Semantic, Orthographic, and Phonological Biases in Humans' Wordle Gameplay

Jiadong Liang, Adam Kabbara, Jiaying Liu, Ronaldo Luo, Kina Kim, Michael Guerzhoy

Division of Engineering Science, University of Toronto

Toronto, Ontario, Canada

esgary.liang@mail.utoronto.ca, adam.kabbara@mail.utoronto.ca,
cindyjy.liu@mail.utoronto.ca, ronaldo.luo@mail.utoronto.ca,
kina19131@gmail.com, guerzhoy@cs.toronto.edu

Abstract

We show that human players' gameplay in the game of Wordle is influenced by the semantics, orthography, and phonology of the player's previous guesses. We compare actual human players' guesses with near-optimal guesses using NLP techniques. We study human language use in the constrained environment of Wordle, which is situated between natural language use and the artificial word association task.

1 Introduction

Wordle is a daily word-guessing game where players attempt to identify an unknown five-letter word within six attempts (Wardle, 2021). Players usually attempt to minimize the number of guesses they make. Players also usually want to maintain a "streak" of having solved the game within at most 6 guesses for several days.

We explore the difference between near-optimal play and human gameplay, which may be influenced by cognitive shortcuts and biases. In order to estimate near-optimal plays, we use the maximum-entropy heuristic. We verify that the heuristic is near-optimal.

In settings where word association is important, humans are influenced by salient past information, a phenomenon known as *priming* in psychology (Schacter and Buckner, 1998). We conjecture that priming effects exist in the game of Wordle as well. Additionally, we conjecture that humans will tend to depart less from previous guesses in order to minimize cognitive load.

We review the prior work on priming in psychology, and in particular, how priming influences future word choice. We then review the optimal strategy in Wordle, as well as heuristics that approximate it. Following this discussion, we introduce our human guess data. We then present our approach to measuring human biases in Wordle

gameplay and demonstrate the systematic differences between human play are near-optimal play.

2 Background: Human Cognitive Processes

Priming is a phenomenon in psychology where past experience influences behavior without the person's explicit knowledge of the influence (Schacter and Buckner, 1998). Priming sometimes manifests in word association. Prior works have demonstrated the grammatical class, semantic meaning and rhyme of the previous (cue) word would influence the later (response) word by humans. This effect has also been explained through models of lexical access in speech production, where residual activation from previously retrieved lexical items can facilitate or bias subsequent retrievals (Levelt, 1989; Roelofs, 1992).

Deese (1962) and De Deyne and Storms (2008) explore the effect of the grammatical class of cue words over the response word.

Steyvers and Tenenbaum (2005) use data collected by Nelson (1999) to demonstrate that the word association network — a graph where two words are connected if they are members of a cue-response pair — is sparse, with each word connected to only 0.44% of other words. They also use data collected by Miller (1995) and Fellbaum (1998), and found that the word network constructed based on semantics of words exhibits sparseness, connectedness, neighboring clustering, and power-law degree distribution. All of these are the same characteristics exhibited in the free association network.

Nelson et al. (1987) demonstrate the effect of rhyme on memory and word association. They ran an experiment where subjects would initially study (read aloud) the cue-target pair of a given rhyme; then 1.5-2 minutes after they finished studying, a meaning-related cue word and its semantic relation

with the target word would be given and the participants would be required to read it aloud and recall the word they studied. In the experiment, cue words that rhyme with many other words would decrease the accuracy of the respondent, regardless of the meaning-related cue word. Through conducting a further experiment that changed all the cue-target pairs studied to be meaning-related and only half to be also rhyme-related, Nelson et al. (1987) showed that the effect of rhyming appears only if the subjects actively attend to it when studying the word pairs.

Matusevych and Stevenson (2018) studied human word association based on word attributes. Luo et al. (2025) demonstrated it is possible to weakly predict human amusement responses to Wordle gameplay using features similar to the ones used in this study.

Related work on word puzzle solving provides further evidence that lexical retrieval and phonological awareness are central to constrained word-generation tasks. Underwood et al. (1994) found that expert crossword solvers outperform intermediate solvers in generating candidate words, solving anagrams, and manipulating morphemic and syllabic components of words. These skills are closely tied to efficient lexical access and morphological processing. These findings suggest that in tasks such as Wordle, players may similarly rely on lexical retrieval strategies and phonological structure to guide their guesses.

3 Background: Wordle solving mechanisms

The objective of Wordle depends on the player — it can be maintaining the streak (i.e., trying not to lose today's game), winning in as few guesses as possible, or even winning the game using funny words.

However, most of the solving mechanisms are designed to optimize objectives regarding the number of guesses, such as minimizing the average number of guesses, minimizing the number of guesses in the worst case, etc. Those mechanisms can be classified into two classes: the exact optimization approach and heuristic approaches. The best approaches based on heuristics achieve results that are only marginally inferior to exact methods. Bertsimas and Paskov (2024) found an optimal and efficient solution for Wordle that minimizes the average number of guesses using dynamic pro-

gramming. They show that the word "SALET" is the best starting guess and the minimum average number of guesses required is 3.421. They demonstrate that their optimal approach never results in a loss (i.e., the algorithm always completes the game within 6 guesses).

Heuristic approaches to Wordle can achieve performance that is very close to optimal. The depth-1 minimax heuristic aims to minimize the number of guesses for the worst case, with search depth of 1. For each guess, it iterates through all possible words in the game and chooses the one that minimizes the size of maximum partition (the amount of possible solutions after the current guess) as the guess. Given the starting guess as "SALET", it is guaranteed to finish the game in 5 guesses, and has the average number of guesses of 3.482 (Cross, 2022). The entropy-based heuristic (also with depth 1) reduces the uncertainty at each step by choosing the guess that decreases (on average) the most number of potential solutions after that guess (Shannon, 1948) (Cross, 2022). It is also guaranteed to complete the game in 6 guesses and have the average number of guesses of 3.432. We use the Doddle[1] implementation to report those results.

4 Data

The human guess data was sourced from Reddit. The machine-generated guesses are obtained using Doddle, an open-source Wordle solver introduced earlier. Although an ideal comparison would be with the optimal model, the Doddle solver was chosen for computational reasons. It's important to note that the performance difference between the exact dynamic programming solution and the heuristic entropy solver is minimal: the exact solution achieves a minimum average of 3.421 guesses, while the heuristic-based solver has an average of 3.482 guesses for its minimax heuristic and 3.432 guesses for its entropy-based heuristic. Doddle's heuristic min-entropy solver we use will be referred to as the *near-optimal* strategy.

4.1 Data collection

The data is collected from the r/Wordle subreddit, where people share their guesses online, contributing to a total of 83,000 data entries (Watchful1, 2023). We extract games posted by people in the

---

[1] https://github.com/CatchemAL/Doddle

commonly-used format on the subreddit using a regular expression.

5 Methods

5.1 Measuring Human Biases

To quantitatively assess the influence of human cognitive biases in Wordle games, human plays are compared to their entropy-based near-optimal counterpart, where five different metrics described below are used to reveal different aspects of human biases (semantic, orthographic, and phonological). For each guess in the data, the metrics below are computed through comparing that guess with the previous one (instead of comparing with all prior guesses) unless otherwise stated.

5.1.1 Levenshtein Distance

The Levenshtein Distance measures the minimum number of edits — insertions, deletions, or substitutions — needed to transform one word into another (Levenshtein, 1966). This feature captures how closely a player's subsequent guesses align with their previous ones in terms of structural similarity. A smaller Levenshtein distance may indicate that the player is selecting guesses that are more similar to their prior attempts, potentially reflecting a reluctance to explore novel letter combinations or a preference for minimizing cognitive effort.

5.1.2 Semantic Distance

The GloVe (Pennington et al., 2014) distance is computed using the negative cosine similarity between GloVe word embedding pairs. Humans may potentially be biased to make guesses that are semantically close to their previous guesses.

5.1.3 Hamming Distance

The Hamming distance between words is the number of locations where the words differ. We conjecture that guesses that are close to previous guesses in the Hamming distance sense are easier to make.

5.1.4 Rhyme

To determine whether two words rhyme or not, their phonic transcription was used. This was achieved using the `pronouncing` library, which provides a phonetic transcription based on the CMU Pronouncing Dictionary (CMU, 2015). Two words are considered to have a *perfect rhyme* if they have matching phonetic endings which include stressed vowels (Per, 2019). We assess whether the guess rhymes with the previous one.

6 Experiments

We compare how human guesses/plays differ systematically from near-optimal plays. We obtain distributions of human plays and near-optimal plays, and compare them. We assess the effect size (difference between the distributions) using Cohen's d, and we computed the p-values based on the t-statistics for the difference between the two distributions.

We analyze separately games starting from different positions. We use the notation $c_g g c_y y c_b b$ and (c_g, c_y, c_b), where the number of "green" guesses (correct letter in the correct place) is denoted with c_g, the number of "yellow" guesses (correct letter in the incorrect place) is denoted with c_y, and the number of letter guesses that are incorrect is c_b.

In Figure 1, we present some observations on the results of our comparison of human play with near-optimal play for specific configurations. We observe that in many, though not all, cases, humans are biased towards their previous guesses. This is particularly pronounced when there is a lot of free choice. This indicates that human gameplay, which combines creativity with optimization, shows more evidence of lack of creativity when there is more free choice. Figures 3 and 4 in the Appendix display the overall patterns: for more constrained positions, differences between near-optimal moves and actual moves tend to be smaller.

We additionally report that the optimal guess rhymes with the previous guess 7.3% of the time, but humans make a guess that rhymes with the previous guess 9.3% of the time (p-value < 0.001).

6.1 Relationship between semantic distance and other measures

For valid Wordle word pairs, the Hamming distance and the Levenshtein distance are strongly related (Pearson-r=0.95 for all pairs, Pearson-r=0.81 for character difference less than 5).

The relationship between character-based differences and semantic distance ($1 - cos(v_a, v_b)$ where v_a and v_b are the GloVe vectors) is weaker: Pearson-r= 0.06 (Figure 2), suggesting that the effects of semantic distance and orthographic distance likely operate separately.

7 Conclusions

Human gameplay in Wordle exhibits a bias toward previous guesses semantically, orthographically, and phonologically (e.g., sharing the last syllable).
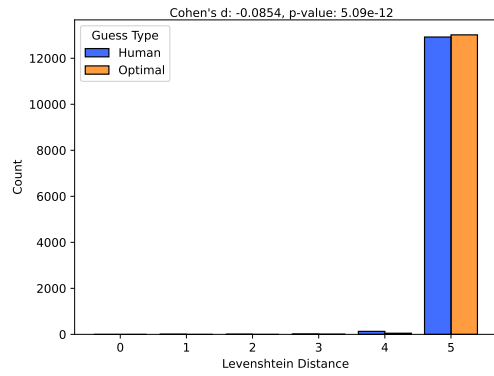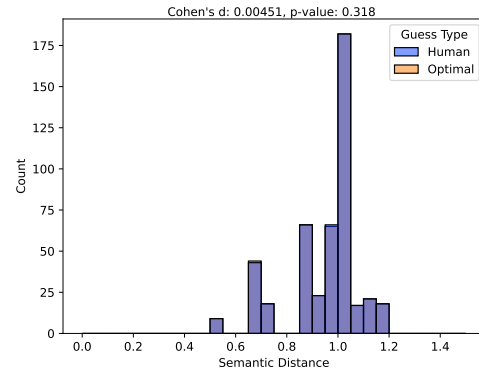
(a) Levenshtein distance between human guesses and near-optimal guesses for 0g0y5b: both choose distance 5 most of the time. Humans suboptimally play letters they know aren't there.
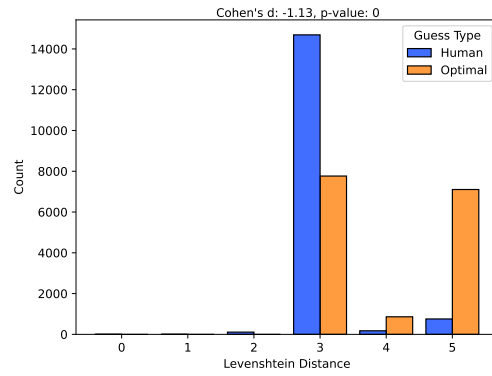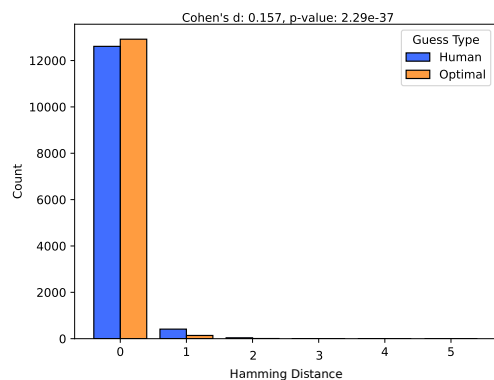
(b) Levenshtein distance between human guesses and near-optimal guesses for 2g0y3b: humans underexplore compared to near-optimal.
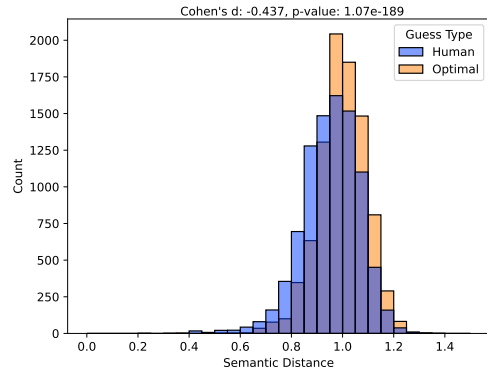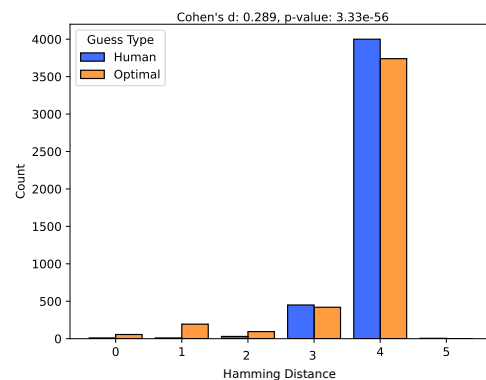
(c) Semantic distance between human guesses and near-optimal guesses for 0g0y5b: humans slightly biased towards underexploring

(d) Semantic distance between human guesses and near-optimal guesses for 3g2y0b: no bias.

(e) Hamming distance between human guesses and near-optimal guesses for 0g0y5b: humans play very obviously suboptimally by reusing characters they know are not there.

(f) Hamming distance between human guesses and near-optimal guesses for 2g2y1b: humans play suboptimally by using new characters.

Figure 1: (c_g, c_y, c_b): the number of "green" guesses (correct letter in the correct place) is denoted with c_g, the number of "yellow" guesses (correct letter in the incorrect place) is denoted with c_y, and the number of letter guesses that are incorrect is c_b.
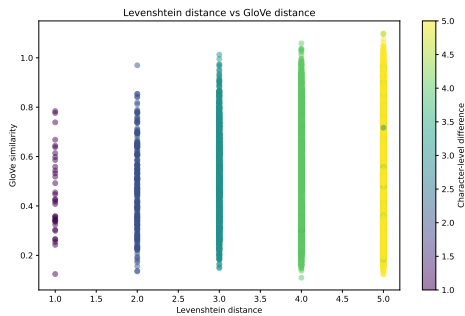
Figure 2: GloVe similarity between guesses vs. Levenshtein Distance

The bias is systematic, indicating human gameplay does not merely randomly deviate from optimal play.

In Wordle, we demonstrate an environment where phenomena like word association can be studied in a setting that is not as varied as natural speech, but where people are not merely performing a task, as in the word association task (cf. Hamilton and Huth (2020)). Environments such as Wordle can be of interest when studying other cognitive patterns.

References

2015. The CMU pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict. Accessed: October 13, 2024.

2019. Exact rhyme. https://literarydevices.net/exact-rhyme/. Literary Devices.

Dimitris Bertsimas and Alex Paskov. 2024. An exact solution to wordle. *Operations Research*.

Alex Cross. 2022. Doddle. *https://github.com/CatchemAL/Doddle*.

Simon De Deyne and Gert Storms. 2008. Word associations: Network and semantic properties. *Behavior research methods*, 40(1):213–231.

James Deese. 1962. Form class and the determinants of association. *Journal of verbal learning and verbal behavior*, 1(2):79–84.

Christiane Fellbaum. 1998. Wordnet: An electronic lexical database. *MIT Press*, 2:678–686.

Liberty S Hamilton and Alexander G Huth. 2020. The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, 35(5):573–582.

Willem J.M. Levelt. 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

V Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Proceedings of the Soviet Physics Doklady*.

Ronaldo Luo, Gary Liang, Cindy Liu, Adam Kabbara, Minahil Bakhtawar, Kina Kim, and Michael Guerzhoy. 2025. Automatically detecting amusing games in wordle. In *Proc. International Conference on Computational Creativity (ICCC)*.

Yevgen Matusevych and Suzanne Stevenson. 2018. Analyzing and modeling free word associations. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 40.

GA Miller. 1995. Wordnet: An on-line lexical database [special issue]. *International Journal of Lexicography*, 3(4).

DL Nelson. 1999. The University of South Florida word association norms. *http://w3.usf.edu/FreeAssociation*.

Douglas L Nelson, Maria Teresa Bajo, and José J Canas. 1987. Prior knowledge and memory: the episodic encoding of implicitly activated associates and rhymes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1):54.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Ardi Roelofs. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42:107–142.

Daniel L Schacter and Randy L Buckner. 1998. Priming and the brain. *Neuron*, 20(2):185–195.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Mark Steyvers and Joshua B Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.

Geoffrey Underwood, Caroline Deihim, and Viv Batt. 1994. Expert performance in solving word puzzles: From retrieval cues to crossword clues. *Applied Cognitive Psychology*, 8(6):531–548.

Josh Wardle. 2021. Wordle. Web. Online game.

Watchful1. 2023. Subreddit comments/submissions 2005-06 to 2023-12.

8 Limitations

We rely on publicly available data voluntarily posted by users on Reddit. In principle, games posted by Reddit users can differ systematically from data sampled randomly from a different population of interest. This is due to selection effects

as well as the fact that the population of Reddit users, and specifically the Wordle subreddit, could differ by demographics and linguistic background, among other things, from other populations of interest. We limit our study to the game of Wordle in English. Our findings may not generalize to other games in other languages.
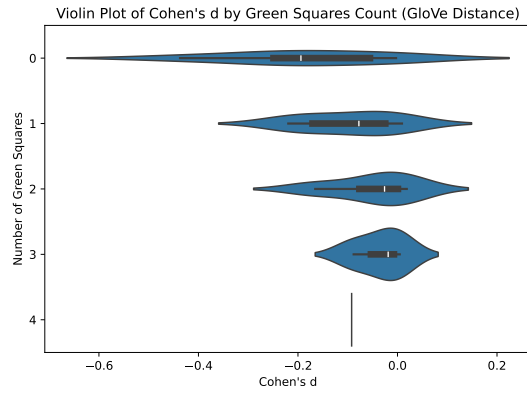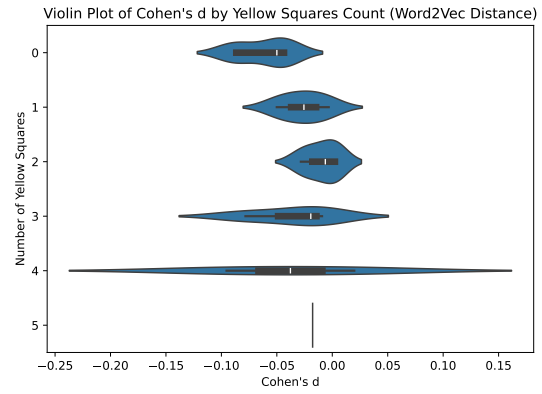
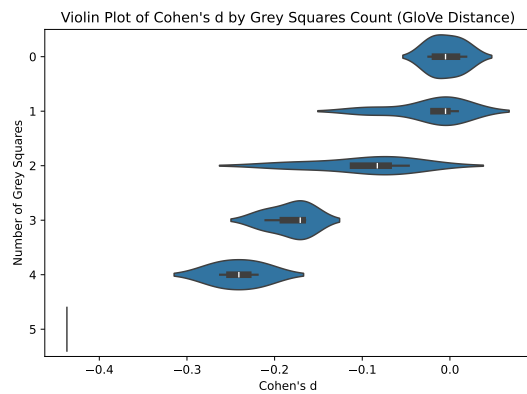(a) Violin plot of Cohen d by *green* squares (GloVe).
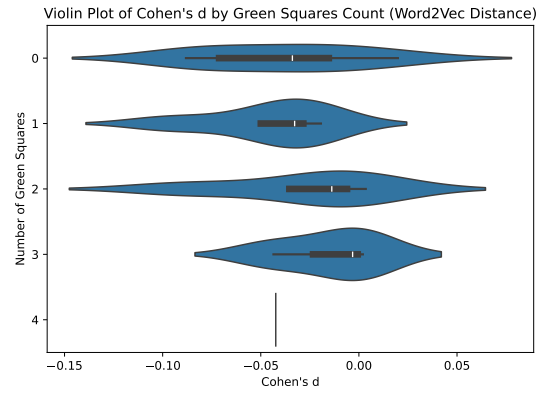
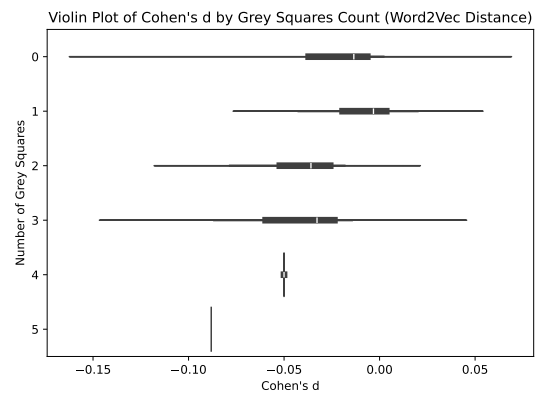(b) Violin plot of Cohen d by *green* squares (Word2Vec).
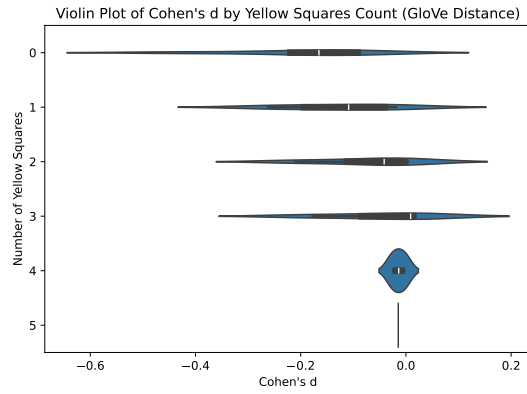
(c) Violin plot of Cohen d by *yellow* squares (GloVe).

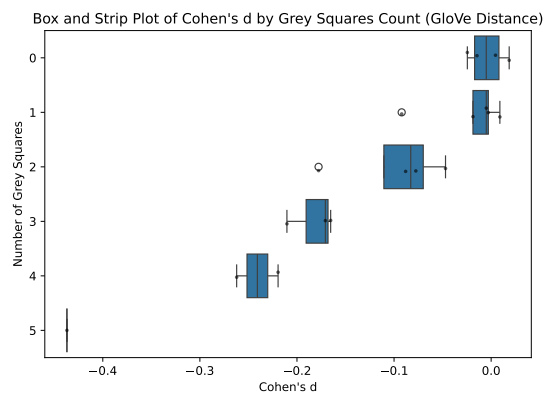(d) Violin plot of Cohen d by *yellow* squares (Word2Vec).

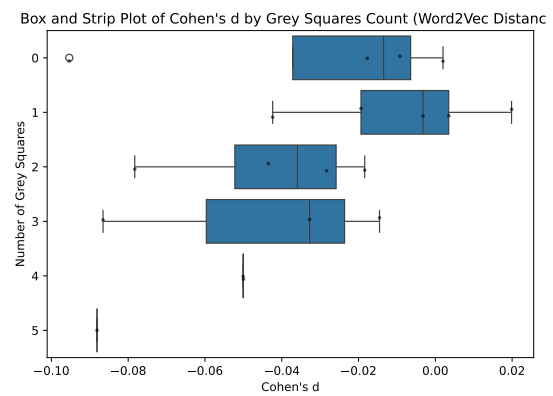(e) Violin plot of Cohen d by *grey* squares (GloVe).

(f) Violin plot of Cohen d by *grey* squares (Word2Vec).

Figure 3: Distances between neighboring guesses, by amount of constraint on guess

(a) Box & strip plot of Cohen d by *grey* squares (GloVe).

(b) Box & strip plot of Cohen d by *grey* squares (Word2Vec).

Figure 4: Distances between neighboring guesses, by amount of constraint on guess