Miscellaneous Interesting Things

Michael J. Challis

Contents

1 Mathematics

A mathematicians job is, by and large, to write proofs and they take pride in writing the most elegant/simple/understandable/short proofs possible. One candidate for the shortest proof is Lander and Parkin (1966) which has a proof of less than fifty words. Before starting this section in earnest consider the theorem and proof below as another possible (albeit slightly tongue in cheek) contender for the title of shortest proof.

Theorem 1.1. There exists a quadrilateral with four equal side lengths and four right interior angles.

Proof. □

1.1 Rationality From Irrationality

The following theorem and proof are cool for a few reasons: (i) it is not immediately obvious that there will exist irrational numbers a and b such that a^b is rational, (ii) the proof is simple (it can fit on three lines), and (iii) the proof is non-constructive. Megill and Wheeler (2019) provide more analysis of this theorem.

Theorem 1.2. There exist irrational numbers a and b such that a^b is rational.

Proof. Let $a = b = \sqrt{2}$. We know $\sqrt{2}$ is irrational.[1] Therefore, if $\sqrt{2}^{\sqrt{2}}$ is rational we are done. So, suppose $\sqrt{2}^{\sqrt{2}}$ is irrational. Now, let $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Then,

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2 \implies a^b \in \mathbb{Q}.$$

□

This proof tells us that one of $\sqrt{2}^{\sqrt{2}}$ and $\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}}$ is rational but not which. (An answer to which is rational with an accompanying explanation is in Appendix A.1.) Another way to prove Theorem 1.2 is to show that e and $\ln(2)$ are both irrational (since $e^{\ln(2)} = 2$). However, proving that $\sqrt{2}$ is irrational is much easier than proving both e and $\ln(2)$ are irrational.

On the topic of $\sqrt{2}$ consider another theorem that seems rather complicated but is, in fact, easy to prove.

Theorem 1.3.

$$\sqrt{2}^{\sqrt{2}^{\sqrt{2}^{\cdot^{\cdot^{\cdot}}}}} = 2.$$

[1] Consider the following proof due to Chaitin (1920) that uses the Fundamental Theorem of Arithmetic: Suppose $\sqrt{2}$ is rational then $\sqrt{2} = a/b$ for some natural numbers a and b. Or, equivalently, $a^2 = 2b^2$. As a and b are integers their squares must have an even number of 2's in their factorization. Therefore, $2b^2$ has an odd number of 2's in its prime factorization. Thus, $a^2 \neq 2b^2$, and therefore $\sqrt{2}$ is irrational. □

Before proving this theorem consider the following aside: raising a number to itself is called a *power tower* (or, less interestingly *tetration*) and is denoted,

$$x \uparrow\uparrow k = \underbrace{x^{x^{\cdot^{\cdot^{\cdot^{x}}}}}}_{k\text{-times}}.$$

This notation is due to Donald Knuth (1976) of LaTeX (technically TeX) fame where $x \uparrow k = \underbrace{x \cdot x \cdots x}_{k\text{-times}}$. Thus, we can restate Theorem 1.3 as,

$$\begin{aligned} \sqrt{2}^{\sqrt{2}^{\sqrt{2}^{\cdot^{\cdot^{\cdot}}}}} &= \sqrt{2} \uparrow\uparrow \infty, \\ &= \lim_{n\to\infty} \sqrt{2} \uparrow\uparrow n, \\ &= 2. \end{aligned}$$

Proof. For ease of exposition let $x \coloneqq \sqrt{2} \uparrow\uparrow \infty$. Then, $x = \sqrt{2}^x$. Solving for x we see that $2 = \sqrt{2}^2$ and $4 = \sqrt{2}^4$ are both possible solutions. Thus, x is either 2 or 4. Now we show that x cannot equal 4 and, therefore, must be 2. Let $a_0 = \sqrt{2}$ and $a_{n+1} = \sqrt{2}^{a_n}$ for all n greater than 0. Clearly, a_0 is less than 2. Suppose that a_k is less than or equal to 2. Then, $a_{k+1} = \sqrt{2}^{a_k}$. We know that the expression $\sqrt{n}^m = 2$ if $n = m = 2$. We also know that n^m is monotonically increasing in both n and m. Therefore, n^m must be less than or equal to 2 for all $n, m \in [0, 2]$. So, a_{k+1} is also less than or equal to 2. We then have that $\lim_{k\to\infty} a_k \leq 2$. So, x cannot be 4 and thus x is 2. □

(Information on the generalizability of Theorem 1.3 is discussed in Appendix A.2.)

1.2 Prime Numbers

> *317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but because it is, because mathematical reality is built that way.*
>
> G.H. Hardy (1940)

Definition 1 (Prime Number)**.** A *prime number* (in $\mathbb{Z}$) is a number that is only divisible by 1 and itself.[2] (Importantly this definition excludes 1.)

Prime numbers have fascinated mathematicians for millennia and continue to do so today.[3] In part because many prime related conjectures are easy to state yet are currently unsolved. For example, the *Golbach Conjecture* which says every positive even integer can be written as the sum of two primes, and the *Twin Prime Conjecture* which says there are infinitely many primes p such that $p + 2$ is also prime.

Here is a Theorem that I self-discovered while trying to go to sleep one night. (I make no claim that I was the first to discover this.)

[2] Note that the notion of a prime can be generalized to fields. So, for example, in the Gaussian integers ($\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$) 5 is not a prime as $5 = (2 + i)(2 - i)$ (Gauss, 1832).

[3] Including: Chebyshev (1852), de la Vallée-Poussin (1896), Euclid (300 B.C.E.), Euler (1744), Fermat (1640), Gauss (1801), Hadamard (1896), and Riemann (1859).

Theorem 1.4. In a base b number system there are infinitely many primes whose last digit is α if and only if α is co-prime with b.

This theorem demonstrates how fundamental primes are — a natural number p is prime no matter how it is expressed. For example, $17_{\text{in base }10} = 11_{\text{in base }16} = 21_{\text{in base }8} = 10001_{\text{in base }2}$ are all prime.

A quick aside: we are about to start talking about numbers in different bases and it feel it necessary to recite a classic number system joke:

> There are 10 types of people in this world: those who understand binary and those who don't!

This joke, in fact, brings up a very fundamental problem with speaking about many different bases — "10" is the 'base' for all base systems *within* that system. That is, if someone grew up counting: 1,2,3,4,5,6,7,8,9,a,10. Then, they would refer to their base system as base 10. So, in this section when when a number is referred to a without referring to what base then the number is in **our** base 10 (i.e., the number of fingers most humans have).

Now we can return to the question at hand. Notice that Theorem 1.4 is true in our normal base 10 system. Given the following numbers: 37417, 65839, 118246, and 295745. You can immediately tell that 118246 is not prime as its last digit is a 6 therefore 2 divides it. Likewise you can tell that 295745 is not prime as 5 divides it. However, it is less easy to tell if 65649 or 65839 are prime.[4] Obviously, if the final digit of a number tells you that it is divisible by something then it cannot be prime but why is it that some digits tell us this and others don't? To ease the proof we first introduce a powerful lemma.

Lemma 1.5 (Dirichlet prime number theorem, 1837)**.** For any two positive co-prime integers a and b, there are infinitely many primes of the form $a + nb$, where n is also a positive integer. In other words, for all $a, b \in \mathbb{Z}_{++}$ such that $\gcd(a, b) = 1$ the sequence,

$$a, a+b,\ a+2b,\ a+3,\ \ldots$$

contains infinitely many primes.

This allows us to, easily, prove Theorem 1.4.

Proof of Theorem 1.4. ($\Longleftarrow$) First we show the only if direction. Fix $b \in \mathbb{N}$, let $\gcd(\alpha, b) = n \neq 1$, and let $\Gamma := \gamma_k \mathbin{+\!\!+} \gamma_{k-1} \mathbin{+\!\!+} \ldots \mathbin{+\!\!+} \gamma_1 \mathbin{+\!\!+} \alpha$ (where $\mathbin{+\!\!+}$ is the concatenation operator). We want to convert Γ into our base 10.

$$\Gamma = \gamma_k b^k + \gamma_{k-1} b^{k-1} + \ldots + \gamma_1 b + \alpha,$$
$$\Gamma = n \cdot \underbrace{\left(\frac{\gamma_k b^k}{n} + \frac{\gamma_{k-1} b^{k-1}}{n} + \ldots + \frac{\gamma_1 b}{n} \right)}_{\substack{\Cap \\ \mathbb{N}}}.$$

Thus, n divides Γ, so, Γ is not prime.

[4]For those interested: $37417 = 17 \cdot 13 \cdot 71$ and 65839 is prime.

($\Longleftarrow$) We now prove the if direction. This follows almost immediately from Lemma 1.5. We know the $\gcd(\alpha, b) = 1$ therefore by Lemma 1.5 the sequence $\{\alpha + nb\}_{n\in\mathbb{N}}$ contains infinitely many primes. Every number in this sequence ends in α in base b. Thus, by Lemma 1.5 there are infinitely many numbers in base b ending in α. □

Theorem 1.6 (Divergence series of the reciprocals of the primes)**.** The sum of the reciprocals of the prime numbers diverges. That is, $\sum_{p \text{ prime}} \frac{1}{p} = \infty$.

The original proof of this theorem is due to Euler (1737). However, the proof below is due to Erdös.

Proof. Let p_i denote the i^{th} prime number. Assume that the sum of the reciprocals of the primes converge. Then there exists a smallest positive integer k such that,

$$\sum_{i=k+1}^{\infty} \frac{1}{p_1} < \frac{1}{2}. \tag{1}$$

For any positive integer x let M_x denote the set of those n in $\{1, 2, \ldots, x\}$ which are not divisible by any prime greater than p_k.

Every n in M_x can be written as $n = m^2 r$ where r is square-free (i.e., the prime factorization of r contains no squares). As there are at most k primes can be part of the prime factorization of r there are at most 2^k possible values of r. Furthermore, there are at most $\sqrt{x}$ possible values for m. Thus, we have an upper bound for $|M_x|$,

$$|M_x| \leq 2^k \sqrt{2}. \tag{2}$$

The remaining $x - |M_x|$ numbers in $\{1, 2, \ldots, x\} \setminus M_x$ are all divisible by a prime greater than p_k. Let $N_{i,x}$ be the set of $n \in \{1, 2, \ldots, x\}$ which are divisible by the i^{th} prime p_i. Then,

$$\{1, 2, \ldots, x\} \setminus M_x = \bigcup_{i=k+1}^{\infty} N_{i,x}.$$

As the number of integers in $N_{i,x}$ is at most x/p_i we get,

$$x - |M_x| \leq \sum_{i=k+1}^{\infty} |N_{i,x}| < \sum_{i=k+1}^{\infty} \frac{x}{p_i}.$$

Then, by Equation (1) we see that

$$\frac{x}{2} < |M_x|. \tag{3}$$

Thus, we have a contradiction because when $x \geq 2^{2k+2}$ the estimates given by equations (2) and (3) cannot both hold since $x/2 \geq 2^k \sqrt{x}$. □

1.3 Pi

The number π is, perhaps, the first irrational number that most students are faced with. Many 7^{th} and 8^{th} graders in the US are taught that the circumference and radius of a circle are $2\pi r$ and πr^2 (respectively). As this is the first interaction that most of us have with π we are often taught that π is the ratio of a circle's circumference C to its diameter d. However, this is not how π is typically defined in mathematics today.[5] Today there are a few common definitions of π (Rudin, 1976).

Definition 2 (π)**.** The number π is:

- the smallest positive zero of the sine function;
- the smallest positive difference between two zeros of $\cos x, \sin x$, and $\tan x$;
- half the fundamental period of each non-zero solution of the differential equation $f'' + f = 0$.

There are *many* more equivalent definitions of π. In fact, π appears in areas of math that seem to have nothing to do with geometry! Consider the following elegant (albeit slowly-converging) formula for π.

Theorem 1.7 (Madhava–Leibniz Series)**.**

$$\frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}.$$

Proof.

$$\begin{aligned}
\frac{\pi}{4} &= \arctan(1), \\
&= \int_0^1 \frac{dx}{1+x^2}, \\
&= \int_0^1 \left(\sum_{k=0}^{n} (-1)^k \cdot x^{2k} + \frac{(-1)^{n+1} \cdot x^{2n+2}}{1+x^2} \right), \\
&= \left(\sum_{k=0}^{n} \frac{(-1)^k}{2k+1} \right) + (-1)^{n+1} \left(\int_0^1 \frac{x^{2n+2}}{1+x^2} dx \right).
\end{aligned}$$

Therefore,

$$0 \le \int_0^1 \frac{x^{2n+2}}{1+x^2}\, dx \le \int_0^1 x^{2n+2}\, dx = \frac{1}{2n+3},$$

and the left hand side goes to zero as n goes to infinity this means,

$$\frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}.$$

□

As with many of the (supposedly) non-geometric areas where π appears when you look under the hood there is something to do with a circle. However, the unusual places π appears are more numerous than just Theorem 1.7. For example, in the next theorem we find π in coprimality. The theorem was originally proposed by Ernesto Cesàro in 1881 and subsequently solved by him in 1883. For a rigorous proof, see Hardy and Wright (1979, Theorem 332). (Note that the proof of Theorem 1.8 hand-waves a lot of stuff, for example, there is no uniform distribution over the integers.)

[5] This is because defining the length of a curved line requires rather involved math and (more importantly) because this definition fails in some non-Euclidean geometries.

Theorem 1.8. The probability of two integers a and b being coprime is $6/\pi^2$.

Proof. We wish to prove, $\mathbb{P}[\gcd(a,b) = 1] = 6/\pi^2$. We can see that for a given prime p,

$$\begin{aligned} \mathbb{P}[p \mid a] = \mathbb{P}[p \mid b] = \frac{1}{p}, \\ \implies \mathbb{P}[p \mid a,b] = \frac{1}{p^2}, \\ \implies \mathbb{P}[p \nmid a,b] = 1 - \frac{1}{p^2}. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{P}[\gcd(a,b) = 1] &= \prod_{p \text{ prime}} \mathbb{P}[p \nmid a,b], \\ &= \prod_{p \text{ prime}} 1 - \frac{1}{p^2}, \\ &= \prod_{p \text{ prime}} \frac{1}{\frac{1}{1-1/p^2}}, \\ &= \frac{1}{\left(1 + \frac{1}{2^2} + \frac{1}{(2^2)^3} + \cdots\right)\left(1 + \frac{1}{3^2} + \frac{1}{(3^2)^2} + \frac{1}{(2^2)^3} + \cdots\right)\cdots}, \\ &= \frac{1}{\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots}, \\ &= \frac{1}{\pi^2/6}, \\ &= 6/\pi^2. \end{aligned}$$

Where the penultimate equality is due to the famous Basel Problem (Euler, 1740). □

Perhaps even more shocking than Theorem 1.8 is *Buffon's Needle Problem* and its relationship with π. The problem was first posed:

> *Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. What is the probability that the needle will lie across a line between two strips?*
>
> Georges-Louis Leclerc, Comte de Buffon (1733)

Theorem 1.9 (Buffon's Needle Problem)**.** If a needle of length ℓ is dropped on paper with equally spaced lines of distance $d \geq \ell$ then the probability that the needle comes to lie in a position where it crosses one of the lines is $p = 2\ell/\pi d$.

Proof (sans calculus). If you drop a needle (of any length) the expected number of crossings is,

$$E = \sum_{k=1}^{\infty} k p_k,$$

where p_n is the probability the needle will have *exactly* n crossings. Then the probability of at least one crossing is,

$$p = \sum_{k=1}^{\infty} p_k.$$

Thus, by our restriction that $d \geq \ell$ we have $E = p$. Consider a needle of length $\ell = x + y$ where x is the "front part" of the needle and y is the "back part" of the needle. Then, the expectation that the needle will cross is,

$$\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y].$$

Then, by induction on n this implies that $\mathbb{E}[nx] = n\mathbb{E}[x]$ for all $n \in \mathbb{N}$. Therefore, $mE\left[\frac{n}{m}x\right] = E\left[m\frac{n}{m}x\right] = \mathbb{E}[nx] = n\mathbb{E}[x]$. So, $\mathbb{E}[rx] = r\mathbb{E}[x]$ for all $r \in \mathbb{Q}$. Moreover, E is clearly monotone in x so $\mathbb{E}[x] = cx$ for some $c = \mathbb{E}[1]$.

Now we endeavor to find the constant c. Suppose we drop a polygonal needle of total length ℓ. Then the number of crossings will be the sum of the numbers of crossings produced by its straight pieces. Hence, the number of crossings is again, $E = c\ell$ by linearity of expectation.

Now, imagine a needle that is a perfect circle C with diameter d and length $x = d\pi$. (This needle will always produce exactly two intersections.) Furthermore, suppose that we inscribe a polygonal needle P_n (the word needle is being used very loosely) inside C and circumscribe a polygonal needle P^n around C. Then, the expected number of intersections must satisfy,

$$\mathbb{E}[P_n] \leq \mathbb{E}[C] \leq \mathbb{E}[P^n].$$

Since both P's are polygons the number of crossings we expect is c times length. While for C there will be exactly two crossings. Thus,

$$c\ell(P_n) \leq 2 \leq c\ell(P^n).$$

Both P_n and P^n approximate C and n goes to infinity. In particular,

$$\lim_{n\to\infty} \ell(P_n) = d\pi = \lim_{n\to\infty} \ell(P^n).$$

So, in the limit we can see that

$$cd\pi \leq 2 \leq cd\pi.$$

So, $c = \dfrac{2}{\pi} \cdot \dfrac{1}{d}$. □

Proof (avec calculus). Suppose the dropped needle has an angle of α away from horizontal. Then, $0 \leq \alpha \leq \frac{\pi}{2}$. This needle has height $\ell \sin(\alpha)$. Clearly, the probability such a needle crosses one of the horizontal lines of distance d is $\ell \sin(\alpha)/d$. Thus, the probability comes from averaging over the possible angles α, as

$$\begin{aligned} p &= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{\ell \sin(\alpha)}{d} d\alpha, \\ &= \frac{2}{\pi} \cdot \frac{\ell}{d} \Big[-\cos(\alpha)\Big]_0^{\pi/2}, \\ &= \frac{2}{\pi} \cdot \frac{\ell}{d}. \end{aligned}$$

□

One of the most interesting places that π appear is in the "Bouncing Balls" problem. For those interested here are a video, a Stack Exchange post, and a link to the original paper.

1.4 The Nearly Perfect Prediction Theorem

This section is inspired by a short paper (with the same name as this subsection) by Joel Hamkins (2025). However, the 'actual' paper is *A Peculiar Connection Between the Axiom of Choice and Predicting the Future* (Hardin & Taylor, 2008).

The goal of this theorem is to guess the value of some function $f : \mathbb{R} \to \mathbb{R}$ (not necessarily continuous) based on the previous values of f. To do this we define a "prediction function" $P : \mathbb{R}^{<\mathbb{R}} \to \mathbb{R}$ on the set of all possible histories $f \upharpoonright t$ for real number t. Then $P(f \upharpoonright t)$ is in $\mathbb{R}$. The theorem says that there exists a strategy P such that P is correct almost everywhere. Or, more specifically, the predictions will be correct everywhere except a countable, nowhere dense set of exceptional points. Perhaps even more surprising: for every number s in $\mathbb{R}$ there will be an interval $(s, s + \varepsilon)$ for which the predictions are correct *and* we can find ε that are arbitrarily large.

To find such a prediction strategy P, fix a well-ordering $\trianglelefteq$ on the set of all functions $f : \mathbb{R} \to \mathbb{R}$ and let P select the $\trianglelefteq$-least function agreeing with the observed data.[6]

Let $f : \mathbb{R} \to \mathbb{R}$ be the function we would like to predict. As time passes there are less functions that P can guess (the class of function P is selecting from is constantly decreasing). So, $s < t$ implies $f^*_s \trianglelefteq f^*_t$, where these are the $\trianglelefteq$-least functions in agreement with the data at those times. Thus, the prediction function climbs higher in the hierarchy $f^*_s \trianglelefteq f^*_t$ at every exceptional point. As the hierarchy is well ordered the exceptional points must be well ordered in the reals.

To predict immediate future values of the function extend the idea of a prediction strategy P to a continuation prediction strategy $F : \mathbb{R}^{<\mathbb{R}} \to \mathbb{R}$ that maps each history $f \upharpoonright t$ to its predicted continuation $F(f \upharpoonright t) = f^*$. Then for almost all $t \in \mathbb{R}$: there exists an $\varepsilon > 0$ such that for every $x < t+\varepsilon$ is is such that $f(x) = F(f \upharpoonright t)(x)$. Once again the continuation prediction's exceptional points will be well ordered in the real numbers. Thus, they once again form a countable, nowhere dense set. So the continuation predictions are almost always correct. Likewise, there will be intervals upon which the prediction is correct everywhere. Moreover, there will be an absolutely least exceptional point r. Thus, the predicted continuations will be correct for the entire interval $(-\infty, r)$. For a more mathematical and in-depth statement and proof of this theorem see Appendix A.3.

This theorem uses the Axiom of Choice (via the Well-Ordering Principle),[7] and therefore, is non-constructive. So, we won't be able to predict the future any time soon! Quoting the final line of Hamkins' paper

> *I am confident that the iridescent mathematical stock traders of the platonic realm are truly making bank.*

[6] That is, given history $f \upharpoonright t$ let $P(f \upharpoonright t) = f^*(t)$ where f^* is the $\trianglelefteq$-least function agreeing with the past data. So, $f^* \upharpoonright t = f \upharpoonright t$.

[7] The Axiom of Choice says that every indexed family $(S_i)_{i\in I}$ of non-empty sets, there exists an indexed set $(x_i)_{i\in I}$ such that $x_i \in S_i$ for every $i \in I$ (Zermelo, 1904).

1.5 The Toilet Paper Problem

This section is dedicated to "The Toilet Paper Problem" which is a mathematical problem published (and solved) in The American Mathematical Monthly by Donald Knuth in 1984. The original version of the paper included section titles: An absorbing barrier, A process of elimination, and Residues (Knuth et al., 1989).

The problem, in non-mathematical terms, is as follows. The toilet paper dispensers in a certain building is designed to hold two rolls and a person can use either roll. There are two kinds of people who use the rest rooms in the building: *big-choosers* and *little-choosers*. A big-chooser always takes a piece of toilet paper from the roll that is currently larger; a little-chooser always does the opposite. However, when the two rolls are the same size, or when only one roll is left, everybody chooses the nearest available roll. (When both rolls are empty everybody has a problem!)

Assume that people enter the toilet stalls independently and at random. With probability p they are big-choosers and probability $q = 1-p$ they are little-choosers. If the janitor supplies a particular stall with two fresh rolls of toilet paper, both of length n, let $M_n(p)$ be the average number of portions left on one roll when the other roll first empties. (We assume that everyone uses the same amount of paper, and that the lengths are expressed in terms of this unit.)

Now, commencing with some math, let $M_n(p)$ be the average number of portions left on one roll when the other roll first empties. The problem is to understand the asymptotic value of $M_n(p)$ for a fixed p as $n \to \infty$.

First, we generalize the problem. Let $M_{mn}(p)$ be the mean number of portions left when one roll empties if there are m units on one roll and n on the other. Thus,

$$\begin{aligned}
M_n(p) &= M_{nn}(p); \\
M_{m0}(p) &= m; \\
M_{nn}(p) &= m; \\
M_{nn}(p) &= M_{n(n-1)}(p), \quad \text{if } n > 0; \\
M_{mn}(p) &= pM_{(m-1)n}(p) + qM_{m(n-1)}(p), \quad \text{if } m > n > 0.
\end{aligned}$$

The value of $M_n(p)$ can be computed for all n from these recurrence relations since there are no pairs (m', n') with $m' < n'$.

The following visualization exercise is given by Knuth:

> It is convenient to visualize the recurrence by drawing certain arcs between adjacent lattice points in the plane, where the arc from (m, n) to $(m-1, n)$ has weight p and from (m, n) to $(m, n-1)$ has weight q, for all $0 < n < m$; the arc from (m, n) to $(n, n-1)$ has weight 1 for all $n > 0$; and there are no other arcs. Then $M_{mn}(p)$ is the sum, over all $k \geq 1$, of k times the sum of the weights of all paths from (m, n) to $(k, 0)$, where the weight of a path is the product of the individual arc weights.
>
> A path that starts at the diagonal point (n, n) must go first to $(n, n-1)$; then it either returns to the diagonal at $(n-1, n-1)$ or goes to $(n, n-2)$, etc. Let c_k be the number of paths from (n, n) to $(n-k, n-k)$ whose intermediate points do not touch the diagonal, and let d_{nk} be the

number of paths from $(n, n-1)$ to $(k, 1)$ whose points do not ever touch the diagonal. A path that starts at (n, n) either returns to the diagonal for the first time at some point $(n-k, n-k)$, or never returns to the diagonal at all.

This visualization gives rise to the following equality,

$$\begin{aligned} M_n(p) &= c_1 p M_{n-1}(p) + c_2 p^2 q M_{n-2}(p) + \cdots + c_{n-1} p^{n-1} q^{n-2} M_1(p) + L_n(p) \\ &= \sum_{0<k<n} c_k p^k q^{k-1} M_{n-k}(p) + L_n(p), \end{aligned}$$

where

$$L_n(p) = \sum_{2 \le k \le n} k d_{nk} p^{n-k} q^{n-1} \qquad \text{for } n \ge 2, \text{ and } \quad L_1(p) = 1.$$

The c_k's are the Catalan numbers and the d_{nk}'s are the numbers that arise in classical ballot problems (see Appendix A.4).

We can discover the required values by observing that d_{nk} is the number of decreasing paths from $(n, n-1)$ to $(k, 1)$ minus the number of decreasing paths from $(n, n-1)$ to $(1, k)$, where a "decreasing path" is any path that decreases either the left component or the right component by unity at each step. This follows because there is a one-to-one correspondence between all decreasing paths from $(n, n-1)$ to $(k, 1)$ that do touch the diagonal and all decreasing paths from $(n, n-1)$ to $(1, k)$; the idea is to reflect the path about the diagonal, starting after the place where it first touches a diagonal point. Since the number of decreasing paths from (a, b) to (c, d) is

$$\binom{a+b-c-d}{a-c} = \binom{a+b-c-d}{b-d}$$

for all $a \ge c$ and $b \ge d$, we have

$$\begin{aligned} d_{nk} &= \binom{2n-k-2}{n-2} - \binom{2n-k-2}{n-1} \\ &= \binom{2n-k-2}{n-2} \cdot \frac{k-1}{n-1}. \end{aligned}$$

Furthermore, $c_{n-1} = d_{n2}$, hence

$$c_n = \binom{2n-2}{n-1} \cdot \frac{1}{n}.$$

Now, letting,

$$M(z) = \sum_{n \ge 1} M_n(p) z^n, \quad \text{and} \quad L(z) = \sum_{n \ge 1} L_n(p) z^n,$$

the recurrence relation for $M_n(p)$ is equivalent to,

$$M(z) - L(z) = q^{-1}\, C(pqz)\, M(z).$$

Then, we can see, with some derivations, that

$$L(z) = z\sum_{k\geq 0} kp^{1-k}\, C(pqz)^{k-1} = \frac{p^2 z}{\big(p = C(pqz)\big)^2}.$$

This allows us to remove the $L(z)$ term from our expression for $M(z)$ and obtain,

$$M(z) = \frac{z}{(1-z)^2}\left(\frac{q - C(pqz)}{q}\right).$$

This, finally, gives us the formula for which we were searching,

$$M_n(p) = n - (n-1)c_1 p - (n-1)c_2 p^2 q - \cdots - 1c_{n-1}p^{n-1}p^{n-1}.$$

There is much more to be said about this problem and for those interested see: The Toilet Paper Problem. We conclude with yet another quote from Knuth, this time from the acknowledgments of this paper:

> *I wish to thank the architect of the computer science building at Stanford University for implicitly suggesting this problem.*

1.6 Putnam Problems

The Putnam Competition is an annual mathematics competition for undergraduates that has a number of interesting problems. This section show some of these questions (with solutions) from throughout the years. For a more detailed catalog of these types of questions see here.

Putnam Problem (B7, 1940)**.** For $n > 8$ let $a = \sqrt{n}$ and $b = \sqrt{n+1}$. Which is greater: a^b or b^a?

(Source: Gleason et al., 1980.)

Solution. Since $a^b = e^{b\ln(a)}$ and $b^a = e^{a\ln(b)}$ the question boils down to,

$$b\ln(a) \overset{?}{\lesseqgtr} a\ln(b).$$

Which is the same as comparing $\ln(a)/a$ and $\ln(b)/b$. Let $f(x) = \ln(x)/x$. Then $f'(x) = 1/x^2 - \ln x/x^2 < 0$ for all $x > e$. Clearly $b > a$ so as $a > e$ (since $e^2 \ll 9$) we have $\ln(a)/a > \ln(b)/b$, therefore, $b\ln(a) > a\ln(b)$ and $a^b > b^a$. □

Putnam Problem (A1, 1985)**.** How many triples (A_1, A_2, A_3) of sets exist with the properties:

(i) $A_1 \cup A_2 \cup A_3 = \{1,2,3,4,5,6,7,8,9,10\}$, and

(ii) $A_1 \cap A_2 \cap A_3 = \varnothing$.

(Source: Kedlaya et al., 2002.)

Solution. This problem is much easier than it may appear. The answer is 60466176. Consider some number $n \in \{1,2,3,4,5,6,7,8,9,10\}$. It must be that $n \in A_1 \cup A_2 \cup A_3$ and $n \notin A_1 \cap A_2 \cap A_3$. So, we can place n in just A_1, just A_2, or just A_3. Additionally, n can be in both A_1 and A_2, A_1 and A_3, or A_2 and A_3. Therefore, there are six possible "places" (speaking loosely) that n can go. Meaning there are $6^{10} = 60466176$ possible triples. □

Putnam Problem (A3, 1998)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f \in \mathcal{C}^3(\mathbb{R})$. Prove that there exists an $a \in \mathbb{R}$ such that $f(a) \cdot f'(a) \cdot f''(a) \cdot f'''(a) \geq 0$.

(Source: Kedlaya et al., 2002.)

Solution. As $f \in \mathcal{C}^3(\mathbb{R})$ the intermediate value theorem implies that $f^n(x)$ must be strictly above or below 0 for $n \in \{0, 1, 2, 3\}$. We can, without loss of generality,[8] assume that $f(x), f'(x) > 0$. Then there are two cases in which $f + f' + f'' + f''' < 0$:

Case 1: $f''(x) > 0$ and $f'''(x) < 0$.

Since $f'''(x) < 0$, the graph of $f'(x)$ is concave down. Thus, the graph of $f'(x)$ lies below its tangent line at $x = 0$. Therefore,

$$f'(x) \leq f'(0) + f''(0) \cdot x \quad \text{and} \quad f'\left(\frac{-f'(0)}{f''(0)}\right) \leq 0 \qquad \lightning.$$

Case 2: $f''(x) < 0$ and $f'''(x) > 0$.

Since $f''(x) < 0$, the graph of $f(x)$ is concave down and hence the graph of $f(x)$ lies below its tangent line at $x = 0$. Thus,

$$f(x) \leq f(0) + f'(0) \cdot x \quad \text{and} \quad f\left(\frac{-f(0)}{f'(0)}\right) \leq 0 \qquad \lightning.$$

Therefore, if must be that either $f''(x), f'''(x) < 0$ or $f''(x), f'''(x) > 0$. □

Putnam Problem (B5, 2006)**.** Let $f : [0, 1] \to \mathbb{R}$ be continuous. Evaluate

$$\max_f \left\{ \int_0^1 \left(x^2 f(x) - x f(x)^2 \right) dx \right\}. \tag{4}$$

(Source: Kedlaya et al., 2002.)

Solution. First, note that

$$x^2 f(x) - x f(x)^2 = -x\big(f(x)^2 - x f(x)\big).$$

Next, factor the quadratic term:

$$f(x)^2 - x f(x) = \left(f(x) - \frac{x}{2}\right)^2 - \frac{x^2}{4}.$$

Thus, the objective function in (4) can be rewritten as

$$\frac{x^3}{4} - x\left(f(x) - \frac{x}{2}\right)^2.$$

It follows that

$$\begin{aligned} \int_0^1 \left(x^2 f(x) - x f(x)^2\right) dx &= \int_0^1 \left[\frac{x^3}{4} - \overbrace{x}^{\geq 0} \underbrace{\left(f(x) - \frac{x}{2}\right)^2}_{\geq 0} \right] dx \\ &\leq \int_0^1 \frac{x^3}{4}\, dx \\ &= 16. \end{aligned}$$

[8] One can replace $f(x)$ with $-f(x)$ if necessary and assume $f(x) > 0$. Likewise, if necessary, by replacing $f(x)$ with $f(-x)$ we can assume $f'(x) > 0$.

Therefore, the value of (4) is 16, and the function that achieves this value is

$$f(x) = \frac{x}{2}.$$

□

Putnam Problem (B3, 2025)**.** Let S be a non-empty subset of $\mathbb{Z}$ with the property:

$$n \in S \implies m \in S \quad \forall m | 2025^n - 15^n.$$

Does $S = \mathbb{Z}$?

(Source: Ullman and Zeitz, 2025.)

Solution. Yes, $S = \mathbb{Z}$. Proceed by strong induction. First note that $1 \in S$ as $S \neq \varnothing$ and $1 | 2025^n - 15^n$ for all n. Now suppose $n > 1$ and $1, \ldots, n-1 \in S$.

Let $n = 3^a 5^b d$ with $a, b \geq 0$ and $\gcd(15, d) = 1$. Define $c := \max(a, b)$. Since d and 135 are coprime we know $135^k \equiv 1 \pmod{d}$ for any k divisible by $\varphi(d)$.[9]

Since $(3^c - 1) \cdot d \geq cd \geq c$, we have

$$\varphi(d) \leq d \leq 3^c d - c \leq n - c,$$

so, $c \leq n\varphi(d)$. Thus, for $k := \varphi(d) \lfloor n/\varphi(d) \rfloor > n\varphi(d) \geq c$, n divides $15^k(135^k - 1) = 2025 - 15^k$. Therefore, $n \in S$. □

1.7 Fun Problems

This section houses the easiest problems of this document arranged in order of difficult. Questions that are particularly hard are marked **[HARD]**. (The full set of solutions are given in Appendix A.5.)

Problem 1. *For what value of n is $\sqrt[n]{n}$ maximized?*

(Solution)

Problem 2. *True of false: for every real number x*

$$e^x + e^{-x} \geq 2?$$

(Prove or disprove the statement.)

(Solution)

Problem 3. *For what (non-zero) function f is $f = \sum_{n=1}^{\infty} f^{(n)}$? (Here $f^{(n)}$ denotes the n^{th} derivative of f.)*

(Solution)

Problem 4. *Solve the integral,*

$$\int_0^1 \left\lfloor \frac{1}{x} \right\rfloor^{-1} dx.$$

[9] Where $\varphi(n)$ is Euler's totient function: $\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$.

(Solution)

Problem 5. [HARD]
Solve the integral,

$$\int_0^\infty \frac{\pi(x)}{x^3 - x}\,dx,$$

Where $\pi(x)$ is the number of primes no greater than x.

(Solution)

2 Philosophy

2.1 Raven Paradox

Figure 1: Rare albino raven. Image courtesy of the National Audubon Society.

The "Raven Paradox" (Hempel, 1945) (o "The Paradox of Indoor Ornithology") says, *very loosely*, that using inductive logic one can provide evidence for the claim "all ravens are black" by observing non-black things that are *not* ravens.[10]

The statement

$$\text{All ravens are black.} \tag{H}$$

can be stated through formal implication as:

$$\text{If something is a raven \textbf{then} it is black.} \tag{H'}$$

Via contraposition[11] H' is equivalent to,

$$\text{\textbf{If} something is \textbf{not} black \textbf{then} it is \textbf{not} a raven.} \tag{C}$$

As H' is true if and only if C is true (any counterfactual world where H' is true C is true, and vice versa).

Clearly, the proposition

$$\text{My friends pet raven is black.} \tag{B}$$

is evidence for the claim H. Then, the proposition

$$\text{The avocado I ate for breakfast this morning was green and not a raven.} \tag{¬B}$$

is evidence for claim C which is equivalent to H' which is the same as H. Thus, while eating breakfast this morning I generate evidence for the ornithological hypothesis of H. Therefore, I can do ornithology without leaving my house (or ever seeing a raven)! (One possible explanation of the paradox is given in Appendix B.1.)

[10] *Inductive logic* (or inductive reasoning) is a method of reasoning in which general conclusions are drawn from specific observations or evidence. Rather than proving statements deductively from first principles, one evaluates how the available evidence increases or decreases the probability that a conclusion is true.

[11] In logic the Law of Contraposition states that $P \to Q$ (if P then Q) is true if and only if $\neg Q \to \neg P$ (not Q then not P) is true.

2.2 The Problem of Grue

In *Fact, Fiction, and Forecast* (Goodman, 1983) a very peculiar "riddle" (the new riddle of induction) is introduced. The riddle continues the work of Hume (1739, 1748) outlining problems of enumerative induction.

Say that an object is *grue* if it is green if observed before time t and blue otherwise. Now, suppose we observe a large number of emeralds to be green before some time t. Then, we can state "all observed emeralds are green." Using an enumerative induction schema[12] we can infer that all emeralds are green. This inference would lead us the believe that if we were to observe an emerald after some time t it would be green. However, it is equally accurate to say that we observed a large number of emeralds to be grue and, thus, that "all observed emeralds are grue." Therefore, we can draw the equally valid inference that all emeralds are grue. This leads to the converse belief that an emerald observed after time t will be blue![13]

Figure 2: Many grue (or are they green?) emeralds. Image courtesy of My Irish Jeweler.

2.3 Epistemic Paradoxes

This section discusses a few epistemic paradoxes. For a *much* more in-depth analysis and discussion on these sort of paradoxes see the SEP entry (Sorensen, 2024).

Epistemology concerns itself with the knowledge. Thus, in what follows we see paradoxes that relate to knowledge. Perhaps the oldest such paradox:

> *You do not believe this sentence.*
>
> John Buridan (c. 1340; trans. 1982)

Clearly, believing the sentence will result in you not believing it and vice versa.

[12]"Enumerative induction is defined as a process where confirming instances increase the probability of other potential instances, indicating that a sufficiently long sequence of confirming cases leads to a high likelihood of subsequent instances being true." (Zabell, 2011)

[13]There are many suggested solutions to this riddle — for example, see Jackson (2019), Norton (2003), Okasha (2001, 2005a, 2005b), and Sober (1988).

For a more in-depth paradox consider a class that meets Monday, Wednesday, and Friday. The teacher declares "there will be a surprise test this week" (this is sometimes called the unexpected hanging paradox but this example is less morbid). A clever student reasons that if it gets to Thursday then she will know that the exam is Friday. However, the test is supposed to be a surprise. Thus, if cannot happen on Friday. If on Tuesday there hasn't been a quiz then she knows it *must* be on Wednesday as a Friday exam would not be a surprise. Hence, the exam cannot be on Wednesday. This leaves only Monday but a similar line of reasoning says that there cannot be an exam on Monday either. So, the student concludes there won't be an exam!

There are many ways that people seek to resolve this paradox, including rejecting the law of bivalence, the KK principle, or the closure principle (Immerman, 2017).[14] Another way that the paradox is attacked is by claiming that the teacher's statement was forced to be false similar to the statement "I am not speaking right now." O'Connor (1948) says the teachers statement "could not conceivably be true in any circumstances." These are classed by Cohen (1950) as Pragmatic paradoxes (i.e., statements falsified by their own utterances). However, the idea that the teachers statement completely falsifies itself does not seem completely true. (If I was a student in this class I would certainly feel that I learned something by hearing her proclamation!)

Weiss (1952) claims that the students argument is flawed as it assumes the teacher is being truthful. He suggests the truth or falsity of the teachers statement can only be known *after* the week as passed. However, this retort is not without its flaws. Quine (1953) dislikes this argument as it rejects bivalence. He instead suggests distinguishing between what the student can know and what is in fact true. He agrees that the student's elimination argument shows the student cannot know in advance that the teacher's announcement is true, but he denies that this undermines the truth of the announcement itself. On Quine's view, the student's ignorance is enough to preserve the surprise: if the student cannot know when the test will occur, then a test on any day would still be unforeseen.

Quine's argument seems counter to intuition though as we would think that the announcement provides some information to the students. His argument also rests upon Hume and Goodman-esque views on the future and our lack of ability to predict it (see Section 2.2).

Quine, in his later work, argues that speaking of "knowing" is too imprecise (e.g., Quine, 1969, 1974, 1990). If the requirement for us to know something is absolute certainty then we can no very little (or perhaps nothing). If we allow "knowing" to include things we are almost sure of then the cutoff becomes arbitrary. He says that in the same way that scientists do not speak of how "big" a thing is but instead speak of the measurement of a thing philosophers should abandon concepts like "knowledge" or "justification." Instead they should look at precise components of belief such as truth and degrees of firmness of belief.[15]

[14]The law of bivalence says that every declarative sentence expressing a proposition has exactly one truth value, either true or false. The KK principle, in essence, says that if one knows p then they know that they know p. The closure principle says that if someone knows p and knows that p entails q then they know q. (We can see that rejection of any of these principles is quite a large concession to make to ameliorate a paradox.)

[15]Following the analogy of science and "bigness" this is akin to thinking of the mass, height, volumes, etc. of an object.

There is much more discussion of this paradox (for more recent literature see Murzi et al., 2021; Williams, 2007, or see Dietzfelbinger, 2026; Earman, 2021 for reviews of the subject).

We now proceed by very briefly presenting one final paradox. Consider the following argument that shows there are unknowable truths. Assume there is a true sentence of the form p but p is not known. Although this sentence is consistent, modest principles of epistemic logic imply that sentences of this form are unknowable:

1.	$K(p \wedge \neg Kp)$	(Assumption)
2.	$Kp \wedge K\neg Kp$	1, Knowledge distributes over conjunction
3.	$\neg Kp$	2, Knowledge implies truth (from the second conjunct)
4.	$Kp \wedge \neg Kp$	2, 3 by conjunction elimination and conjunction introduction
5.	$\neg K(p \wedge \neg Kp)$	1, 4 *Reductio ad absurdum*

This implies that $p \wedge \neg Kp$ is unknowable. The foregoing argument is sufficient to refute even claims as weak as "empirical propositions are knowable." (Stephenson, 2015)

2.4 The Hardest Logic Puzzle Ever

The following problem is known as "The Hardest Logic Puzzle Ever" and was originally published in The Harvard Review of Philosophy (Boolos, 1996). The problem's core set up is credited to Raymond Smullyan and John McCarthy. The question is:

> Three gods A, B, and C are called, in no particular order, True, False, and Random. True always speaks truly, False always speaks falsely, but whether Random speaks truly or falsely is a completely random matter. Your task is to determine the identities of A, B, and C by asking three yes–no questions; each question must be put to exactly one god. The gods understand English, but will answer all questions in their own language, in which the words for yes and no are *da* and *ja*, in some order. You do not know which word means which.

You may ask more than one question to one of the gods; the second and third questions you choose to ask may be functions of the question(s) preceding them and Random should be thought of as flipping a fair coin in her head that determines if she answers falsely or truly.

I will preceded with the solution in the same manner as Boolos (1996). First we consider three simpler puzzles:

(I) Noting their locations, I place two aces and a jack face down on a table, in a row; you do not see which card is placed where. Your problem is to point to one of the three cards and then ask me a single yes/no question, from the answer to which you can, with certainty, identify one of the three cards as an ace. If you have pointed to one of the aces, I will answer your question truthfully. However, if you have pointed to the jack, I will answer your question yes or no, completely at random.

(II) Suppose that, somehow, you have learned that you are speaking not to Random but to True or False — you don't know which — and that whichever god you're talking to has condescended to answer you in English. For some reason, you need to know whether Dushanbe is in Kirghizia or not. What one yes/no question can you ask the god from the answer to which you can determine whether or not Dushanbe is in Kirghizia?

(III) You are now quite definitely talking to True, but he refuses to answer you in English and will only say *da* or ja. What one yes/no question can you ask True to determine whether or not Dushanbe is in Kirghizia?

The solutions to each of these problems is given below, however, the reasoning is relegated to Appendix B.2.

(I) Point to the middle card and ask "Is the left card an ace?" If the answer is yes, choose the left card. If the answer is no, choose the right card.

(II) Ask the god "Are you True iff[16] Dushanbe is in Kirghizia?" Given this question you can know with certainty if Dushanbe is in Kirghizia.

(III) Ask True, "Does *da* mean yes iff Dushanbe is in Kirghizia?"

Now we return to The Hardest Logic Puzzle Ever. First you need to find a god who you know is not Random.

1. Ask A: "Does *da* mean yes iff you are True iff B is Random?"

If you get the response *da* then C is either True of False and if you get the answer *ja* then B is either True or false (again, further reasoning will be provided in Appendix B.2). Next ask a question of either B or C whoever you determined is not Random. For ease of exposition suppose it is B.

2. Ask B "Does *da* mean yes iff Rome is in Italy?"

True will answer *da* and false will answer ja. Thus, we now know if B is True or False. The final question will be, once again, to B.

3. Ask B: "Does *da* mean yes iff A is Random?"

If B is True and the their response is *da* then: A is Random, B is True, and C is False. If B is True and answers *ja* then A is false, B is True, and C is Random. If B is False and their response is *da* then A is True, B is False, and C is Random. If B is False and answers *ja* then A is Random, B is False, and C is True.

This puzzle's solution has a quite direct connection to the proof of Theorem 1.2 because both require the Law of Excluded Middle. The Law of Excluded Middle states that for any proposition X either X is true or X is not true (but not both). This is implicitly used when we write proofs by contradiction as we did in the proof of Theorem 1.2 but it is also used here. Without this assumption we would not be able to solve this problem, nor would we be able to reason about many

[16] There will be a lot of use of the biconditional "if and only if" (i.e., $\iff$) so it will be shorten it to the abbreviation "iff."

things in our day-to-day lives. However, Intuitionists mathematicians deny the law of excluded middle. (They can't do proof by contradiction!)

A weaker law is the Law of Non-contradiction which states both X and not X cannot both be true. It is more widely accepted (although there are truth systems where it is not accepted, see: Catuṣkoṭi). This slight digression serves two puroposes. First learning about Catuṣkoṭi lead me to read *An Introduction to Non-Classical Logic* (Priest, 2008) which is a truly wonderful book that I highly recommend to anyone who is reading this, and secondly because it gives cause to quote Ibn Sīnā (Avicenna).

> *Anyone who denies the law of non-contradiction should be beaten and burned until he admits that to be beaten is not the same as not to be beaten, and to be burned is not the same as not to be burned.*
>
> Ibn Sīnā (c. 1020; trans. 2005)

References

Avicenna. (1000). *Al-shifa': Al-ilahiyyat (Metaphysics).*

Avicenna. (2005). *The metaphysics of the healing* (M. E. Marmura, Trans.). Brigham Young University Press.

Barry, C. (2014). Who sharpened occam's razor. *Irish Philosophy.*

Bertrand, J. (1887). Solution d'un problème. *Comptes Rendus de l'Académie des Sciences de Paris*, *105*, 369.

Boolos, G. (1996). The hardest logic puzzle ever. *The Harvard Review of Philosophy*, *6*(1), 62–65.

Buffon, C. d., Georges-Louis Leclerc. (1733). Original sources on buffon's needle problem [Histoire de l'Acad. Roy. des Sciences (1733), pp. 43–45; Histoire naturelle, générale et particulière, Supplément 4 (1777), p. 46].

Buridan, J. (1982). *John buridan on self-reference: Chapter eight of buridan's Sophismata* (G. E. Hughes, Ed. & Trans.). Cambridge University Press.

Cesàro, E. (1881). Question proposée 75. *Mathesis: recueil mathématique à l'usage des écoles spéciales et des établissements d'instruction moyenne*, *1*, 184.

Cesàro, E. (1883). Question 75 (solution). *Mathesis: recueil mathématique à l'usage des écoles spéciales et des établissements d'instruction moyenne*, *3*, 224–225.

Chaitin, G. (1920). *Meta math!: The quest for omega.* Vintage.

Chebyshev, P. (1852). Mémoire sur les nombres premiers. *Journal de Mathématiques Pures et Appliquées*, *17*, 366–390.

Cohen, L. J. (1950). Mr. o'connor's" pragmatic paradoxes". *Mind*, *59*(233), 85–87.

de la Vallée-Poussin, C. J. (1896). Recherches analytiques sur la théorie des nombres premiers. *Annales de la Société Scientifique de Bruxelles*, *20*, 183–256.

Dietzfelbinger, M. (2026). Dismantling the surprise test "paradox". https://doi.org/10.48550/arXiv.2601.19418

Dirichlet, P. L. (1837). Beweis des satzes, dass jede unbegrenzte arithmetische progression, deren erstes glied und differenz ganze zahlen ohne gemeinschaftlichen factor sind, unendlich viele primzahlen enthält. *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, *45*, 81.

Earman, J. (2021). A user's guide to the surprise exam paradoxes. https://philsci-archive.pitt.edu/19303/

Euclid. (300 B.C.E.). *Elements* [Book IX, Proposition 20].

Euler, L. (1737). Variae observationes circa series infinitas [English translation: "Various observations concerning infinite series"]. *Commentarii Academiae Scientiarum Petropolitanae*, *9*, 160–188.

Euler, L. (1740). De summis serierum reciprocarum [Presented in 1735]. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *7*, 123–134.

Euler, L. (1744). Variae observationes circa series infinitas. *Commentarii academiae scientiarum Petropolitanae*, *9*, 160–188.

Fermat, P. d. (1640). *Correspondance de fermat et mersenne* [Introduced Fermat primes].

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis*, *63*(3), 194–199.

Fitelson, B., & Hawthorne, J. (2010). How bayesian confirmation theory handles the paradox of the ravens. In *The place of probability in science: In honor of ellery eells (1953-2006)* (pp. 247–275). Springer.

Gauss, C. F. (1801). *Disquisitiones arithmeticae.* F. Vieweg.

Gauss, C. F. (1832). Theoria residuorum biquadraticorum, commentatio secunda. *Commentationes Societatis Regiae Scientiarum Gottingensis*, *7*.

Gelfond, A. (1934). Sur le septième problème de hilbert. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, (4), 623–634.

Gleason, A. M., Greenwood, R. E., & Kelly, L. M. (1980). *The william lowell putnam mathematical competition: Problems and solutions, 1938–1964*. Mathematical Association of America.

Good, I. J. (1967). The white shoe is a red herring. *The British Journal for the Philosophy of Science*, *17*(4), 322–322.

Goodman, N. (1983). *Fact, fiction, and forecast*. Harvard University Press.

Hadamard, J. (1896). Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, *24*, 199–220.

Hamkins, J. D. (2025). The nearly perfect prediction theorem. *Notices of the American Mathematical Society*, *72*(3), 308–309.

Hardin, C. S., & Taylor, A. D. (2008). A peculiar connection between the axiom of choice and predicting the future. *The American Mathematical Monthly*, *115*(2), 91–96.

Hardy, G. H. (1940). *A mathematician's apology* (1st). Cambridge University Press.

Hardy, G. H., & Wright, E. M. (1979). *An introduction to the theory of numbers*. Oxford university press.

Hempel, C. G. (1945). Studies in the logic of confirmation (i.) *Mind*, *54*(213), 1–26.

Hosiasson-Lindenbaum, J. (1940). On confirmation. *The Journal of Symbolic Logic*, *5*(4), 133–148.

Hume, D. (1739). *A treatise of human nature*. Oxford University Press.

Hume, D. (1748). *An enquiry concerning human understanding*. Oxford University Press.

Immerman, D. (2017). Question closure to solve the surprise test. *Synthese*, *194*(11), 4583–4596.

Jackson, A. (2019). How to solve hume's problem of induction. *Episteme*, *16*, 157–174.

Kedlaya, K. S., Poonen, B., & Vakil, R. (2002). *The william lowell putnam mathematical competition 1985–2000: Problems, solutions, and commentary*. Mathematical Association of America.

Knuth, D. E. (1976). Mathematics and computer science: Coping with finiteness: Advances in our ability to compute are bringing us substantially closer to ultimate limitations. *Science*, *194*(4271), 1235–1242.

Knuth, D. E. (1984). The toilet paper problem. *The American Mathematical Monthly*, *91*(8), 465–470.

Knuth, D. E., Larrabee, T., & Roberts, P. M. (1989). *Mathematical writing*. Cambridge University Press.

Lander, L., & Parkin, T. (1966). Counterexample to euler's conjecture on sums of like powers.

Megill, N., & Wheeler, D. A. (2019). *Metamath: A computer language for mathematical proofs*. Lulu. com.

Murzi, J., Eichhorn, L., & Mayr, P. (2021). Surprise, surprise: Kk is innocent. *Thought: A Journal of Philosophy*, *10*(1), 4–18.

Norton, J. D. (2003). A material theory of induction. *Philosophy of Science*, *70*(4), 647–670.

O'Connor, D. J. (1948). Pragmatic paradoxes. *Mind*, *57*, 358–359.

Okasha, S. (2001). What did hume really show about induction? *The Philosophical Quarterly*, *51*(204), 307–327.

Okasha, S. (2005a). Bayesianism and the traditional problem of induction. *Croatian Journal of Philosophy*, *5*(14), 181–194.

Okasha, S. (2005b). Does hume's argument against induction rest on a quantifier-shift fallacy? *Proceedings of the Aristotelian Society*, *105*, 237–255.

Priest, G. (2008). *An introduction to non-classical logic: From if to is*. Cambridge university press.

Quine, W. V. O. (1953). On a so-called paradox. *Mind*, *62*(245), 65–67.

Quine, W. V. O. (1969). *Ontological relativity and other essays*. Columbia University Press.

Quine, W. V. O. (1974). *Roots of reference*. Open Court.

Quine, W. V. O. (1990). *Pursuit of truth* [Revised edition published in 1992]. Harvard University Press.

Riemann, B. (1859). Über die anzahl der primzahlen unter einer gegebenen grösse. *Monatsberichte der Berliner Akademie*.

Rinard, S. (2014). A new bayesian solution to the paradox of the ravens. *Philosophy of Science*, *81*(1), 81–100.

Rudin, W. (1976). Principles of mathematical analysis. *3rd ed.*

Schaffer, J. (2015). What not to multiply without necessity. *Australasian Journal of Philosophy*, *93*(4), 644–664.

Schneider, T. (1934). Transzendenzuntersuchungen periodischer funktionen. I. Transzendenz von potenzen. *Journal für die Reine und Angewandte Mathematik*, *172*, 65–69.

Sober, E. (1988). *Reconstructing the past: Parsimony, evolution and inference*. MIT Press.

Sorensen, R. (2024). Epistemic paradoxes. In E. N. Zalta & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy* (Fall 2024). https://plato.stanford.edu/archives/fall2024/entries/epistemic-paradoxes/

Stephenson, A. (2015). Kant, the paradox of knowability, and the meaning of experience. *Philosophers' Imprint*, *15*(17), 1–19.

Ullman, D. H., & Zeitz, P. (2025). The eighty-fifth william lowell putnam mathematical competition. *The American Mathematical Monthly*, *132*(8), 723–736. https://doi.org/10.1080/00029890.2025.2532297

Weiss, P. (1952). The prediction paradox. *Mind*, *61*(242), 265–269.

Whitworth, W. A. (1878). Arrangements of m things of one sort and n things of another sort under certain conditions of priority. *Messenger of Mathematics*, *8*, 105–114.

Williams, J. N. (2007). The surprise exam paradox: Disentangling two reductios. *Journal of Philosophical Research*, *32*, 67.

Zabell, S. L. (2011). Carnap and the logic of inductive inference. In *Handbook of the history of logic* (pp. 265–309, Vol. 10). Elsevier.

Zermelo, E. (1904). Beweis, dass jede menge wohlgeordnet werden kann: Aus einem an herrn hilbert gerichteten briefe. *Mathematische Annalen*, *59*(4), 514–516.

A Mathematical Appendix

A.1 Which Number is Rational?

Theorem A.1 (Gelfond-Schneider)**.** For any algebraic numbers[17] a and b such that $a \notin \{0, 1\}$ and $b \notin \mathbb{Q}$ a^b is a transcendental number.[18] (Gelfond, 1934; Schneider, 1934).

Then by Theorem A.1 we know that $\sqrt{2}^{\sqrt{2}}$ is irrational, implying $\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}}$ is rational. In contrast to the proof of Theorem 1.2 the proof of Theorem A.1 is *not* simple.

A.2 Generalizability of Theorem 1.3

Theorem 1.3 has a few nice properties. Namely, that the only number involved is 2. This rasies the natural question is this a feature of the number 2? The square root operator?[19] Or, both? We investigate this question further here. (Most of the things presented here is the work of others that I have collected and/or restated. However, what follows in this section are proofs of my own making.)

Theorem A.2. The infinite tetration,

$$\sqrt{x}^{\sqrt{x}^{\sqrt{x}^{\cdot^{\cdot^{\cdot}}}}} = x, \tag{5}$$

if and only if $x \in \{1, 2\}$.

Proof. Clearly $x = 1$ satisfies Equation (5). However, this is not particularly interesting. Now we consider the general case,

$$\begin{aligned} & y = \sqrt{x}^{\sqrt{x}^{\sqrt{x}^{\cdot^{\cdot^{\cdot}}}}} \\ \implies & y = \sqrt{x}^{y} \\ \implies & y = (x)^{y/2}. \end{aligned}$$

Assuming that the left hand side of Equation (5) exists (if $x = 0$ no solution exists). Then, letting $y = x$,

$$\begin{aligned} x &= x^{x/2} \\ \ln(x) &= \left(\frac{x}{2}\right)\ln(x). \end{aligned}$$

Which is only true for $x \in \{1, 2\}$. ☐

We can generalize this further.

[17]An algebraic number is a number that is a root of a non-zero polynomial in one variable with rational coefficients.

[18]That is, a^b is a real (or complex) number that is not algebraic (i.e., not the root of a non-zero polynomial with rational coefficients).

[19]As $\sqrt{x}$ is really just $x^{1/2}$.

Theorem A.3. The infinite tetration,

$$\left(x^{\frac{a}{b}}\right)^{\left(x^{\frac{a}{b}}\right)^{\left(x^{\frac{a}{b}}\right)^{\cdot^{\cdot^{\cdot}}}}} = x, \tag{6}$$

is only true if $x = \frac{b}{a}$ *(*or, $x = 1$*)*.

Proof. If Equation (6) is true then,

$$\begin{aligned} x &= \left(x^{\frac{a}{b}}\right)^{\left(x^{\frac{a}{b}}\right)^{\left(x^{\frac{a}{b}}\right)^{\cdot^{\cdot^{\cdot}}}}} \\ x &= \left(x^{\frac{a}{b}}\right)^{x} \\ x &= x^{x\left(\frac{a}{b}\right)} \\ \ln(x) &= x\left(\frac{a}{b}\right)\ln(x). \end{aligned} \tag{7}$$

The log function is monotonic, therefore, Equation (7) holds if and only if $x = \left(\frac{b}{a}\right)$ or $x = 1$. ☐

A.3 A Peculiar Connection Between the Axiom of Choice and Predicting the Future

The key intuition behind this result is given in Hardin and Taylor (2008):

> We often model systems that change over time as functions from the real numbers $\mathbb{R}$ (or a subinterval of $\mathbb{R}$) into some set S of states, and it is often our goal to predict the behavior of these systems. Generally, this requires rules governing their behavior, such as a set of differential equations or the assumption that the system (as a function) is analytic. With no such assumptions, the system could be an arbitrary function, and the values of arbitrary functions are notoriously hard to predict. After all, if someone proposed a strategy for predicting the values of an arbitrary function based on its past values, a reasonable response might be, "That is impossible. Given any strategy for predicting the values of an arbitrary function, one could just define a function that diagonalizes against it: whatever the strategy predicts, define the function to be something else." This argument, however, makes an appeal to induction: to diagonalize against the proposed strategy at a point t, we must have already defined our function for all $s < t$ in order to determine what the strategy would predict at t. So, the argument would only be valid if $\mathbb{R}$ were well-ordered, but $\mathbb{R}$ is emphatically not well-ordered.

In this section some of the work of Hardin and Taylor (2008) is recreated (almost one-to-one). Fix sets S and T, with $|S| \geq 2$, and a binary relation $\trianglelefteq$ on T. In what follows we describe the "μ-strategy".

Definition 3 (Scenarios)**.** Let ${}^{T}S$ be the set of all functions from T to S. Call the elements of ${}^{T}S$ "scenarios". For each $t \in T$, define the equivalence relation $\approx_t$ on ${}^{T}S$ by $f \approx_t g$ if and only if$f$ and g agree on $\trianglelefteq t = \{s \in T \mid s \trianglelefteq t\}$, the set of $\trianglelefteq$-predecessors of t.

Definition 4 (Strategy)**.** Let $\mathcal{O} = \{[v]_t \mid v \in {}^T S \text{ and } t \in T\}$. A "strategy" is a function $g : \mathcal{O} \to {}^T S$ such that $g([v]_t) \in [v]_t$ for any $v \in {}^T S$ and $t \in T$. Fix a well-ordering $\preceq$ of ${}^T S$. The μ-strategy is the strategy $\mu : \mathcal{O} \to {}^T S$ defined by letting $\mu([f]_t)$ be the $\preceq$-least element of $[f]_t$. Let $\mu([v]_t) := \langle v \rangle_t$.

One possible interpretation of this ordering of functions (given by Hardin and Taylor, 2008) is that $f \prec g$ means f is *simpler* than g and then the μ-strategy is akin to Occam's razor.[20]

Let $v \in {}^T S$ be the true scenario and

$$W_0 = \{t \in T \mid \langle v \rangle_t(t) \neq v(t)\}.$$

So, W_0 is the set of points where the μ-strategy is wrong.

Lemma A.4. If $\lhd$ is transitive, then W_0 is well founded in $\lhd$ (i.e., it has no infinitely descending $\lhd$-chains).

Corollary A.5. If $T = \mathbb{R}$ and $\lhd$ is $<$, then W_0 is countable, measure sero, and nowhere dense.

What the above corollary says is that, if we model the universe as a function from the real numbers into some set of states, then the μ-strategy correctly predicts the present from the past of a full measure.

To show the μ-strategy is quite good at predicting (some of) the future retain the notation from above and let,

$$W_1 = \{t \in T \mid \langle v \rangle_t \neq \langle v \rangle_{t'} \text{ whenever } t \lhd t'\}, \tag{8}$$

be the set of instants where the μ-strategy's guess differs from all later guesses.

Theorem A.6. For $t \in \mathbb{R}$, say that μ-strategy "guesses well" at t if there exists an $\varepsilon > 0$ such that $\langle v \rangle_t$ and v agree on $[t, t + \varepsilon)$. Then the set of $t \in \mathbb{R}$ where the μ-strategy guesses well has full measure.

Proof. Take a $t \in \mathbb{R} \setminus W_1$. By the definition of W_1, there must be some $t' > t$ such that $\langle v \rangle_t = \langle v \rangle_{t'}$. For any $u < t'$, we have

$$\langle v \rangle_t(u) = \langle v \rangle_{t'}(u) = v(u),$$

so $\langle v \rangle_t$ and v agree on $(-\infty, t')$; in particular, $\langle v \rangle_t$ and v agree on $[t, t + \varepsilon)$ where $\varepsilon = t' - t$. So the μ-strategy guesses well on $\mathbb{R} \setminus W_1$, which has full measure by Corollary A.5 (with W_1 in place of W_0). ☐

A.4 More on Toilet Paper

The Catalan numbers are named after Eugène Catalan but were first discovered by Minggatu (a Mongolian astronomer and mathematician). They are an infinite sequence of natural numbers that often appear in counting problems.

[20] Occam's razor is often cited as "*Entia non sunt multiplicanda præter necessitatem.*" Or, "*Entities must not be multiplied beyond necessity.*" (Barry, 2014; Schaffer, 2015). It is commonly used to say something along the lines of — whatever is the simplest solution must be the correct one.

Definition 5 (Catalan numbers)**.** The Catalan numbers are the numbers

$$1, 1, 2, 5, 14, 132, 429, 1430, 4862 \ldots$$

where the n^{th} number is given by,

$$C_n = \frac{(2n)!}{(n+1)!\, n!}, \quad \text{for } n \geq 0.$$

There are a number of interpretations of the Catalan numbers. For example,

(i) C_n is the number of "Dyck words" of length $2n$.[21]

(ii) A convex polygon with $n + 2$ sides can be cut into triangles by connecting vertices with non-crossing line segments. The number of triangles formed is n and they can be formed in C_n different ways.

(iii) C_n is the number of permutations of $\{1, \ldots, n\}$ that avoids permutations with three consecutive increasing terms.

(iv) C_n is the number of ways to form a "mountain range" with n upstrokes and n downstrokes that all stay above a horizontal line.

Bertrand's ballot theorem says that in an election where candidate A receives p votes and candidate B receives q votes with $p > q$, what is the probability that A will be strictly ahead of B throughout the count under the assumption that votes are counted in a randomly picked order? The solution is given by Bertrand (1887) and Whitworth (1878) to be, $(p - q)/(p + q)$.

A.5 Solutions to Fun Problems (Section 1.7)

Solution to Problem 1. (Back to Problem)

We wish to maximize $\sqrt[n]{n}$. Clearly,

$$\arg\max \left\{ \sqrt[n]{n} \right\} = \arg\max \left\{ \frac{1}{n} \ln n \right\},$$

and this expression is maximized when its first derivative is zero:

$$\begin{aligned} \frac{d}{dn} \frac{\ln n}{n} &= 0, \\ \iff \frac{1}{n^2} &= \frac{\ln n}{n^2}, \\ \iff \ln n &= 1. \end{aligned}$$

Thus, the solution is that $\boxed{n = e}$.

Solution to Problem 2. (Back to Problem)

The answer is $\boxed{\text{true}}$. This can easily be seen by looking at the graph of $e^x + e^{-x}$ given in Figure 3. However, the more exciting part of this question is how you solve it. I will provide three ways to solve the question. First, the way that I initially did it. Second, a more pedagogical way. Thirdly, a method using a more

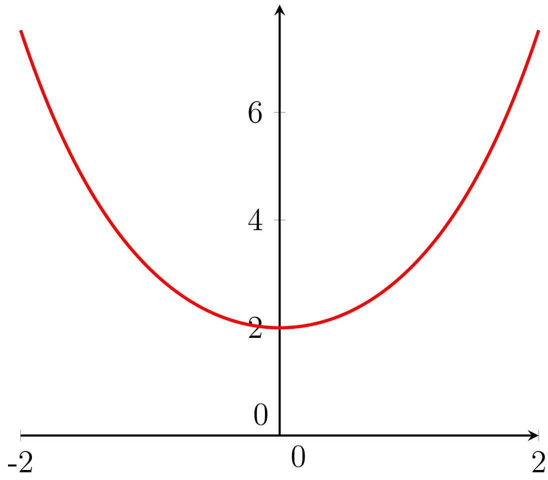

Figure 3: Plot of the graph of $e^x + e^{-x}$.

general inequality. Finally, we will see that a much more general version of the statement of the problem is true.
First note that $e^x + e^{-x}$ equals $e^y + e^{-y}$ for $y = -x$, thus, we can restrict our attention to $x \in \mathbb{R}_+$. We see that for $x = 0$ it is given that $e^0 + e^0 = 2$. Now, consider the function $f = e^x + e^{-x}$. Then, it suffices to show that $f' \geq 0$ for all $x \geq 0$. Since $f' = e^x - e^{-x} \geq 0$ holds for all $x \geq 0$ it must be that f is non-decreasing.[22] Therefore, it is true that $e^x + e^{-x} \geq 2$ for all x.

A more pedagogical way of answering this is to define f as $e^x + e^{-x}$. Then notice that,

$$\begin{aligned} f' &= e^x - e^{-x}, \quad \text{and} \\ f'' &= e^x + e^{-x}. \end{aligned}$$

Then we see that $f'(x) = 0$ at $x = 0$ and that $f''(0) > 0$ implying that f is convex so 0 is a global minimum. Thus, the inequality must hold for all x.

An alternative solution to this question is to use the AM-GM inequality which says that the arithmetic mean of a list of non-negative real numbers is greater than or equal to the geometric mean of the same list. That is, for a list $x_1, \ldots, x_n$ of non-negative real numbers

$$\frac{1}{n} \sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i \right)^{1/n}.$$

[21] A Dyck word is a string of X's and Y's such that no initial segment of the string has more Y's than X's. It is similar to a formalization of the rules of parenthesis.

[22] This is because $e^x - e^{-x} \geq 0 \iff e^x \geq e^{-x}$. Since $e^x > 0$, we may multiply both sides by e^x to obtain $e^{2x} \geq 1$. But if $x \geq 0$, then $2x \geq 0$, so since the exponential function is increasing, $e^{2x} \geq e^0 = 1$. Therefore $e^x - e^{-x} \geq 0$ for all $x \geq 0$.

Then, given the list $\{e^x, e^{-x}\}$ we have $\text{AM} = (e^x + e^{-x})/2$ and $\text{GM} = \sqrt{e^x \cdot e^{-x}}$. Thus,

$$\begin{aligned} \text{AM} &\geq \text{GM} \\ \frac{1}{2}\left(e^x + e^{-x}\right) &\geq \sqrt{e^x \cdot e^{-x}} \\ \implies e^x + e^{-x} &\geq 2. \end{aligned}$$

In fact, the AM-GM inequality provides a much stronger theorem.

Theorem A.7. For any strictly positive number α the inequality $\alpha^x + \alpha^{-x} \geq 2$ holds for any x.

Proof. Let α be greater than zero and $x \in \mathbb{R}$. Then the list $\{\alpha^x, \alpha^{-x}\}$ is a list of non-negative real numbers and we can use the AM-GM inequality. We can see that the geometric mean of this list is 1. Moreover, the arithmetic mean of the list is $(\alpha^x + \alpha^{-x})/2$. Then, by the AM-GM inequality:

$$\begin{aligned} \text{AM} &\geq \text{GM} \\ \frac{1}{2}\left(\alpha^x + \alpha^{-x}\right) &\geq 1 \\ \implies \alpha^x + \alpha^{-x} &\geq 2. \end{aligned}$$

□

Solution to Problem 3. (Back to Problem)

We would like an f such that

$$f = f' + f'' + f''' + \cdots$$

A reasonable guess is the function e^x as we know $e^x = \frac{d^n}{dx^n} e^x$ for any natural number n. However, we can see that if $f = e^x$ then $\sum_{n=1}^{\infty} f^{(n)}$ diverges. Thus, we clearly also need a function such that the coefficients generated by differentiating are summable. With this in mind consider $f = e^{x/2}$. Then, $f^{(n)} = \frac{1}{2^n} e^{x/2}$. Since $\sum_{n=0}^{\infty} \frac{1}{2^n} = 2$ implies that $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$ we have that

$$\begin{aligned} \sum_{n=1}^{\infty} f^{(n)} &= \sum_{n=1}^{\infty} \frac{1}{2^n} e^{x/2}, \\ &= e^{\frac{x}{2}} \sum_{n=1}^{\infty} \frac{1}{2^n}, \\ &= e^{x/2}. \end{aligned}$$

More generally, we can say $\boxed{f = c \cdot e^{x/2} \text{ for some } c \in \mathbb{R} \setminus \{0\}}$.

Solution to Problem 4. (Back to Problem)

Notice that,

$$\int_0^1 \left\lfloor \frac{1}{x} \right\rfloor^{-1} dx = \sum_{n=1}^{\infty} \int_{1/(n+1)}^{1/n} \left\lfloor \frac{1}{x} \right\rfloor^{-1} dx.$$

For $x \in \left(\frac{1}{n+1}, \frac{1}{n}\right]$, we have $n \leq 1/x < n+1$, so $\left\lfloor \frac{1}{x} \right\rfloor = n$. Hence

$$\begin{aligned}\int_0^1 \left\lfloor \frac{1}{x} \right\rfloor^{-1} dx &= \sum_{n=1}^{\infty} \int_{1/(n+1)}^{1/n} \frac{1}{n}\, dx \\ &= \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{1}{n} - \frac{1}{n+1}\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2(n+1)}.\end{aligned}$$

Using partial fractions,

$$\frac{1}{n^2(n+1)} = -\frac{1}{n} + \frac{1}{n^2} + \frac{1}{n+1}.$$

Therefore,

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n^2(n+1)} &= \sum_{n=1}^{\infty} \left(\frac{1}{n+1} - \frac{1}{n}\right) + \sum_{n=1}^{\infty} \frac{1}{n^2} \\ &= -1 + \frac{\pi^2}{6}.\end{aligned}$$

So

$$\int_0^1 \left\lfloor \frac{1}{x} \right\rfloor^{-1} dx = \boxed{\frac{\pi^2}{6} - 1}.$$

Solution to Problem 5. (Back to Problem)

Let p_n be the n^{th} prime,

$$f(x) := x^3 - x, \quad \text{and} \quad S := \int_0^{\infty} \frac{\pi(x)}{f(x)}\, dx.$$

Then,

$$\begin{aligned}S &= \int_{p_1}^{p_2} f(x)^{-1}\, dx + 2\int_{p_2}^{p_3} f(x)^{-1}\, dx + 3\int_{p_3}^{p_4} f(x)^{-1}\, dx + 4\int_{p_4}^{p_5} f(x)^{-1}\, dx + \cdots \\ &= \sum_{n=1}^{\infty} n \int_{p_n}^{p_{n+1}} f(x)^{-1}\, dx.\end{aligned}$$

As the integral of f^{-1} is $\frac{1}{2}\ln\left|\frac{x^2-1}{x^2}\right| + C$, we get that

$$S = \sum_{n=1}^{\infty} \frac{n}{2} \ln\left(\frac{p_n^2 p_{n+1}^2 - p_n^2}{p_n^2 p_{n+1}^2 - p_{n+1}^2}\right).$$

Simplifying the term inside the log:

$$\frac{p_n^2 p_{n+1}^2 - p_n^2}{p_n^2 p_{n+1}^2 - p_{n+1}^2} = \frac{p_n^2(p_{n+1}^2 - 1)}{p_{n+1}^2(p_n^2 - 1)} = \frac{1 - 1/p_{n+1}^2}{1 - 1/p_n^2}.$$

Now, define $a_n \coloneqq \frac{1}{2}\ln\left(1-\frac{1}{p_n^2}\right)$. Then

$$S=\sum_{n=1}^{\infty} n\left(a_{n+1}-a_n\right).$$

Using summation by parts

$$\sum_{n=1}^{N} n(a_{n+1}-a_n)=Na_{N+1}-\sum_{n=1}^{N} a_n.$$

As n goes to infinity a_n goes to zero. Thus,

$$S=-\sum_{n=1}^{\infty} a_n=-\frac{1}{2}\sum_{n=1}^{\infty}\ln\left(1-\frac{1}{p_n^2}\right).$$

Using the identity $\sum \ln b_n = \ln \prod b_n$, we convert the sum into a product:

$$S=-\frac{1}{2}\ln\prod_p\left(1-\frac{1}{p^2}\right).$$

Using Euler's product formula for the Riemann zeta function,

$$\zeta(s)=\prod_p\frac{1}{1-p^{-s}}\qquad (s>1),$$

we have

$$\prod_p\left(1-\frac{1}{p^2}\right)=\frac{1}{\zeta(2)}=\frac{6}{\pi^2}.$$

Therefore,

$$\begin{aligned} S &= -\frac{1}{2}\ln\left(\frac{6}{\pi^2}\right)\\ &= \frac{1}{2}\ln\left(\frac{\pi^2}{6}\right)\\ &= \boxed{\ln\left(\frac{\pi}{\sqrt{6}}\right)}. \end{aligned}$$

B Philosophical Appendix

B.1 More on the Raven Paradox

One possible answer to the Raven paradox is given by Hosiasson-Lindenbaum (1940). It claims that the discovery of a green avocado does, *ever so slightly*, confirm the claim H. If the following theorem is true then the paradox is solved.

Theorem B.1 (Raven Theorem)**.** If $\Pr(\neg R \mid \neg B)$ is very high and $\Pr(\neg B \mid H) = \Pr(\neg B)$, then $\Pr(H \mid \neg R \wedge \neg B)$ is slightly larger than $\Pr(H)$.

The first assumption $\Pr(\neg R \mid \neg B)$ is not particularly controversial. Most things are not ravens and conditional on not being black even less things are ravens. The second assumption, however, is controversial. It seems that conditioning on our hypothesis should *decrease* the probability of things in the universe not being black hence $\Pr(\neg B \mid H) < \Pr(\neg B)$. For more discussion of this issue see, for example, Fitelson (2003), Fitelson and Hawthorne (2010), Good (1967), and Rinard (2014)

B.2 Reasoning for The Hardest Logic Puzzle Ever (Section 2.4)

The reasoning for the three preliminary puzzles is as follows.

(I) Whether the middle card is an ace or not you are certain to find an ace by choosing the left card if you hear yes and the right card if you hear no. If the middle card is an ace then the answer is truthful and then the left card is an ace if "yes" and the right card is an ace if "no." If the middle card is a jack then both of the other cards are aces so the choice of left or right is irrelevant.

(II) By asking the biconditional question there become four possible states of the world:

   (a) The god is True and Dushanbe is in Kirghizia $\implies$ you get answer "yes."

   (b) The god is True and Dushanbe is not in Kirghizia $\implies$ you get answer "no."

   (c) The god is False and Dushanbe is in Kirghizia $\implies$ you get answer "yes."

   (d) The god is False and Dushanbe is not in Kirghizia $\implies$ you get answer "no."

   Thus, you get "yes" if Dushanbe is in Kirghizia and "no" otherwise.

(III) This solution is similar to above. By asking a biconditional question there are four possibilities:

   (a) *Da* means yes and Dushanbe is in Kirghizia $\implies$ True says *da*.

   (b) *Da* means yes and Dushanbe is not in Kirghizia $\implies$ True says *ja* (meaning no).

(c) *Da* means no and Dushanbe is in Kirghizia $\Longrightarrow$ True says *da* (meaning no).

(d) *Da* means no and Dushanbe is not in Kirghizia $\Longrightarrow$ True says *ja*.

So, you get the answer *da* if Dushanbe is in Kirghizia and *ja* if not.

Now we enumerate the logic underpinning the three questions required to solve our main puzzle. Each of the questions utilizes the logic of the corresponding smaller puzzle enumerated above.

1. If A is True or False and you get the answer *da*, then B must be Random, therefore, C is either True or False. If A is True or False and the answer is *ja* then B is not Random, therefore, B is either True or False. (If A is Random then it does not matter if we choose B or C!) Thus, no matter what A is if you get the answer *da* then C is either True or False and if you get the answer *ja* then B is either True or False.

2. For this question True will answer *da* and False will answer *ja*. Thus, B is either True or False.

3. Repeating the logic provided in the body of the text suppose B is True. Then if you get the answer *da* then A is Random, and therefore B is True, and C is False, and you are done; but if you get the answer *ja*, then A is not Random, so A is False, B is true, C is Random, and you are again done. Suppose B is False. Then if you get the answer *da*, then since B speaks falsely, A is not Random, and therefore A is True, B is False, C is Random, and you are done; but if we get *ja*, then A is Random, and thus B is False, and C is True, and you are again done.

For those wondering: Dushanbe is the capital of *Tajikistan* and is not in Kirghizia (Kyrgyzstan).