Wordle, The Dancing Men Code, and Quantum Computers

Or Hilton Cubitt vs Quantum Qubits

By Max Magee[1]

This article was originally published in the Bootmakers of Toronto publication, Canadian Holmes Volume 48 Number 1 – Winter 2024/2025. It was sparked in response to messages on the Hounds of the Internet email list-serve[2] by Karen Murdock's (a.k.a. May Blunder) original post and Janis Robinson's reply[3] about canonical starting words for Wordle—there are quite a few Sherlockians Wordle players on FB and X/Twitter, I've noticed.

Martin Gardner was a prolific annotator of literature (I just picked up his Annotated Alice[4]), likely inspiring the great Sherlockian Annotators (Baring-Gould, Klinger, and others). He was also a famous mathematical puzzler and Sherlockian who worked with (and was friends with) Dana Richards, another noted Sherlockian and Gardner historian, to construct and publish puzzles. I observe that there is significant overlap between people who like words, like puzzles, like word puzzles, and who enjoy mysteries (particularly well-written, well-constructed mysteries, like the Sherlock Holmes tales). A discussion of the connection between Gardner, John Bennett Shaw, and Vincent Starrett was an interesting read from the archives of the Friends of the Collections at the University of Minnesota (see the link in the notes).[5]

---

[1] Max holds a B.S. in Engineering Mechanics and Astronautics from the UW-Madison, an M.B.A. from UH-Clear Lake, and works as a Software Development Manager at a software firm in Madison, Wisconsin. His home scion is the Notorious Canary-Trainers of Madison, Wisconsin.

[2] https://www.sherlockian.net/sharing/hounds/

[3] Karen M. wrote "*Being a Sherlockian, I wondered what words in Canonical titles would be good Wordle words. Taking the 4-letter J. F. Christ abbreviations for the titles and expanding those to 5 letters, I compiled the following list, in Doubleday order. I plan to use these over the next ten days.*

STUDY
NOBLE
BERYL
FINAL
EMPTY
BLACK
HOUND
DYING
LIONS"

Janis mentioned that a cousin won't reveal her starter word and asked *"What better way to solve a puzzle (mystery) than starting with these?"*

[4] https://www.arvindguptatoys.com/arvindgupta/annotated-alice.pdf

[5] https://static.lib.umn.edu/pdf/holmes/v13n4.pdf

For those who are unfamiliar, Wordle is a free word guessing game—or more accurately a letter-guessing game—in which you have six attempts to guess the secret, target word.[6] The mechanics are similar to the old-style board games Mastermind or Battleship (you gain knowledge from both hits and near-misses). The target word is always a five-letter, relatively common, English word (almost never esoteric or rare). Even words with repeated letters seem generally to be avoided, and when they occur, the Wordle community usually complains loudly. See the list of the *worst* Wordle words that were "too hard!" in the notes.[7] I'd be remiss if I didn't mention the original developer of the Wordle, Josh Wardle, the Welsh software developer living in Brooklyn, who built the game to amuse his partner, Palak Shah and hosted it online during the COVID-19 pandemic. Shah helped him curate the original list of 2,500 English target words.[8] In early 2022, Wordle was bought by the New York Times and now is hosted by them.

Jim Hawkins, (an invested member of the Baker Street Irregulars, known as BSI), in particular, has recently been embellishing his solved puzzles with graphics and comments and posting them to Facebook. There are quite a few, mostly Sherlockian, folks who commiserate with comments on those posts with particularly difficult words or close calls.

I'm a long-time Wordle player myself. Well-known Sherlockians Ray Betzner, Ray Riethmeier, Regina Stinson, George Vanderberg, Ron Fish, Rich Krisciunas, and even Michele Lopez of Italy chime in regularly—one of the fun parts of playing Wordle is shouting your answer into the void and seeing others respond with their own. Sharing your solution without spoiling the ending (notice that no letters are shared, just the number of steps and which letters were guessed and contained within the target word but that were out of place, 🟨, or in exactly the right spot in the word, 🟩) has always been built into the game, so this is an *intentionally* social game. For example, after solving this morning's (5 July, 2024) Wordle, it *told* me to share it by copying the following to my clipboard:

Wordle 1,112 5/6

⬛🟩⬛🟩⬛
⬛🟩⬛🟩⬛
⬛🟨🟨⬛⬛
🟩🟩🟩🟩⬛
🟩🟩🟩🟩🟩

Some players like to use ADIEU to optimize discovering vowels early, others have a specific word they enjoy that may not be optimal for the game but gives them some thrill as a first guess. There exist many naïve analyses[9] of optimal starting words (not focused on the

[6] https://www.nytimes.com/games/wordle/index.html
[7] https://gamerant.com/nytimes-hardest-wordle-words-too-difficult/
[8] https://ca.style.yahoo.com/inside-real-life-love-story-193221045.html
[9] https://parade.com/1326658/jessicasager/wordle-tips-tricks-starting-words/

corpus of target words, but rather the letter frequency discussed below). There are also targeted retroactive analyses,[10] tracking which words *would* have been the best to use all along—that feels less helpful though (the two best starting words would've been CRANE and SAUCY, by the way).

Smart play in Wordle, like Wheel of Fortune, Scrabble, hangman, and other letter-guessing word games, is aided by an understanding of letter frequency in a particular language (English, in this case). The Canonical connections are clear when we think about the Dancing Men and even the Red Circle codes (although there are some issues with assuming patterns in one language apply to other languages with that second story in particular[11]).

In the The Adventure of the Dancing Men (1903), Holmes even lays out the way that one might approach a substitution cypher (basically guessing letters based on frequency of their occurrence in the language as a whole). This approach is somewhat mitigated by the small sample size of a note left in chalk or a few words on a slip of paper, and even more so in a 5 letter sequence like the words in a Wordle puzzle. As soon as one letter is solved, the rest of the initial frequency analysis needs to be thrown out the window with a view to guessing the *specific* word at hand—Holmes notes that in this passage:

same, so it was reasonable to set this down as E. It is true that in some cases the figure was bearing a flag and in some cases not, but it was probable from the way in which the flags were distributed that they were used to break the sentence up into words. I accepted this as a hypothesis, and noted that E was represented by



"But now came the real difficulty of the inquiry. The order of the English letters after E is by no means well marked, and any preponderance which may be shown in an average of a printed sheet may be reversed in a single short sentence. Speaking roughly, T, A, O, I, N, S, H, R, D, and L are the numerical order in which letters occur; but T, A, O, and I are very nearly abreast of each other, and it would be an endless task to try each combination until a meaning was arrived

[10] https://wordsrated.com/best-wordle-starting-words/
[11] See Don Yates, B.S.I.'s article, A Final Illumination on the Lucca Code in the BSJ v6 no.3, July 1956

For instance, once you know where one letter is or even one of the letters (not knowing where exactly it fits in the target), you can eliminate whole classes of words. The inverse is also true, in that the Wordle "clues" are the individual verification steps which also *eliminate* possible letters. In the example above, letters that were guessed but not in the target word are black boxes, and when guessing in the game (see Jim's embellished picture which follows), the background of any letter which is not contained within the target word remains black when the guess is submitted (it would become highlighted by one of the other colors if it was correct).

Also consider letter *pairs*: you often find u with q and hardly ever find q without u. H in the second character strongly implies T, S, or C in the previous (although, uncommonly, G or other letters may be there):

> Herbert S. Zim, in his classic introductory cryptography text Codes and Secret Writing, gives the English letter frequency sequence as "ETAON RISHD LFCMU GYPWB VKJXZQ", the most common letter pairs as "TH HE AN RE ER IN ON AT ND ST ES EN OF TE ED OR TI HI AS TO", and the most common doubled letters as "LL EE SS OO TT FF RR NN PP CC".[12]

Note that the assertion of letter frequency both by Zim and Holmes above are not the same, and depending on what source *corpus* you use for measuring the frequency, they will vary somewhat—for example, *word* frequency in language is important when discussing frequency of the *letters*—for instance "the," "a/an/and," and plural *nouns* (trailing -s or -es) are very common in news sources and literature. However, when it comes to word-guessing games, the frequency of the *word* is not applicable, just the frequency of letters occurring within any given word (one way of describing this is *frequency in dictionary* words). A secondary issue is *which* English language are we considering? In a Wordle, for instance, COLOUR wouldn't fit, but COLOR certainly would. I'm afraid the New York Times is trying to take *u* out of the picture, this means that its frequency is generally considerably lower than a dictionary analysis which includes both spellings.

This is one of the earliest frequency analyses of some of the most common letters (notice, it matches up pretty well with Zim's numbers)—made by Samuel Morse, who counted the number of occurrences of each letter in a printer's type set.

> 12,000 E
> 9,000 T
> 8,000 A, I, N, O, S
> 6,400 H
> 6,200 R
> 4,400 D
> 4,000 L
> 3,400 U[13]

---

[12] https://en.wikipedia.org/wiki/Letter_frequency

[13] The phrase "etaoin shrdlu" were the first two rows of characters for many versions of hot-type typesetting machines. That phrase was sometimes included in real newspaper columns accidentally, because an operator had intended to throw a line with a typo away (they had to complete the line so they used those two rows of gibberish), but it was inadvertently included in the final drafts.
See https://www.merriam-webster.com/dictionary/etaoin%20shrdlu

Looking at the dictionary entry analysis, you get very different numbers (note that if THE, IS, and plural nouns are removed, certain letter frequencies like T, S and H fall drastically)

E 11.1607% 56.88
A 8.4966% 43.31
R 7.5809% 38.64
I 7.5448% 38.45
O 7.1635% 36.51
T 6.9509% 35.43
N 6.6544% 33.92
S 5.7351% 29.23
L 5.4893% 27.98
C 4.5388% 23.13
U 3.6308% 18.51
D 3.3844% 17.25

Source for both is a cryptography primer from Notre Dame, linked below[14]

However, in the aggregate, choosing a starting word with E or A, and some of the most frequently used consonants (RTNSL—that letter combination should look familiar to Wheel of Fortune viewers) is a good initial choice. As mentioned, the way humans solve puzzles diverges from frequency analysis quickly, once they've got a letter or two solved/in the appropriate place or even eliminated. That's when letter frequency takes a back seat to language analysis.

To illustrate that, let me take a step back and walk through a recent puzzle with one of the most common "good" starting words. I'm going to use one of Jim Hawkins' recent beautifully-embellished puzzles and its solution, hopefully this doesn't spoil anything for you—consider July 2nd's puzzle:

[14] https://www3.nd.edu/~busiforc/handouts/cryptography/letterfrequencies.html



Notice that Jim started with the guess word *ADIEU* (the first line) and then next guessed *VIRAL* (again, a good start with ADIEU, lots of vowels, although I would note that he also starts with SLATE or AISLE in other puzzle solutions he's posted, as the fancy strikes him—find him on Facebook to see more examples). His choice of VIRAL seems non-standard as a second guess: it would be an unlikely/poor first guess, based on frequency analysis, as we've seen. However, it is a fantastic guess, as related to the correct (A and I) letters and the likely letters he had left to choose from. Note that although V was a less-than-optimal frequency choice, R and L are particularly high-frequency letters. By the second guess, he's found four of five target letters and the positions of one (plus he's eliminated two locations for the letter I (capital "i" not lowercase L).

By his third guess, which, at first blush, doesn't seem terribly helpful for finding the proper letter, notice that he's eliminated any possible location for the I (capital "i"), which leaves

only two possible places for the L, and he's both identified the second to last letter, N, and knows it must be in one of three spots.

The final guesses are narrowed to three combinations INLA_, IL_AN, I_LAN and the letters to choose from are narrowed from 26 initially, to 19 (with many of them, like Z, X, or Q out of the question).

In particular, this process strikes me as much more similar to the way that Holmes solved the Dancing Man/Abe Slaney's cypher, than rote frequency analysis. He had other knowledge (ELSIE's name, repeated), and I'm sure folks like Glen Miranker picked up on the similarities with the way that the German naval ENIGMA cypher was cracked during World War II. More about cracking the ENIGMA after one final note about word games. In the Wordle, there's a built-in incentive to take wild swings. Your record of wins and losses and how many tries it took you to win, is all tracked. Therefore, if you think you may guess the target word in two or three guesses, there's a bit of a gambling mentality at play, too: You risk not gaining any useful information/wasting your guess, but you may get it in only two guesses, an occasion to be proud of!

Anyway, back to the case of the ENIGMA. Although we may think of the ENIGMA as one machine, the phrase really refers to a series of different but related electro-mechanical encryption devices.[15] Most variations had three rotors with slightly different construction—the first three generations had 3 rotors, but the fourth generation (by coincidence) had an added 4th rotor—which encrypted data through a series of electrical and mechanical connections that reconfigured themselves, and which made its output very difficult to decode without a corresponding machine with the correct settings to decrypt the message. Theoretically, the ENIGMA was uncrackable by brute force/guess-and-check attacks at the time, if used properly, though occasionally the encoder would not reset the tumblers daily or used repeating header messages, which allowed signalmen and code-breakers[16] (who were intercepting their encoded

[15] https://www.cryptomuseum.com/crypto/enigma/m3/
[16] https://blogs.bl.uk/european/2018/01/polish-mathematicians-and-cracking-the-enigma.html

messages) to back off the encryption scheme more quickly.[17] In addition, captured machines and code books greatly aided the Allied effort to break the encryption schemes.[18]

But whenever there's a human in the loop, and very smart people are your adversaries, there are bound to be security flaws and they will be exploited. For instance, a repeated three-letter block in the header/preface as a verification of the correct encoding, repeated phrases of Wetterkurzschlüssel/Short Code Weather Reports, the phrase *nothing to report* "KEINEBESONDERENEREIGNISSE," and spelling out numbers like "eins" allowed much faster codebreaking attempts to be verified or discarded quickly by using those as cribs to verify output matching expected message content.[19] Certainly, we need to thank Alan Turing[20] and the other mathematicians, machinists who constructed the codebreaking *bombe* machines, and code-breakers for the Allies (men and women) for their pioneering work on breaking ENIGMA and other wartime codes in WWII.[21]

Fast-forward to today, and many of the most popular and secure (usable) encryption algorithms in the world use what's known as elliptical curve cryptography[22]—I had to qualify that statement with *usable* because, to achieve the same level of security (and difficulty to decrypt using brute force methods), the time it takes to encrypt keys and size of the returned data is quite fast and small, compared to other algorithms, because of some quirks to how the algorithm works. Unfortunately though, quantum decryption algorithms, which rely upon the

---

[17] Glen Miranker, BSI, who reviewed this paper adds:
Don't underestimate the genius/organizational/recordkeeping cleverness that allowed the Bletchley Park (BP) codebreakers to know when a particular message was likely to have such *hints* in it. The overall organization of the BP codebreaking effort is massive, fiendishly clever and impressive. This industrialization of the process is largely due to Gordon Welchman. Let us not forget by mid-1943 over 80,000 messages per month were being cracked. The scale of the problem over and above the cracking of a single message is staggering.

Also, little noted, is the difficulty of cracking the Naval Enigma was only in small part due to the fourth rotor. The Naval procedures for encipherment pretty much prevent operator errors. For example, the setting of the message or outer key (the initial position of the rotors). In all services, other than the Navy, the operator picked the message key,

In the Navy, the operator picked 3 letters and then looked them up in a code book (basically a dictionary). This then gave the operator three different letters. If their operator picked poorly, e.g. "ASD", he would not find an entry in the codebook. The codebook also specified a second set of three letters. These two triads of codebook provided message specific letters would then be added together in a special way to get the 4 letter message key to set the machine for encipherment.

Finally, little noted, the Kriegsmarine maintained a special unit that monitored their own traffic, looking for operational errors!

[18] https://www.theguardian.com/world/2017/oct/20/enigma-code-u-boat-u559-hms-petard-sebag-montefiori

[19] https://www.101computing.net/enigma-crib-analysis/

[20] https://www.mub.eps.manchester.ac.uk/science-engineering/2018/11/28/cracking-stuff-how-turing-beat-the-enigma/

[21] https://www.smithsonianmag.com/history/how-women-codebreakers-wwii-helped-win-war-180965058/

[22] https://avinetworks.com/glossary/elliptic-curve-cryptography/

physical models and characteristics of quantum computers[23] can decrypt what was once very difficult to back out using traditional computational efforts. This means that other, post-quantum[24] (or quantum-decryption-*resistant*) methods may soon be preferred.

Most of the secure/TLS traffic over the internet and in bitcoin wallets, for example, is encrypted using some form of elliptic curve cryptography (or at least an initial *key exchange* which relies upon elliptic curve cyphers). Researchers are still trying to understand what algorithms are best to protect against quantum decryption. In 2022, for example,[25] one of the four finalist algorithms—all developed and being tested for resistance to quantum-codebreaking—from the (American) National Institutes of Standards and Technology (NIST) was cracked using a NON-quantum computing algorithm—the result of which, was a successful 10 minute attack[26] on an algorithm that theoretically should've taken longer than the existence of humans on Earth to decrypt. Although finding prime factors of large numbers is quite computationally expensive using traditional electronic computers, quantum computers[27] are especially good at finding prime factors of large numbers quickly.[28] The difficulty of finding those is one of the things we've relied on to keep our data safe for a long time. We'll need to use other (already existing, but slower-performing) algorithms if we want to achieve quantum-proof encryption.

An interesting quote from one of the researchers who developed the algorithm that was assumed secure until it was shown easily hackable:

> *There is a lot of deep mathematics which has been published in the mathematical literature but which is not well understood by cryptographers. I lump myself into the category of those many researchers who work in cryptography but do not understand as much mathematics as we really should.*

David Jao, a professor at the University of Waterloo and co-inventor of $IKE/SIKE[29]

So keep at it with the Wordles, folks, you'll never know when your country or the world might need your skills to win a war or to secure your latest eBay transaction! Fun fact: they

---

[23] https://quantumzeitgeist.com/quantum-computing-breakthrough-could-crack-ecc-cryptography-exposing-internet-secrets-claims-psiquantum-researcher/

[24] https://en.wikipedia.org/wiki/Post-quantum_cryptography

[25] https://www.wired.com/story/new-attack-sike-post-quantum-computing-encryption-algorithm/

[26] https://thehackernews.com/2022/08/single-core-cpu-cracked-post-quantum.html $IKEp217 was cracked in 6 minutes, the alternate candidate for NIST inclusion, $IKEp434, took about an hour on a single core Xeon processor released in 2013

[27] https://blog.cloudflare.com/nist-post-quantum-surprise

[28] https://www.science.org/content/article/surprising-and-supercool-quantum-algorithm-offers-faster-way-hack-internet-encryption

[29] Ibid in 28

screened potential code-breaking candidates for crossword puzzle skills.[30] And, especially for Sherlockians who may not be interested in word or letter puzzles, I'd point out *Murdle*,[31] which, as soon as you see the matrix, should strike any Cluedo (Clue in American English) player as very similar to the process of elimination needed to excel at that game.

You may have noticed that I'm using the British English version of the word *cypher* (rather than the American spelling, *cipher*) throughout this article to prove the point that even if you think you know the letter frequency, it all goes out the window sometimes. If I used the French spelling, *Le Chiffre*,[32] Bond, James Bond (see, even he messes up letter frequency with his weird way of introducing himself), might try to take me out.[33]

---

[30] https://roomescapeartist.com/2017/02/11/solve-this-crossword-in-less-than-12-minutes-and-you-could-have-been-a-wwii-codebreaker/

[31] https://murdle.com/

[32] https://en.wikipedia.org/wiki/Le_Chiffre

[33] A letter frequency analysis of this article (exclusive of notes) follows:

Letter	Count	Percentage
E	1715	12.6%
T	1245	9.2%
O	1016	7.5%
A	996	7.3%
I	941	6.9%
S	920	6.8%
N	910	6.7%
R	904	6.6%
H	656	4.8%
L	638	4.7%
D	474	3.5%
C	462	3.4%
U	427	3.1%
M	335	2.5%
W	306	2.3%
G	293	2.2%
F	290	2.1%
Y	280	2.1%
P	247	1.8%
B	164	1.2%
K	129	1.0%
V	121	0.9%
Q	54	0.4%
Z	41	0.3%
X	26	0.2%
J	19	0.1%