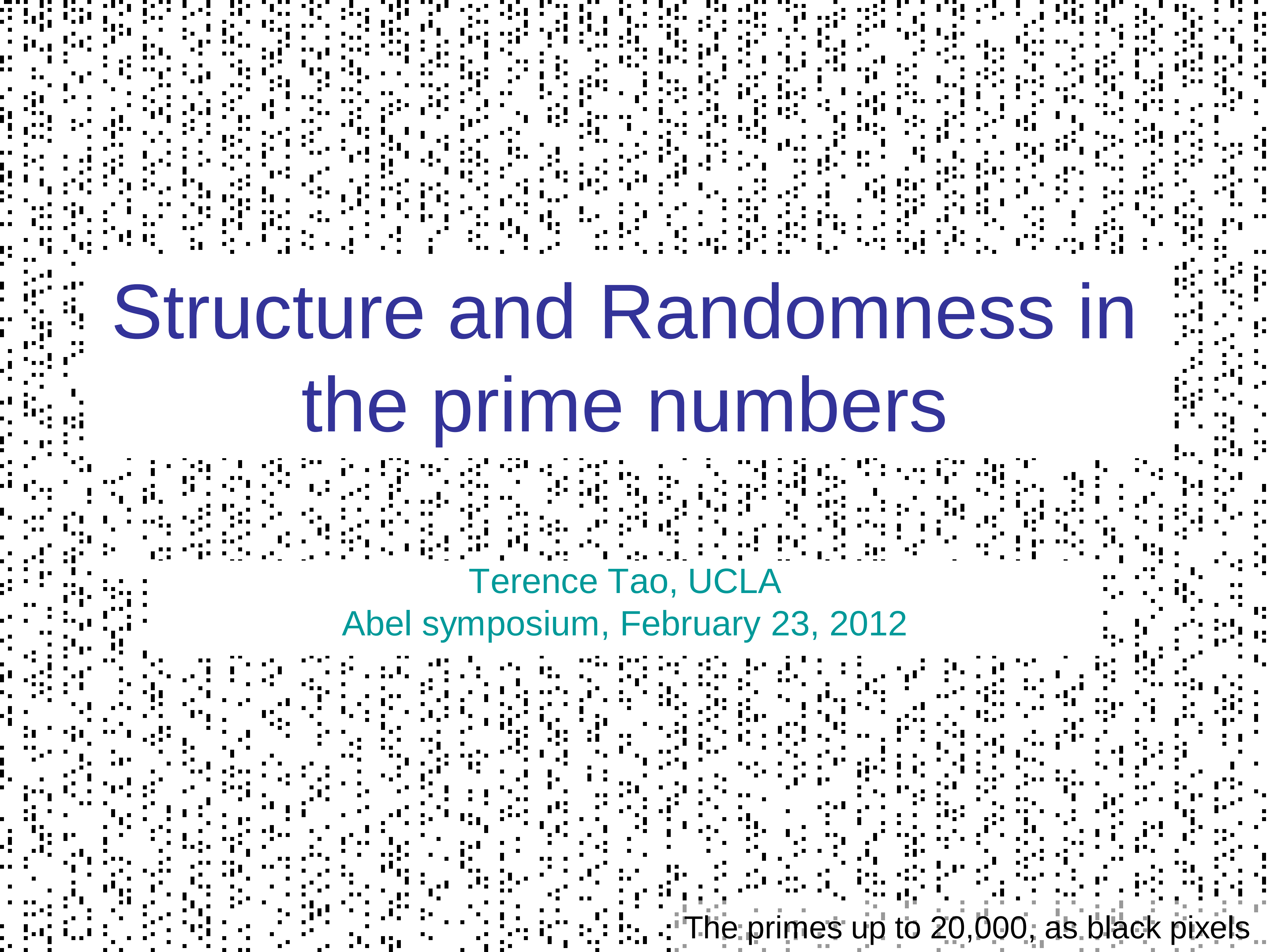
Structure and Randomness in the prime numbers

Terence Tao, UCLA

Abel symposium, February 23, 2012

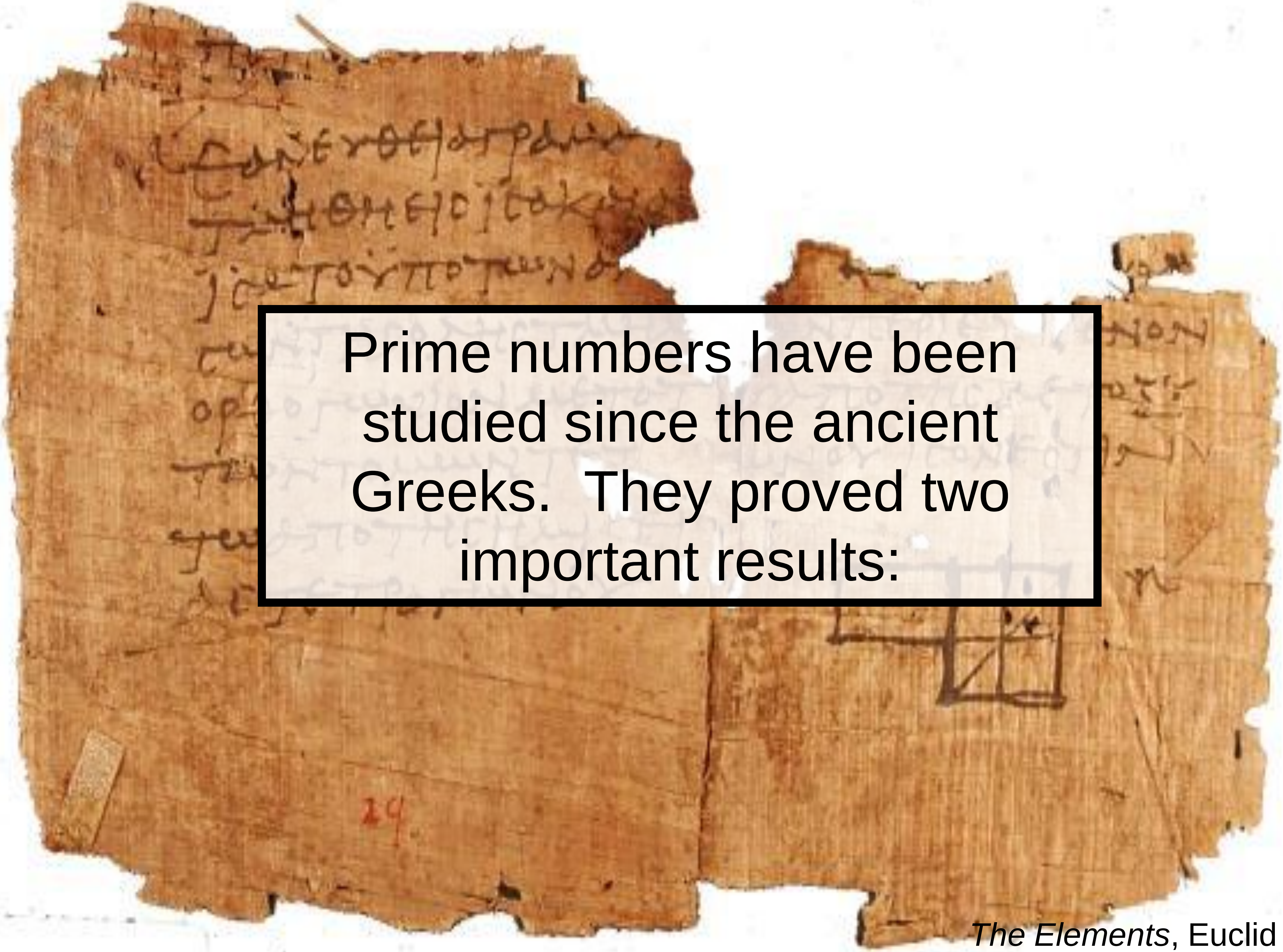


The primes up to 20,000, as black pixels

2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97 101
103 107 109 113 127 131 137 139 149 151 157 163 167 173 179 181 191 193
197 199 211 223 227 229 233 239 241 251 257 263 269 271 277 281 283 293
307 311 313 317 331 337 347 349 353 359 367 373 379 383 389 397 401 409
419 421 431 433 439 443 449 457 461 463 467 479 487 491 499 503 509 521
523 541 547 557 563 569 571 577 587 593 599 601 607 613 617 619 631 641
643 647 653 659 661 673 677 683 691 701 709 719 727 733 739 743 751 757
761 769 773 787 797 809 811 821 823 827 829 839 853 857 859 863 877 881
883 887 907 911 919 929 937 941 947 953 967 971 977 983 991 997 1009
1013 1019 1021 1031 1033 1039 1049 1051 1061 1063 1069 1087 1091 1093
1097 1103 1109 1117 1123 1129 1151 1153 1163 1171 1181 1187 1193 1201
1213 1217 1223 1229 1231 1237 1249 1259 1277 1279 1283 1289 1291 1297
1301 1303 1307 1319 1321 1327 1361 1367 1373 1381 1399 1409 1423 1427
1429 1433 1439 1447 1451 1453 1459 1471 1481 1483 1487 1489 1493 1499
1511 1523 1531 1543 1549 1553 1559 1567 1571 1579 1583 1597 1601 1607
1609 1613 1619 1621 1627 1637 1657 1663 1667 1669 1693 1697 1699 1709
1721 1723 1733 1741 1747 1753 1759 1777 1783 1787 1789 1801 1811 1823
1831 1847 1861 1867 1871 1873 1877 1879 1889 1901 1907 1913 1931 1933
1949 1951 1973 1979 1987 1993 1997 1999 2003 2011 2017 2027 2029 2039
… $2^{43,112,609}-1$ (GIMPS, 2008) …

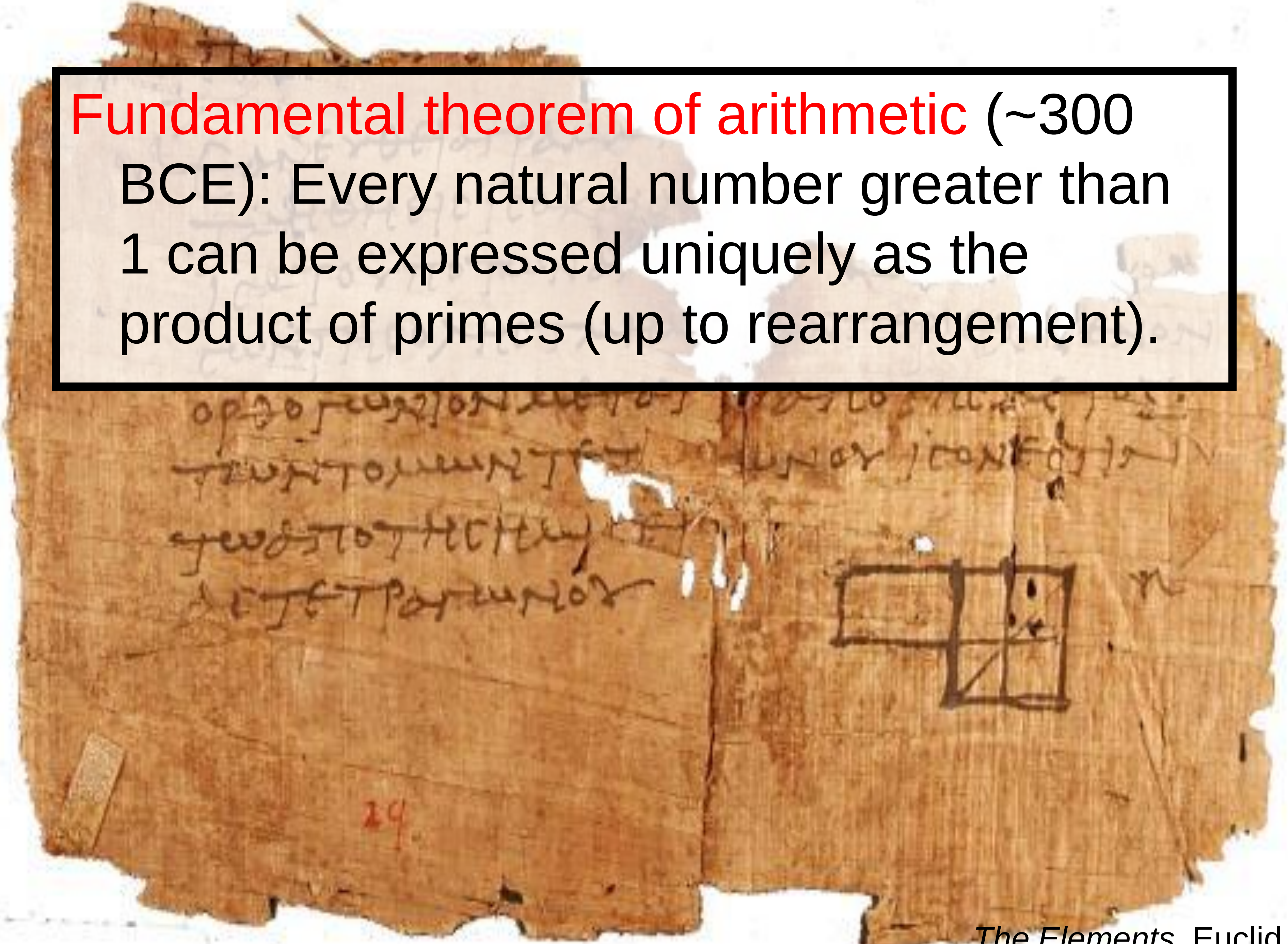
A prime number is any natural number greater than 1, which cannot be factored as the product of two smaller numbers.

Prime numbers have been studied since the ancient Greeks. They proved two important results:

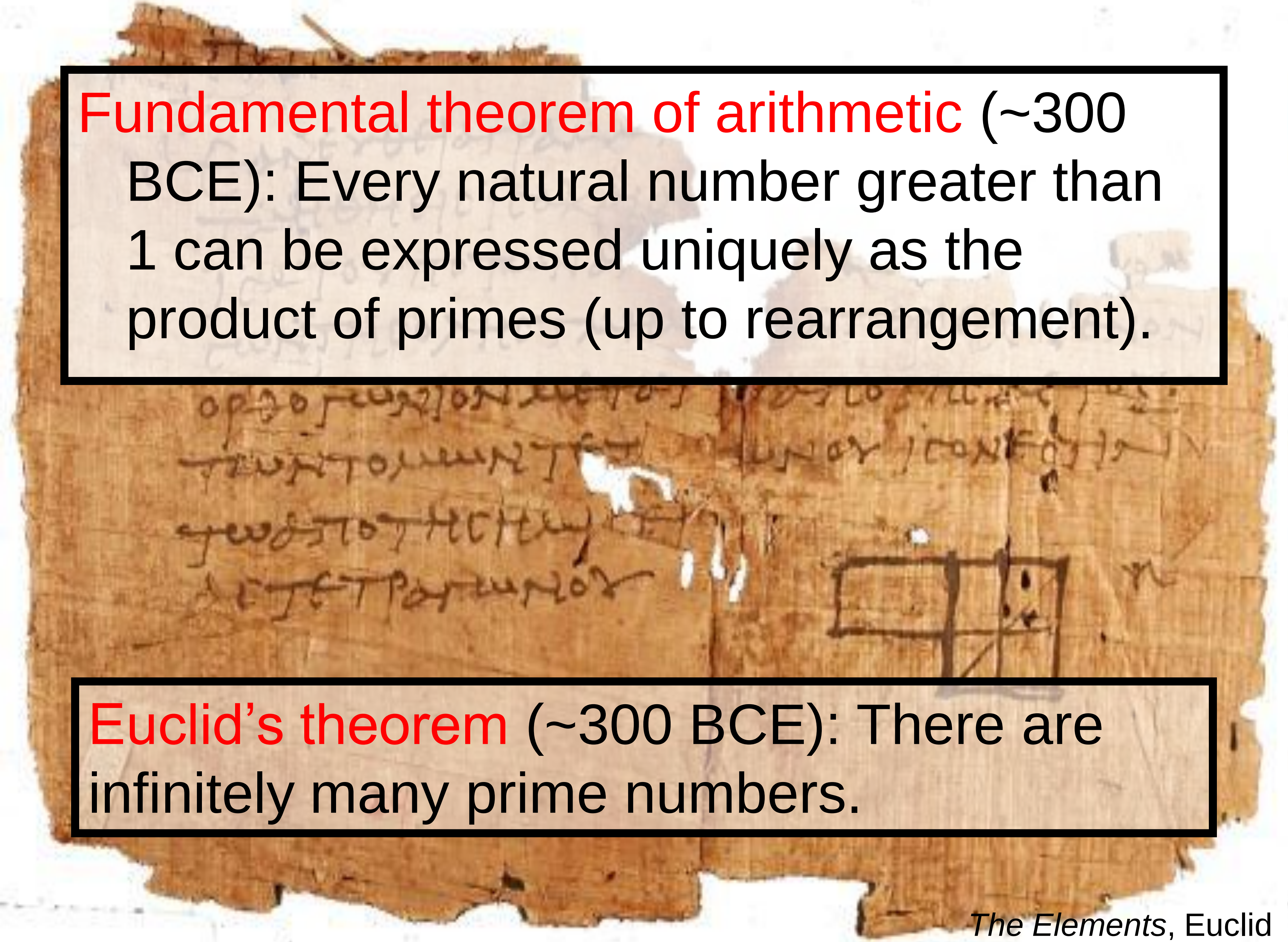


The Elements, Euclid

Fundamental theorem of arithmetic (~300 BCE): Every natural number greater than 1 can be expressed uniquely as the product of primes (up to rearrangement).



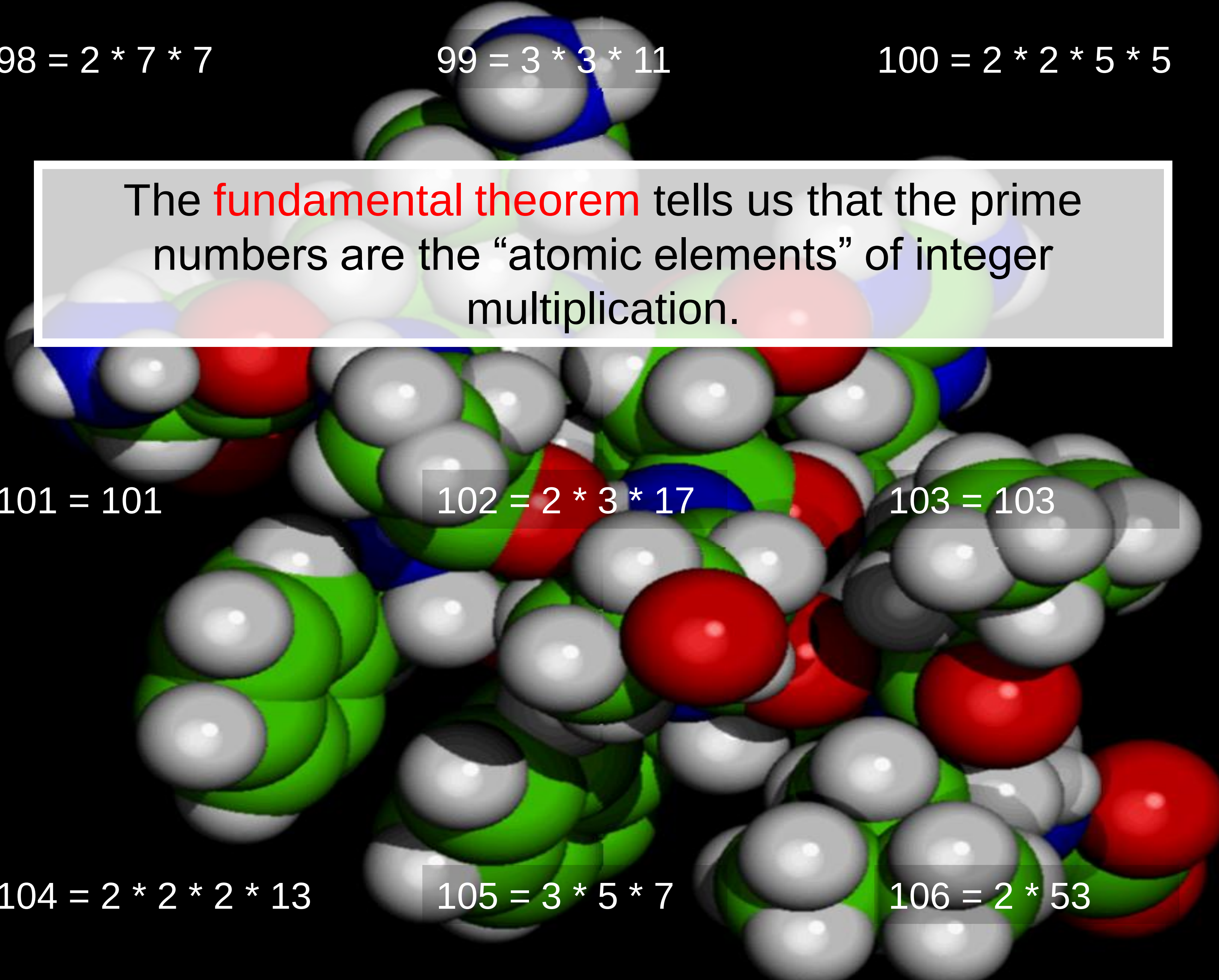
The Elements, Euclid



Fundamental theorem of arithmetic (~300 BCE): Every natural number greater than 1 can be expressed uniquely as the product of primes (up to rearrangement).

Euclid’s theorem (~300 BCE): There are infinitely many prime numbers.

The Elements, Euclid



98 = 2 * 7 * 7

99 = 3 * 3 * 11

100 = 2 * 2 * 5 * 5

The fundamental theorem tells us that the prime numbers are the “atomic elements” of integer multiplication.

101 = 101

102 = 2 * 3 * 17

103 = 103

104 = 2 * 2 * 2 * 13

105 = 3 * 5 * 7

106 = 2 * 53

98 = 2 * 7 * 7

99 = 3 * 3 * 11

100 = 2 * 2 * 5 * 5

The fundamental theorem tells us that the prime numbers are the “atomic elements” of integer multiplication.

101 = 101

102 = 2 * 3 * 17

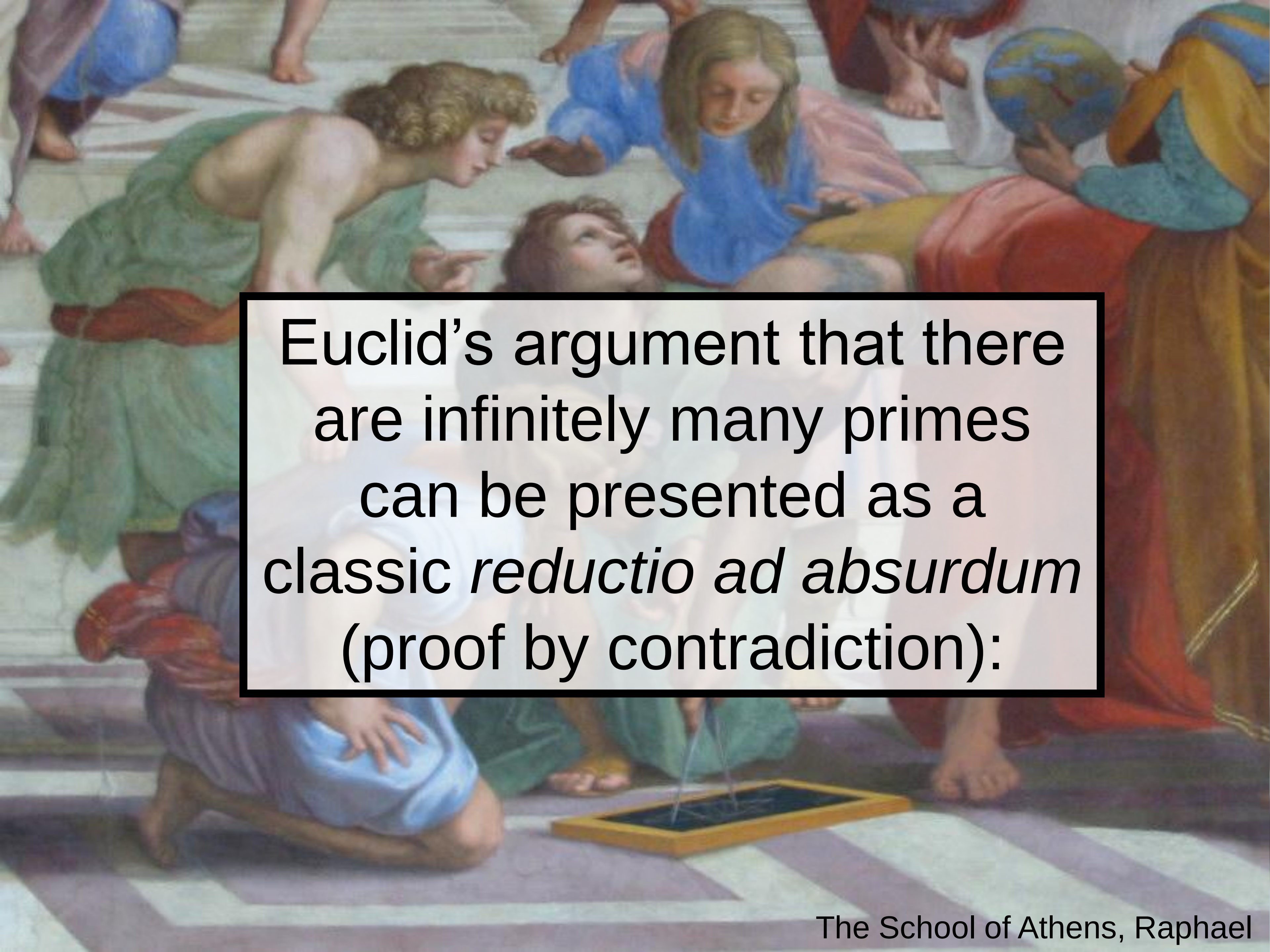
103 = 103

It is because of this theorem that we do not consider 1 to be a prime number.

104 = 2 * 2 * 2 * 13

105 = 3 * 5 * 7

106 = 2 * 53

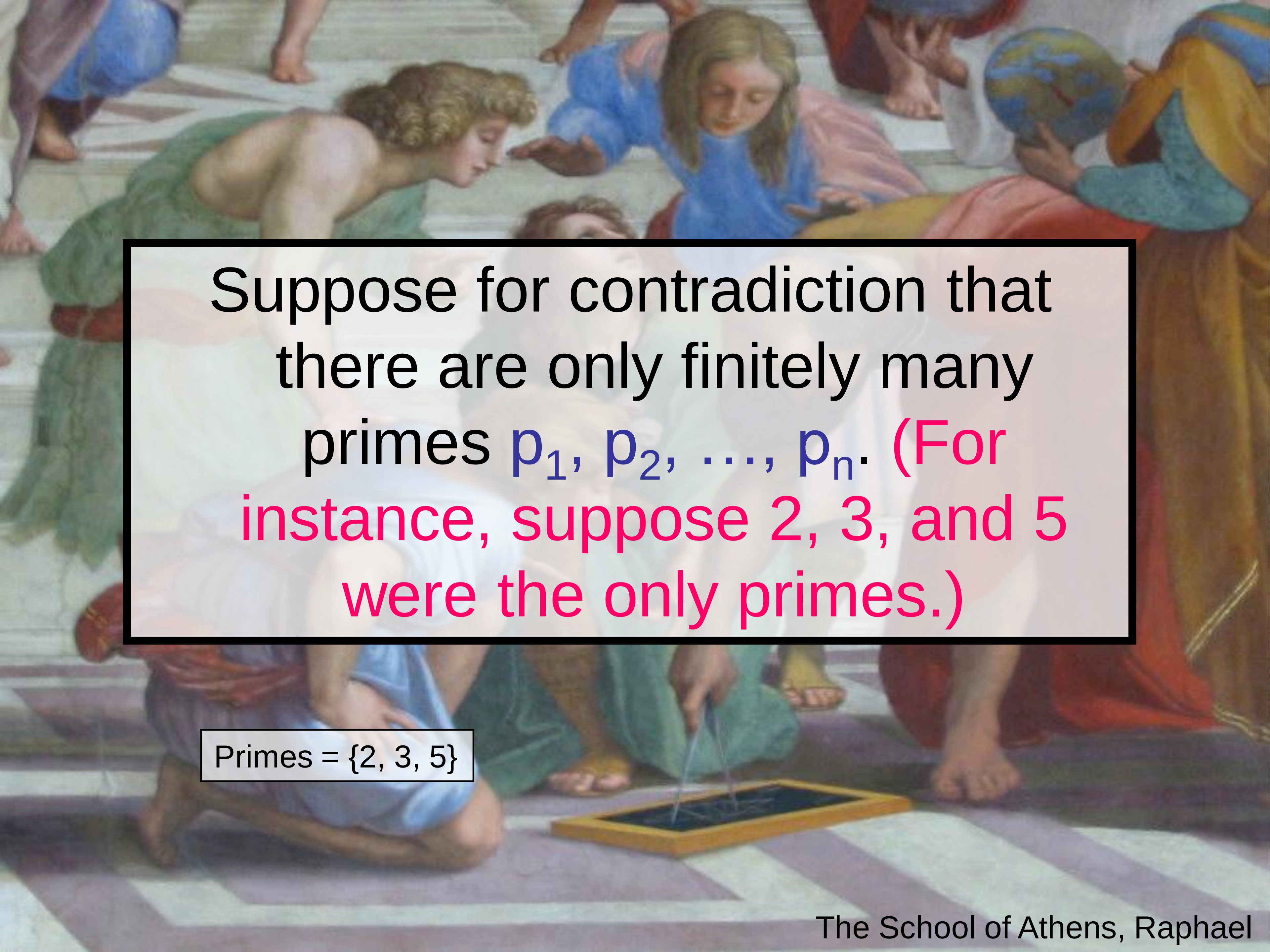


Euclid’s argument that there are infinitely many primes can be presented as a classic *reductio ad absurdum* (proof by contradiction):

The School of Athens, Raphael

Suppose for contradiction that there are only finitely many primes $p_1, p_2, \ldots, p_n$. (For instance, suppose 2, 3, and 5 were the only primes.)

Primes = {2, 3, 5}

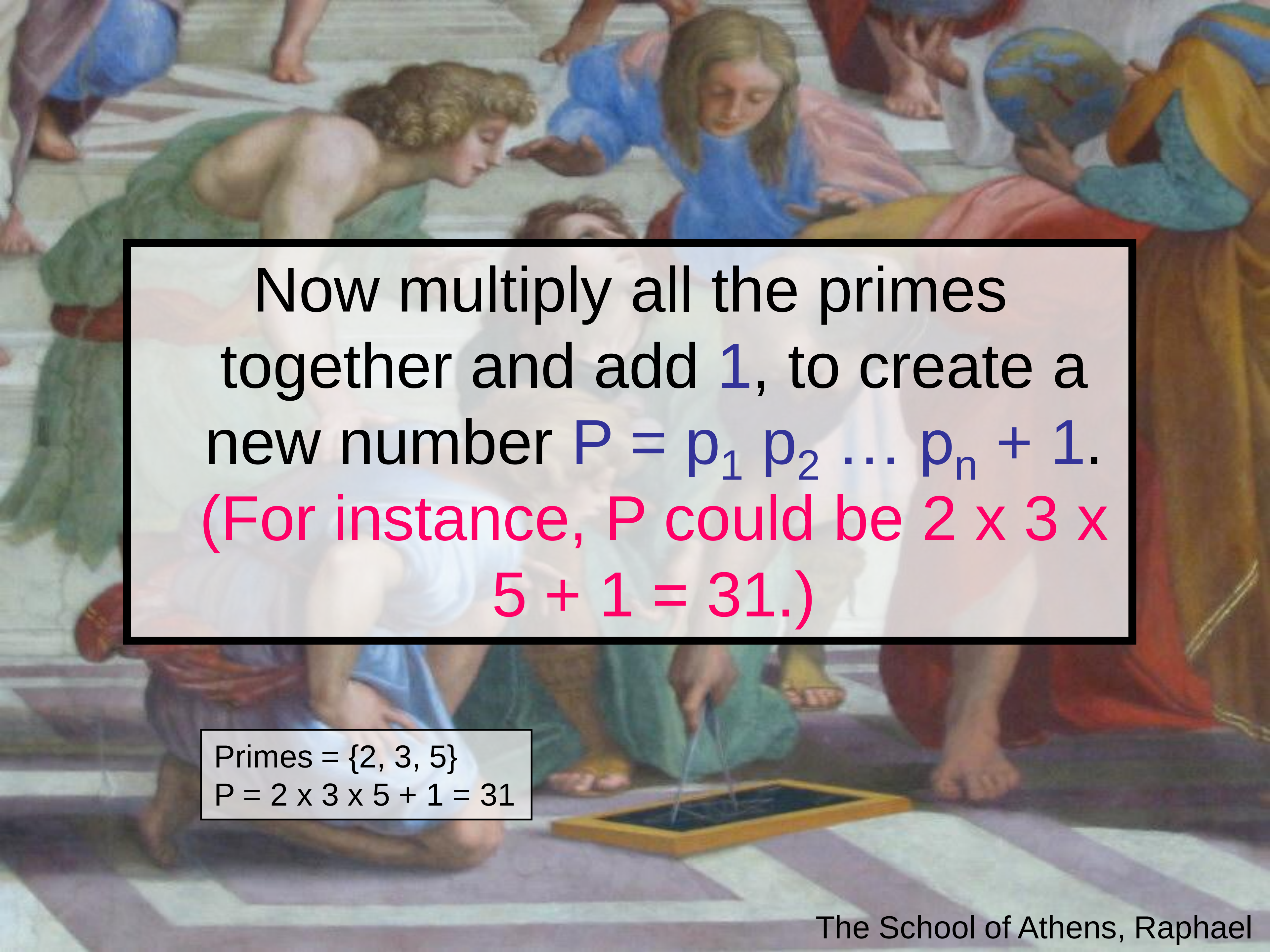


The School of Athens, Raphael

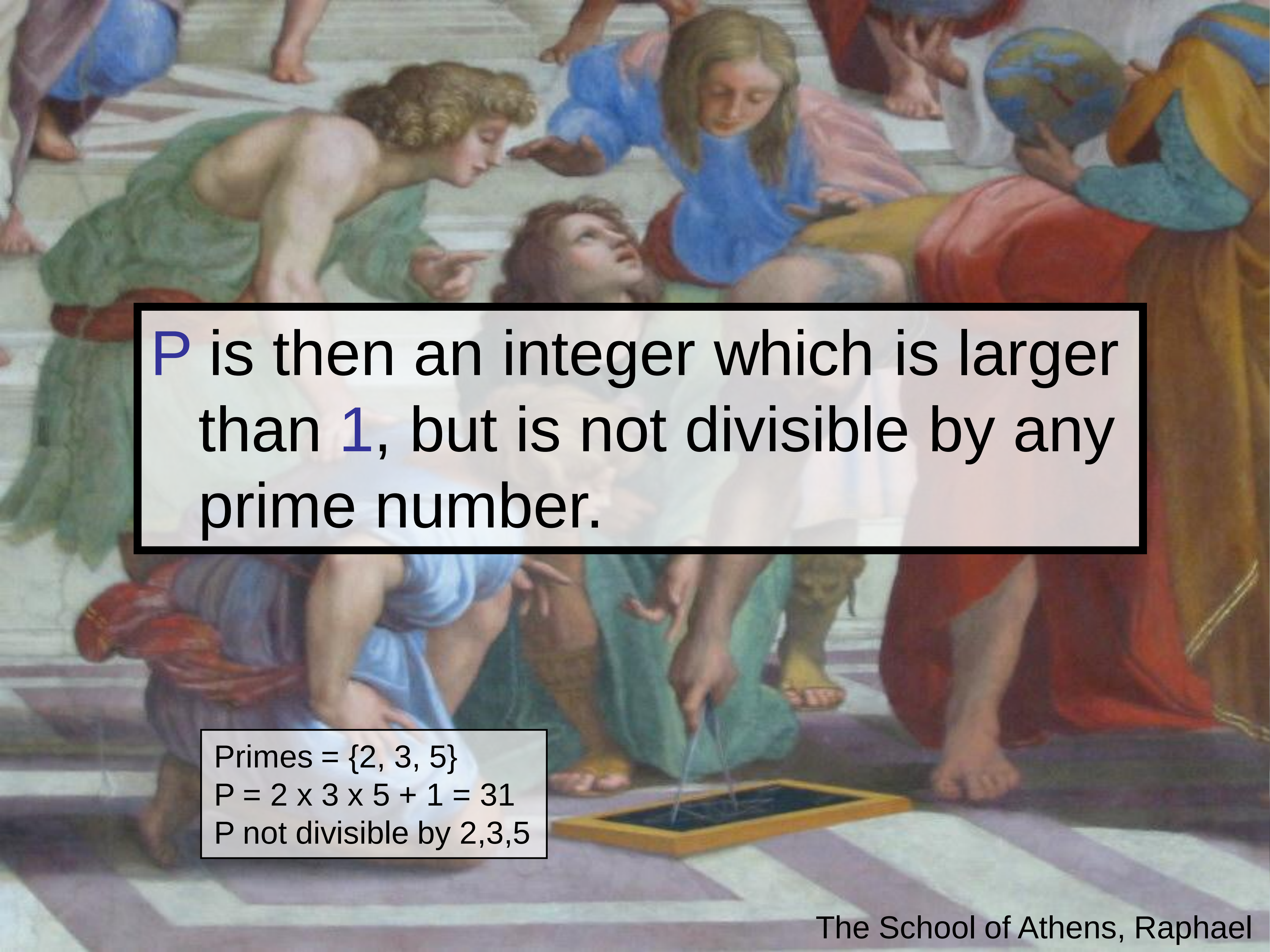
Now multiply all the primes together and add 1, to create a new number $P = p_1 \, p_2 \, \ldots \, p_n + 1$.

(For instance, P could be 2 x 3 x 5 + 1 = 31.)

Primes = {2, 3, 5}
P = 2 x 3 x 5 + 1 = 31



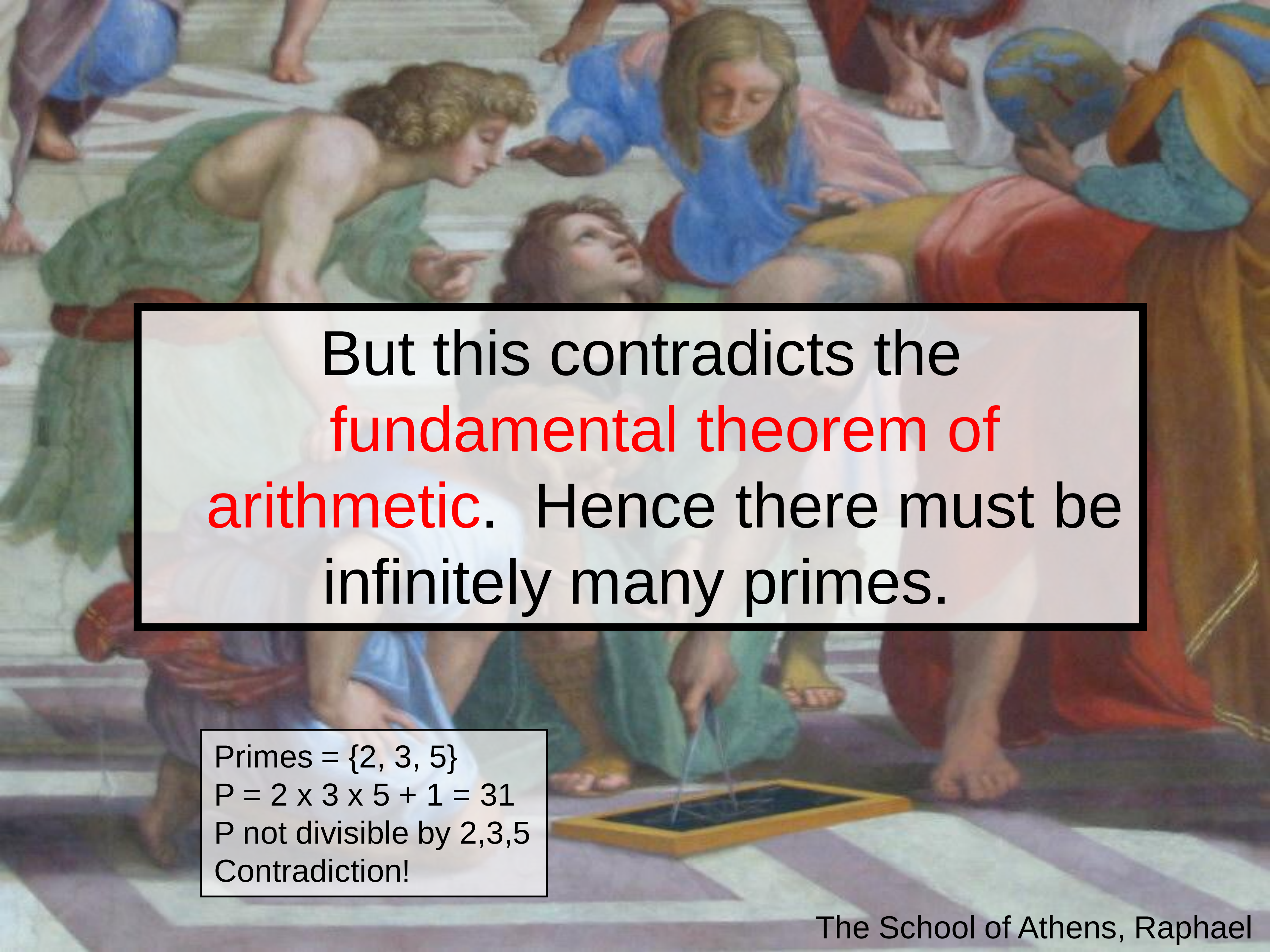
The School of Athens, Raphael



P is then an integer which is larger than 1, but is not divisible by any prime number.

Primes = {2, 3, 5}
P = 2 x 3 x 5 + 1 = 31
P not divisible by 2,3,5

The School of Athens, Raphael



But this contradicts the fundamental theorem of arithmetic. Hence there must be infinitely many primes.

Primes = {2, 3, 5}
P = 2 x 3 x 5 + 1 = 31
P not divisible by 2,3,5
Contradiction!

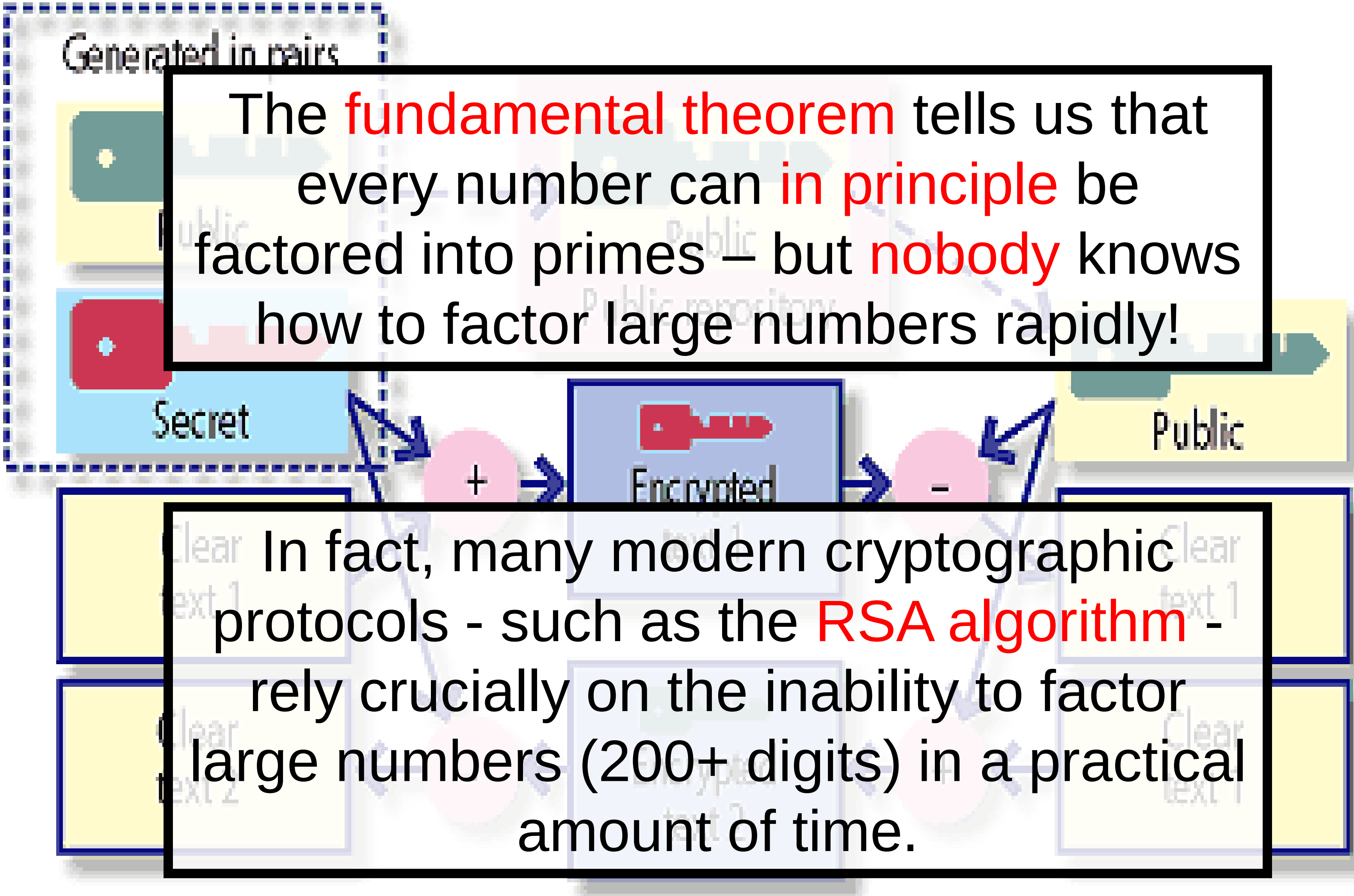
The School of Athens, Raphael

“*Reductio ad absurdum*, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game ”.

(G.H. Hardy, 1877-1947)

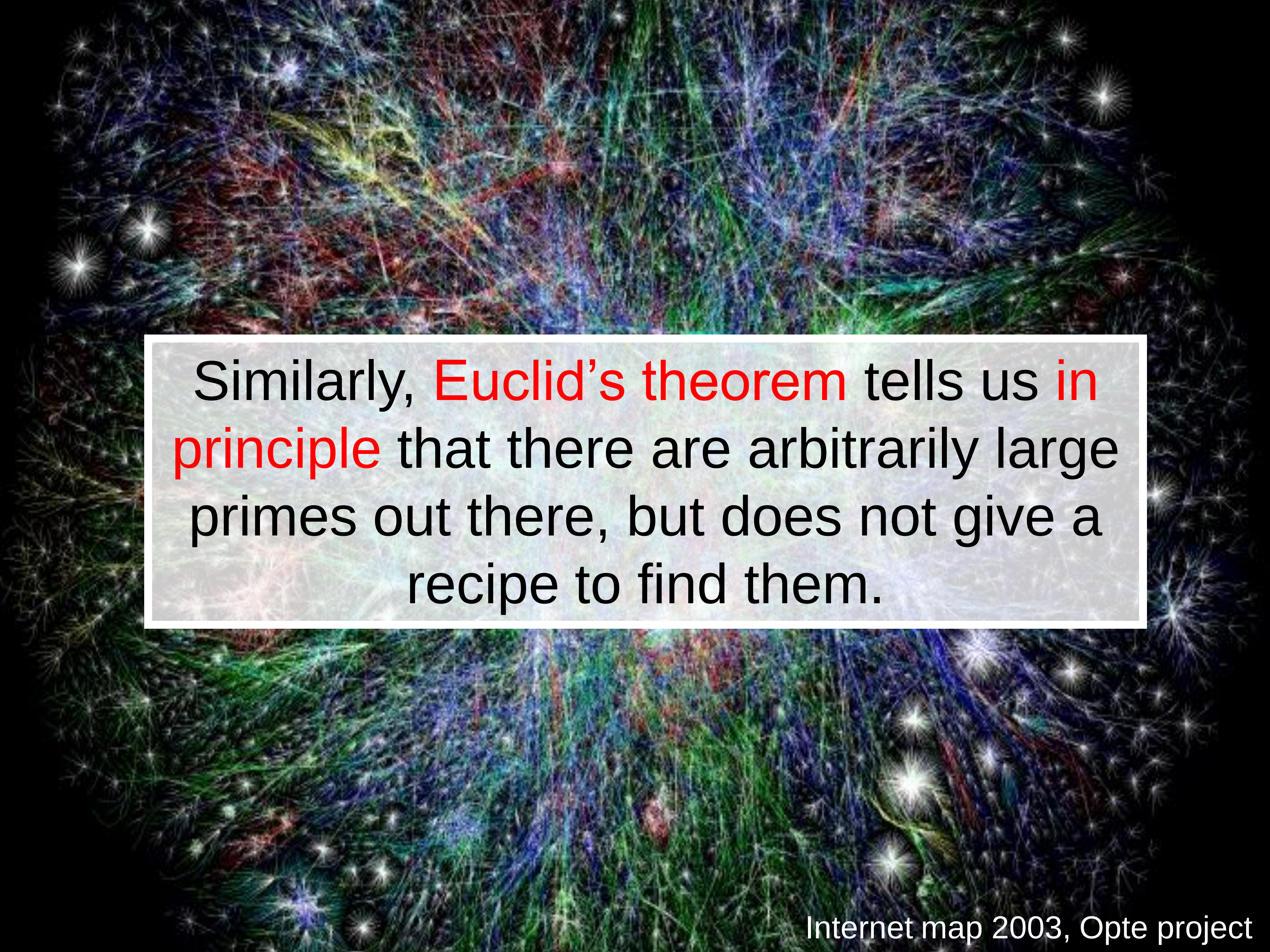
The fundamental theorem tells us that every number can in principle be factored into primes – but nobody knows how to factor large numbers rapidly!

In fact, many modern cryptographic protocols - such as the RSA algorithm - rely crucially on the inability to factor large numbers (200+ digits) in a practical amount of time.



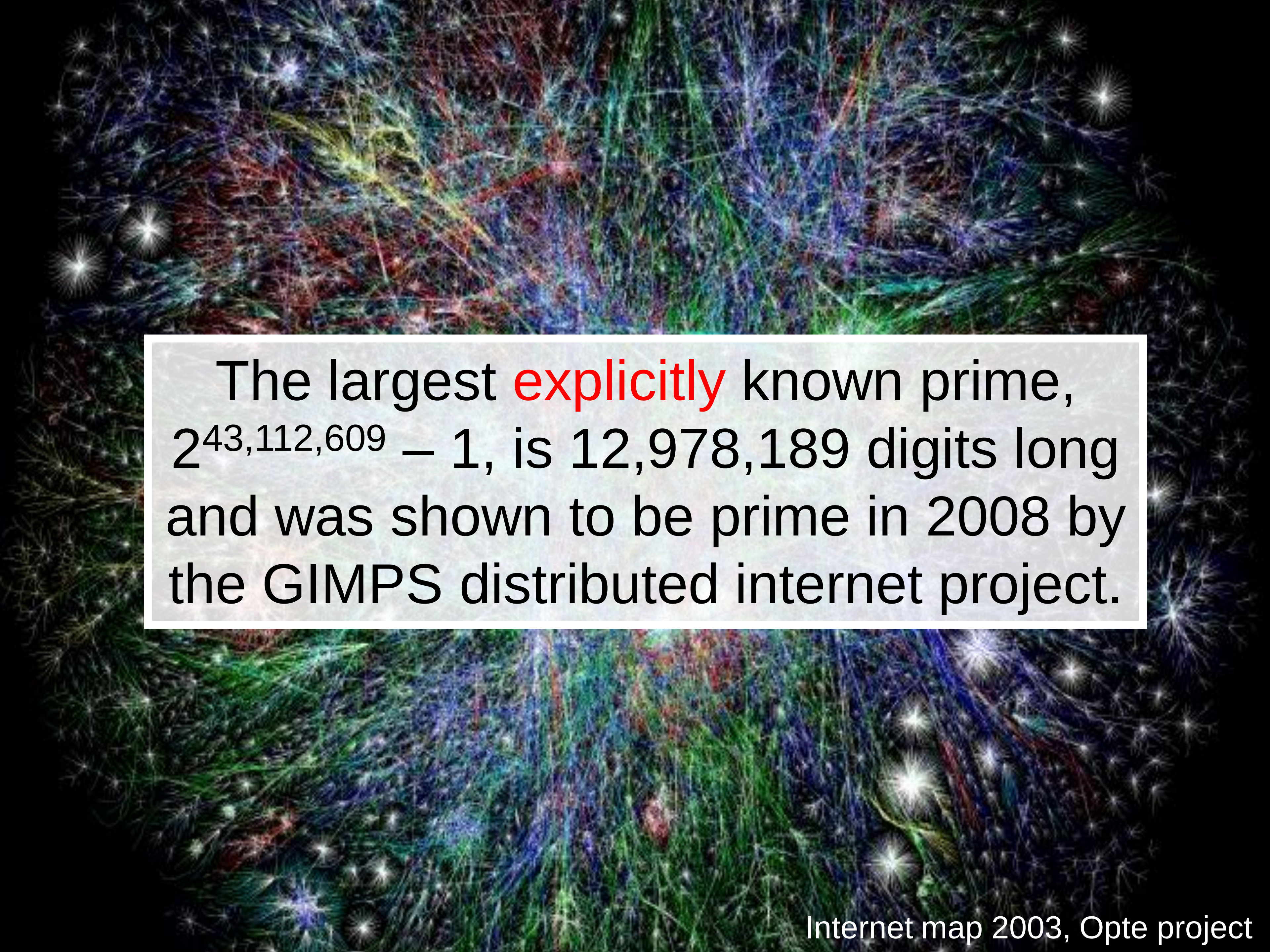
Public key encryption

Similarly, Euclid’s theorem tells us in principle that there are arbitrarily large primes out there, but does not give a recipe to find them.



Internet map 2003, Opte project

The largest explicitly known prime, $2^{43,112,609} - 1$, is 12,978,189 digits long and was shown to be prime in 2008 by the GIMPS distributed internet project.



Internet map 2003, Opte project

(3, 5), (5, 7), (11, 13), (17, 19), (29, 31), (41, 43), (59, 61), (71, 73), (101, 103), (107, 109), (137, 139), (149, 151), (179, 181), (191, 193), (197, 199), (227, 229), (239, 241), (269, 271), (281, 283), (311, 313), (347, 349), (419, 421), (431, 433), (461, 463), (521, 523), (569, 571), (599, 601), (617, 619), (641, 643), (659, 661), (809, 811), (821, 823), (827, 829), (857, 859), (881, 883), (1019, 1021), (1031, 1033), (1049, 1051), (1061, 1063), (1091, 1093), (1151, 1153), (1229, 1231), (1277, 1279), (1289, 1291), (1301, 1303), (1319, 1321), (1427, 1429), (1451, 1453), (1481, 1483), (1487, 1489), (1607, 1609), …

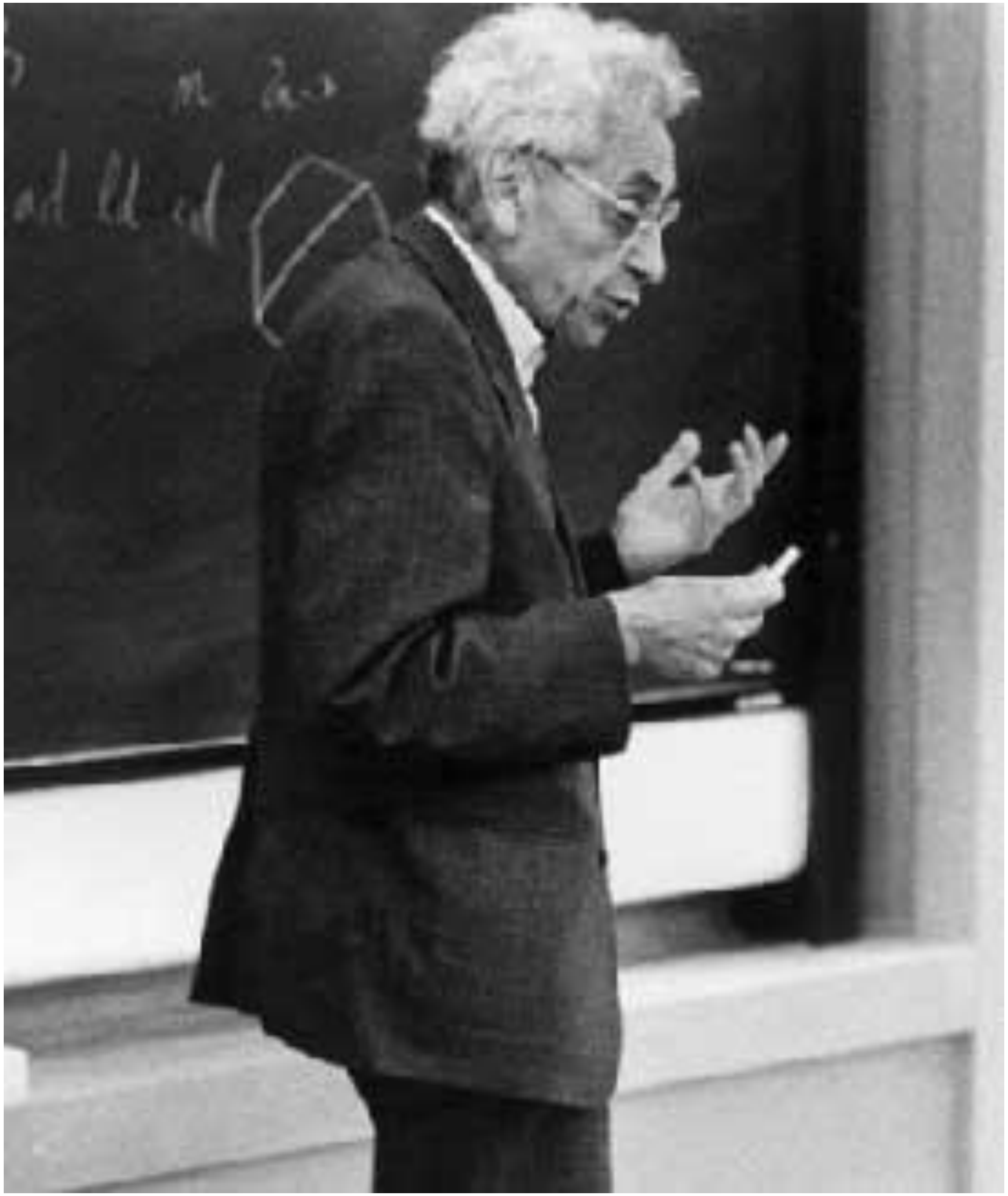
…, (3,756,801,695,685 x $2^{666,669}$ $\pm$1) [Winslow et al. 2011], ¿…?

Indeed, the prime numbers seem to be so "randomly" distributed that it is often difficult to establish what patterns exist within them. For instance, the following conjecture remains unproven:

(3, 5), (5, 7), (11, 13), (17, 19), (29, 31), (41, 43), (59, 61), (71, 73), (101, 103), (107, 109), (137, 139), (149, 151), (179, 181), (191, 193), (197, 199), (227, 229), (239, 241), (269, 271), (281, 283), (311, 313), (347, 349), (419, 421), (431, 433), (461, 463), (521, 523), (569, 571), (599, 601), (617, 619), (641, 643), (659, 661), (809, 811), (821, 823), (827, 829), (857, 859), (881, 883), (1019, 1021), (1031, 1033), (1049, 1051), (1061, 1063), (1091, 1093), (1151, 1153), (1229, 1231), (1277, 1279), (1289, 1291), (1301, 1303), (1319, 1321), (1427, 1429), (1451, 1453), (1481, 1483), (1487, 1489), (1607, 1609), …

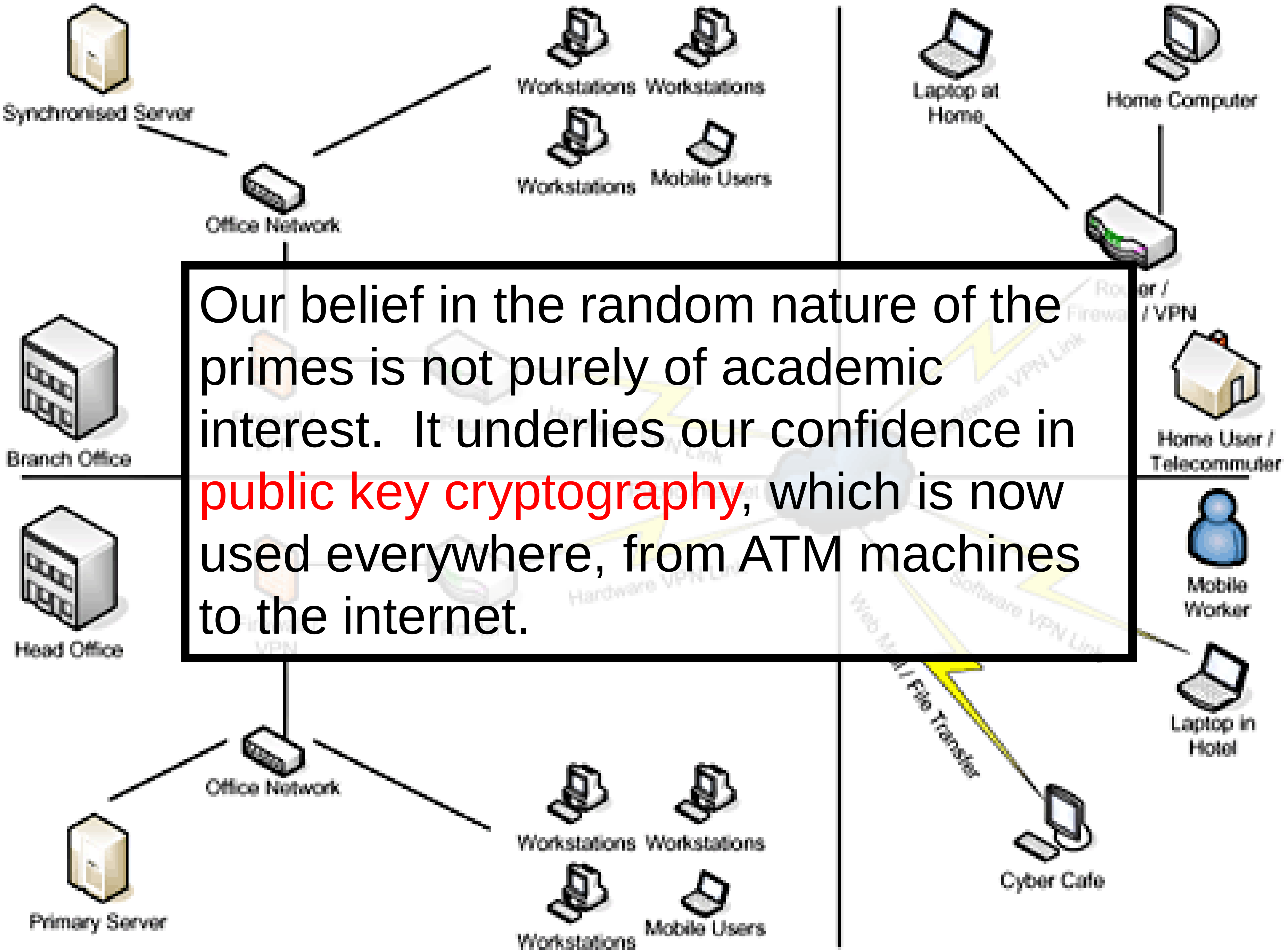
Twin prime conjecture (? ~300 BCE ?): There exist infinitely many pairs p, p+2 of primes which differ by exactly 2.

…, (3,756,801,695,685 x $2^{666,669}$ $\pm$1) [Winslow et al. 2011], ¿…?



“God may not play dice with the universe, but something strange is going on with the prime numbers”.

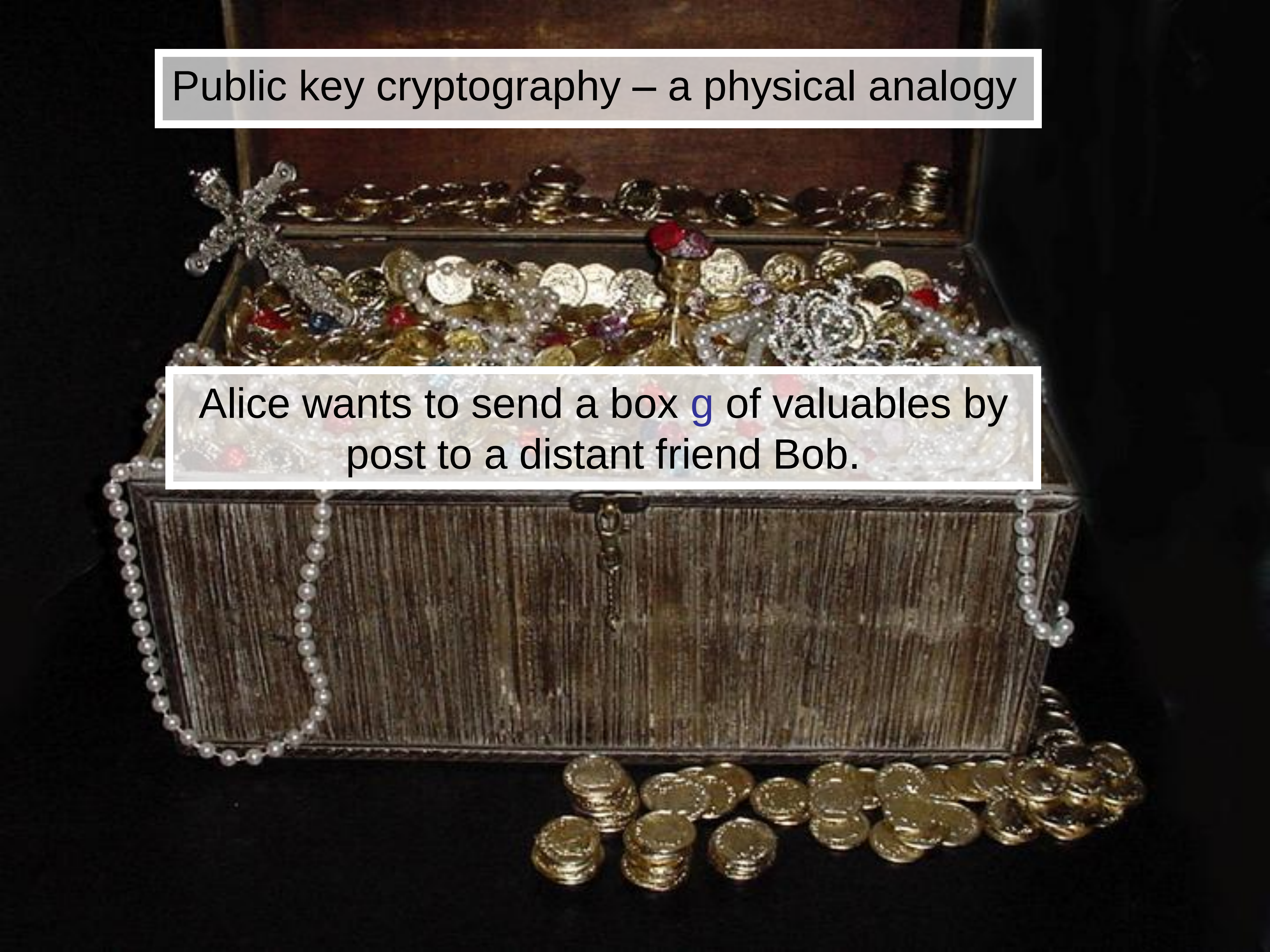
(Paul Erdős, 1913-1996)



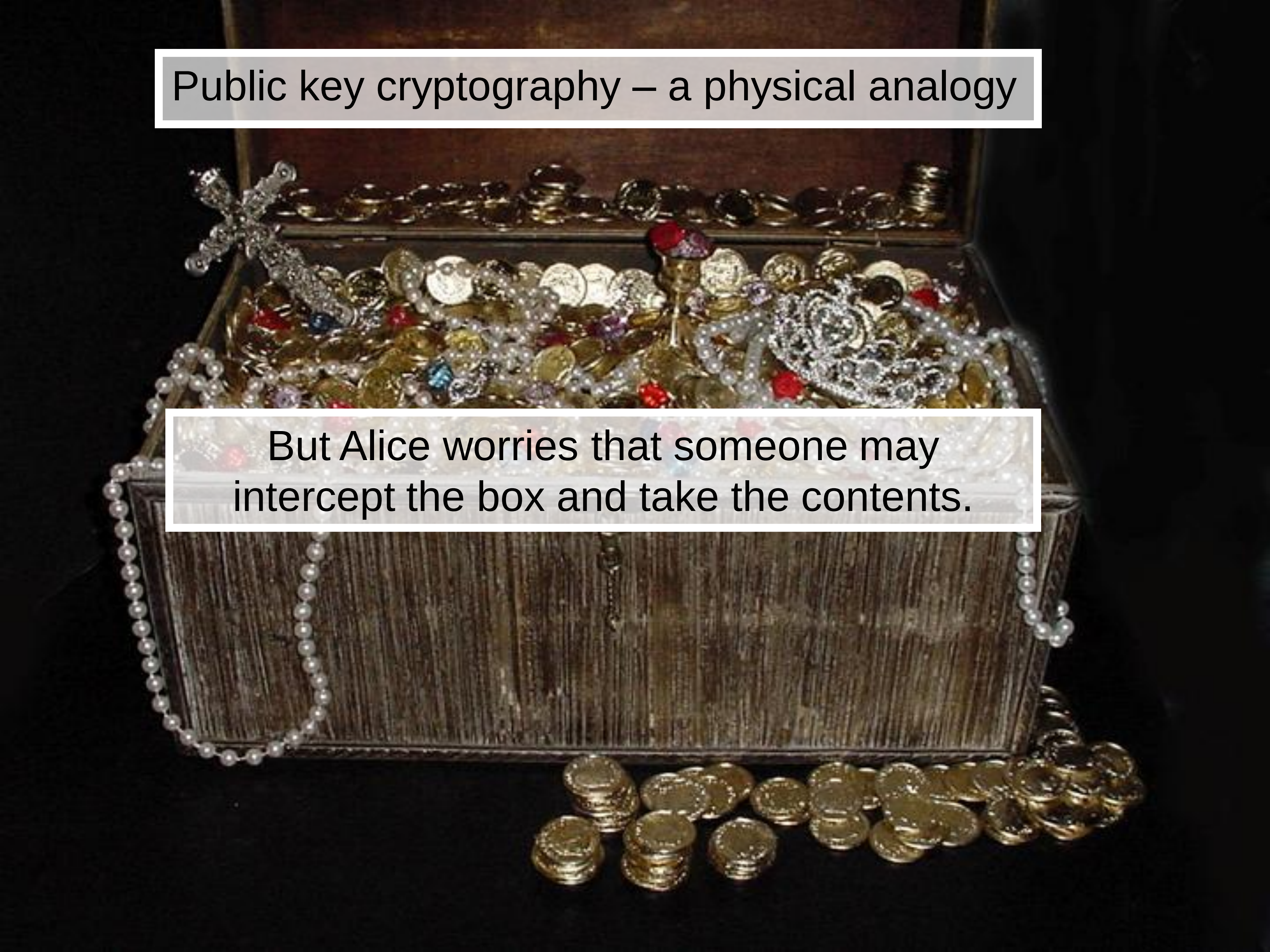
Our belief in the random nature of the primes is not purely of academic interest. It underlies our confidence in public key cryptography, which is now used everywhere, from ATM machines to the internet.

Public key cryptography – a physical analogy

Alice wants to send a box g of valuables by post to a distant friend Bob.

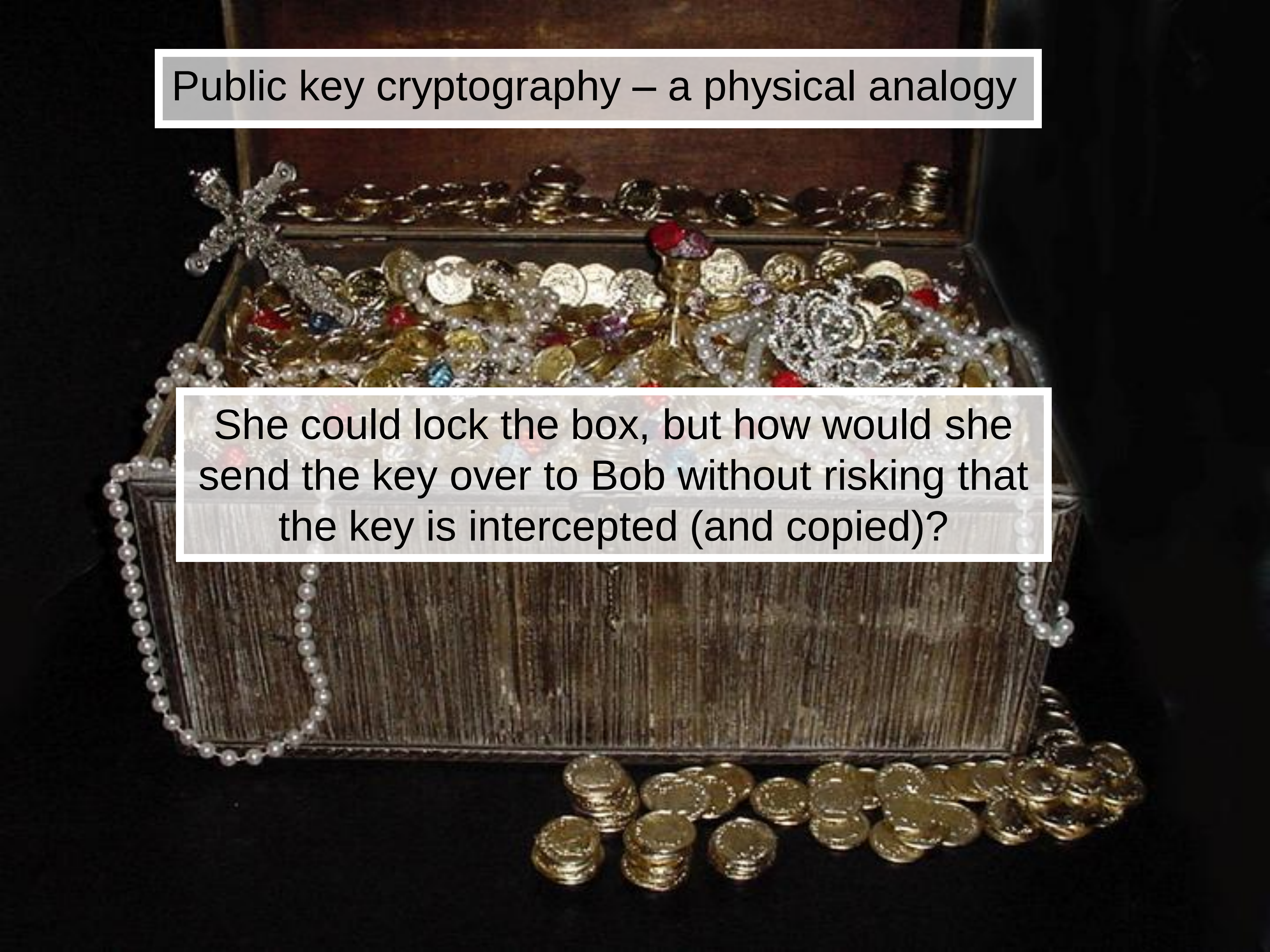


Public key cryptography – a physical analogy



But Alice worries that someone may intercept the box and take the contents.

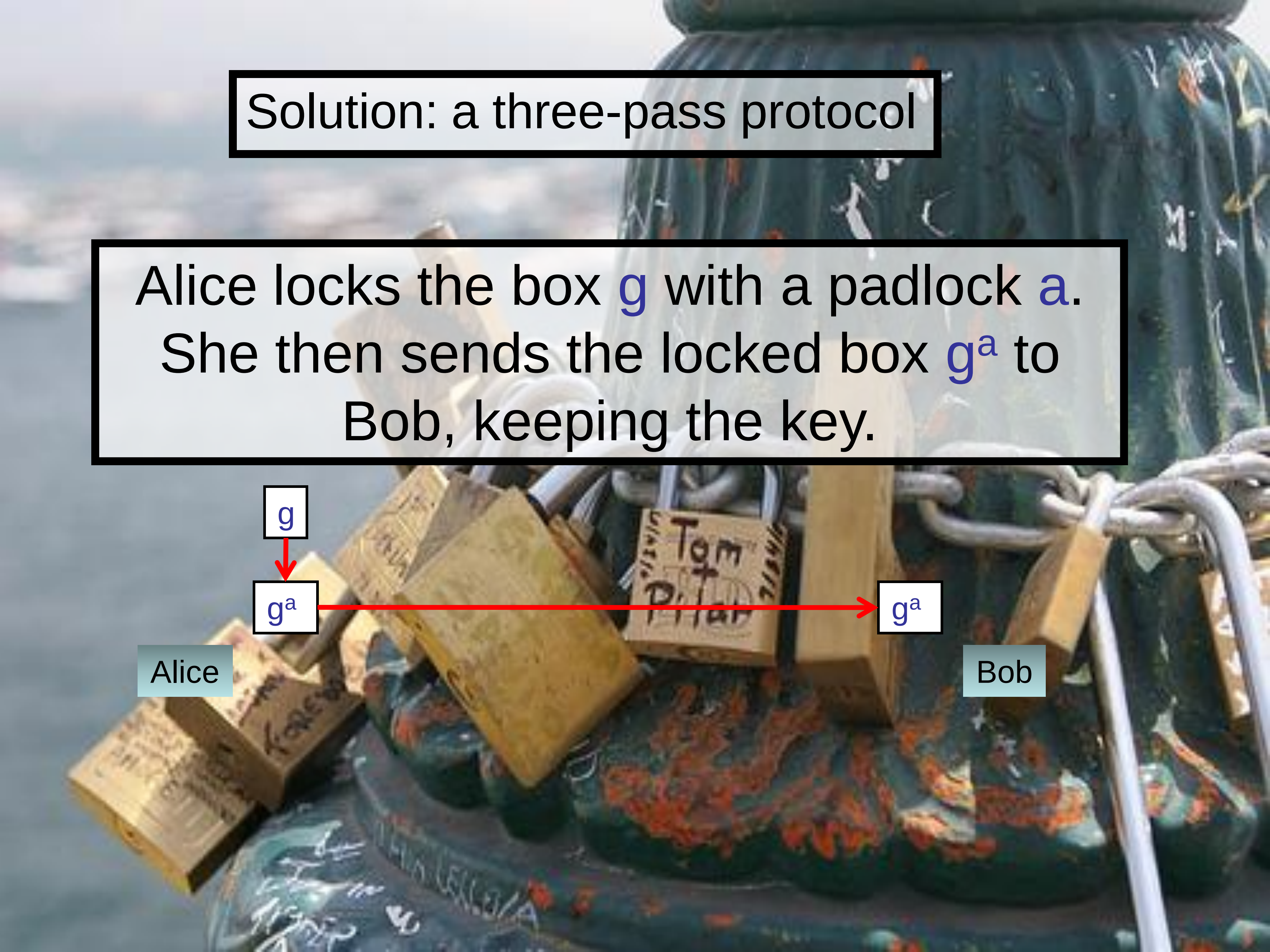
Public key cryptography – a physical analogy



She could lock the box, but how would she send the key over to Bob without risking that the key is intercepted (and copied)?

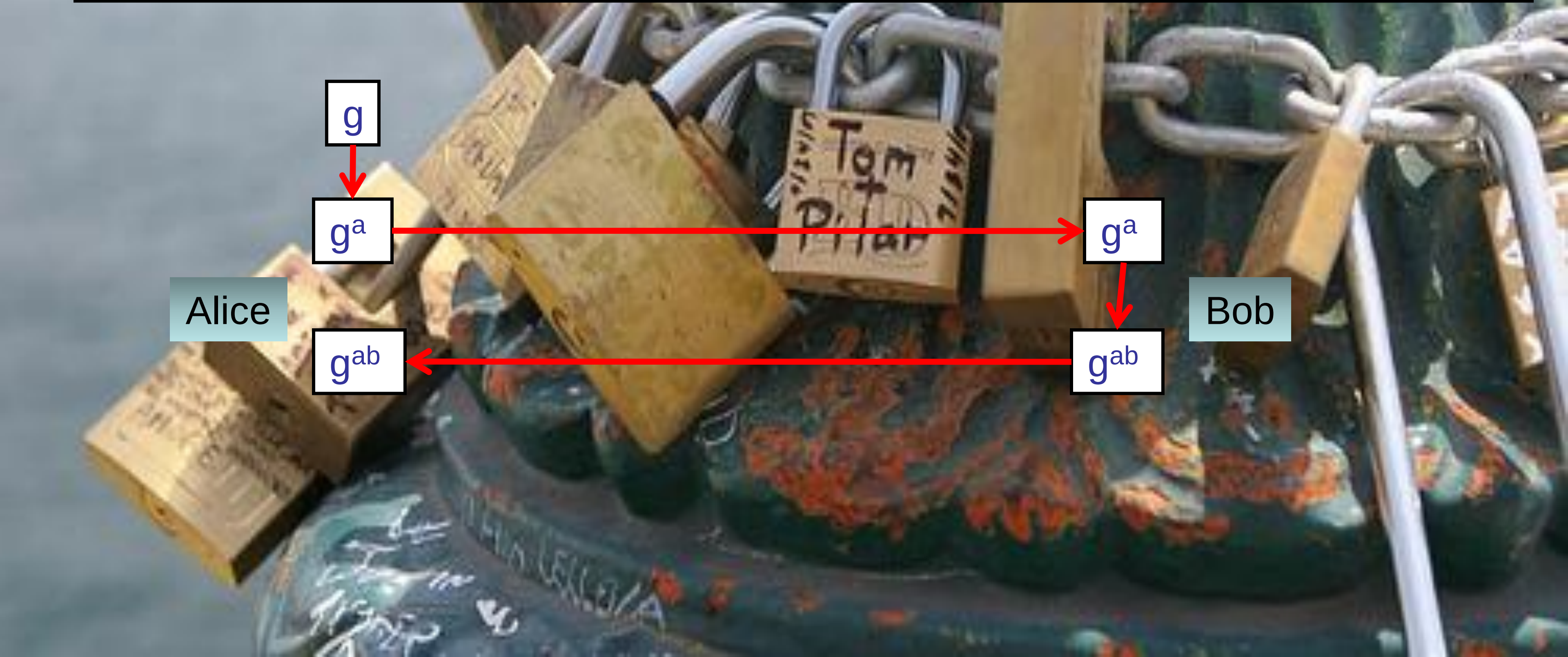
Solution: a three-pass protocol

Alice locks the box g with a padlock a. She then sends the locked box g^a to Bob, keeping the key.



Solution: a three-pass protocol

Bob cannot unlock the box… but he can put his own padlock b on the box. He then sends the doubly locked box g^{ab} back to Alice.

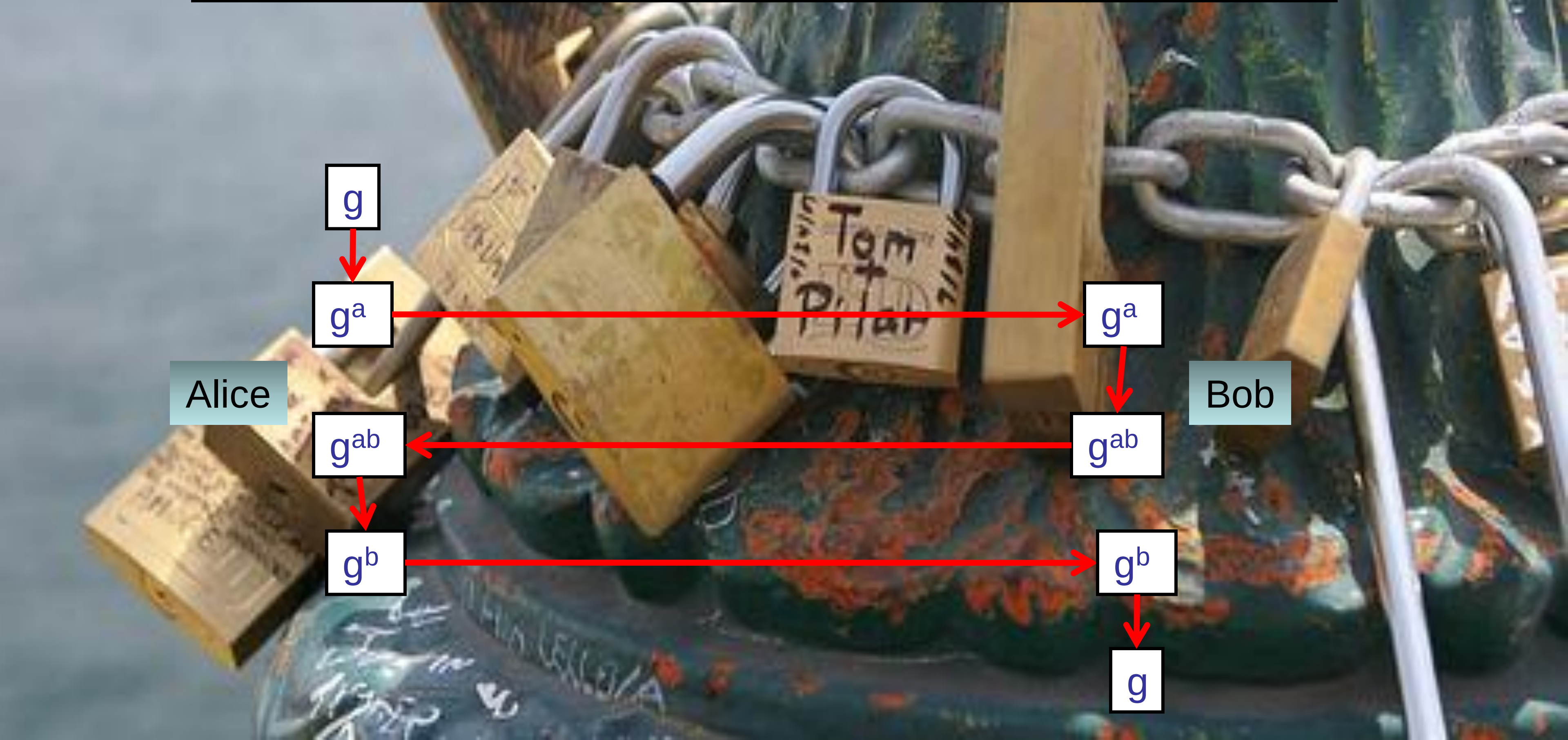


Solution: a three-pass protocol

Alice can't unlock Bob's padlock… but she can unlock her own! She then sends the singly locked box g^b back to Bob.

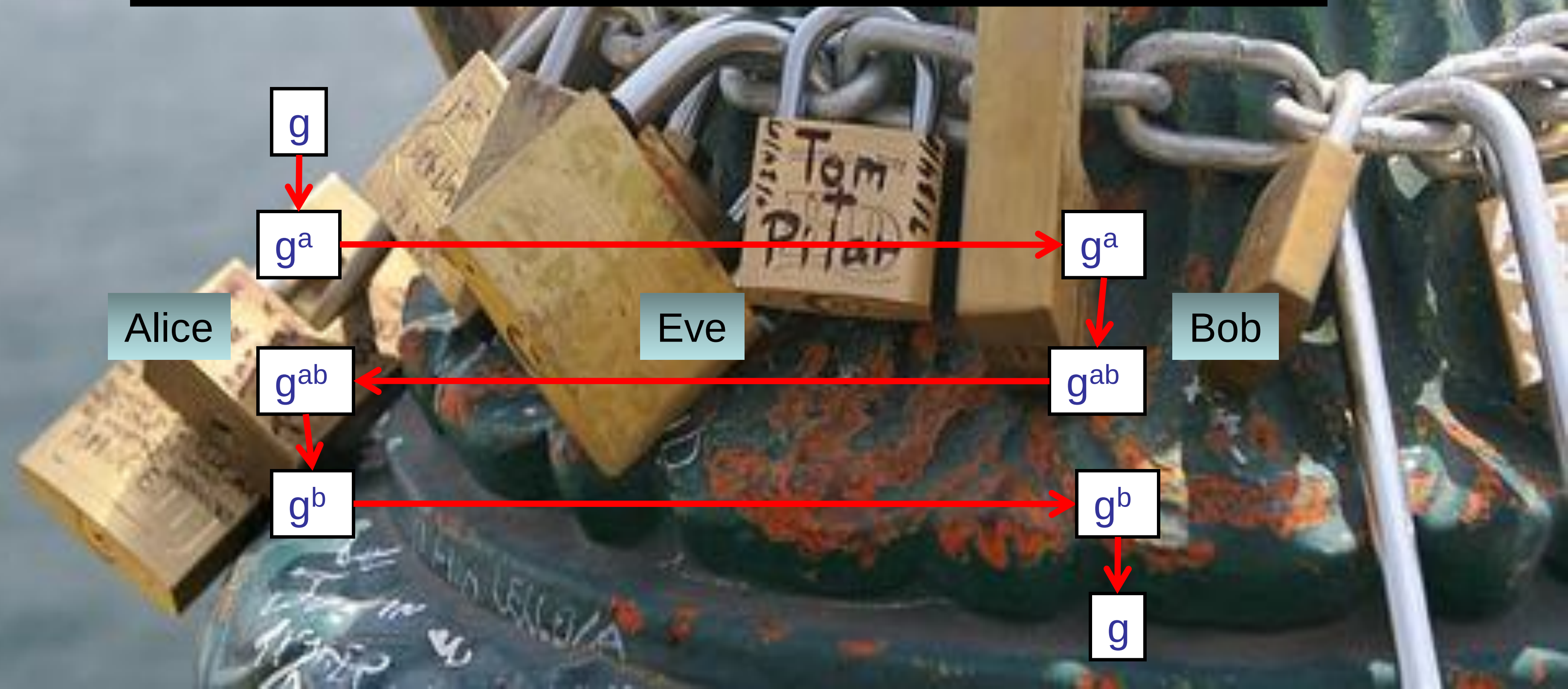
Solution: a three-pass protocol

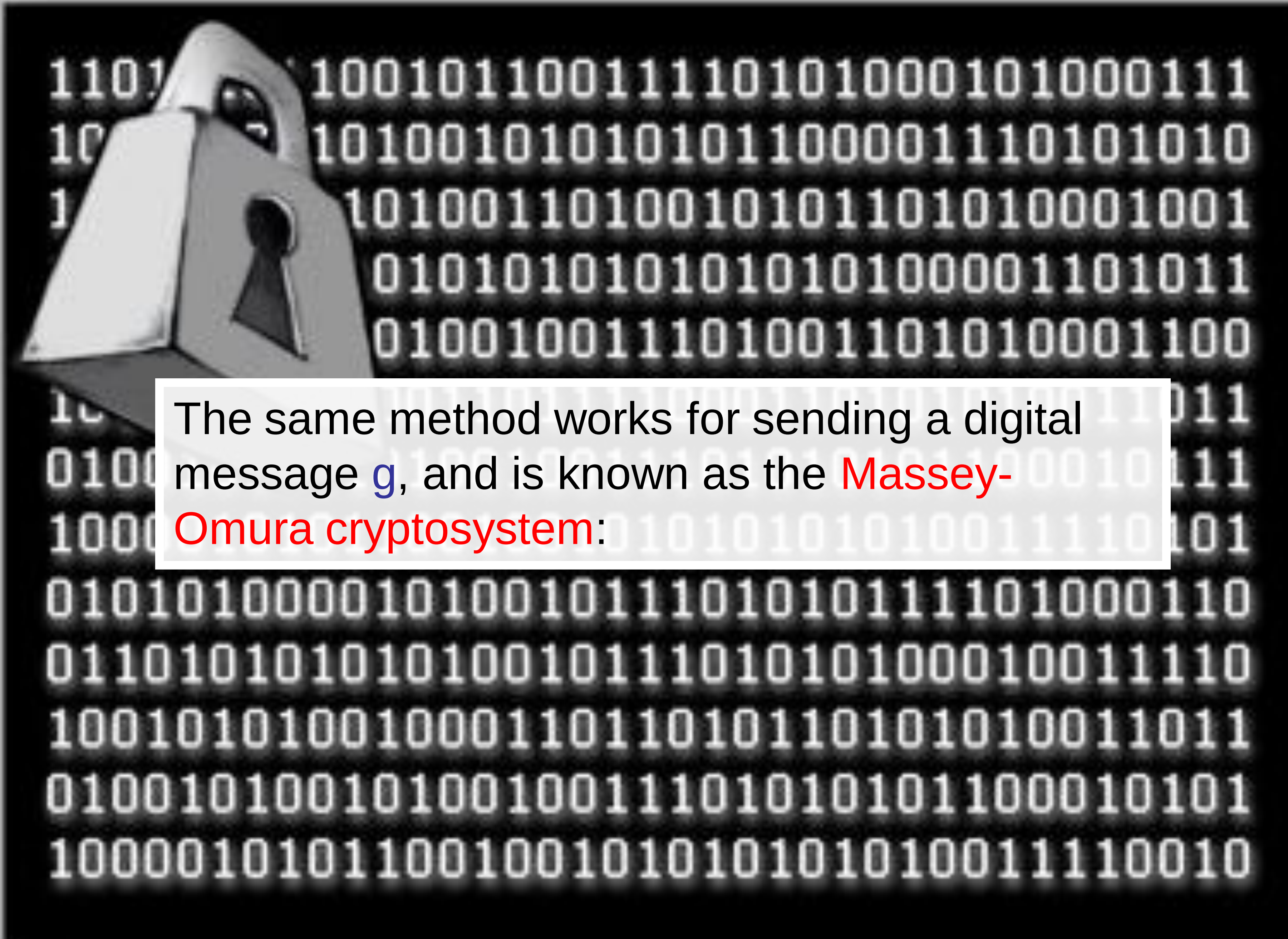
Bob then unlocks his own lock and opens the box.



Solution: a three-pass protocol

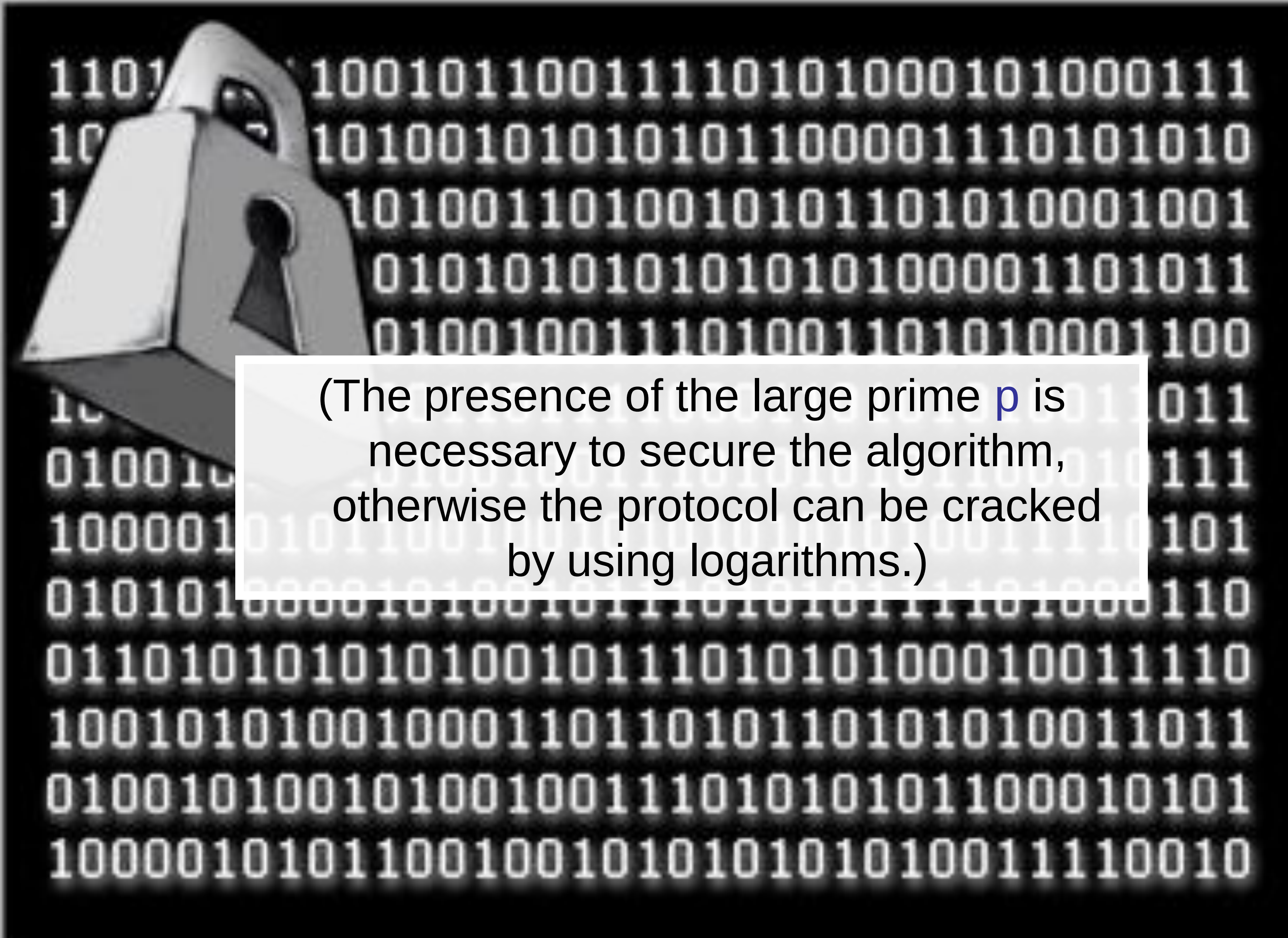
An eavesdropper would see the locked boxes g^a, g^{ab}, g^b… but never the unlocked box g.



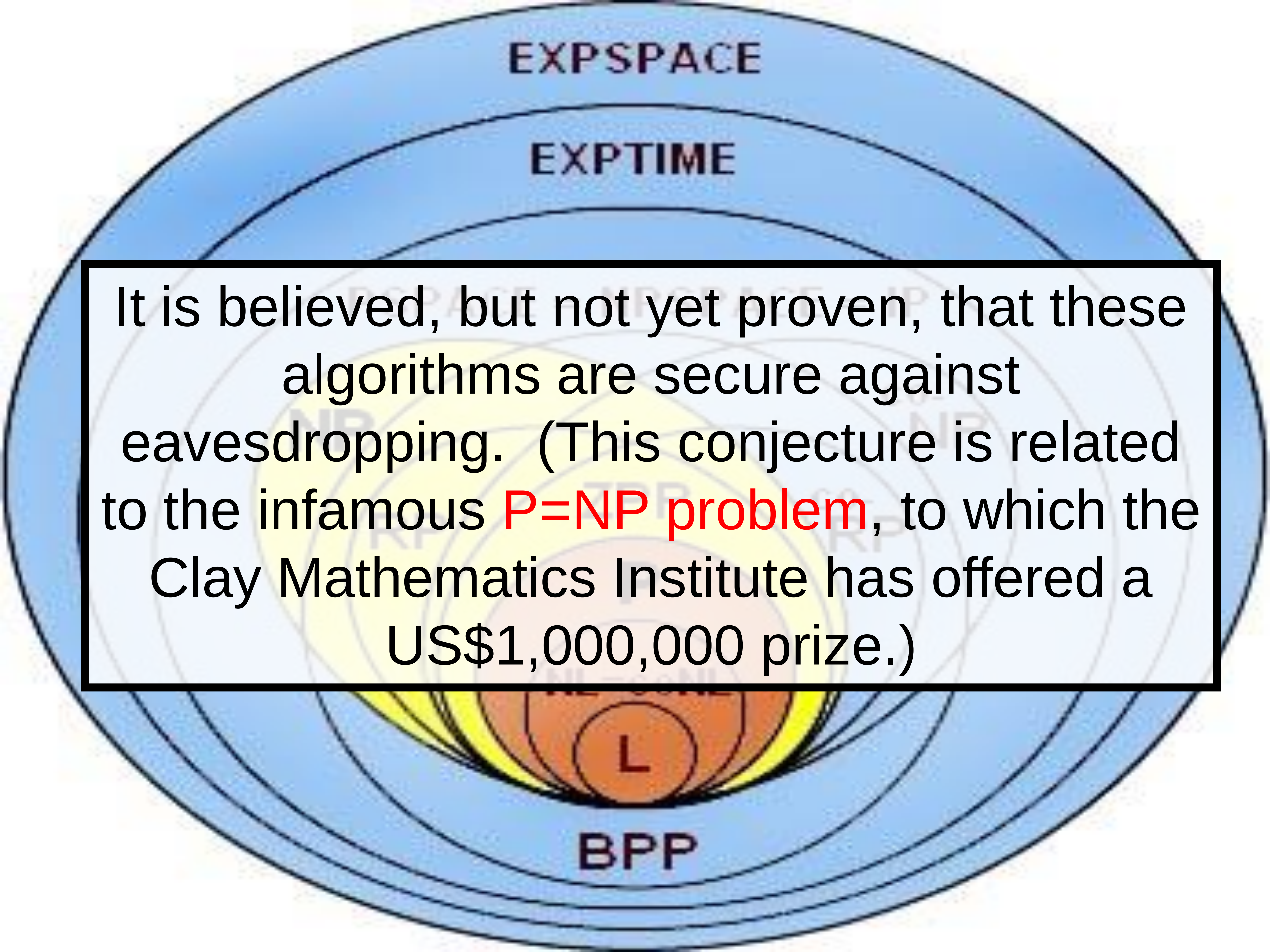


The same method works for sending a digital message g, and is known as the Massey-Omura cryptosystem:

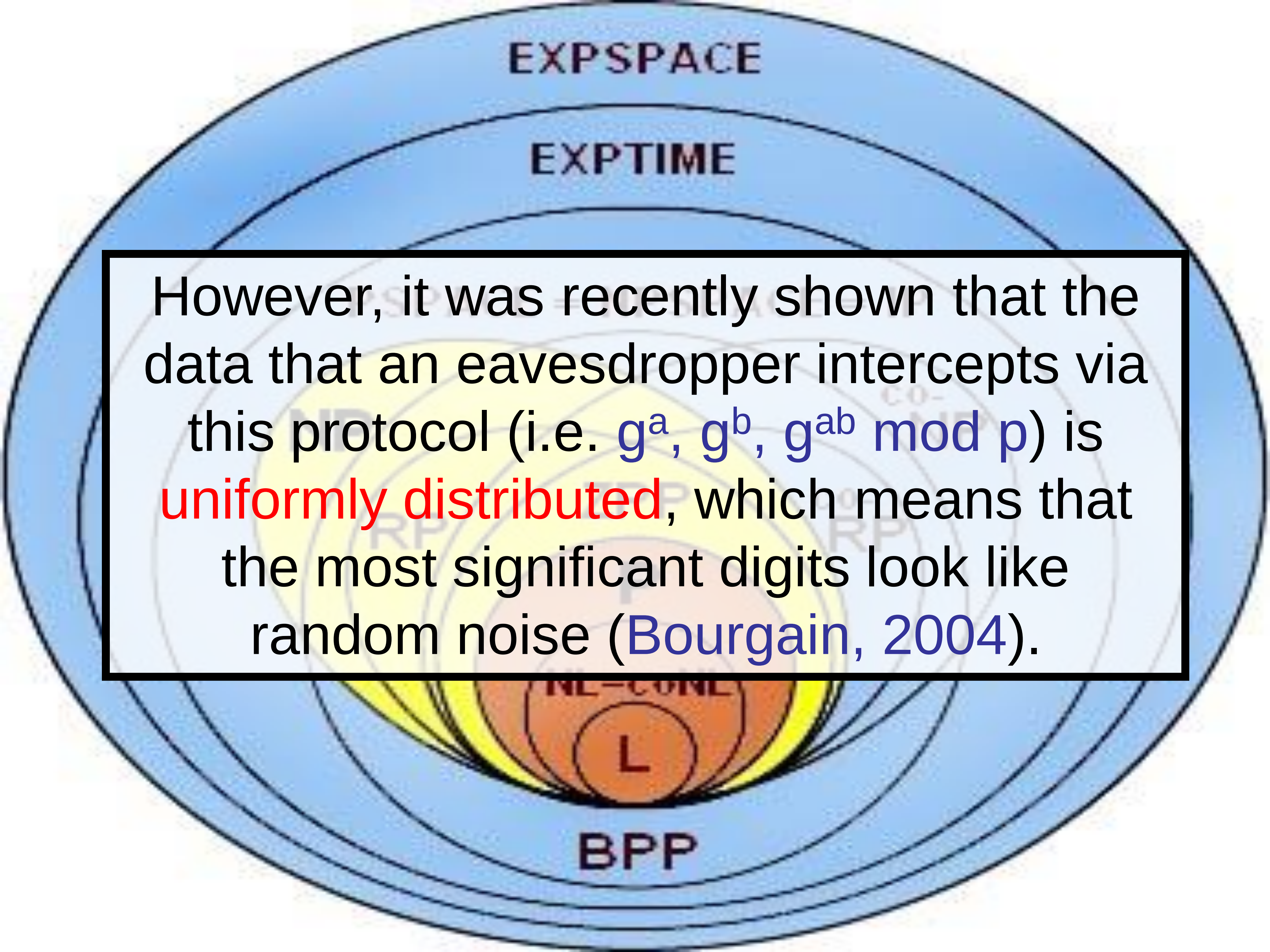
1. Alice and Bob agree (publicly) on a large prime p.
2. Alice "locks" g by raising it to the power a for some secretly chosen a. She then sends $g^a \bmod p$ to Bob.
3. Bob "locks" the message by raising to his own power b, and sends $g^{ab} \bmod p$ back to Alice.
4. Alice takes an a^{th} root to obtain $g^b \bmod p$, which she sends back to Bob.
5. Bob takes a b^{th} root to recover g.



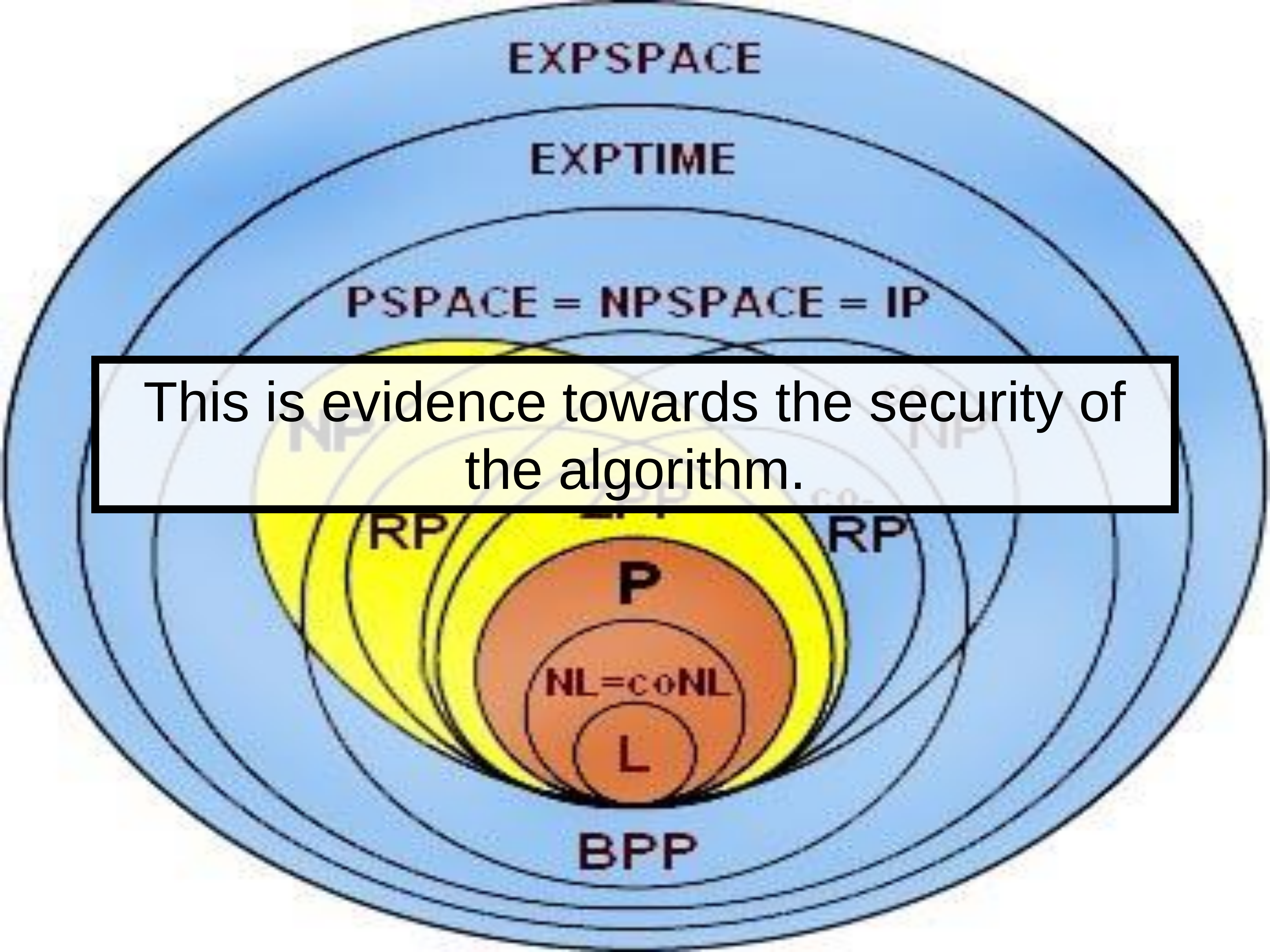
(The presence of the large prime p is necessary to secure the algorithm, otherwise the protocol can be cracked by using logarithms.)



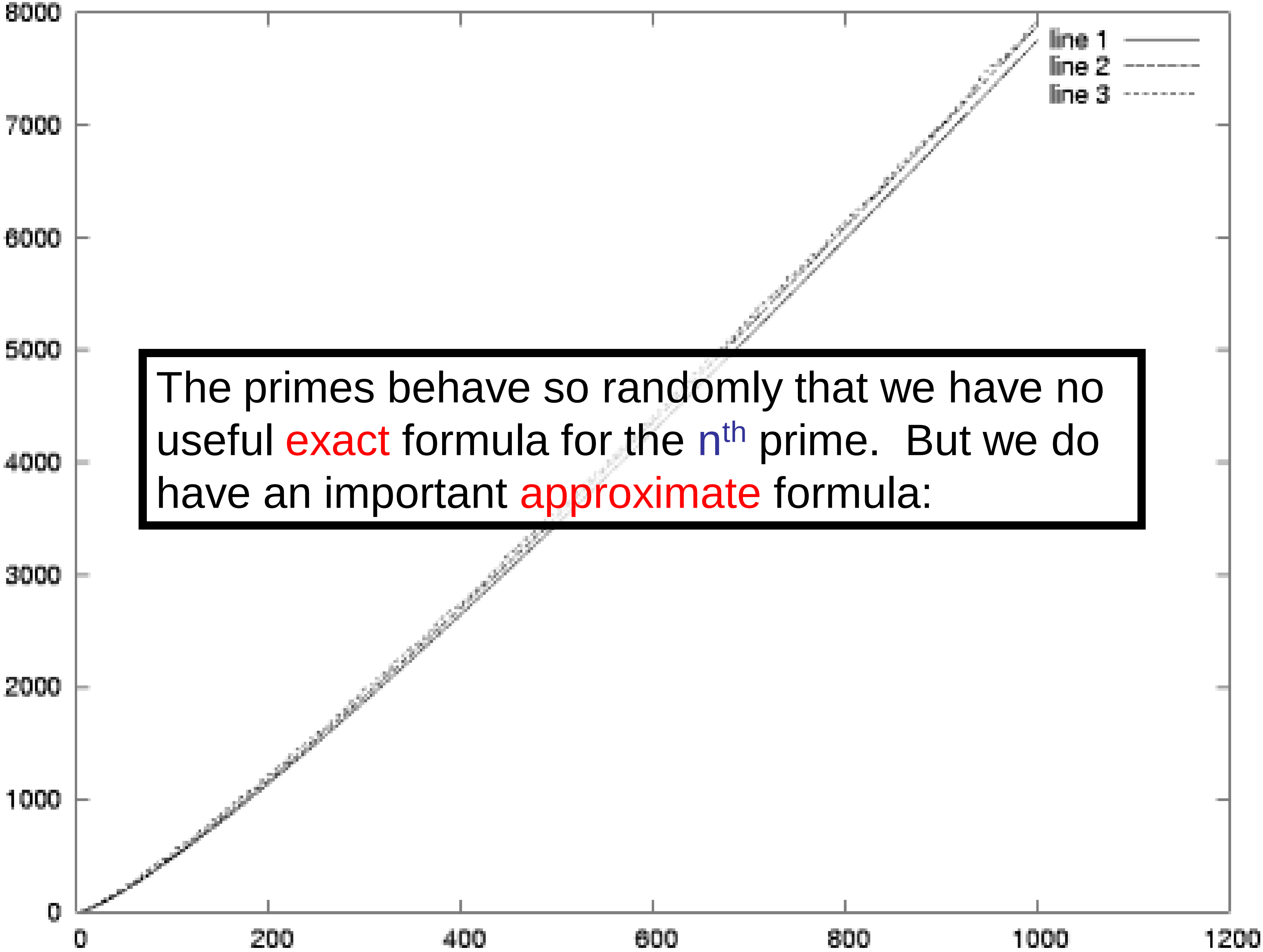
It is believed, but not yet proven, that these algorithms are secure against eavesdropping. (This conjecture is related to the infamous P=NP problem, to which the Clay Mathematics Institute has offered a US$1,000,000 prize.)



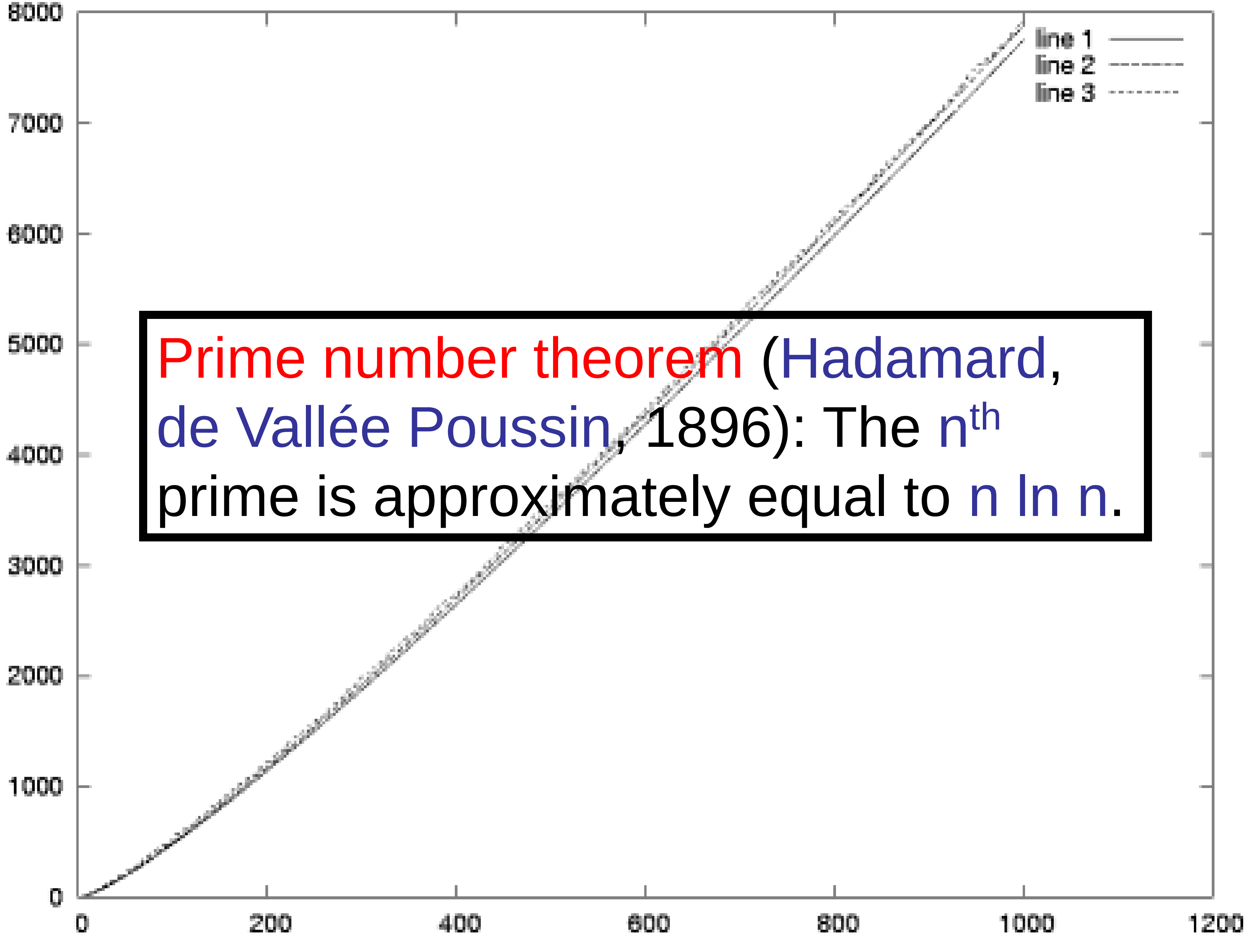
However, it was recently shown that the data that an eavesdropper intercepts via this protocol (i.e. g^a, g^b, $g^{ab} \bmod p$) is uniformly distributed, which means that the most significant digits look like random noise (Bourgain, 2004).

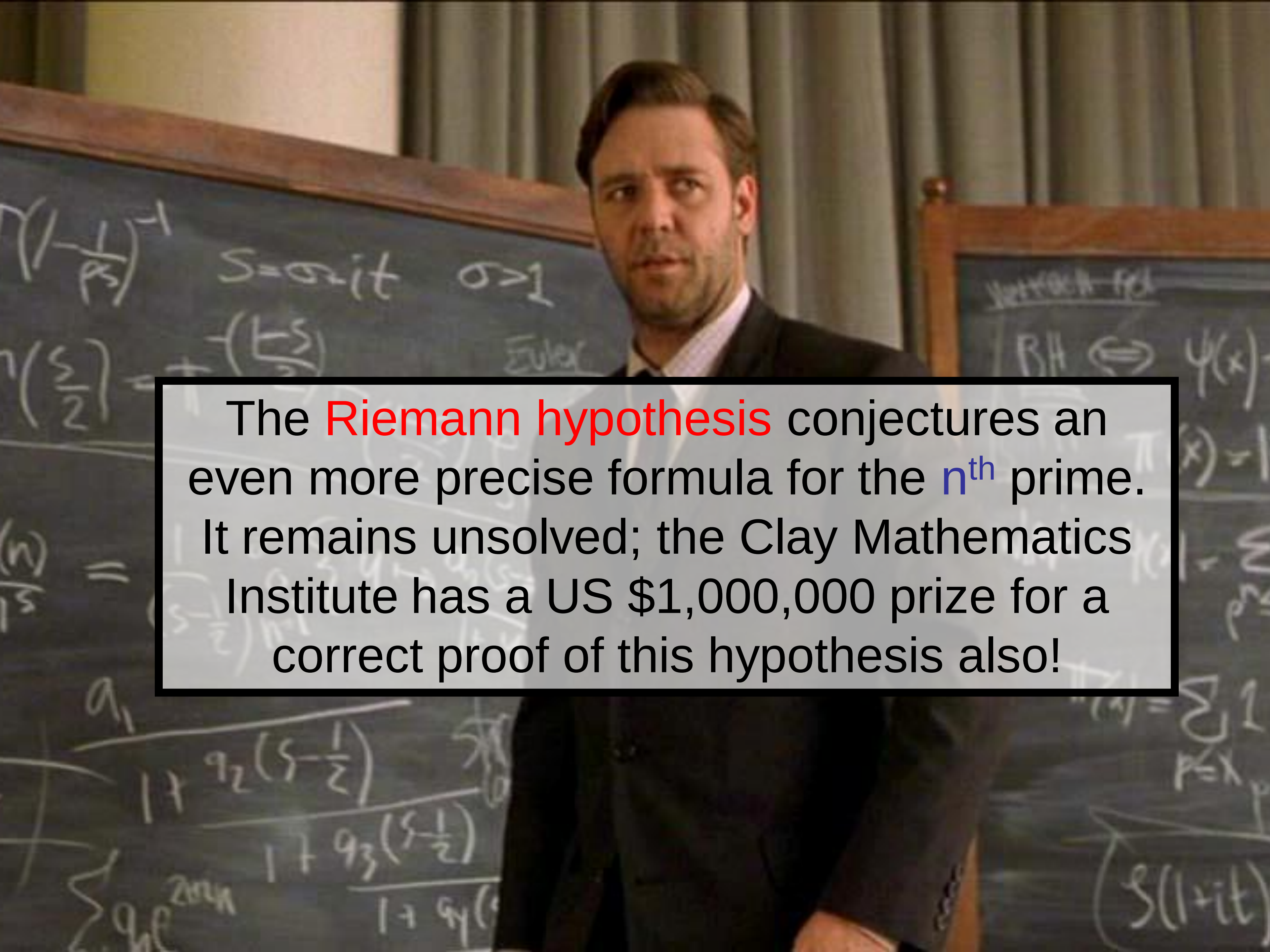


This is evidence towards the security of the algorithm.

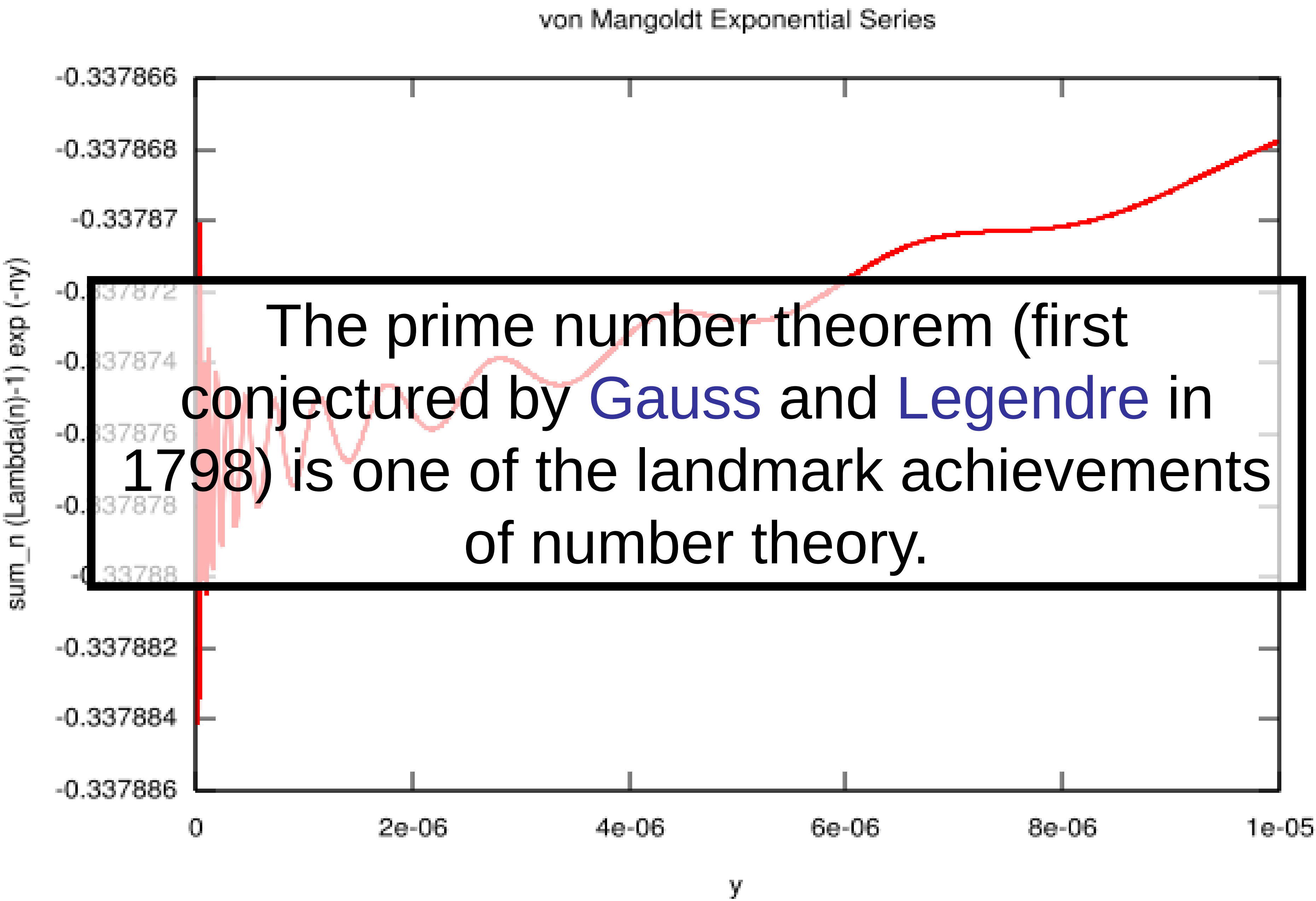


The primes behave so randomly that we have no useful exact formula for the n^{th} prime. But we do have an important approximate formula:

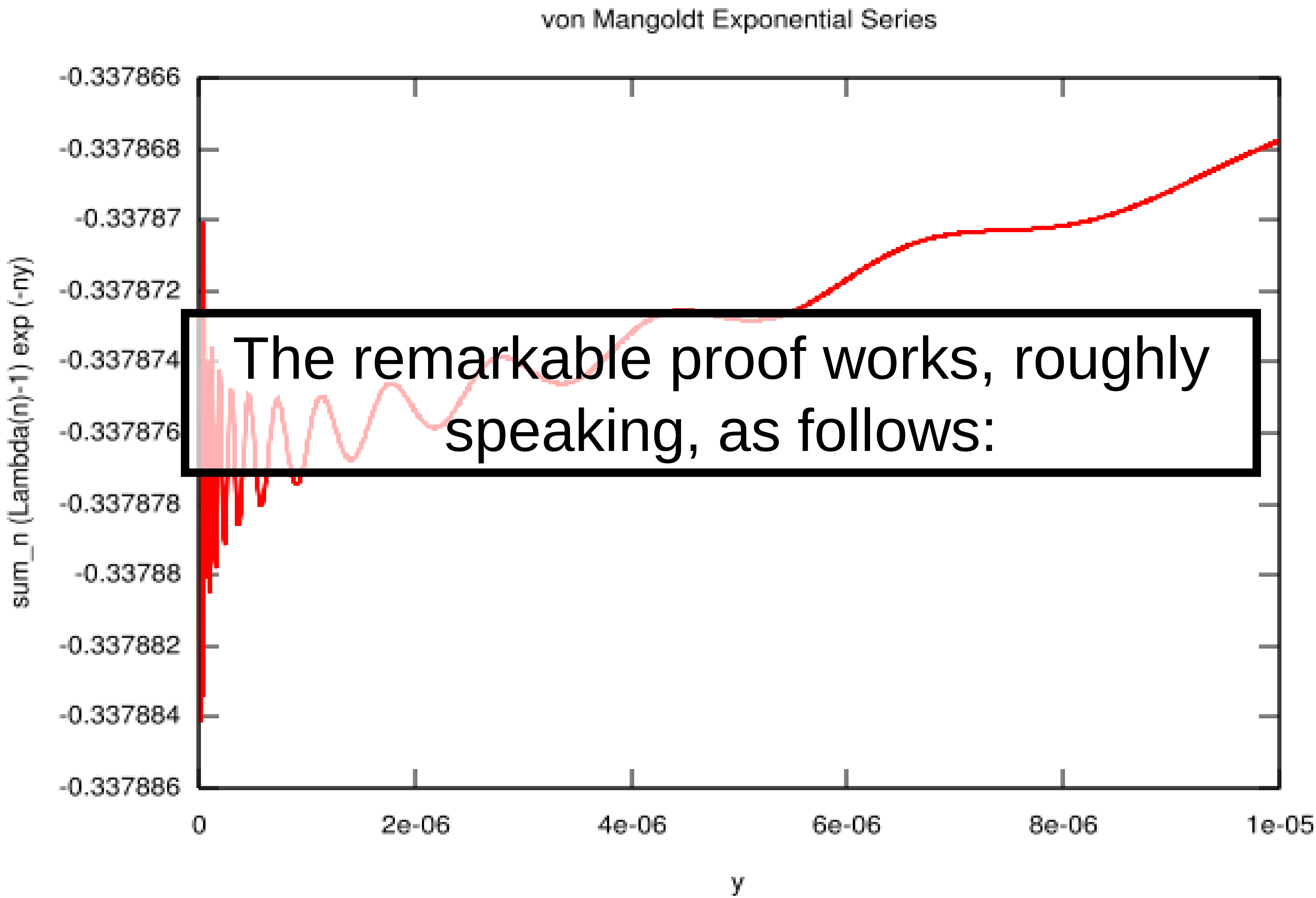

8000
7000
6000
5000
4000
3000
2000
1000
0
0
200
400
600
800
1000
1200
line 1
line 2
line 3
Prime number theorem (Hadamard, de Vallée Poussin, 1896): The nth prime is approximately equal to n ln n.



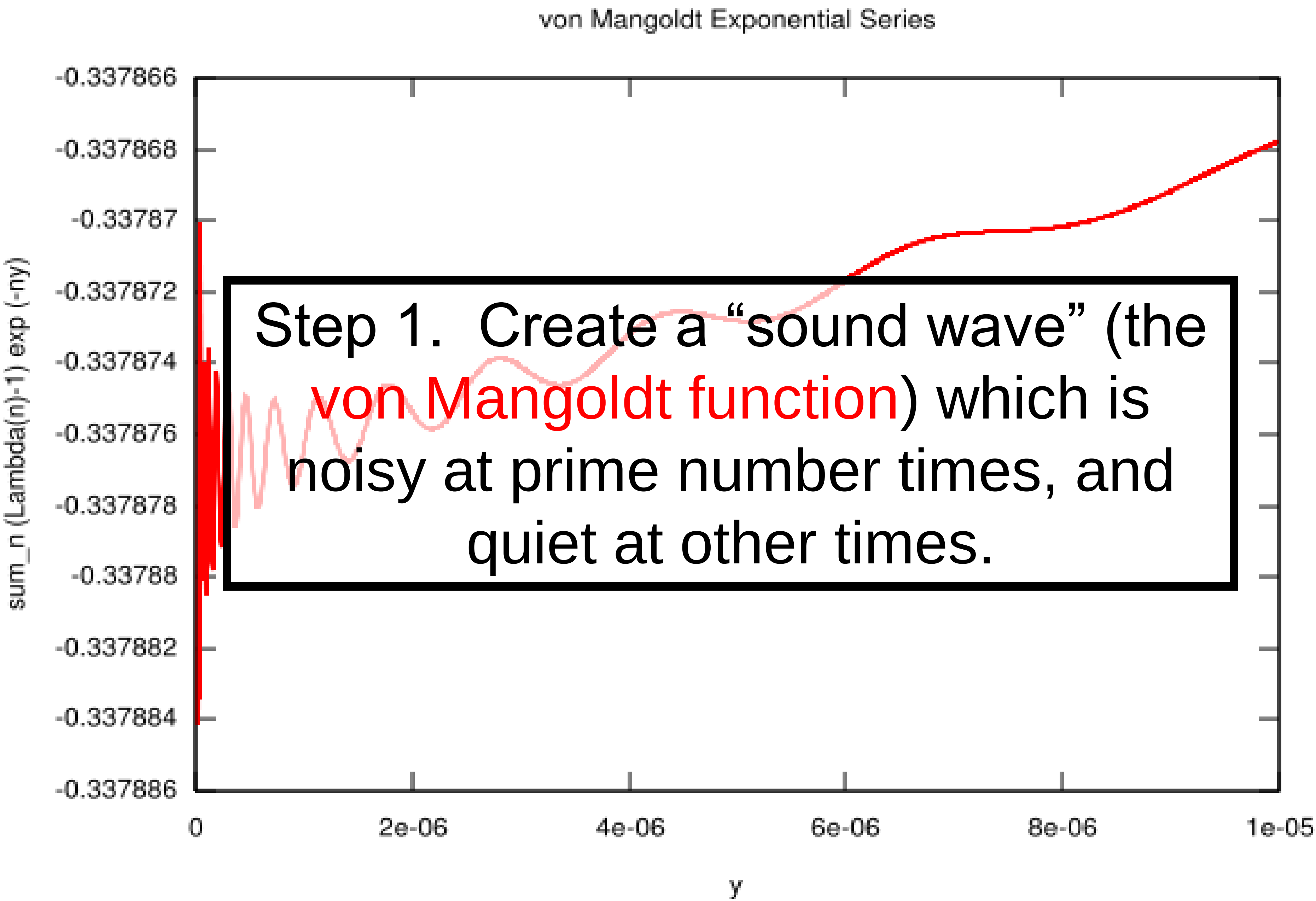
The Riemann hypothesis conjectures an even more precise formula for the n^{th} prime. It remains unsolved; the Clay Mathematics Institute has a US $1,000,000 prize for a correct proof of this hypothesis also!

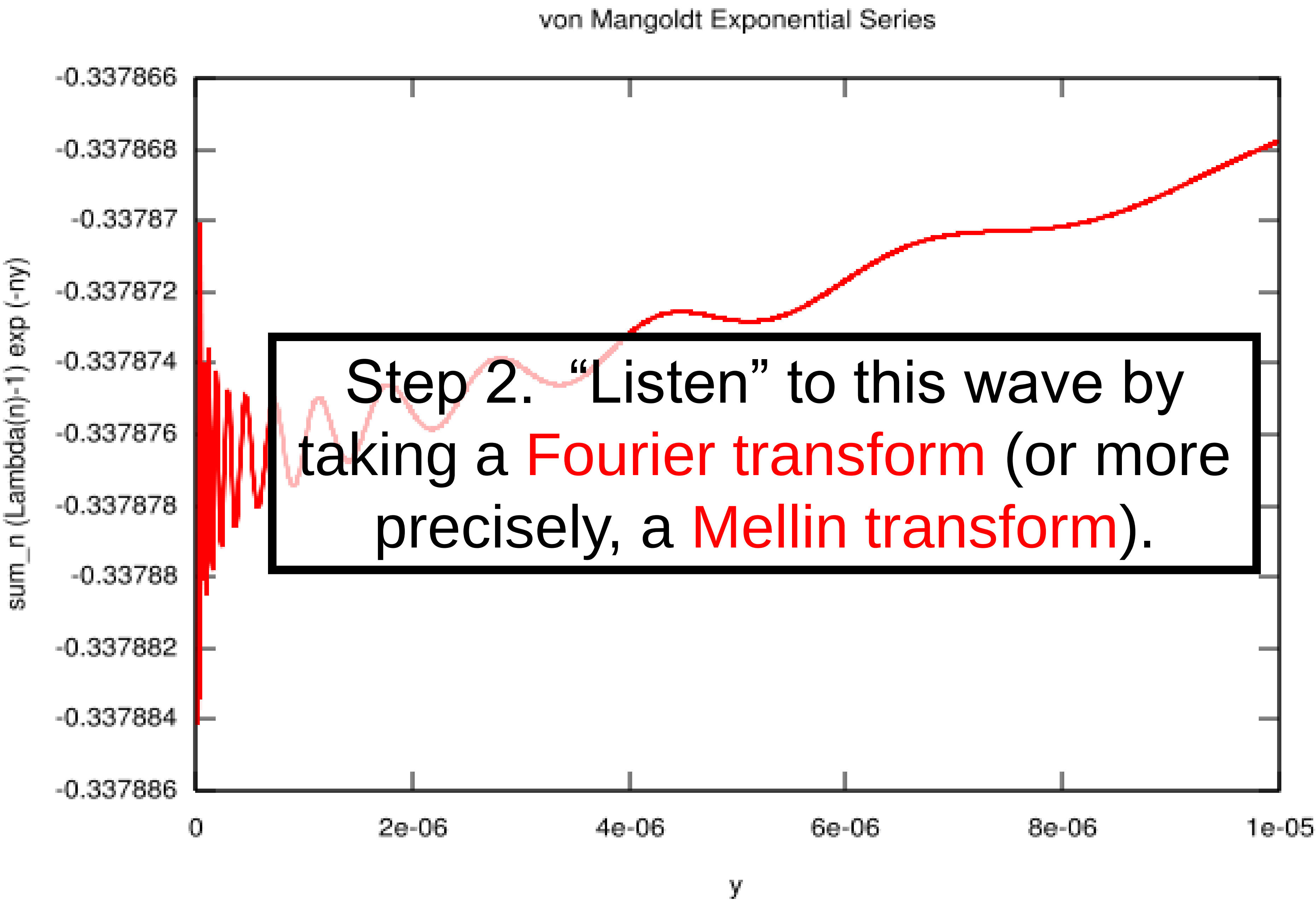


The prime number theorem (first conjectured by Gauss and Legendre in 1798) is one of the landmark achievements of number theory.

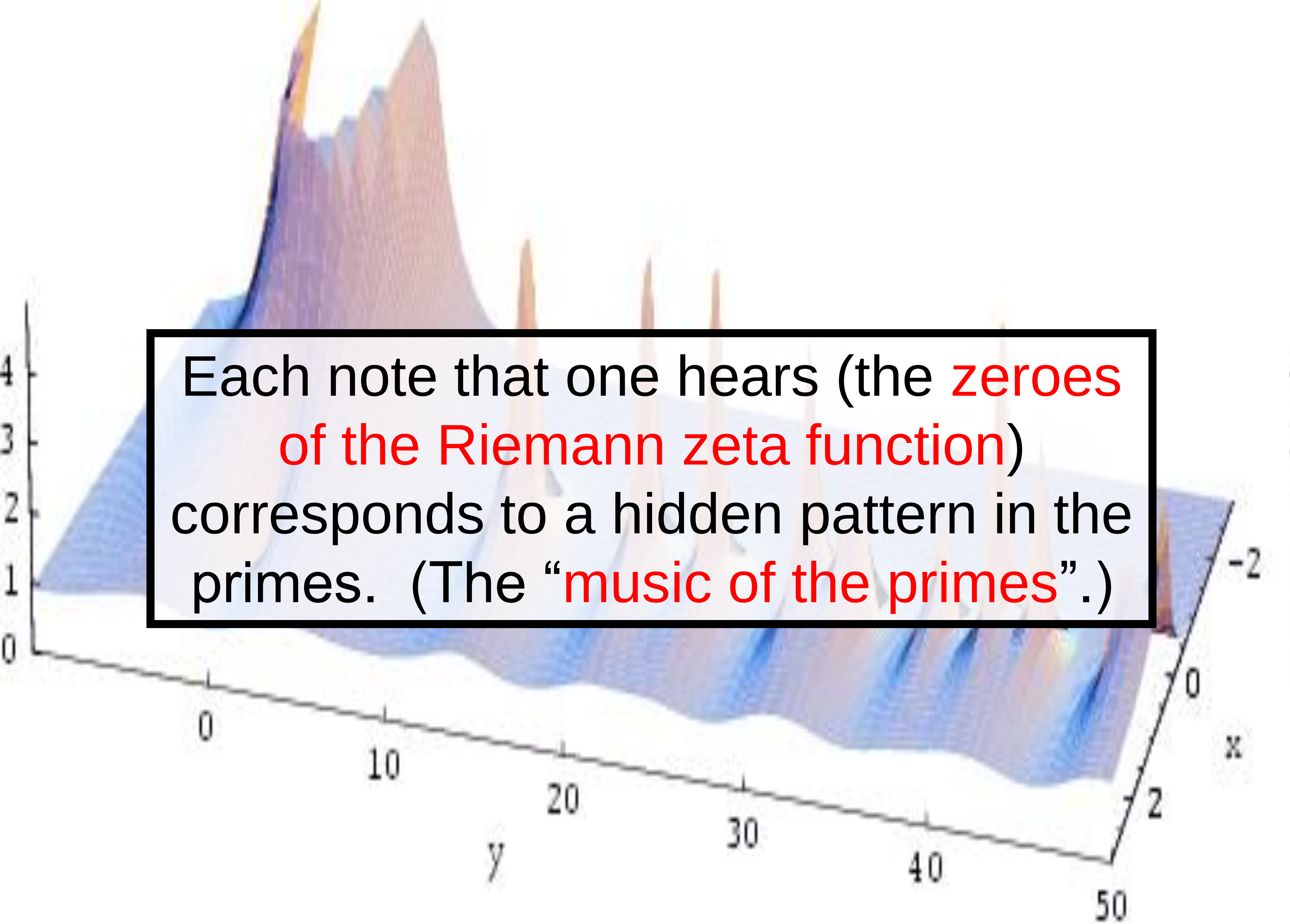


The remarkable proof works, roughly speaking, as follows:


von Mangoldt Exponential Series
sum_n (Lambda(n)-1) exp (-ny)
y
-0.337866
-0.337868
-0.33787
-0.337872
-0.337874
-0.337876
-0.337878
-0.33788
-0.337882
-0.337884
-0.337886
0
2e-06
4e-06
6e-06
8e-06
1e-05
Step 1. Create a “sound wave” (the von Mangoldt function) which is noisy at prime number times, and quiet at other times.

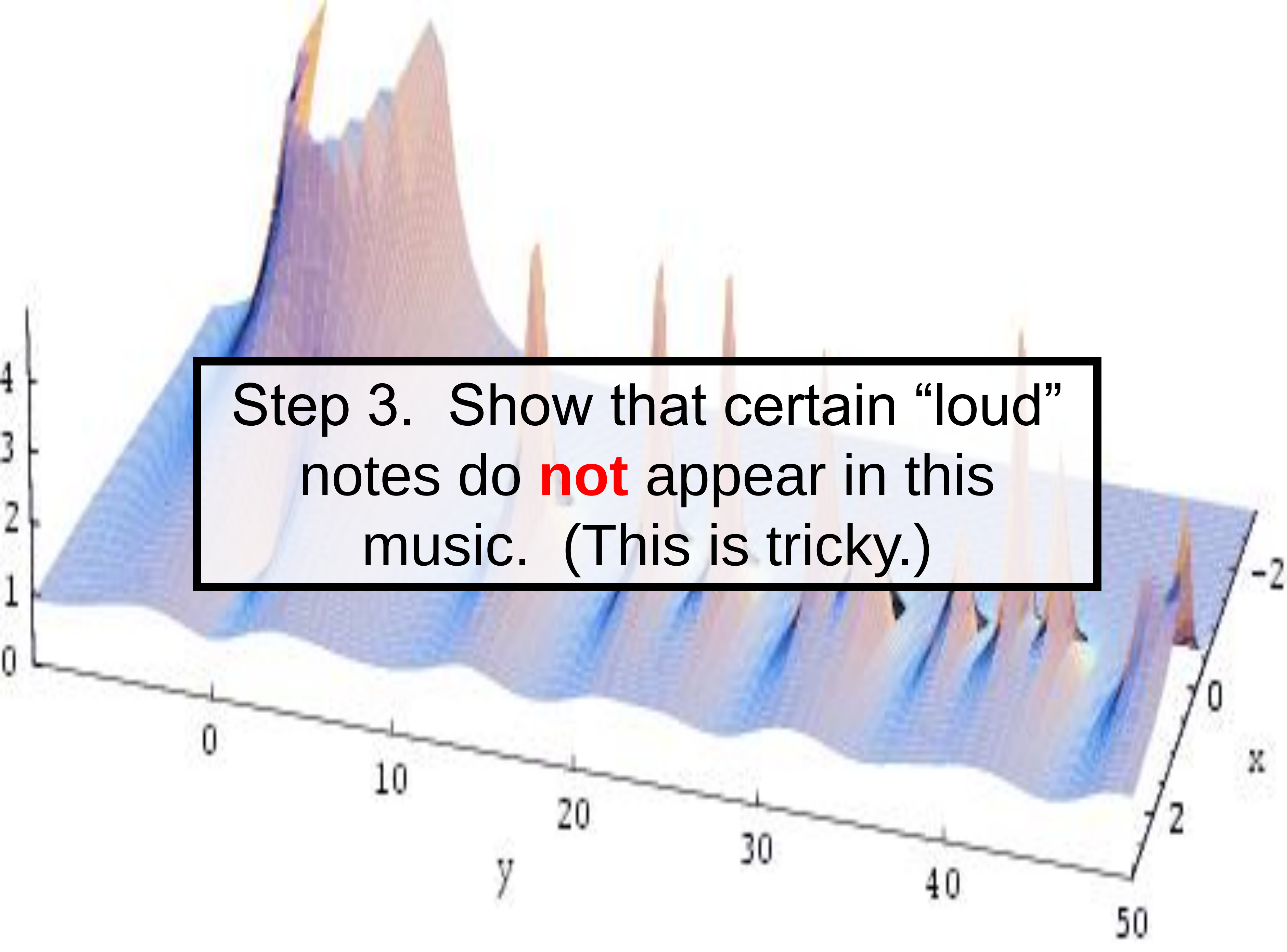


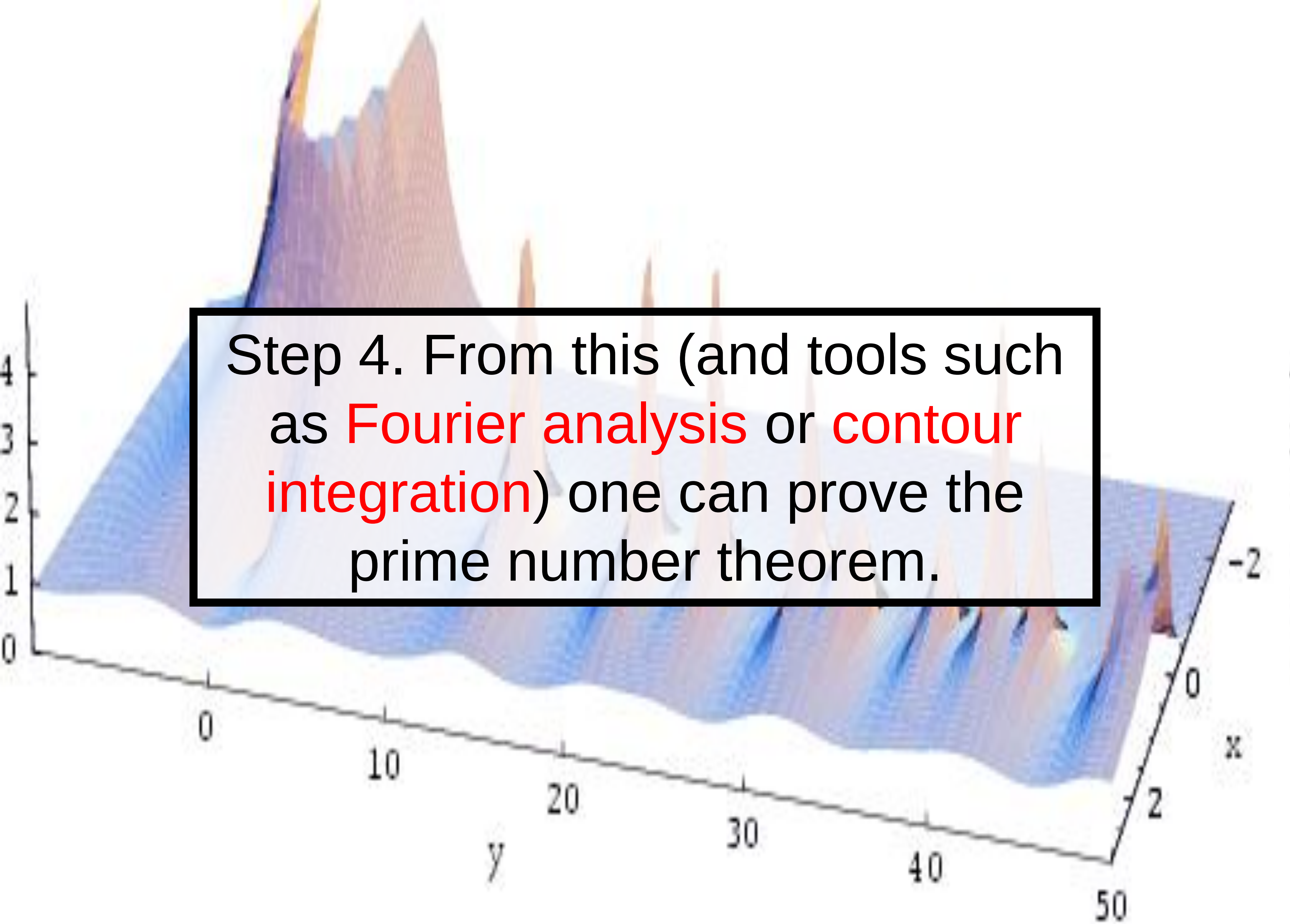
Step 2. "Listen" to this wave by taking a Fourier transform (or more precisely, a Mellin transform).



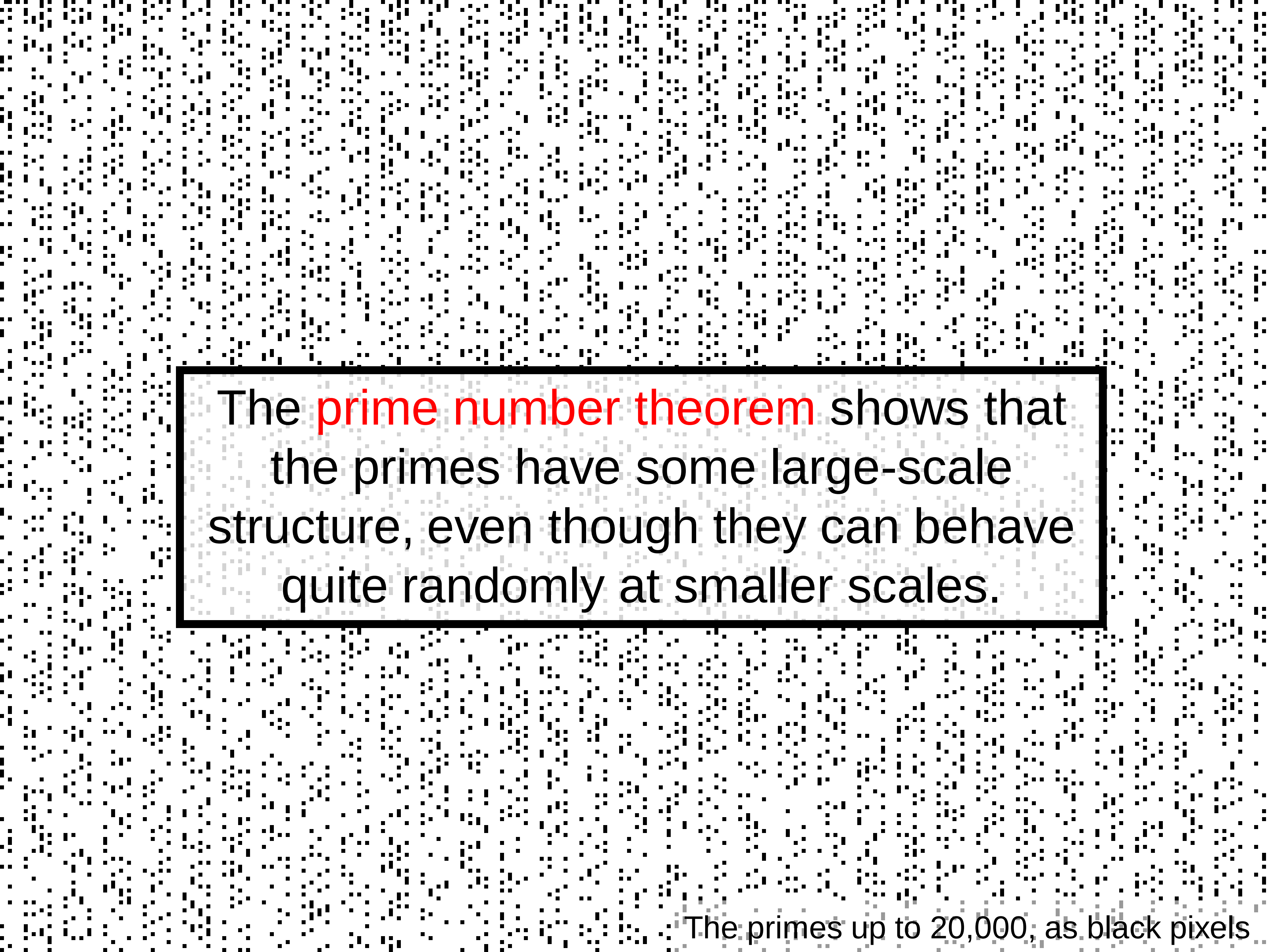
Each note that one hears (the zeroes of the Riemann zeta function) corresponds to a hidden pattern in the primes. (The “music of the primes”.)

Step 3. Show that certain “loud” notes do **not** appear in this music. (This is tricky.)



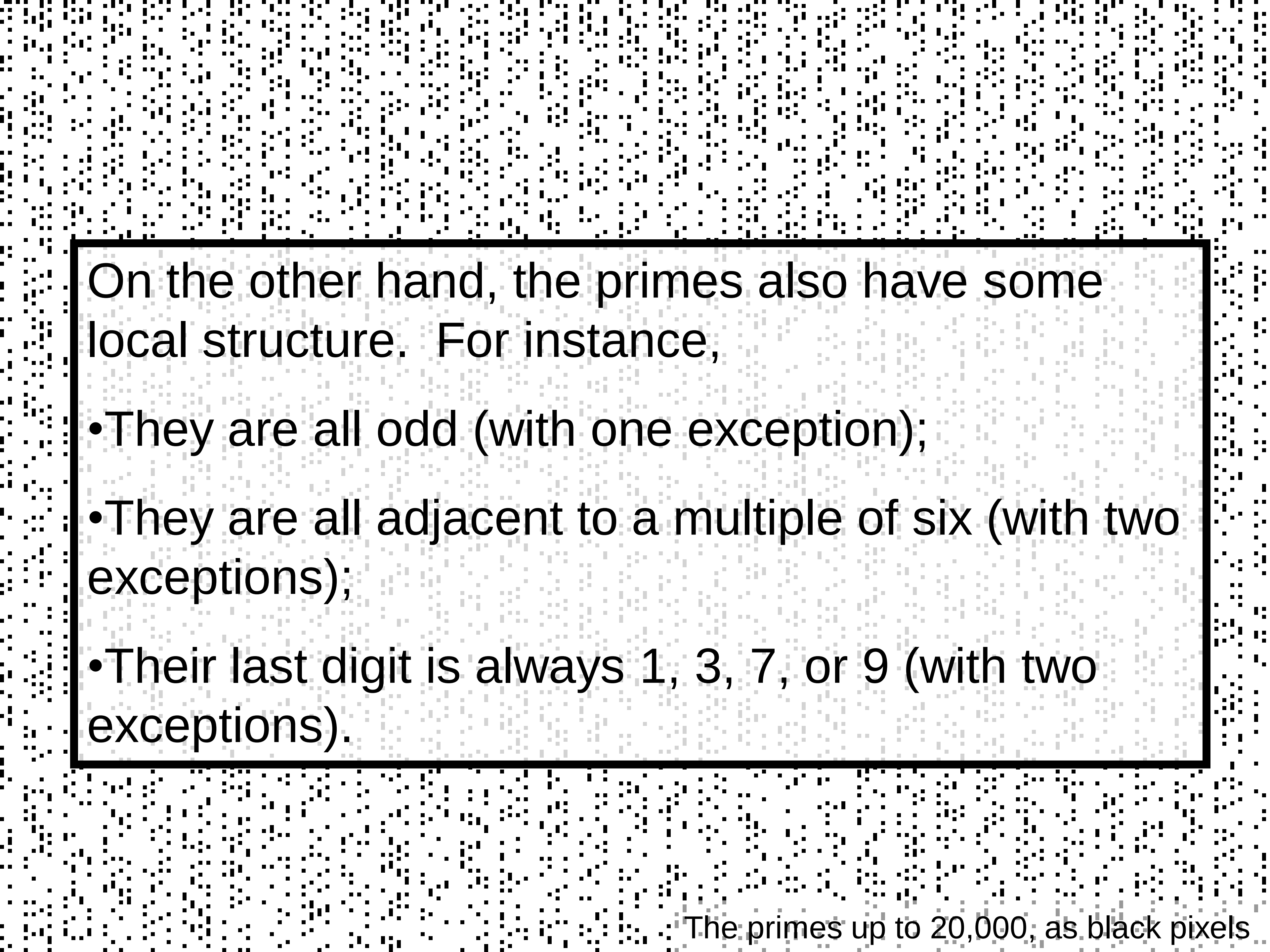


Step 4. From this (and tools such as Fourier analysis or contour integration) one can prove the prime number theorem.



The prime number theorem shows that the primes have some large-scale structure, even though they can behave quite randomly at smaller scales.

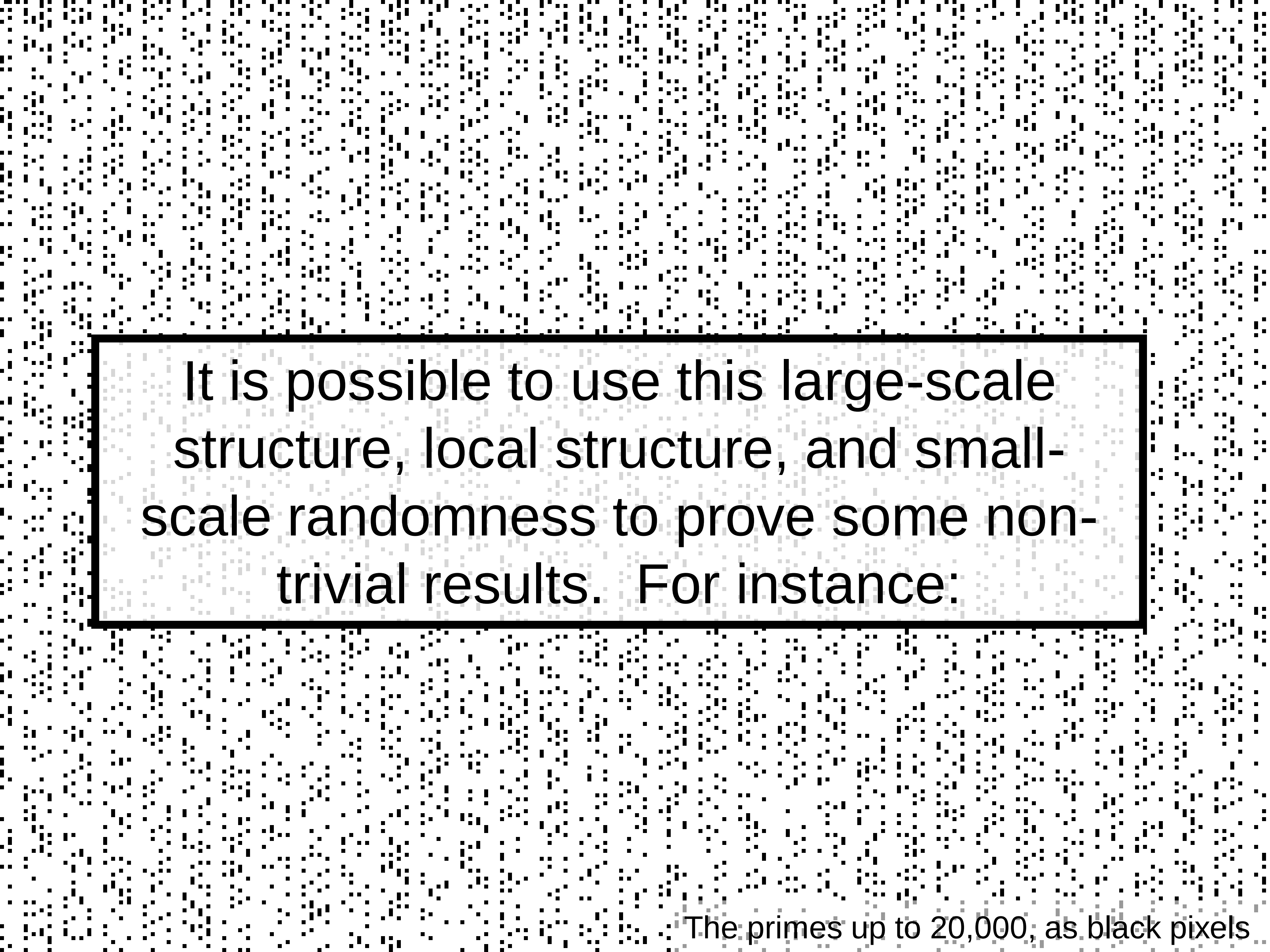
The primes up to 20,000, as black pixels



On the other hand, the primes also have some local structure. For instance,

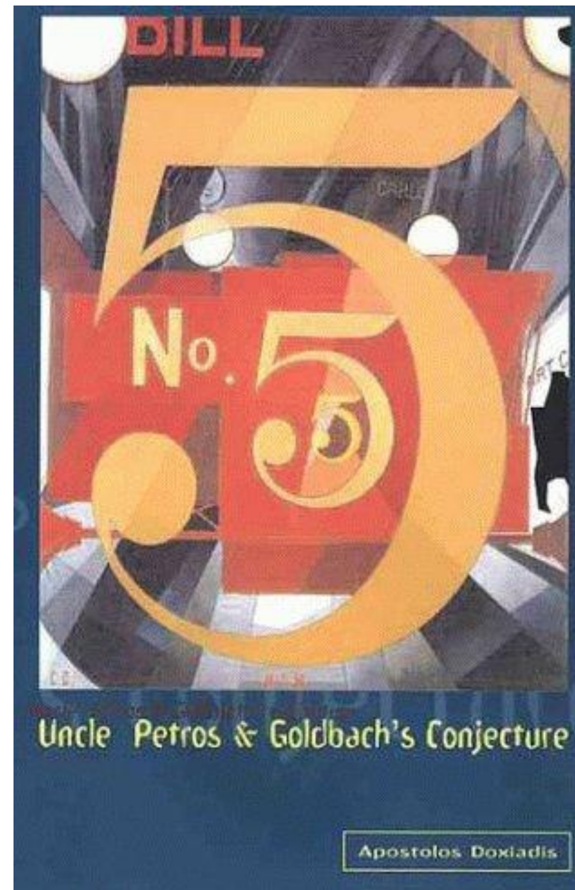
- They are all odd (with one exception);
- They are all adjacent to a multiple of six (with two exceptions);
- Their last digit is always 1, 3, 7, or 9 (with two exceptions).

The primes up to 20,000, as black pixels



The primes up to 20,000, as black pixels

Vinogradov's theorem (1937): every sufficiently large odd number n can be written as the sum of three primes.

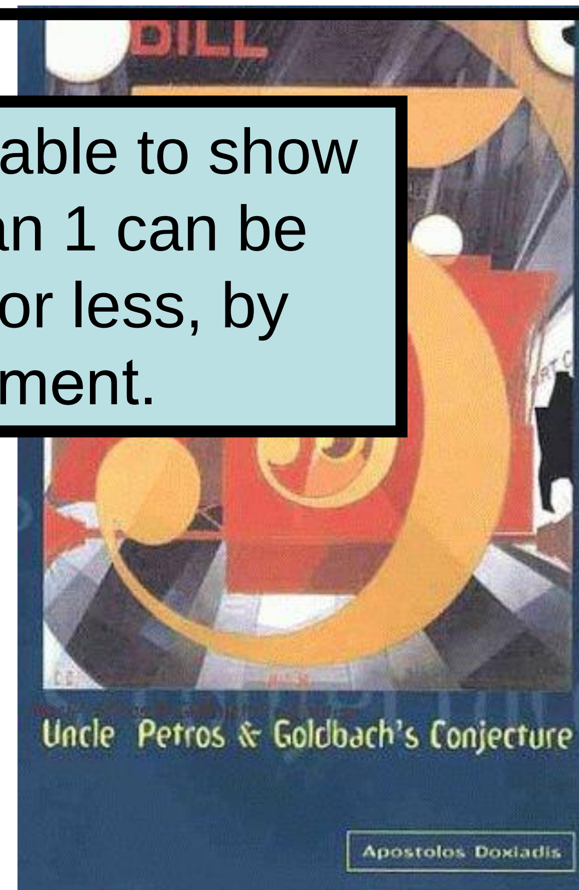


In 1742, Christian Goldbach conjectured that in fact every odd number n greater than 5 should be the sum of three primes. This is currently only known for n larger than 10^{1346} (Liu-Wang, 2002) and less than 10^{20} (Saouter, 1998).

Vinogradov's theorem (1937): every sufficiently large odd number n can be written as the sum of three primes.

News flash: On 31 Jan 2012, I was able to show that every odd number greater than 1 can be written as the sum of five primes or less, by modifying Vinogradov's argument.

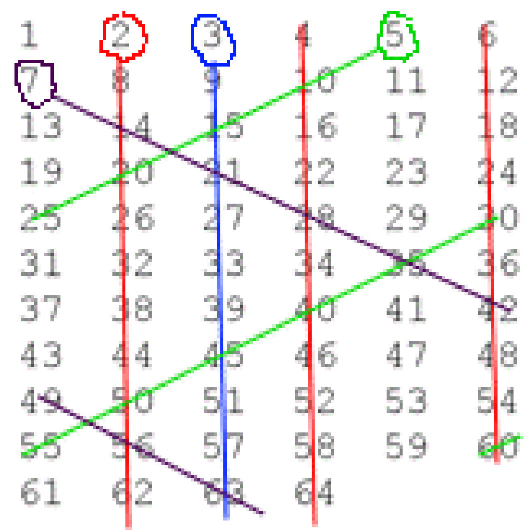
In 1742, Christian Goldbach conjectured that in fact every odd number n greater than 5 should be the sum of three primes. This is currently only known for n larger than 10^{1346} (Liu-Wang, 2002) and less than 10^{20} (Saouter, 1998).



Chen's theorem (1966). There exists infinitely many pairs p, p+2, where p is a prime, and p+2 is either a prime or the product of two primes.

This is the best partial result we have on the twin prime conjecture. The proof uses an advanced form of sieve theory.

The Sieve of Eratosthenes, n=1 to 64



2

2,3

3,5,7

5,11,17,23

5,11,17,23,29

7,37,67,97,127,157

7,157,307,457,607,757

199, 409, 619, 829, 1039, 1249, 1459, 1669

199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879

199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089

Green-Tao theorem (2004). The prime numbers contain arbitrarily long arithmetic progressions.

110437, 124297, 138157, 152017, 165877, 179737, …, 249037

…

56,211,383,760,397 + 44,546,738,095,860n, n=0,…,22 (Frind et al., 2004)

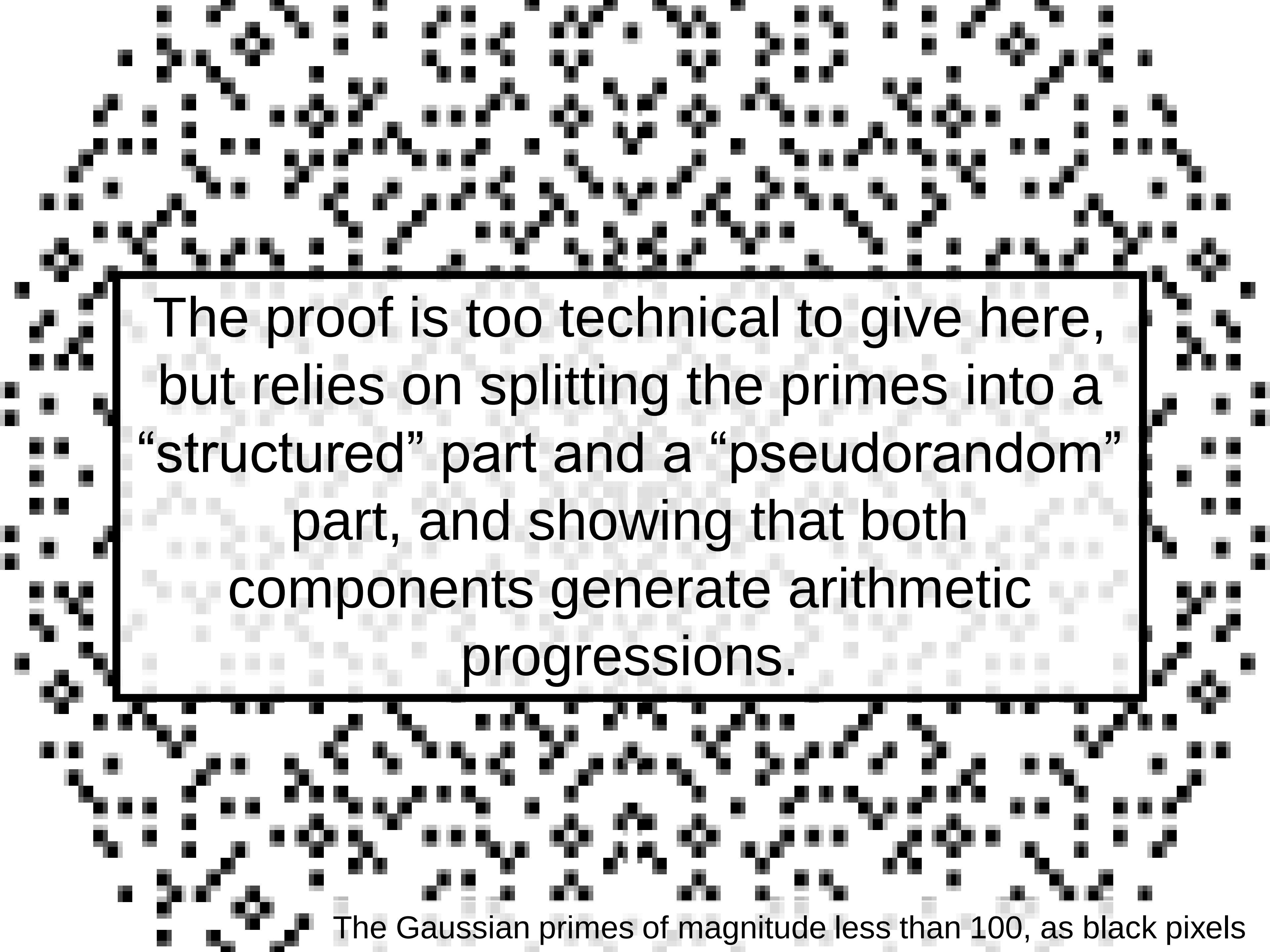
468,395,662,504,823 + 45,872,132,836,530n, n=0,…,23 (Wroblewski, 2007)

6,171,054,912,832,631 + 81,737,658,082,080n, n=0,…,24 (W.-Chermoni, 2008)

43,142,746,595,714,191 + 528,323,403,597,990n, n=0,…,25 (Perichon-W.-Reynolds 2010)

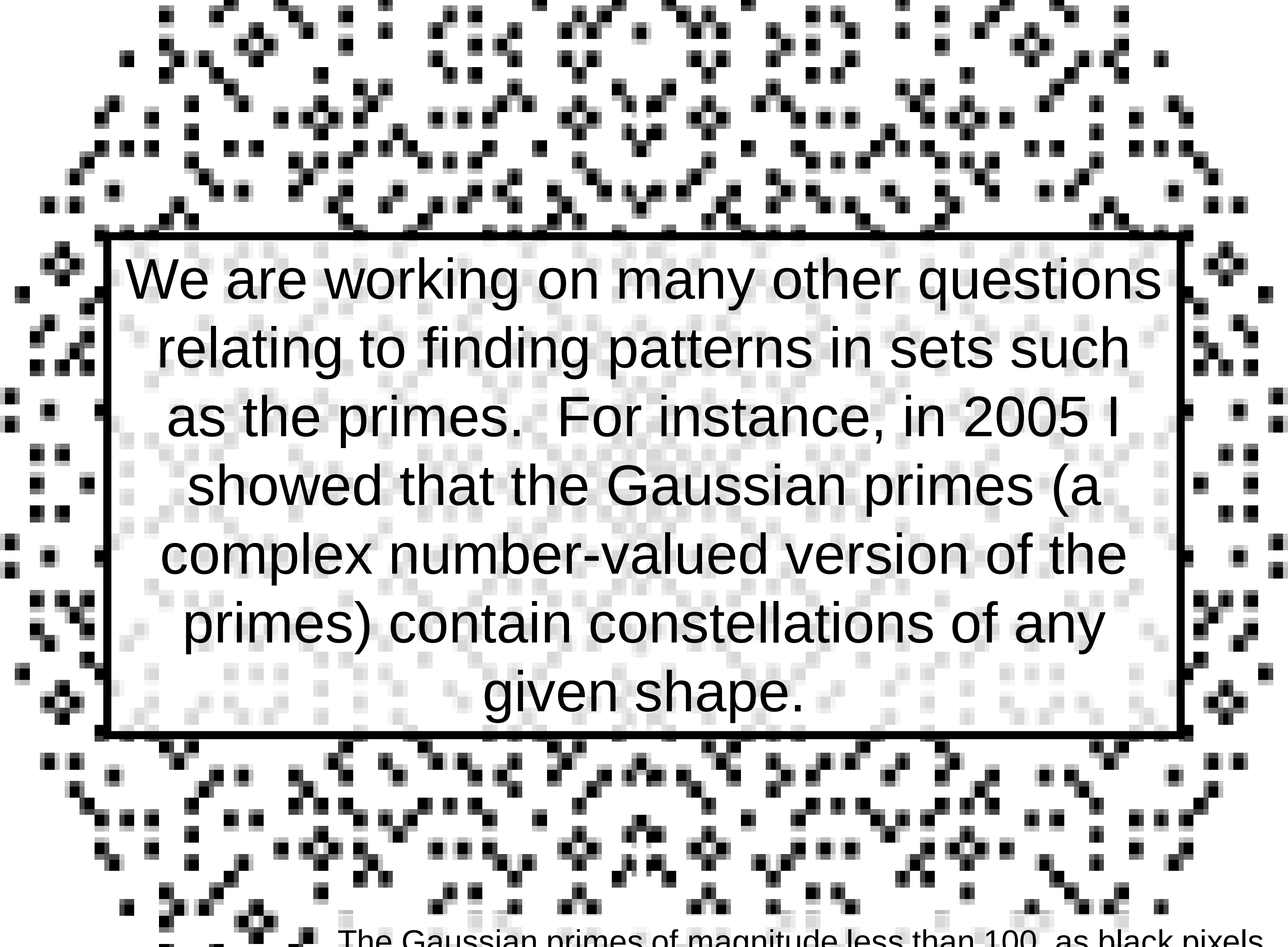
…

The proof is too technical to give here, but relies on splitting the primes into a “structured” part and a “pseudorandom” part, and showing that both components generate arithmetic progressions.



The Gaussian primes of magnitude less than 100, as black pixels

We are working on many other questions relating to finding patterns in sets such as the primes. For instance, in 2005 I showed that the Gaussian primes (a complex number-valued version of the primes) contain constellations of any given shape.



The Gaussian primes of magnitude less than 100, as black pixels

In 2010, Ben Green, Tamar Ziegler, and I used tools from ergodic theory to “count” many types of patterns in the primes.

An irrational orbit on a torus

For instance, we can calculate the number of arithmetic progressions of length 4 of primes from 1 to N to asymptotically be

$$0.4764\ldots * N^2 / \log^4 N$$

as N goes to infinity.

An irrational orbit on a torus

But there are still many, many unsolved problems in the subject to work on!

Twin prime conjecture

Riemann hypothesis

Prime tuple

Cramer’s conjecture

GUE hypothesis

Density conject

Schinzel Hypothesis H

Goldbach conjecture

Landau’s conjecture

Lindeho

Artin’

Elliott-Halberstam conjecture