9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Is luck random?

Keywords: Luck, chance, randomness, twitter, statistical inference

Extended Abstract

Some people are lucky and others are not. Luck is often attributed to a combination of readiness and random chance [3] — *"fortune favors the prepared mind"* — Louis Pasteur. However, we wonder, how random is luck really? Are some people luckier than others in a way inconsistent with randomness? And in particular, does some notion of skill seem to mediate it in a substantial way [1]?

In this work, we set out to evaluate the randomness of luck using a corpus of 15 million Wordle[1] tweets [2]. Wordle is an online word game in the style of Mastermind in which, each day, players try to guess a new 5 letter word within 6 turns and aided by hints. Players can only enter words included in a corpus of 2315, 5-letter words. A unique property of Wordle tweets is that they follow a strict format and indicate if a user solved the puzzle, and if so, the number of guesses they took, and the hints they received, as shown in Figure 1. The tweets do not show exact letters of guesses or the correct word, but online sources have aggregated the correct words for each day since the launch of the game, as well as the entire corpus of words that can be submitted [2]. At a high-level, the hints shown in the tweets indicate if a letter is not in the answer (⬜), in the answer but in the wrong place (🟨), or in the correct place (🟩).. Accordingly, using the hits shown in a Wordle tweet, its possible to impute information about what a player knew at a given turn, and an upper bound on the number of possible words in the corpus that the hints help the player eliminate. For example, in Figure 4 the hint from the second turn helps the player understand that only words with the same second letter, with the third and fifth letter in explicitly different positions, and not including the first and fourth letter could be the correct word. While this approach does not benefit from as much insight as the player has — we can't accurately estimate the content of the ⬜ hint letters from the tweets alone — it does allow for establishing a bound on the chance the player experiences.

With the properties of Wordle tweets in mind, we define *luck* of a given attempt as the ratio of the number of words from the corpus that satisfy the current hint and the number of words from the corpus that satisfy the previous hint, as formalized by Equation 1, for the i^{th} person's r^{th} turn, and given a function a that computes the number of possible words that satisfy a hint, h. To minimize the influence of ⬜ hint letters, we opt to focus only on the penultimate turn in games that the player managed to guess the correct word in the final turn — wherein the number of words that satisfies the final turn's hints is definitely 1 — which we formalize with Equation 2. While the first guess also offers a clean comparison, many players use strategically selected first words to maximize their learning on the first turn, so luck on this turn is biased.

$$l_{ir} = 1 - \frac{a(h_i^{(r)})}{a(h_i^{(r-1)})} \qquad (1)$$

$$l_i = 1 - \frac{1}{a(h_w)} \qquad (2)$$

[1] `nytimes.com/games/wordle`

We define *skill* as the average comparative strength of one player over the average player, which is computed as the z-score of guesses a player takes to guess a particular word correctly.

We filtered the corpus to exclude tweets in which the player never guessed the correct word and to only include twitter users who posted at least 10 Wordle tweets — to get a stronger estimate of skill for all users. We additionally removed users who guessed the correct answer on the first turn, which we took as an indication that they may be cheating.

Figure 2 shows striking similarity between the empirical distribution of luck and a matched normal distribution , suggesting in a course fashion, that luck is largely random. We additionally conducted a Kolmogorov-Smirnov (KS) test, showing that the empirical distribution differs significantly from a normal distribution ($p < 0.001$), suggesting that luck is not just random. This of course comes with a sizable caveat that when considering 15 million units, meaningless fluctuations can be found to be significant, and subjectively, the empirical data appear to track a normal distribution closely. While some of the twitter users have incredibly consistent luck, the number demonstrating this phenomenon don't tend to exceed those expected by chance. Notably, the game design of Wordle leads to a situation where average luck is not at some obvious midpoint, e.g. centered around 0 or $\frac{1}{2}$, and this current analysis assumes that the empirical mean of luck scores is a valid baseline — by contrast, its possible that those who tweet about Wordle are more or less lucky on average, however the size of the data, and a simple sampling bias test suggest this is not the case (see Figure 4 for details).

Figure 3 shows a weak and non-predictive relationship ($R^2 = 0.036$) between our indices of skill and luck across all players. In other words, skill is not usefully correlated with luck, resulting in a situation in which the adage attributed to Pasteur may not hold true — luck actually doesn't seem to favor anyone all that much.

Our analysis suggests that luck is substantially random, and even with a large dataset, few people are so lucky as to beat the statistically expected distribution of luck as chance. There are, of course, many caveats with this style of analysis, including the substantial limitations of the information contained in Wordle tweets alone, and the likely case that players under report failures and low scoring plays. Further, we have not yet conducted analysis of what simulated players might predict — are humans leveraging luck to the best of their ability, or is there some systematic strength or weakness in their strategies as might be predicted by [1]. As a final point, although we think the indices we have established are reasonable, there is certainly room for a wide range of alternative indices in this style of analysis, for example *luck* that integrates over all turns, not just the penultimate one, and *skill* that holds meaning without aggregation across users or words.

References

[1] Takahiro Komai, Hiroaki Kurokawa, and Song-Ju Kim. "Human Randomness in the Rock-Paper-Scissors Game". In: *Applied Sciences* 12.23 (2022). DOI: `10.3390/app122312192`.

[2] Robert Lesser. *Wordle, 15 Million Tweets Later*. 2022. URL: `https://observablehq.com/@rlesser/wordle-twitter-exploration` (visited on 11/15/2022).

[3] "Success and luck in creative careers". In: *EPJ Data Science* 9.9 (2020), pp. 1127–2193. DOI: `https://doi.org/10.1140/epjds/s13688-020-00227-w`.

Figure 1: The anatomy of a Wordle tweet: a single line describing the word number (which entails the date that the play occurred on), as well as a numerical representation of the number of turns taken as a fraction. This is followed by lines of colored square emoji indicating if a letter is not in the answer (⬜), in the answer but in the wrong place (🟨), or in the correct place (🟩).

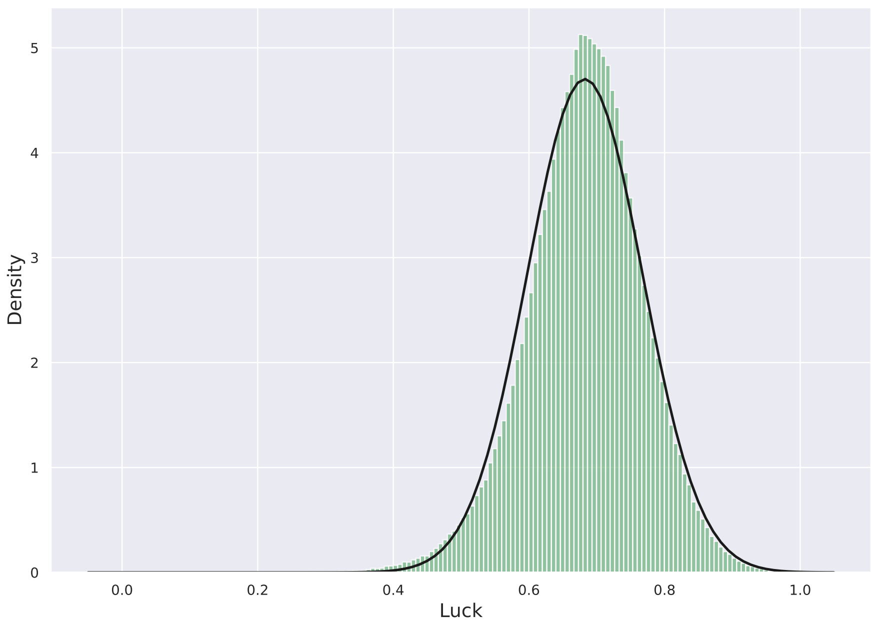

Figure 2: An empirical distribution of luck on the level of tweets across the entire filtered population, and a normal distribution with the same μ and σ as the empirical data.

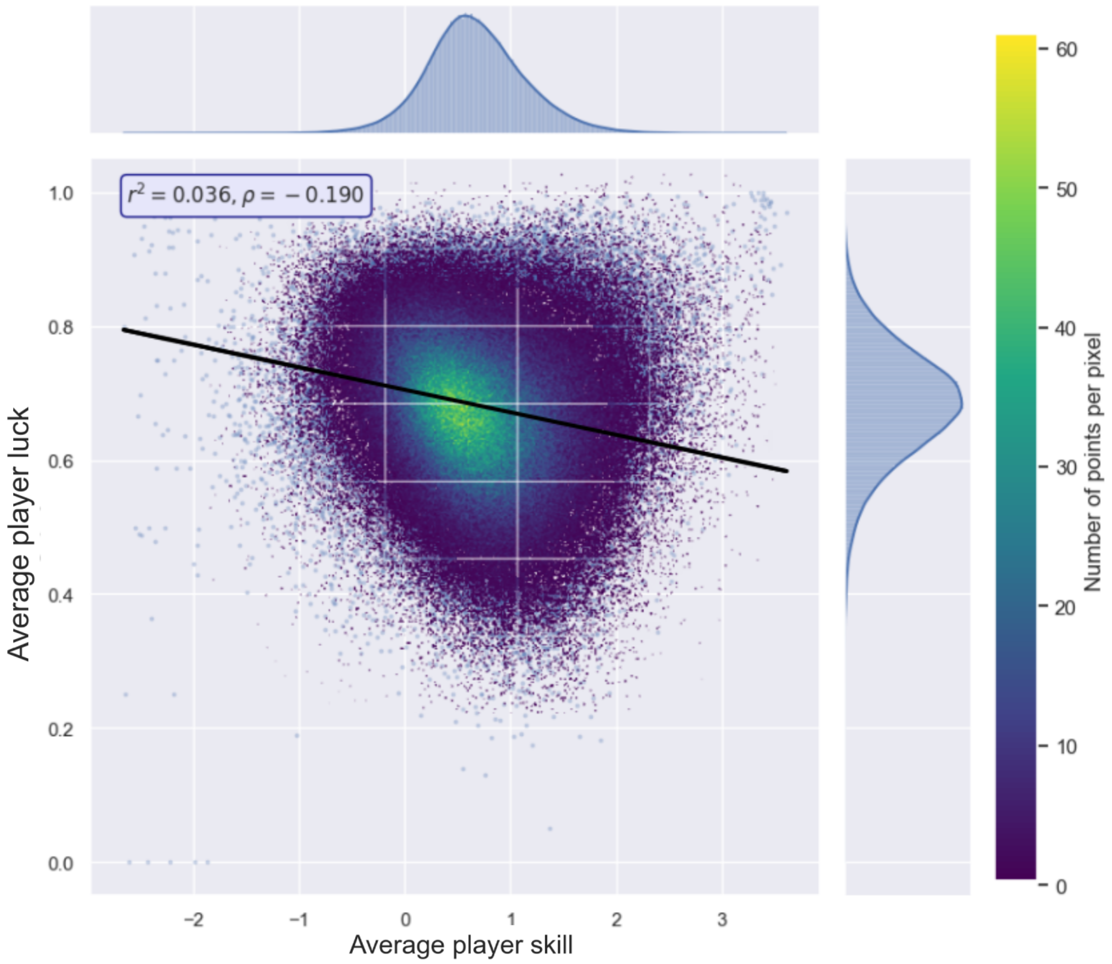

Figure 3: Density of player level average skill and luck shown with marginals and a linear fit.

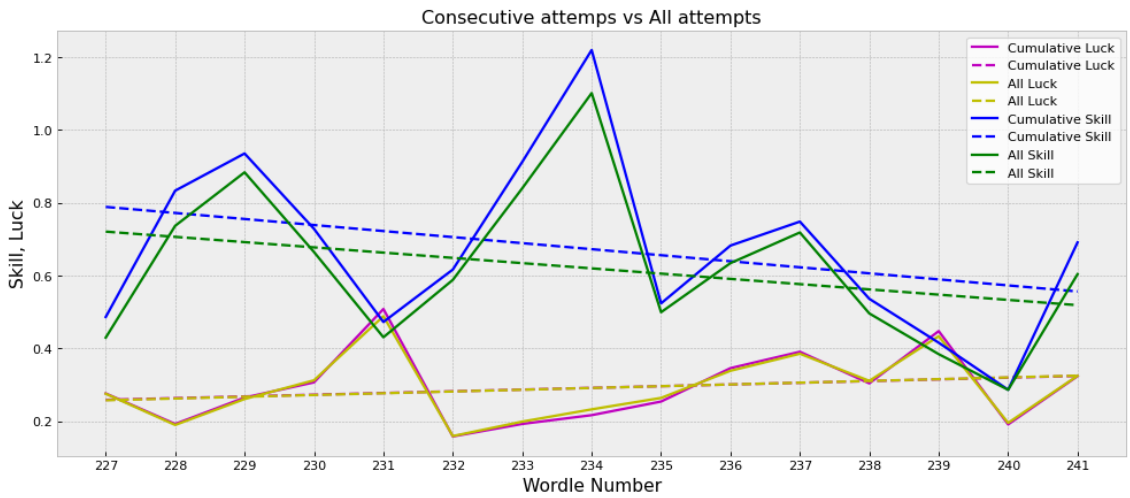

Figure 4: Comparison of indicies between players who posted tweets for every day in the 2022/02/01–2022/02/16 date window and those who did not post a tweet every day in that time. 95%CI are shown but not visible due to the data size. While these pairs of lines are statistically distinct, they track identical trends across populations.