a good bit of this as is evidenced in the second feat, writing clear coherent compositions. What has not been done yet is to *combine* (i) programs knowing a lot about language with (ii) ears and eyes plus their accompanying interpretation programs of speech and images, and (iii) arms ands legs to move and hold things plus their programs for doing this, i.e. put all the parts together, let the machine loose and see what happens. If you can pull this off, you'll likely have a truly smart robot. Maybe it will turn out that something big is still missing in our algorithms. But, as far as I have seen, studying these questions, it all seems to be within reach.

Try to imagine a world in which robots can replace human workers in almost every job. A first reaction is "great": if wealth is even partially spread around, no one will have to work very much so they will be free for travel and playing sports and need never worry about food and shelter. What a paradise! But look more closely: a stable population means no growth is possible and, aside from a cadre of engineers and doctors, there are no jobs. Unfortunately, it is built into our adult psyches to want to *do something*, to be a *productive part of a society.* If a large proportion of society is unneeded, they will loose their self-respect and then what? Will we feel inferior if the robot can do so many things much better than we can and has taken away so many jobs? Such a society has never existed but it could come to pass.

And, going further, suppose the robots have the ability to talk to you, one on one, seemingly exactly the same way another human would. Sounds like great fun and certainly good for lonely people. But you'll also begin to ask questions: you'd want to know what *motivates* the robot, can you trust it, does it have emotions (or understand yours)? The robot must certainly have been given drives by its programmer but maybe, with all its knowledge and eventually experience, it will express these drives in unexpected ways. This is called the "alignment problem", aligning the robots goals with those of its handlers. Goethe's famous story, told in his poem *Die Zauberlehrling* (The Sorcerer's Apprentice), describes one way it might go wrong. And maybe some cultures, especially some religions, will declare making such robots illegal or immoral. Already, AI assisted spammers are mimicking human bloggers with uncanny accuracy to subvert social media. As in every other internet advance, people are slow to recognize how criminals are going to employ them.

Or maybe we will discover a way to partner organic humans with robots that makes a super-being, neither a child of nature nor an algorithm. How this will work out is a gigantic unknown, but it is likely to begin happening in my grandchildren's lifetimes. An exciting and scary adventure indeed.

v. Playing God with the Genome

Turning to the second block buster, recall that the biochemistry of the body consists in its DNA, its proteins and a few other molecules like fats that fill the body. Many people assumed that when DNA and its code for producing proteins were discovered, we had solved all the basic problems of biochemistry. Actually, that was just the beginning. We need to discover why DNA produces what proteins when, what are all the things each of these proteins do in all different types of cells in the body, and through what sorts of complex chemical reactions. And how the cells coordinate with each other in the exquisite dance we call *gestation*. For example, no one has a clue where on the genome is the information that says we should have *five* fingers, not four or six. But there is good reason to expect that all this is going to be worked out within a few generations because literally hundreds of well funded research labs around the world are working full time on it.

It's interesting that Freeman Dyson, who viewed the Paris accords on climate change as a step in the wrong direction, has written that the most plausible solution to the CO_2 problem is to create mutated trees that gobble up CO_2, making some chemically stable compounds that can be buried or used in other ways, [Dys08]. He writes there "I consider it likely that we shall have 'genetically engineered carbon-eating trees' within twenty years, and almost certainly within fifty years" and that, widely planted, they could cut the CO_2 in the atmosphere by half in 50 years. Controlling climate is going to be expensive, whether it is done by huge economic shifts or by massive projects to remove carbon from the atmosphere. I have believed for some time that this will happen when a sufficiently huge climate related catastrophe occurs; but I once made a prediction – the trigger will be when sea level rise plus hurricanes destroy a major part of the super-wealthy's mansions that line the coast of Florida. But, having watched how virtually all devastated houses on ocean beaches are rebuilt, maybe a little stronger but still sitting ducks if any substantial part of Greenland melts, I wonder if there is any limit to the irrational optimism of human nature.

Exciting news: a recent advance has transformed the theoretical study of genes into a branch of medical *engineering*, namely the invention of the tool called CRISPR/Cas9, or simply Crispr. Every living cell more advanced than a bacteria manufactures numerous enzymes with which it manipulates and sometimes corrects gene sequences. Now scientists have created another such molecule: this one crawls along the genome looking for a precise sequence of the four bases G, A, T and C. When it finds it, it replaces the next base with another that you can choose, in other words it *edits* your genome. Applications of this tool replacing bases causing disease are an obvious application and are now a hot area of work. Catherine Feuillet of the French National Institute for Agricultural Research reacted on learning of Crispr by saying "Oh my God, we have a tool. We can put breeding on steroids" (quoted in the New York Times, 6/26/22).

Who said one should stop at curing diseases when there are so many genetically controlled things about our lives and bodies that we wish were better? For starters, living longer would be nice. In Babylonian times, the mythical Gilgamesh, haunted by his fear of

death, went on the first quest for the secret of eternal life. Given that tortoises live several hundred years, there seems no obvious reason why humans can't. Genetic modifications for this are sure to become a hot potato for Crispr or its successors. Stronger, smarter, more beautiful children, why not? Of course, it's likely to be expensive but, once offered, people will do anything so that their children won't miss the boat.

Let's be honest: this has a name, it's called *eugenics*. Since Hitler, eugenics has been considered totally unacceptable, a form of racism and utterly beyond the pale. I know what I am saying here is highly disagreeable to many people but I'm sorry, I'm just trying to be logical. Interestingly enough, for instance, Plato advocated eugenics strongly (see *The Republic*, book V) as did many Victorian scientists like Galton. And it is not just longer healthier life that we desire, we want, for example, to avoid anti-social criminal behavior. I think it is indisputable that dogs have been successfully bred to express a whole smorgasbord of adult characteristics including being loving, obedient, aggressive, etc. Different breeds have highly inheritable unique characteristics. This convinces me that many personality characteristics of the human adult are also strongly influenced by certain genes. A study of the gene variations between different dog breeds could well lead to identifying particular genes that affect personality, e.g. affecting, as with dogs, loving, obedient, and aggressive tendencies. If this is done and Crispr is harnessed, parents could also use this knowledge to improve the chances that their children have all sorts of desirable personalities. And certainly, once a small group of people perfects itself in this or any other way, it will prefer to inbreed. Considering the likelihood of space exploration, one line night be bred to live in low gravity environments in space and will relinquish the possibility of returning to earth. Aha, so much for mere skin color differences, now *homo sapiens* can really divide into multiple species. Phew, now we ask, is this going to be a utopia or a dystopia? This is surely the opening of *Pandora's Box* and, just as surely, its temptations are likely to overcome our scruples. This was said best by Oppenheimer: "When you see something that is technically sweet, you go ahead and do it and you argue about what to do about it only after you have had your technical success. That is the way it was with the atomic bomb." Pandora's box has already been opened a crack. Yet another truly frightening challenge for my grandchildren and great-grandchildren!

vi.   Unknowns

There are some huge possibilities that we can imagine but are hard put to guess if they will transpire.[4] There are engineering things for which we know the science but, like airplanes in the 19th century, do not see clearly the technology. Taming fusion is one: find a way to contain what is, in effect, a miniature sun. If this ever succeeds, we will have huge amounts of energy at our disposal and what we do with this is impossible to predict. Another is quantum computing: maintaining an atomic size superposition of states free

---

[4]We exclude here Donald Rumsfeld's famous category, "the unknown unknowns"!

from decoherence while also controlling it. This would put immense computing power in our hands. A third is constructing space elevators. If this can be done, space will become exploitable on an immense scale. All three of these are technically feasible

Arguably, the biggest unknown of all is the existence and potential contact with *extraterrestrials*. Personally, I believe there are such "beings" and, in fact, that the earth was seeded by extraterrestrial micro-organisms of some sort about 4 billion years ago, a theory known as "panspermia." I think our difficulty in discussing the possibility of life of some sort, not on earth, is connected to our difficulty in discussing where we think mankind is heading in the third millennium – and beyond. We are in the middle of such huge change that we can't even formulate what we hope for in the year 3000 CE. So why would we have the vaguest idea of what a culture with a billion years of history would be doing? Even the creation of a coherent timeline for a galactic culture is impossible if our galaxy is explored at speeds approaching that of light, due to general relativity, see my paper [**E-2021**] . The two movies "Contact" and "2001," with a bit of wisdom, threw up their hands over what their human explorers found.

Well, the old curse "May you live in interesting times" certainly applies to everyone living today. The future will be both fascinating and terrifying and I think that standing back to imagine the really big picture – population, space, robots, genes – of where humanity is going is worthwhile. This picture, for me, makes the future seem very daunting even though I won't be there to see it. But human nature, besides its struggle for power, control and individual preservation, has another side containing love, empathy, cooperation. Everyone who stops to think about it surely realizes this side must somehow become humanities guiding star in navigating the huge challenges I have just sketched.

Author's Bibliography

I. Books

Bk-1964 *Lectures on Curves on Surfaces*, (with George Bergman), Princeton University Press, 1964.

Bk-1965 *Geometric Invariant Theory*, Springer-Verlag, 1965; 2nd enlarged edition, (with John Fogarty), 1982; 3rd enlarged edition, (with Frances Kirwan and John Fogarty), 1994.

Bk-1967 *The Red Book of Varieties and Schemes*, mimeographed notes from the Harvard Mathematics Department 1967, reprinted in Lecture Notes in Mathematics 1348, Springer-Verlag 1988 and enlarged in 1999 with contributions from Enrico Arbarello.

Bk-1970 *Abelian Varieties*, (with C. P. Ramanujam), Oxford University Press, 1st edition 1970, 2nd edition 1974, reprinted and enlarged, with contributions from Yuri Manin, by the Tata Institute of Fundamental Research and the AMS.

Bk-1973 *Toroidal Embeddings I* (with George Kempf, Finn Knudsen and Bernard Saint-Donat), Lecture Notes in Mathematics 339, Springer-Verlag 1973.

Bk-1975a *Curves and their Jacobians*, University of Michigan Press, 1975, reprinted as part of the 'Red Book', 1999.

bk-1975b *Smooth Compactifications of Locally Symmetric Varieties* (with Avner Ash, Michael Rapoport, Yung-Shen Tai), Lie Groups: History Frontiers and Applications, Vol. 4, Math. Sci. Press 1975, reprinted by Cambridge University Press, 2010.

Bk-1976 *Algebraic Geometry I: Complex Projective Varieties*, Springer-Verlag, New York, 1976.

Bk-1982/91 *Tata Lectures on Theta* (with C. Musili, Madhav Nori, Peter Norman, Emma Previato and Michael Stillman), Birkhauser-Boston, Part I, 1982, Part II, 1983, Part III, 1991.

Bk-1993 *Filtering, Segmentation and Depth*, (with Mark Nitzberg and Takahiro Shiota), Springer Lecture Notes in Computer Science **662**, 1993.

Bk-1999 *Two and Three dimensional Patterns of the Face*, (with Peter Giblin, Gaile Gordon, Peter Hallinan and Alan Yuille), AKPeters, 1999.

Bk-2002 *Indra's Pearls* (with Caroline Series and David Wright), Cambridge University Press, 2002.

Bk-2004 *Selected Papers on the Classification of Varieties and Moduli Spaces*, Springer-Verlag, 2004.

Bk-2010a *Selected Papers Volume II: On Algebraic Geometry, including Correspondence with Grothendieck* (edited by Amnon Neeman, Ching-Li Chai and Takahiro Shiota) Springer-Verlag, 2010

Bk-2010b *Pattern Theory, the Stochastic Analysis of Real World Signals*, (with Agnes Desolneux), AKPeters/CRC Press/Taylor&Francis, 2010.

Bk-2015 *Algebraic Geometry II* (with Tadao Oda), Hindustan Book Agency, 2015.

II. Papers in Algebraic Geometry: 1959-1982

A-1961a Topology of Normal Singularities and a Criterion for Simplicity, *Publ. de l'Institut des Hautes Etudes Scientifiques*, 1961, pp. 5-22.

A-1961b Pathologies of Modular Geometry, *Amer. J. of Math.*, 1961, pp. 339-342.

A-1961c An Elementary theorem in Geometric Invariant Theory, *Bull. Amer. Math. Soc.*, 1961, pp. 483-487.

A-1962a Further Pathologies in Algebraic Geometry, *Amer. J. of Math.*, 1962, pp. 642-648.

A-1962b The Canonical Ring of an Algebraic Surface, an appendix to a paper by Oscar Zariski, *Ann. of Math.*, **76**, 1962, pp. 612-615.

A-1963a Topics in the Theory of Moduli, (in Japanese), *Sugaku*, 1963.

A-1963b Projective Invariants of Projective Structures, *International Congress of Mathematicians, Stockholm 1962*, 1963, pp. 526-530.

A-1964 Two Fundamental Theorem on Deformations of Polarized Varieties (with T. Matsusaka), *Amer. J. of Math.*, 1964, pp. 668-684.

A-1965a A Remark on Mordell's Conjecture, *Amer. J. of Math.*, 1965, pp. 1007-1016.

A-1965b Picard Groups of Moduli Problems, in *Arithmetic Algebraic Geometry*, (Proc. of a Conference in Purdue, 1963), Harper and Row, 1965.

A-1966a On the Equations Defining Abelian Varieties I, II, III, *Inventiones Mathematicae*, 1966, **1**, pp.287-384; 1967, **3**, pp. 75-135 and pp. 215-244.

A-1966b Families of Abelian Varieties, in *Proc. of Symposium in Pure Math.*, **9**, Amer. Math. Soc., 1966.

A-1967a Pathologies III, *Amer. J. of Math.*, 1967, **89**, pp. 94-104.

A-1967b Abelian Quotients of the Teichmuller Modular Group, *Journal d'Analyse*, 1967, **28**, pp. 227-244.

A-1968a Deformations and Liftings of Finite Commutative Group Schemes (with F. Oort), *Inventiones Mathematicae*, 1968, **5**, pp. 317-334.

A-1968b Periods of Moduli Spaces of Bundles on Curves (with P. Newstead), *Amer. J. of Math.*, 1968, **90**, pp. 1200-1208.

A-1969a Enriques' Classification of Surfaces in Char. p, I, in *Global Analysis (papers in honor of K. Kodaira)*, Spencer and Iyanaga editors, U. of Tokyo Press, 1969, pp. 325-339.

A-1969b Bi-extensions of Formal Groups, in *Algebraic Geometry*, Oxford University Press, 1969, pp. 307-322.

A-1969c A Note on Shimura's paper "Discontinuous Groups and Abelian Varieties", *Math. Annalen*, 1969, **181**, pp. 345-351.

A-1969d Rational Equivalences of 0-cycles on Surfaces, *J. of Math. of Kyoto Univ.*, 1969, **9**, pp. 195-204.

A-1969e The Irreducibility of the Space of Curves of Given Genus (with P. Deligne), *Publ. Math. de l'I.H.E.S.*, 1969, **36**, pp. 75-109.

A-1970 Varieties Defined by Quadratic Equations, in *Questions on Algebraic Varieties*, C.I.M.E., 1969, publ. by Editioni Cremonese, 1970.

A-1971a 6 Appendices to *Algebraic Surfaces*, by O. Zariski, 2nd edition, Springer-Verlag, 1971.

A-1971b Theta Characteristics of an Algebraic Curve, *Annales de l'Ecole Norm. Sup.*, 1971, pp. 181-192.

A-1971c A Remark on Mahler's Compactness Theorem, *Proc. Amer. Math. Soc.*, 1971., **28**, pp. 289-194.

A-1971d The Structure of the Moduli Spaces of Curves and Abelian Varieties, *Actes du Congress Int. du Math.*, Nice, 1970; Gauthier-Villars, 1971.

A-1972a An Analytic Construction of Degenerating Curves Over Complete Local Rings, *Composito Math.*, 1972, **24**, pp. 129-174

A-1972b An Analytic Construction of Degenerating Abelian Varieties over Complete Rings, *Composito Math.*, 1972, **24**, pp. 239-272.

A-1972c Some Elementary Examples of Unirational Varieties which are not Rational (with M. Artin), *J. London Math. Soc.*, 1972, **25**, pp. 75-95.

A-1972c Introduction to the Theory of Moduli (with K. Suominen), in *Algebraic Geometry, Oslo 1970*, F. Oort editor, Wolters-Noordhoff, 1972, pp. 171-222.

A-1973a Introduction to *Oscar Zariski: Collected Works*, vol. I, MIT Press, 1972 and vol. II, MIT Press, 1973.

A-1973b A Rank 2 Vector Bundle on $\mathbf{P}^4$ with 15,000 Symmetries (with G. Horrocks), *Topology*, 1973, **12**, pp. 63-81.

A-1973c An Example of a Unirational 3-fold which is not Rational, *Accad. Naz. dei Lincei*, 1973.

A-1973d A Remark on the Paper of M. Schlessinger, in *Complex Analysis, 1972*, Rice University Studies, **59**, 1973, pp. 113-117.

A-1974 Prym Varieties I, in *Contributions to Analysis*, Academic Press, 1974, pp. 325-350.

A-1975a A New Approach to Compactifying Locally Symmetric Varieties, in *Discrete Subgroups of Lie Groups*, (Proc. of International Colloquium Bombay, 1973), Oxford University Press, 1975, pp. 211-224.

A-1975b Matsusaka's Big Theorem (with D. Lieberman), in *Algebraic Geometry, Arcata 1974*, AMS Proc. of Symposia in Pure Math., **29**, 1975, pp. 513-530.

A-1975c The Self-Intersection Formula and the "Formule-Clef" (with A.T. Lascu and D.B. Scott), *Math. Proc. Camb. Phil. Soc.*, 1975, **78**, pp. 117-123.

A-1975d Pathologies IV, Amer. J. of Math., 1975, **97**, pp. 847-849.

A-1976a Hilbert's 14th Problem - The Finite Generation of Subgroups such as Rings of Invariants, in *Proc. of a Conference on Hilbert's Problems*, (Dekalb, 1974), Amer. Math. Society, 1976.

A-1976b The projectivity of the moduli space of stable curves, I: Preliminaries on "det" and "Div", *Math. Scand.*, **1**, 1976, pp.19-55.

A-1976c Enriques Classification of Surfaces in Char. p, III (with E. Bombieri), *Invent. Math.*, 1976, **35**, pp. 197-232.

A-1977a Enriques Classification of Surfaces in Char. p, II (with E. Bombieri, in *Complex Analysis and Algebraic Geometry*, Baily and Shioda editors, Cambridge Univ. Press, 1977, pp. 23-42.

A-1977a Stability of Projective Varieties, Monographie No. 24 de *L'Enseignement Math.*, **23**, 1977, pp. 39-110.

A-1977b Hirzebruch's Proportionality Theorem in the non-compact case, *Invent. Math.*, 1977, **42**., pp. 239-272.

A-1977c An Algebro-Geometric Construction of Commuting Operators and of Solutions to the Toda Lattice Equation, Korteweg de Vries Equation and Related Non-Linear Equations, in *Proc. of the Int. Symp. on Alg. Geom.*, (Kyoto, 1977), Kinokuniya, Tokyo 1978, pp.115-153.

A-1978a Some Footnotes to the Work of C.P. Ramanujam, in *C.P. Ramanujam– A Tribute*, Studies in Math. No 8, Tata Institute of Fundamental Research, 1978, pp.247-262.

A-1978b The Work of C.P.Ramanujam in Algebraic Geometry, in *C.P. Ramanujam– A Tribute*, Studies in Math. No 8, Tata Institute of Fundamental Research, 1978, pp.8-10.

A-1978c An Instinct for the Key Idea (with John Tate), *Science*, 1978, **202**, pp. 737-739.

A-1979a An Algebraic Surface with K ample, $(K^2) = 9, p_g = q = 0$, *Amer. J. of Math.*, 1979, **101**, pp. 233-244.

A-1978b The Spectrum of Difference Operators and Algebraic Curves (with Pierre Van Moerbeke), *Acta Math.*, 1979, **143**, pp. 93-154.

A-1982 On the Kodaira Dimension of the Moduli Space of Curves (with Joe Harris and an appendix by William Fulton), *Inv. Math.*, 1982, pp.23-88.

A-1983a On the Kodaira Dimension of the Siegel Modular Variety, in *Algebraic Geometry - Open Problems*, (Ravello, 1982), *Lecture Notes in Mathematics 997*, Springer-Verlag 1983, pp.348-375.

A-1983b Towards an Enumerative Geometry of Moduli Space of Curves, in *Arithmetic and Geometry*, edited by M. Artin, J. Tate, Birkhauser-Boston, 1983, pp.271-326.

A-1984 A Stratification of the Null Cone via the Moment Map (with Linda Ness), *Amer. J. of Math.*, **106**, 1984, pp.1281-1329.

A-1990 Foreword for non-mathematicians, in *The Unreal Life of Oscar Zariski*, by Carol Parikh, Academic Press, 1990.

A-1993 What can be Computed in Algebraic Geometry? (with Dave Bayer), in *Computational Algebraic Geometry and Commutative Algebra*, ed. D.Eisenbud & L.Robbiano, Camb Univ. Press, 1993, pp. 1-48.

III. Research in Computer Vision: 1983-2007

V-1984 The Representation of Shape (with A. Latto and J. Shah), in *Proceedings of the 1984 IEEE Workshop on Computer Vision*, pp. 183-191, 1984.

V-1985 Boundary Detection by Minimizing Functionals I (with J. Shah), in *Image Understanding 1989*, Ablex Press, preliminary version in *1985 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 1985.

V-1987 The Problem of Robust Shape Descriptors, in *Proc of 1st IEEE International Conference on Computer Vision* (ICCV), 1987, pp.602-606.

V-1989 Optimal Approximations of Piecewise Smooth Functions and Associated Variational Problems (with J. Shah), *Comm. in Pure and Appl. Math.*, 1989, **42**, pp.577-685.

V-1990 The 2.1D Sketch (with M. Nitzberg), in *Proc. of 3rd IEEE International Conference on Computer Vision* (ICCV), 1990, pp.138-144.

V-1991a Parametrizing Exemplars of Categories, *J. Cognitive Neuroscience*, 1991, **3**, pp. 87-88.

V-1991b Mathematical Theories of Shape: do they model perception?, in *Proc. Conference 1570, Soc. Photo-optical & Ind. Engineers*, 1991, pp. 2-10.

V-1992a Texture Segmentation by Minimizing Vector-Valued Energy Functionals: the coupled membrane model (with Tai Sing Lee and Alan Yuille), *Proc. European Conf. Comp. Vision, 1992*, Lecture Notes in Computer Science 588, pp. 165-173.

V-1992b A Bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions, (with P. Belhumeur), *Proc. IEEE Conf. Comp. Vision and Pattern Recognition, 1992* (CVPR), pp. 506-512.

V-1993 Elastica and Computer Vision, in *Algebraic Geometry and its Applications*, ed. C. Bajaj, Springer-Verlag, 1993, pp. 507-518.

V-1994a Commentary on Grenander & Miller "Representations of Knowledge in Complex Systems", *Proc. Royal Stat. Soc.*, 1994.

V-1994b Pattern Theory: a Unifying Perspective, in *Proceedings 1st European Congress of Mathematics*, Birkhauser-Boston, 1994. Revised version in *Perception as Bayesian inference*, ed. D.Knill and W.Richards, Cambridge Univ. Press, 1996, pp. 25-62.

V-1994c Chordal completions of planar graphs (with F.R.K. Chung), *J. of Combinatorics*, **62**, 1994, pp.96-106.

V-1994d The Bayesian Rationale for Energy Functionals, in *Geometry-Driven Diffusion in Computer Vision*, Bart Romeny editor, Kluwer Academic, 1994, pp. 141-153.

V-1995 The Statistical Description of Visual Signals, *ICIAM 95* ed. K.Kirshgassner, O.Mahrenholtz & R.Mennicken, Akademie Verlag, 1996.

V-1996 Review of *Variational Methods in image segmentation*, by J-M Morel & S. Solimini, *Bull. Amer. Math. Soc.*, **33**, 1996, 211-216.

V-1997a Learning generic prior models for visual computation (with S.C.Zhu), *Proc. IEEE Conf. Comp. Vision and Pattern Rec.* 1997, 463-469, Comp Sci Press.

V-1997b Minimax Entropy Principle and its Application to Texture Modeling (with S.C.Zhu and Y.N.Wu), *Neural Computation*, **9**, 1997, 1627-60.

V-1997c Prior Learning and Gibbs Reaction-Diffusion (with Song Chun Zhu), *IEEE Trans. Patt. Anal. and Mach. Int.*, **19**, 1997, 1236-50.

V-1998 FRAME: Filters, Random Field and Maximum Entropy, (with S.C.Zhu and Y.Wu), *Int. J. Comp. Vis.*, **27**, 1998.

V-1999 The Statistics of Natural Images and Models (with J.Huang), *Proc. IEEE Conf. Comp. Vision and Pattern Rec.* 1999, pp.541-547, Comp Sci Press.

V-2000 Statistics of range images (with Jinggang Huang and Ann Lee), *Proc. IEEE Conf. Comp. Vision and Pattern Rec.* 2000, pp. 324-331, Comp Sci Press.

V-2001a Stochastic Models for Generic Images (with Basilis Gidas), *Quarterly Appl. Math.*, **59**, 2001, pp.85-111.

V-2001b Occlusion models for natural images: A statistical study of a scale-invariant dead-leaves model, (with Ann Lee and Jinggang Huang), *Int. J. Computer Vision*, **41**, 2001, pp. 35-59.

V-2003 The Nonlinear Statistics of High-contrast Patches in Natural Images (with Ann Lee and Kim Pedersen), *Int. J. Comp. Vision*, **54**, 2003, pp.83-103.

V-2006 Empirical Statistics and Stochastic Models for Visual Signals, in *Brain and Systems: New Directions in Statistical Signal Processing*, ed. by S.Haykin, J.Principe, T.Sejnowski, and J.McWhirter, MIT Press, 2006.

V-2007 A stochastic grammar of images (with Song-Chun Zhu), *Foundations and Trends in Computer Graphics and Vision*, **2**, 2007, pp. 259-362

V-2010 *Pattern Theory, the Stochastic Analysis of Real World Signals*, (with Agnes Desolneux), AKPeters/CRC Press/Taylor&Francis, 2010.

IV. Research on Geometry of Shape Spaces: 2001-2012

S-2001 Surface evolution under curvature flow (with Conglin Lu and Yan Cao), Special Issue on Partial Differential equations in Image Proc. Comp. Vision and Comp. Graphics, *Journal of Visual Communication and Image Representation*, 2001.

S-2002 Geometric Structure Estimation of Axially Symmetric Pots from Small Fragments(with Yan Cao), in *Proc. of Int. Conf. on Signal Processing, Pattern Recognition, and Applications*, Crete, 2002.

S-2004 2D-Shape Analysis using Conformal Mapping (with Eitan Sharon), *Int. J. of Comp. Vision* **70**, 2006, pp.55-75; preliminary version in *Proc. IEEE Conf. Comp. Vision and Patt. Rec.* , 2004.

S-2005 Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms (with Peter Michor), *Documenta Mathematica*, **10**, 2005.

S-2006a Riemannian Geometries on Spaces of Plane Curves (with Peter Michor), *J. of the European Math. Society*, **8**, 2006, pp.1-48.

S-2006b Stuff It! Review of *Introduction to Circle Packing: The Theory of Discrete Analytic Functions* by Kenneth Stephenson, *The American Scientist*, bf 94, 2006.

S-2007 An overview of the Riemannian metrics on spaces of curves using the Hamiltonian Approach, (with Peter Michor), *Applied and Computational Harmonic Analysis*, **23**, 2007, pp. 74-113.

S-2009 A metric on shape space with explicit geodesics, (with Laurent Younes, Peter Michor and Jayant Shah), *Rendiconti Lincei – Matematica e Applicazioni*, **19**, 2009, p. 2557.

S-2012a Sectional Curvature in terms of the Co-Metric and with Applications to the Riemannian Manifolds of Landmarks (with M. Micheli and P. Michor), *SIAM J. on Imaging Sciences*, **5**, 2012,, pp. 394-433.

2012b The Geometry and Curvature of Shape Spaces, talk with summary in *De Giorgi Colloquium 2009* published in Edizioni della Normale by the Scuola Normale Superiore, Pisa, 2012.

S-2013a Sobolev Metrics on Diffeomorphism Groups and the Derived Geometry of Spaces of Submanifolds (with Mario Micheli and Peter Michor), *Izvestiya RAN, Math. Series*, **77**, 2013, pp. 109-138.

S-2013b On Euler's equation and 'EPDiff' (with Peter Michor), *Journal of Geometric Mechanics.* **5**, 2013, pp. 319-344.

S-2013c A zoo of difeomorphism groups on $\mathbb{R}^n$, (with Peter Michor), *Annals of Global Analysis and Geometry*, **44**, 2013, pp. 529-540.

S-2014 Geodesic completeness for Sobolev metrics on the space of immersed plane curves (with Martins Bruveris and Peter Michor), *Forum of Math., Sigma*, **1**, 2014.

V. Research in the Biology and Psychology of Vision: 1983-2004

B-1987 Discriminating Figure from Ground: the role of edge detection and region growing (with S. Kosslyn, L. Hillger and R. Herrnstein), *Proc. Nat. Acad. Sci.*, 1987, **84**, pp.7354-7358.

B-1989 Teaching Pigeons an Abstract Relational Rule: Insideness (with R. Herrnstein, W. Vaughan and S. Kosslyn), *Perception and Psychophysics*, 1989, **46**, pp. 56-64.

B-1991 On the Computational Architecture of the Neocortex, I: The role of the thalamo-cortical loop, *Biological Cybernetics*, 1991, **65**, pp.135-145; II: The role of cortico-cortical loops, *Biological Cybernetics*, **66**, pp. 241-251.

B-1994 Neuronal Architectures for Pattern-theoretic Problems, in *Large Scale Neuronal Theories of the Brain*, MIT Press, 1994, pp. 125-152.

B-1995a Thalamus, in *The Handbook of Brain Theory and Neural Networks*, M. Arbib editor, MIT Press, 1995.

B-1995b Neural correlates of boundary and medial axis representations in primate striate cortex, (with T.S.Lee, K.Zipser & P.H.Schiller), ARVO abstract, 1995.

B-1997a Issues in the mathematical modeling of cortical functioning and thought, in *The Legacy of Norbert Wiener: A Centennial Symposium*, ed. D.Jerison et al, Amer. Math. Society, 1997, pp. 235-260.

B-1997b Visual Search and Shape from Shading Modulate Contextual Processing in Macaque Early Visual Cortices, (with T.S.Lee, R.Romero, A.Tobias & T.Moore), *Neuroscience Abstract*, 1997.

B-1997c The Role of V1 in Shape Representation (with Tai Sing Lee, Song Chun Zhu & Victor Lamme), *Computational Neuroscience*, ed. Bower, Plenum Press, 1997.

B-1998 The Role of Primary Visual Cortex in Higher Level Vision (with T.S.Lee, R.Romero and V.Lamme), *Vision Research*, **38**, 1998, 2429-2454.

B-1999 Thalamus, in *MIT Encyclopedia of the Cognitive Sciences*, MIT Press, 1999.

B-2001 Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency (with Tai Sing Lee, C.Yang, R.Romero), *Nature Neuroscience*, **5**, 2002, 589-597.

B-2003 Hierarchical Bayesian Inference in the Visual Cortex, (with Tai Sing Lee), *Journal of the Optical Society of America*, **20**, 2003, 1434-1448.

B-2004a Modeling and Decoding Motor Cortical Activity using a Switching Kalman Filter, (with W.Wu, M.Black, Y.Gao, E.Bienenstock, J.Donoghue), *IEEE Trans. on Biomed. Eng.* , **51**, pp. 933-942, 2004.

B-2004b Movement Direction Decoding using Fast oscillation in Local Field Potential and Neural Firing, (with Wei Wu, W.Truccolo, M.Saleh and J.Donoghue), *13th Computational Neuroscience Meeting*, 2004.

B-2005 Minds must unite: It's time for experimentalists to stop ignoring computational modelers (with David Donoho and Bruno Olshausen), 'Opinion' section, *The Scientist*, June 6, 2005.

VI. Math Education, History of Math, Popular Math, Reviews, Obituaries

E-1'986 Oscar Zariski: 1899-1986, *Notices of the Amer. Math. Society*, **33**, 1986, pp.891-894.

E-1995 Contributor to *Multi-variable Calculus*, The Calculus Consortium based at Harvard, Wiley, 1995.

E-1997 Calculus Reform – For the Millions, *Notices Amer. Math. Soc.*, May 1997, pp. 559-563.

E-1998 Trends in the Profession of Mathematics, *Mitteilungen der Deutsche Mathemtiker Verein* REF, 1998.

E-2000 The Dawning of the Age of Stochasticity, in *Mathematics: Frontiers and Perspectives*, edited by V.Arnold, M.Atiyah, P.Lax and B.Mazur, AMS, 2000.

E-2005 Mathematics in the Near East, some personal observations, *Notices of the AMS*, May 2005.

E-2006 Mathematics Belongs in a Liberal Education, *The Arts and Humanities in Higher Education*, **5**, 2006, pp.21-32.

E-2008a Henri's Crystal Ball (with Phil Davis), *Notices of the AMS*, **55**, 2008, pp. 458-466.

E-2008b Why I am a Platonist, *Newsletter of European Math. Soc.*, December 2008, pp. 27-30.

E-2008c The Wolf Prize and Supporting Palestinian Education, *Notices of the AMS*, **55**, 2008, pp.919 & 1368.

E-2009 Intelligent Design Found in the Sky with $p < 0.001$, Newsletter of the Swedish Math Society (Svenska matematikersamfundet Medlemsutskicket), Feb. 2009, pp. 64-70.

E-2010a Review of *Mathematics in India* by Kim Plofker, *Notices of the AMS*, 2010, pp. 385-390.

E-2010b Passages to India, *Mathematics Intelligencer, Hyderabad edition*, 2010, pp.51-55.

E-2010c What's so Baffling about Negative Numbers – a Cross-Cultural Comparison, in *Studies in the History of Indian Mathematics*, C. S. Seshadri editor, Hindustan Book Agency (distr. in US by AMS), 2010.

E-2010d What Should a Mathematics Professional Know About Mathematics, Newsletter of the Swedish Math Society (Svenska matematikersamfundet Medlemsutskicket), May. 20010, pp. 5-14.

E-2011a Intuition and Rigor and Enriques's Quest, *Notices of the AMS*, **58**, 2011, pp. 250-260.

E-2011b How to Fix Our Math Education (with Sol Garfunkel), OpEd contribution to *New York Times*, August 25, 2011.

E-2012a "Yu laid out the lands": georeferencing the Chinese Yujitu [Map of the Tracks of Yu] of 1136 (with Alexander Akin), *Cartography and Geographic Information Science Journal*, 2012.

E-2012b Foreword to *The Best Writing on Mathematics 2012*, edited by Mircea Pitici, Princteon Univ. Press, 2012.

E-2014 Appreciations of the work of Alexander Grothendieck: (i) My Introduction to Schemes and Functors, in *Alexander Grothendieck: A Mathematical Portrait*, ed. L. Schneps, International Press, 2014; (ii) Alexander Grothendieck (1928-2014) (with John Tate), *Nature*, **517**, 15 Jan. 2015; (iii) Alexandre Grothendieck, 1928-2014, sketch of some of his mathematical work (with Michael Artin, Allyn Jackson and John Tate), *Notices of the Amer. Math. Soc.*, **63**, 2016.

E-2016 Assessing the accuracy of ancient eclipse predictions, to appear *Proc. Takebe Conference*, Advanced Studies in Pure Math,. Math. Soc. Japan, 2016.

E-2019 Thoughts on Consciousness, *Journal of Cognitive Science*, **20**, 2019, pp. 251-279.

E-2021 Ruminations on Cosmology and Time, *Notices of the AMS*, November 2021, vol. 68, pp. 1715–1725.

Bibliography

[Ack16]   Jennifer Ackerman. The Genius of Birds. Penguin Press, 2016.

[ADJ⁺16]   Douglas Arnold, Guy David, David Jerison, Svitlana Mayboroda, and Marcel Filoche. Effective confining potential of quantum states in disordered media. Physical Review Letters, 2016.

[Amm99]   Sarasvati Amma. Geometry in Ancient and Medieval India. Motilal Banarsidass, Delhi, 1999.

[AV64]   Michael Artin and Jean-Louis Verdier. Seminar on 'Etale cohomology of number fields, 1964. Thanks to John Milne, available online at https://www.jmilne.org/math/Documents/woodshole.pdf.

[AvBG⁺11] A. Auersperg, A. von Bayern, G. Gaidon, L. Huber, and A. Kacelnik. Flexibility in problem solving and tool use of kea and new caledonian crows in a multi access box paradigm. PLOS ONE, 2011.

[AWW91]   Pascal Auscher, Guido Weiss, and Victor Wickerhauser. Local sine and cosine bases of coifman and meyer. In C. K. Chui, editor, Wavelets – a tutorial. Academic Press, 1991.

[AZM⁺10]   Srdjan Antic, Wen-Liang Zhou, Anna Moore, Shaina Short, and Katerina Ikonomu. The decade of the dendritic nmda spike. Journal of Neuroscience Research, 88:2991–3001, 2010.

[Bel62]   John Bell. On the Einstein-Podolsky-Rosen paradox. Physics Physique Физика, 1:195–200, 1962.

[Bel90]   John Bell. Against 'measurement'. Physics World, pages 33–40, August 1990.

[BH74]   J. J. Bastian and K. J. Harrison. Subnormal weighted shifts and asymptotic properties of normal operators. Proceedings of the American Mathematical Society, 42, 1974.

[Bis20]   Christopher Bishop. Weil-petersson curves, conformal energies, β-numbers, and minimal surfaces. Under preparation for publication, 2020.

[Bla14]   Bruce Blausen. Blausen 0114 brainstemanatomy.png, 2014. From Wikimedia Commons, Medical gallery of Blausen Medical 2014, WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010, ISSN 2002-4436.

[Bri16]   Jean Bricmont. Making Sense of Quantum Mechanics. Springer, 2016.

[BT24]   Stefan Banach and Alfred Tarski. Sur la décomposition des ensembles de points en parties respectivement congruentes. Fundamenta Mathematicae, pages 244–277, 1924.

[Bur17]   Stephen Buranyi. Is the staggeringly profitable business of scientific publishing bad for science?, 2017.

[BZV$^+$19]   Irwan Bello, Barret Zoph, Ashish Vaswani, Jonathon Shlens, and Quoc Le. Attention augmented convolutional networks, 2019. Online at arXiv: 1904.09925v5.

[Cal73]   John Calhoun. Death squared: The explosive growth and demise of a mouse population. Proceedings of the Royal Society of Medicine, 66:80–86, 1973.

[Car92]   J. L. Carr. Harpole and Foxberrow, General Publishers. Gardner Books, 1992.

[CC15]   Marie-Claire Cammaerts and Roger Cammaerts. Are ants (hymenoptera formicidae) capable of self recognition? Journal of Science, 5:521–532, 2015.

[CG21]   Al Cuoco and Paul Goldenberg. Computational thinking in mathematics and computer science: What programming does to your head. Journal of Humanistic Mathematics, 11, 2021.

[Cla54]   Alexis Clairaut. Sur l'orbite apparente du soleil autour de la terre. Mèmoires de Mathèmatiques et de la Physique, Acadèmie des Sciences, pages 521–561, 1754.

[CM91]   Ronald Coifman and Yves Meyer. Remarques sur l'analyse de fourier à fenêtre. Comptes Rendus de l'Academie de Science Paris, serie 1, 312:259–261, 1991.

[Com81]   Bernard Comrie. Language Universals and Linguistic Typology: Syntax and Morphology. University of Chicago Press, 1981.

[Con74]   Hyla Stuntz Converse. The agnicayana rite: Indigenous origin? History of Religions, pages 81–95, 1974.

[Cul82]   Christopher Cullen. An eight century table of tangents. Chinese Science, 1:1–33, 1982.

[Cuo10]   Al Cuoco. CME Project: Algebra 1. Pearson, CME Project, 2010.

[Cuo19]   Al Cuoco. Calculating the monthly payments on a loan, 2019. Available online at `https://go.edc.org/essays`.

[Dar72]   Charles Darwin. The Expression of the Emotions in Man and Animals. 1872.

[Deh14]   Stanislas Dehaene. Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts. Viking, 2014.

[Deh23]   Stanislas Dehaene. Cours : Quel code neural pour les représentations mentales?, 2023.

[Des54]   René Descartes. La Geometrie. Dover publications, 1954. A facsimile of the 1637 French text plus an English translation.

[DHH⁺05]   C. M. Dawson, H. L. Haselgrove, A. P. Hines, D. Mortimer, M. A. Neilsen, and T. J. Osborne. Quantum computing and polynomial equations over $\mathbb{Z}_2$. Quantum Information Computing, 2005.

[Dia05]   Jared Diamond. Collapse. Viking Press, 2005.

[Div18]   P. P. Divakaran. The Mathematics of India. Springer, 2018.

[DJJ91]   Ingrid Daubechies, Stéphane Jaffard, and Jean-Lin Journé. A simple wilson orthonormal basis with exponential decay. SIAM Journal on Mathematical Analysis, pages 554–572, 1991.

[DKSZ96]   A. I. Dyachenko, E. A. Kuznetsov, M. D. Spector, and V. E. Zakharov. Analytical description of the free surface dynamics of an ideal fluid. Physics Letters A, 221:73–79, 1996.

[Don09]   Wendy Doniger. The Hindus. Viking/Penguin, 2009.

[dS77]   Benedictus de Spinoza. Ethica Ordine Geometrico Demonstrata. 1677. Available free online in Project Gutenberg as 'The Ethics', `https://www.gutenberg.org/files/3800/3800-h/3800-h.htm`.

[dW19]   Frans de Waal. Mama's Last Hug. W. W. Norton, 2019.

[DYQ11]   Jia Du, Laurent Younes, and Anqi Qiu. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. Neuroimage, 56:162–173, 2011.

[Dys81]   Freeman Dyson. Disturbing the Universe. Basic Books, 1981.

[Dys08]   Freeman Dyson. The question of global warming. New York Review of Books, 2008. Issue June 12.

[Ecc90]   John Eccles. A unitary hypothesis of mind-brain interaction in the cerebral cortex. Proceedings of the Royal Society London, B240:433–451, 1990.

[Ell99]   Willis D. Ellis. A source book of Gestalt psychology. Psychology Press, 1999.

[EM56]   Andrzej Ehrenfeucht and Andrzej Mostowski. Models of axiomatic theories admitting automorphisms. Fundamenta Mathematicae, 1956.

[Ens83]   V. Enss. Asymptotic observables on scattering states. Communications in Math. Physics, 89:245–268, 1983.

[EP77]   John Eccles and Karl Popper. The Self and its Brain. Routledge, 1977.

[EPR35]   Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? Physical Review, 47:777–780, 1935.

[Eul60]   Leonhard Euler. Recherches sur la courbure des surfaces. Mémoires de l'Académie des Sciences de Berlin, pages 119–143, 1760.

[Eve09]   Daniel Everett. Don't sleep, There are snakes: Life and Language in the Amazonian Jungle. Vintage, 2009.

[Fey85]   Richard Feynman. Surely you're joking Mr. Feynman. W. W. Norton, 1985.

[FHMV95]   R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. Reasoning about Knowledge. MIT Press, 1995.

[Fol08]   Gerald Folland. Quantum Field Theory: A Tourist Guide for Mathematicians. American Mathematical Society, 2008.

[Fre86]   Christopher Freiling. Axioms of symmetry: throwing darts at the real line. Journal of Symbolic Logic, 51:190–200, 1986.

[Fri07]   Joran Friberg. Amazing Traces of a Babylonian Origin in Greek Mathematics. World Scientific, 2007.

[Fri11]   Harvey Friedman. Maximal invariant cliques and incompleteness, 2011. Online at `https://u.osu.edu/friedman.8/foundational-adventures/download able-manuscripts/`, manuscript 81. You can find all Friedman's recent work in preprint form here.

[Frö19]   Jürg Frölich. A brief review of the eth- approach to quantum mechanics, 2019. ArXiv:1905.06603v2.

[Frö22]   Jürg Frölich. Irreversibility and the arrow of time, 2022. ArXiv:2202.04619v1.

[FS99]   Susan Forman and Lynn Steen. Beyond eighth grade: Functional mathematics for life and work. University of California at Berkeley, 1999. Available online at `https://eric.ed.gov/?id=ED434271`.

[FTL09]   Winrich Freiwald, Doris TsAo, and Margaret Livingston. A face feature space in the macaque temporal lobe. Nature Neuroscience, pages 1187–1196, 2009.

[FV63]   Richard Feynman and Frank Vernon. The theory of a general quantum system interacting with a linear dissipative system. Annals of Physics, 24:118–173, 1963.

[Gar83]   Howard Gardner. Frames of Mind: The Theory of Multiple Intelligences. Basic Books, 1983.

[Går05]   Lars Gårding. Encounters with science: Dialogues in five parts (1). In Perspectives in Analysis, Mathematical Physics Studies, Vol.27. Springer, 2005.

[GK09]   Loren Graham and Jean-Michel Kantor. Naming Infinity. Harvard University Press, 2009.

[GM07]   Ulf Grenander and Michael Miller. Pattern Theory. Oxford University Press, 2007.

[Gre81]   Ulf Grenander. Lectures in Pattern Theory I, II and III, Pattern Analysis, Pattern Synthesis, Regular Structures. Springer, 1981. Three parts published starting in 1976.

[Gre12]   Ulf Grenander. A Calculus of Ideas: A Mathematical Study of Human Thought. World Scientific, 2012.

[Gro86]   Alexander Grothendieck. Recoltes et semailles. `https://www.quarante-deux.org/archives/klein/prefaces/Romans_1965-1969/Recoltes_et_semailles.pdf`, 1986. Online; distribution restricted by his family.

[GS16]   Peter Godfrey-Smith. Other Minds: the Octopus, the Sea and the Deep Origins of Consciousness. Farrar, Straus and Giroux, 2016.

[GZ15]   M. Giustina and A. Zeilinger. Significant-loophole-free test of bell's theorem with entangled photons. Physical Review Letters, 115, 2015.

[Haa96]   Rudolf Haag. Local Quantum Physics. Springer, 1996.

[Hai12]   Jonatan Haidt. The Righteous Mind. Pantheon, 2012.

[Har07]   Guershom Harel. The dnr system as a conceptual framework for curriculum development and instruction. In R. Lesh, J. Kaput, and E. Hamilton, editors, Foundations for the Future in Mathematics Education. Lawrence Erlbaum Inc., 2007.

[HCL+21]   Irina Higgins, Le Chang, Victoria Langston, Demis Hassabis, Chistopher Summerfield, Doris Tsao, and Matthew Botvinick. Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. Nature Communications, 2021.

[Hei58]   Werner Heisenberg. The Physicist's Conception of Nature. Harcourt, Brace, 1958.

[HHM98]   Deborah Hughes-Hallett and William McCallum. Calculus: Single and Multivariable. Wiley, 2nd edition, 1998.

[Hil21]   Fiona Hill. There Is Nothing For You Here: Finding Opportunity in the Twenty-First Century. Mariner Books, 2021.

[HM82]   Joe Harris and David Mumford. On the kodaira dimension of the moduli space of curves. Inventiones Mathematicae, 67:23–86, 1982.

[HS00]   W. Hunziker and I. M. Sigal. The quantum n-body problem. Journal Math. Physics, 41:3448–3510, 2000.

[HYB+14]   Marc Hauser, Charles Yang, Robert Berwick, Ian Tattersall, Michael Ryan, Jeffrey Watumull, Noam Chomsky, and Richard Lewontin. The mystery of language evolution. Frontiers of Psychology, 2014.

[HZ09]   Feng Han and Song Chun Zhu. Bottom-up/top-down image parsing by attribute graph grammar. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 59–73, 2009.

[Imh09]   Annette Imhausen. Traditions and myths in the historiography of egyptian mathematics. In Oxford Handbook of the History of Mathematics, pages 781–800. Oxford University Press, 2009.

[Jan63]   H. W. Janson. History of Art. Thames and Hudson, 1963.

[Jec97]   Thomas Jech. Set Theory. Spinger, 1997.

[Kan80]   Gaetano Kanisza. Grammatica del Vedere. Il Mulino, Bologna, Italy, 1980. French translation, La Grammaire du Voir, Diderot, 1997.

[Kan03]   Akihiro Kanamori. The Higher Infinite. Springer, 2003.

[Kat73]   Tosio Kato. Continuity of the map $s \rightarrow |s|$ for linear operators. Proceedings of the Japan Academy, 49:157–160, 1973.

[Kle79]   Felix Klein. Vorlesungen uber die Entwicklung der Mathematik im 19.Jahrhundert. Springer-Verlag, reprint edition, 1979.

[KM78]   Akihiro Kanamori and Menachem Magidor. The evolution of large cardinal axioms in set theory. In Higher Set Theory, number 669 in Lecture Notes in Mathematics, 1978.

[KM97]   Andreas Kriegl and Peter Michor. The Convenient Setting of Global Analysis. American Mathematical Society, 1997.

[Kol33]   Andrey Kolomogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. 1933. English translation published by Chelsea Books in 1956 and reprinted many times.

[Kol14]   Elizabeth Kolbert. The Sixth Extinction. Henry Holt and Company, New York City, 2014.

[Kri64]   Saul Kripke. Transfinite recursion on admissible ordinals. Journal of Symbolic Logic, pages 161–162, 1964.

[Kus09]   Sergei Kushnarev. Teichons: Solitonlike geodesics on universal teichmüller space. Experimental Mathematics, 18:325–336, 2009.

[Lan15]   Nick Lane. The Vital Question: Energy, Evolution, and the Origins of Complex Life. W. W. Norton, 2015.

[LCD⁺87]   A. J. Leggett, S. Chakravart, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zweiger. Dynamics of the dissipative two-state system. Review Modern Physics, 59, 1987.

[LDG⁺17]   Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection, 2017. Online at arXiv: 1612.03144v2.

[L.H29]   L.Haggerty. What a two-and-a-half-year-old child said in one day. Journal of Genetic Psychology, pages 75–85, 1929.

[Lie84]   Philip Lieberman. <u>The Biology and Evolution of Language</u>. Harvard University Press, 1984.

[Lin97]   C. C. Lin. Almost commuting self-adjoint matrices and applications. <u>Fields Institute Communications</u>, 13, 1997.

[LL65]   L. D. Landau and E. M. Lipschitz. <u>Quantum Mechanics: Non-Relativistic Theory</u>. Pergamon Press, 1965.

[LLC$^+$22] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows, 2022. Online at arXiv: 2103.14030v2.

[LMW$^+$22] Zhuand Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Carrell, and Saining Xie. A convnet for the 2020s, 2022. Online at arXiv: 2201.03545v2.

[Mal98]   Thomas Robert Malthus. <u>An Essay on the Principle of Population</u>. William Goodwin, London, 1798.

[Man]   Manu. Manusmrti. English translation by George Bühler , online at `http://www.sacred-texts.com/hin/manu.htm`.

[Man20]   Daniel Mansfield. Perpendicular lines and diagonal triples in old babylonian surveying. <u>Journal of Cuneiform Studies</u>, 72:87–99, 2020.

[Man21]   Daniel Mansfield. Plimpton 322: A study of rectangles. <u>Foundations of Science</u>, 2021.

[Mar82]   David Marr. <u>Vision</u>. W. H. Freeman and Co., 1982. Reprinted by MIT Press.

[Mat00]   Richard Mattessich. <u>The Beginnings of Accounting and Accounting Thought</u>. Garland Publishing, 2000.

[MB21]   ManuelBlum and Lenore Blum. A theoretical computer science perspective on consciousness. <u>Journal of Artificial Intelligence and Consciousness</u>, 8:1–42, 2021.

[McC20]   Colin McCann. <u>Apeirogon</u>. Random House, 2020.

[MCCD13]   Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013. Online: arXiv:1301.3781v3.

[McI71]   Alan McIntosh. Counterexanple to a question on commutators. <u>Proceedings of the American Mathematical Society</u>, 29:337–340, 1971.

[McK19]   Ian McKewan. Machines like Me. Nan A. Talese, 2019.

[MD19]   Earl Miller and Robert Desimone. Charles gordon gross (1936–2019). Neuron, 2019.

[Mel16]   Bartlett Mel. Toward a simplified model of an active dendritic tree. In Dendrites. Oxford University Press, 2016.

[Mer07]   Bjorn Merker. Consciousness without a cerebral cortex: A. challenge for neuroscience and medicine. Behavioral and Brain Sciences, 30:63–134, 2007.

[Met87]   Nicholas Metropolis. The beginning of the monte carlo method. Los Alamos Science, Special Issue:125–130, 1987.

[Met09]   Thomas Metzinger. The Ego Tunnel: the Science of Mind and the Myth of the Self. Basic Books, 2009.

[MH19]   Chris Manning and John Hewitt. A structural probe for finding syntax in word representations, 2019. Online: arXiv: 1803.00188.

[Mit67]   John Mitchell. An inquiry into the probable parallax, and magnitude, of the fixed stars, from the quantity of light which they afford us, and the particular circumstances of their situation. Philosophical Transactions, 57:234–264, 1767.

[MJ78]   Colin McEvedy and Richard Jones. Atlas of World Population History. Puffin-Penguin, London, 1978.

[Mon45]   Piet Mondrian. Plastic Art and Pure Plastic Art. Wittenborn and Company, 1945.

[Mon15]   Sy Montgomery. The Soul of an Octopus. Simon & Schuster, 2015. You must see also her lecture "https://www.youtube.com/watch?v=_N2yDf7_1oc.

[Mor12]   Masanori Morishita. Knots and Primes, an Introduction to Arithmetic Topology. Springer, 2012.

[Mos09]   Yiannis Moschovakis. Descriptive Set Theory, 2nd edition. American Mathematical Society, 2009.

[MSW02]   David Mumford, Caroline Series, and David Wright. Indra's Pearls. Cambridge University Press, 2002.

[MTY02]   Michael Miller, Alain Trouvé, and Laurent Younes. On the metrics and euler-lagrange equations of computational anatomy. Annual Reviews Biomedical Engineering, 4:375–405, 2002.

[Mum30]   W. Bryant Mumford. Malangali school. Atrica, 3:265–292, 1930.

[MVC⁺14]   Nikola Markov, Julien Vezoli, Pierre Chameau, Arnold Falchier, René Quilo-
           dran, and Shimon Ullman et al. Anatomy of hierarchy: Feedforward and feed-
           back pathways in macaque visual cortex. Journal of Comparative Neurology,
           522:225–259, 2014.

[OM19]   Tsukane Ogawa and Mitsuo Morimoto. Mathematics of Takebe Katahiro and
           History of Mathematics in East Asia. Mathematical Society of Japan, 2019.

[Osb10]   Alfred Osborne. Nonlinear Ocean Waves and the Inverse Scattering Transform.
           Academic Press, London, UK, 2010.

[Pan04]   Jaak Panksepp.   Affective Neuroscience: The Foundations of Human and
           Animal Emotions. Oxford University Press, 2004.

[PB04]   Jaak Panksepp and Lucy Biven. The Archeology of Mind: Neuroevolutionary
           Origins of Human Emotions. W. W. Norton, 2004.

[Pea89]   Guiseppe   Peano.   'Arithmetices   principia,   nova   methodo
           exposita.   Fratres   Bocca,   Rome,   1889.   Avaliable   at
           https://archive.org/details/arithmeticespri00peangoog.

[Pea09]   Judea Pearl. Causality: Models, Reasoning and Inference. Cambridge Uni-
           versity Press, 2009.

[PH77]   Jeff Paris and Leo Harrington. A mathematical incompleteness in peano arith-
           metic. In The Handbook of Mathematical Logic. North-Holland, Amsterdam,
           1977.

[PKS17]   Mark Penney, Dax Koh, and Robert Spekkens. Quantum circuit dynamics via
           path integrals: Is there a classical action for discrete time paths? New Journal
           of Physics, 2017.

[Pre85]   Emma Previato. Hyperelliptic quasi-periodic and soliton solutions of the non-
           linear schrödinger equation. Duke Mathematical Journal, 52:329–377, 1985.

[PSY⁺13]   Mingtao Pei, Zhangzhang Si, Benjamin Yao, Y. Jia, and Song-Chun Zhu.
           Learning and parsing video events with goal and intent prediction. Computer
           Vision and Image Understanding, pages 1203–1548, 2013.

[Pto11]   Claudius Ptolemy. Geography of Claudius Ptolemy. Cosimo Classics, 2011.
           Translated by Edward Luther Stevenson originally in 1932, reprinted often.

[Rap03]   Michael Rappenglueck. BBC News, 2003. Broadcast on Jan. 21, 2003.

[RF18]   Gabriele Radnikow and Dirk Feldmeyer. Layer- and cell type-specific modulation of excitatory neuronal activity in the neocortex. Frontiers in Neuroanatomy, 12, 2018.

[RG06]   Ben Rudiak-Gould. The sum-over-histories approach to quantum computing, 2006. ArXiv: quant-ph/0607151.

[Rob02]   Eleanor Robson. Words and pictures: New light on plimpton 322. American Mathematical Monthly, 109, 2002.

[Ros16]   Alex Rosenberg. Why you don't know your own mind. `https://www.nyti mes.com/2016/07/18/opinion/why-you-dont-know-your-own-mind.html`, 2016. OpEd piece in the New York Times, July 18, 2016.

[SB83]   S. N. Sen and A.K. Bag. The Sulbasutras. Indian National Science Academy, 1983.

[SB92]   Denise Schmandt-Besserat. Before Writing: from Counting to Cuneiform. Univ. Texas Press, 1992.

[Sch11]   Maximillan Schlosshauer. Elegance and Enigma, The Quantum Interviews. Springer, 2011.

[Sch12]   Peter Scholze. Perfectoid spaces. Publications de l'Institut des Hautes Études Scientifique, 116, 2012.

[Sha80]   Igor Shafarevich. The Socialist Phenomenon. Harper and Row, 1980.

[Sha90]   Igor Shafarevich. Russophobia. https://apps.dtic.mil/dtic/tr/fulltext/u2/a335121.pdf, 1990. Online English translation.

[Sho94]   Peter Shor. Algorithms for quantum computation: discrete logarithms and factoring. In Proceedings 35th Annual Symposium on Foundations of Computer Science, pages 124–134. IEEE Computer Society Press, 1994.

[Sig03]   Laurence Sigler. Fibonacci's Liber Abaci. Springer, 2003. Latin original appeared in 1202 by the author whose proper name is Leonardo de Pisa.

[Sil71]   Jack Silver. Some applications of model theory in set theory. Annals of Mathematical Logic, pages 45–110, 1971.

[Sil80]   Ruth Silcock. Albert John out Hunting. Viking Kestral Picture Books, 1980.

[Sim10]   Stephen Simpson. Subsystems of Second Order Arithmetic. Cambridge University Press, Cambridge, UK, 2010.

[SN15]   L. K. Shaim and S. W. Nam. Significant-loophole-free test of bell's theorem with entangled photons. Physical Review Letters, 115, 2015.

[Sol67]   Robert Solovay. A nonconstructible δ_3^1 set of integers. Transactions American Math Society, 127:50–75, 1967.

[Sol70]   Robert Solovay. A model of set-theory in which every set of reals is lebesgue measurable. Annals of Mathematics, pages 1–56, 1970.

[Sol71]   Robert Solovay. Real-valued measurable cardinals. In Axiomatic Set Theory. American Math Society, 1971.

[SSH16]   Nelson Spruston, Greg Stuart, and Michael Hausser. Principles of dendritic integration. In Dendrites. Oxford University Press, 2016.

[Sta96]   Frits Staal. Ritual and Mantras, Rules without Meaning. Motilal Banarsidass, 1996.

[Swe92]   Frank Swetz. The Sea Island Mathematical Manual. Pennsylvania State University Press, 1992.

[Tho08]   Sarah Thornton. Seven Days in the Art World. W.W.Norton, 2008.

[Tol97]   Eckhart Tolle. The Power of Now. Namaste Publishing, 1997.

[vN55]   John von Neumann. Mathematical Foundations of Quantum Mechanics. Princeton University Press, 1955. English translation of the German 1932 book.

[VSP+17]   Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Aidan Gomez, Likasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. Online: arXiv: 1706.03762.

[Wei12]   Ulrich Weiss. Quantum Dissipative Systems. World Scientific, 2012.

[Whi15]   Susan Whitfield. Life along the Silk Road. University of California Press, 2015.

[Wig62]   Eugene Wigner. Symmetries and Reflections. Ox Bow Press, 1962.

[Wol08]   Tom Wolfe. The Painted Word. Farrar, Straus and Giroux, 2008.

[Woo80]   Harry Woolf. Some Strangeness in the Proportion: A Centennial Symposium to Celebrate the Achievements of Albert Einstein. Addison-Wesley, 1980. Report of the Einstein Centennial Symposium held 4-9 March 1979 at Princeton, New Jersey, edited by Harry Woolf.

[Zak68]   V. E. Zakharvov. Stability of periodic waves of finite amplitude on the surface
of a deep fluid. Zhurnal Prikladnoi Mekhaniki i Tekhnicheskoi Fiziki, 9:86–94,
1968.

[ZDE02]   V. E. Zakharov, A. I. Dyachenko, and O. E.Vasilyev. New method of numerical
simulation of a non-stationary potential flow of incompressible fluid with a free
surface. European J. of Mechanics B - Fluids, 21:283–291, 2002.

[Zim46]   Heinrich Zimmer. Myths and Symbols in Indian Art and Civilization. Prince-
ton University Press, 1946.

[ZRBA14]   Semir Zeki, John Romaya, Dionigi Benincasa, and Michael Atiyah. The expe-
rience of mathematical beauty and its neural correlates. Frontiers in Human
Neuroscience, 8:1–12, 2014.

[Zur91]   Wojciech Zurek. Decoherence and the transition from quantum to classical.
Physics Today, 1991.