Wordle as a thermodynamic process

Drew Goldstein

April 2023

Abstract

The objective of this paper is to show the connection between thermodynamics and information theory by treating the Wordle as a thermodynamic process. Wordle is a game in which the player tries to guess a random 5-letter word, starting with no information, by narrowing down the possibilities with each guess. When a player makes a guess, each letter in their guess is assigned a color depending on its relation to the correct word: a green letter means that it appears in the answer in the same place, a yellow letter means that it appears somewhere else in the word, and a black letter means that it is not in the answer at all. In this paper, I will show how we can use these rules to treat the game as a thermodynamic process.

Contents

1 Introduction

1.1 Entropy in thermodynamics

In thermodynamics, entropy is a measurable quantity related to a system's disorder, or uncertainty. For the purpose of this paper, I will be focusing on microstates of macroscopic systems because of its proximity to information theory.

The following derivation follows the textbook [12].

Suppose that there is a system of N different microstates. Let's also divide these into macrostates, each one containing n_i microstates. Two conditions for this system are:

$$\Sigma_i n_i = N, \tag{1}$$

$$\Sigma_i \frac{n_i}{N} = \Sigma_i P_i = 1. \tag{2}$$

Equation (1) states that the sum of the microstates in each macrostate is equal to the total number of microstates in the system.

Equation (2) states that the probability P_i of finding the the system in the i-th macrostate is given by the number of microstates in the ith macrostate divided by the total number of microstates in the system, and the sum of this probability over all macrostates equals 1.

The expression for entropy of a system in a particular macrostate is given by

$$S_i = k_B \ln n_i, \tag{3}$$

and the expectation value of S_i is the is the entropy associated with the ability to exist in different microstates:

$$S_{micro} = \langle S_i \rangle = \Sigma_i P_i S_i. \tag{4}$$

We can then say that the total entropy for the system is

$$S_{tot} = k_B \, \Sigma \ln n_i = k_B \, \ln N. \tag{5}$$

Now, measuring each particular microstate in the system can be incredibly difficult when N is sufficiently large. For example, one mole of an ideal gas in contact with a thermal reservoir has $6{\cdot}10^{23}$ molecules which can exist in many possible microstates each. Without easily accessible information to each microstate, we cannot measure S_{tot} nor S_{micro} directly.

However, we know that S_{tot} is equal to the sum of measured entropy S, which is associated with the freedom to exist in different macrostates, and the entropy S_{micro}. In other words,

$$S_{tot} = S + S_{micro}. \tag{6}$$

Hence,

$$S = S_{tot} + S_{micro}$$

$$S = k_B \left(\ln N - \Sigma_i P_i \ln n_i\right)$$

$$S = k_B \, \Sigma_i P_i (\ln N - \ln n_i)$$

$$S = -k_B \, \Sigma_i P_i \ln P_i, \tag{7}$$

where we have used $\ln N - \ln n_i = -\ln \frac{n_i}{N} = -\ln P_i$. Equation (7) is called the **Gibbs' expression for entropy**.

1.2 Entropy and information

In order to understand the relation between entropy and information, I will present a few statements:

1. I have a birthday.
2. My birthday is in April.
3. My birthday is on the 24th of a month.

The first statement has the least amount of information because everybody has a birthday– this statement has a probability of $P_1 = 1$ (or a $P_1^{\times} = 0$ of not being the case). The second statement contains more information than the first because it provides a specific detail about my birthday – it has a probability $P_2 = \frac{1}{12}$. The third statement contains the most information because there are only 12 days in the year that qualify it. It has a probability of $P_3 = \frac{12}{365}$.

These statements illustrate entropy in information theory: The less likely an event is to occur, the more valuable a message about it is; this informational value is also referred to as a message's surprisal [1]. In other words, as a message's surprisal increases, the probability of its likelihood decreases.

We can derive a function for this surprisal in terms of bit units, the unit for information in computers.

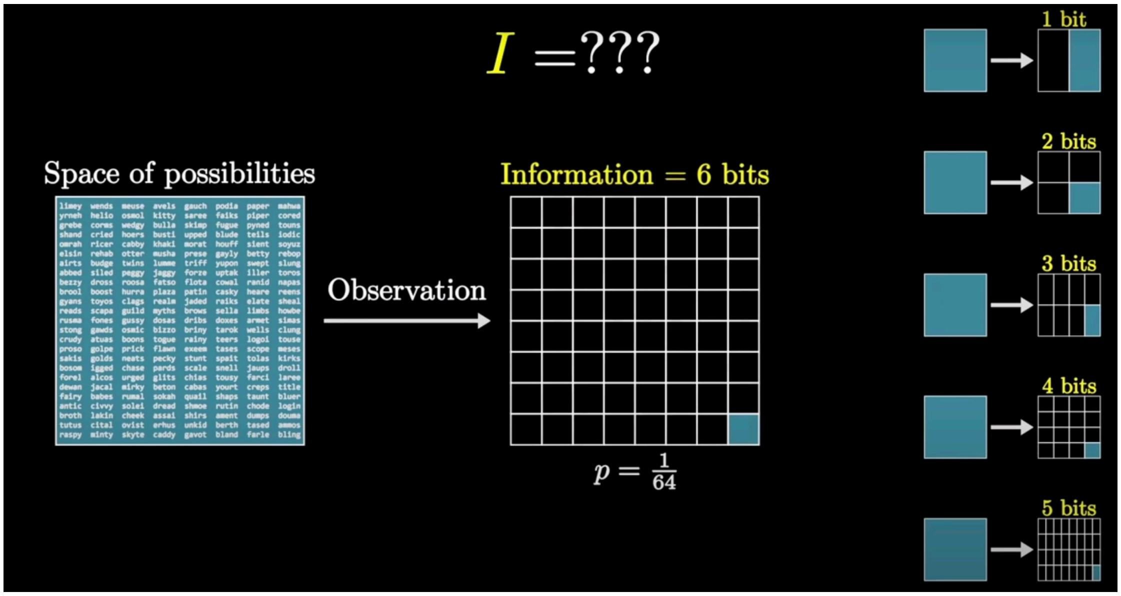

Figure 1: Screenshot from 3Blue1Brown's YouTube video [2].

The above figure illustrates the relationship between an observation and the quantity of bits it contains: If an observation cuts the space of possibilities in half, it would contain 1 bit of information; if it cuts the space of possibilities to a quarter, it would contain 2 bits of information. The figure shows that each time the space of possibilities are cut in half, there is 1 more bit of information (entropy is additive!). Mathematically, this is given by

$$P = \left(\frac{1}{2}\right)^{Q},$$

where Q is bits of information. When we rearrange this equation,

$$\frac{1}{P} = 2^{Q}$$

$$\log_2 \frac{1}{P} = Q$$

$$Q = -\log_2 P. \tag{8}$$

Thus, we have an expression for the information content of a message. If we have a set of messages each with probability P_i, the average information from the set is then

$$\langle Q \rangle = S = \Sigma_i P_i Q_i = -\Sigma_i P_i \log_2 P_i. \tag{9}$$

The above expression is known as **Shannon entropy**. This expression quantifies the information we expect to gain when we take a specific measurement, or the amount of uncertainty a measurement has before we measure it.

When we rewrite equation (9) as

$$S = -k\,\Sigma_i P_i \log_2 P_i,$$

where k = 1, we see that the formula for Shannon entropy is exactly the same as the formula for Gibbs' expression entropy in thermodynamics if we take k = k_B and $\log_2$ = ln. Thus, we can relate the two concepts to better understand thermodynamic entropy as a measure of uncertainty of a system, given a limited amount of information.

With this connection in mind, I will treat the optimization of Wordle as a thermodynamic process.

2 Process

There are a total of 12,953 5-letter words in the English language. However, we don't see many of these words in books or in everyday life, such as 'sepoy', 'zooea', or 'towze.' So, out of this list of words, let's say there 2,309 words that are likely to be used as Wordle answers.

I will refer to specific color patterns with notation P_{ABCDE}, where each letter refers to a color of one letter: G ≡ green, Y ≡ yellow, and B ≡ black. So P_{YYBBG} refers to the probability of the pattern yellow, yellow, black, black, green.

In the following sections, I will discuss how I may approach Wordle as a thermodynamical process in order to optimize the game.

2.1 Starting guess

From these 2,903 possible words, how can we decide which word is the best first guess? What qualifies a word as a 'good' guess or a 'bad' guess? This is where our connection between thermodynamic entropy and Shannon entropy becomes important.

Let's pick a random word as our first guess and analyze it: 'train'.

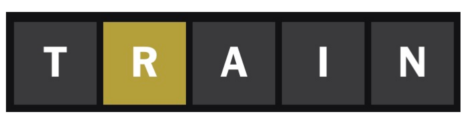

Figure 2: Opening with 'train', we know that the answer contains an 'R' somewhere not in the second letter, and does not contain a 'T', 'A', 'I', nor 'N'. (ANSWER: 'joker') [4]

Running 'train' as a first guess through a simple program I wrote [8], 164 words out of the list of 2309 words match the 5 conditions [6]. Because there are truly 12,953 words that are allowed as Wordle answers, the probability of this pattern is $P_{BYBBB} = \frac{164}{12{,}953}$ = 0.012661. This gives 6.303 bits of information (8). However, 'train' as an opening guess is not always guaranteed to give me 6.303 bits of information. After running my program again with a different Wordle answer, this is the pattern it returned:

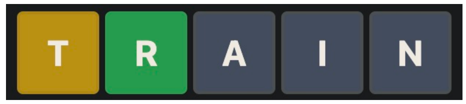

Figure 3: Opening with 'train', we know that the answer contains an 'R' in the second letter, a 'T' in the third, fourth, or fifth letter, and does not contain 'A', 'I', nor 'N'. (ANSWER: 'erect') [4]

In this case, only 14 words out of the 2,309 possible words match all 5 conditions [7]. The probability of this pattern is $P_{YGBBB} = \frac{14}{12953} = 0.0012$, and its associated information content is Q = 9.85 bits (8).

I could guess it again with a different Wordle answer and return a much longer list of possible answers with significantly less information content, but it doesn't answer the question: How do we determine if it is a good guess?

A better question we may ask: how much information do we expect to gain from an opening guess? We can calculate the expected information content, or Shannon entropy, of a guess by averaging the information content each combination of colors will give (since there are 5 independent letters with three possible colors each: $3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243$ total combinations). 3Blue1Brown wrote a code which does exactly this [3]: by averaging the information content contained in all 243 color combinations, their code calculated the Shannon entropy of 'train' to be $S_{train} = 5.553$ bits ([9], line 147).

In other words, if I guess 'train' as a first word, I can expect to gain 5.553 bits of information, but I may gain more or less upon measurement (demonstrated by figures 2 and 3, where we gained 6.303 and 9.85 bits of information, respectively).

For the following data, I supplemented 3Blue1Brown's program with some of my own code to pair each word with its associated Shannon entropy and sort them in reverse order based on their entropies [5]. These are the best 10 first-guesses based on their expected information content:

	Word	$\langle Q \rangle$
1.	'soare'	5.8852
2.	'roate'	5.8848
3.	'raise'	5.8783
4.	'reast'	5.8677
5.	'raile'	5.8652
6.	'slate'	5.8558
7.	'salet'	5.8360
8.	'crate'	5.8352
9.	'irate'	5.8328
10.	'trace'	5.8304

Table 1: The top 10 words with their associated Shannon entropy [9].

So, out of all 2,309 possible words, 'soare' has the greatest Shannon entropy; in other words, it contains the *potential* for the greatest amount of information content, but this also includes the greatest amount of uncertainty. Since our goal is to gain the most information possible in the shortest amount of turns, it's up to the player to determine if the risk of uncertainty is worth its potential benefit.

It is important to note that this isn't a definitive best first-guess solution: there are methods we may take to further optimize our score. For example, we may choose to calculate the entropy of a guess based on the frequency of each letter; we may also choose to calculate the entropies for the first two guesses to find which specific pair of words gives the greatest additive Shannon entropy (1Blue3Brown actually included this is their code [3]). However, for the purpose of this project, I will choose 'soare' as an optimum opening guess.

2.2 After the first guess

The process after the first guess follows the same physics: look at the list of possible words, and find the Shannon entropy of each word to determine which next guess may contain the most information content. After running the program using the word 'soare' as a first guess, the list was reduced from 2,309 words to 40 words [11], giving Q = 5.85 bits of information content. It returned the following pattern:

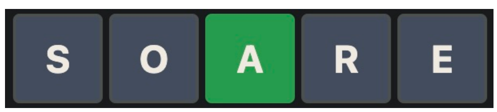

Figure 4: Opening with 'soare', we got a pattern which tells us that there is an 'A' in the third letter of the answer, and no 'S','O','R', or 'E'. (ANSWER: 'aback') [4]

Again using 3Blue1Brown and I's joint code [5], these are the top 3 next guesses based on their associated Shannon entropy.

	Word	$\langle Q \rangle$
1.	'clink'	4.596
2.	'plink'	4.496
3.	'clint'	4.444

Table 2: Best 3 second-guesses and their associated Shannon entropies. [10]

Notice that none of the three words match the condition from the first guess ('A' in the third letter). In fact, the 9th best second-guess is the first word in the list which matches the condition: 'clank' with Shannon entropy $\langle Q \rangle = 4.237$. This is no accident: Since we know there is an 'A' in the third letter, any next guess that matches the condition would provide information about one less letter than a word with letters we haven't tried yet. The probability of 'A' being in the third letter of the answer is P = 1, which gives us no information content. Of course, we could attempt to guess the answer on the second try by using one of the 40 possible words, but each word has an equally small probability P = $\frac{1}{40}$ = 0.025 of being correct, so this is a big risk that may not be worth taking.

Figure 5: Using 'clink' as a second guess, we now know that there is also a 'K' in the last letter, a 'C' either in the second or fourth letter, and no 'L', 'I', or 'N'. (ANSWER: 'aback') [4]

Using 'clink' as a second-guess returned Q = 3.73 bits of information and narrowed the list from 40 to 3 possible words: 'aback', 'quack', and 'whack', with corresponding Shannon entropies $\langle Q \rangle = 1.584$, 1.584, and 0.918, respectively. Now, guessing one of the three words poses a significantly smaller risk than guessing one of the forty possible words in the second turn. Since 'quack' and 'whack' have equal Shannon entropy, they each have the same amount of uncertainty and guessing one or the other is the difference between a score of 3 or 4.

3 Discussion

In this paper, I have demonstrated how we can link thermodynamics and information theory by treating Wordle as a thermodynamic process.

We can treat all the possible guesses as microstates of a turn, where there are 6 total macrostates, one representing each turn. When we calculate the Shannon entropy contained in each possible guess, we are finding the expected information we would gain from each constituent. Shannon entropy also relates to the uncertainty, or surprisal, of a specific microstate; we would gain much more information after guessing a word with letters we haven't tried than guessing a word that definitely meets the conditions of previous guess(es). In other words, it is more advantageous to guess the word with the highest Shannon entropy.

References

[1] https://en.wikipedia.org/wiki/Entropy_(information_theory)

[2] https://www.youtube.com/watch?v=v68zYyaEmEA&t=766s&ab_channel=3Blue1Brown

[3] https://github.com/3b1b/videos/tree/master/_2022/wordle

[4] https://mywordle.strivemath.com/

[5] https://github.com/ultraluckyclover/Wordle_project/blob/main/3Blue1Brown_borrowed.py

[6] https://github.com/ultraluckyclover/Wordle_project/blob/main/train_joker.txt

[7] https://github.com/ultraluckyclover/Wordle_project/blob/main/train_erect.txt

[8] https://github.com/ultraluckyclover/Wordle_project/blob/main/Wordle_drew.py

[9] https://github.com/ultraluckyclover/Wordle_project/blob/main/entropies_sorted_first.txt

[10] https://github.com/ultraluckyclover/Wordle_project/blob/main/entropies_sorted_second.txt

[11] https://github.com/ultraluckyclover/Wordle_project/blob/main/narrowed_down_after_soare.txt

[12] Blundell, Stephen. "Concepts in thermal physics." Oxford ; Oxford University Press, 2010.