Analyses of Predicted Successful Communication Based on Target Words and Relationship Types

Oishani Bandopadhyay

December 2025

1 Introduction

In this report, I conduct two levels of analysis, using a database with data about players playing a communication game similar to Taboo. The SCARFS Database contains a corpus of English nominal referring expressions (NREs) produced in this communication game, and the original publication producing the dataset analyzes outcomes of the communication game based on linguistic features of the target word (Trott et al., 2023). I reproduce the analyses in this publication, focusing on the effects of Age of Acquisition, Concreteness, and Frequency of the target word on the outcomes of the game (Won vs Lost and Won vs Out of time) (Trott et al., 2023). I further expand my analysis to look at the interaction of concreteness with the part of speech feature of the target word, and with the relationship between the pair of participants playing (whether they were friends or strangers). I find significant effects of the linguistic features on outcomes of the game, and effects of the interaction terms for parts of speech and pair types with concreteness, depending on which outcomes were compared.

2 Data

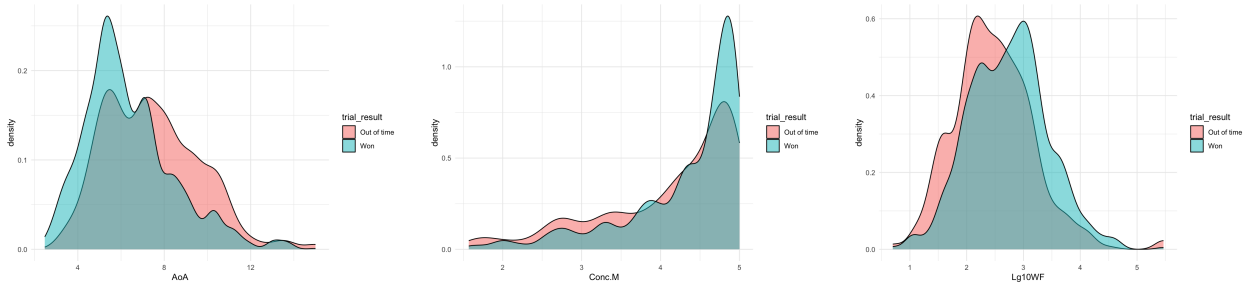
The SCARFS dataset was created by having pairs of participants, who were either friends or strangers, play a communication game similar to Taboo (Trott et al., 2023). Participants were recorded, and their speech was transcribed, followed by the identification of noun phrases in their utterances, and then tagging with form tags to get full noun phrases and other forms of NREs (Trott et al., 2023).

For the initial analyses replicating the linguistic analyses in the original publication, I followed the procedure of merging multiple relevant datasets (Trott et al., 2023). Using the behavioral results dataset from the SCARFS database, the Brysbaert concreteness norms dataset, the SUBTLEX English frequency data, the Kuperman Age of Acquisition Norms dataset, and a Taboo dataset with measures of average distances between target and Taboo words using Glove embeddings, I successfully created a merged dataset for analysis (Trott et al., 2023; Brysbaert et al., 2014; Brysbaert and New, 2012; Kuperman et al., 2012; Pennington et al., 2014). This merged dataset had 3,731 rows, then filtered to only keep Nouns, Verbs, and Adjectives, leaving 3,702 rows. The mean concreteness of the target words in this dataset was 4.24, and the median concreteness was 4.59.

I further analyzed the dataset from the SCARFS database containing only full noun phrases (Trott et al., 2023). This dataset had 4,701 rows, each with a head noun used in an NRE to refer to a card. There were 987 Lost, 1,070 Out of Time, and 2,644 Won outcomes. Losing and running out of time were different outcomes, since losing meant that participants used a taboo word, whereas out of time means that participants could not converge on the target word in time (Trott et al., 2023). The dataset was quite evenly split between Friends and Strangers, with 50.9% of the pairs being friends, and 49.1% being strangers. Adjectives and Verbs were 6.25% and 12.8% of the part of speech tags for the target words respectively, with 80.9% of the words being tagged as Nouns. The mean concreteness in this dataset was 4.25, similar to the mean concreteness in the merged dataset above.

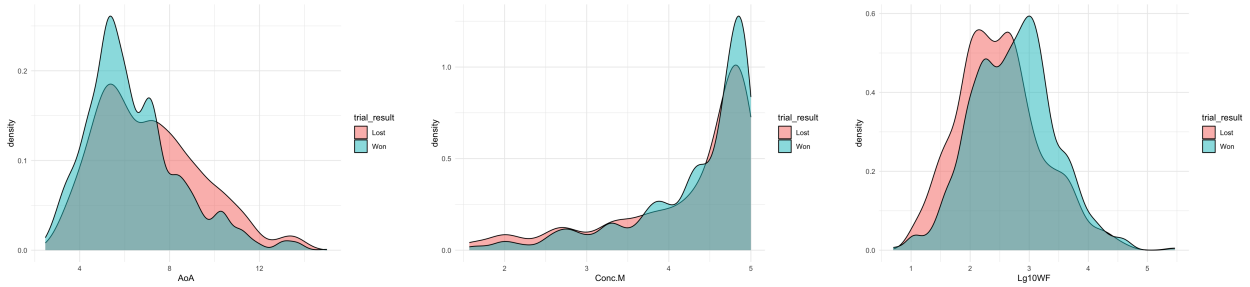
3 Visualizations

First, I plot the outcomes Won vs Out of Time, replicating 2 plots from the analyses in the original paper looking at the density of each outcome across Log Frequency and Age of Acquisition of the target word (Trott et al., 2023). Then, I add my own visualization for the density of Won vs Out of Time across Concreteness. I also create similar visualizations across all three linguistic variables for the outcomes Won vs Lost. This is a meaningful distinction as mentioned in Section 2, since Out of Time is failing to come to the right conclusion in time, whereas Lost is using a taboo word. So, observing each pair of outcomes individually was more useful for observing the analyses in the original paper (Trott et al., 2023).



(a) Figure A (b) Figure B (c) Figure C

Figure 1: Fig. A shows the distribution of Won vs Out of time outcomes across the Age of Acquisition, Fig. B shows the same outcomes across Concreteness, Fig. C shows the same outcomes across Frequency, with each of the linguistic features referring to the target word.



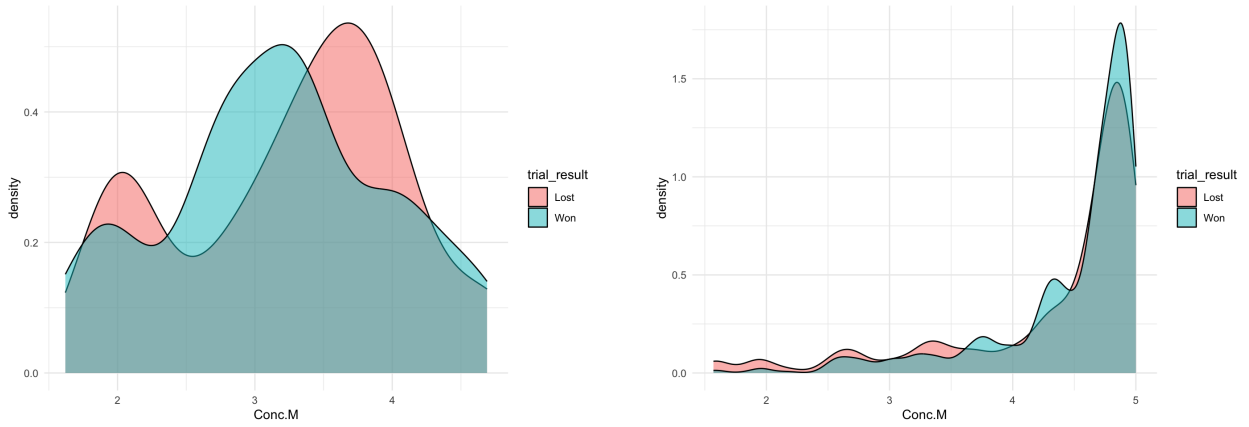
(a) Figure A (b) Figure B (c) Figure C

Figure 2: Fig. A shows the distribution of Won vs Lost outcomes across the Age of Acquisition, Fig. B shows these outcomes across Concreteness, Fig. C shows these outcomes across Frequency, with each of the linguistic features referring to the target word.

From these visualizations, it can be seen that the outcome Won seems to be more densely distributed near the lower range of Age of Acquisition. The outcome Won is also seen to be more densely distributed around high concreteness and high frequency words than the Out of time outcome in Figure 1 and the Lost outcome in Figure 2. So, for comparisons of both Win vs Out of time outcomes and Wins vs Lost outcomes, lower age of acquisition, higher concreteness, and higher frequency seems to predict more Wins.

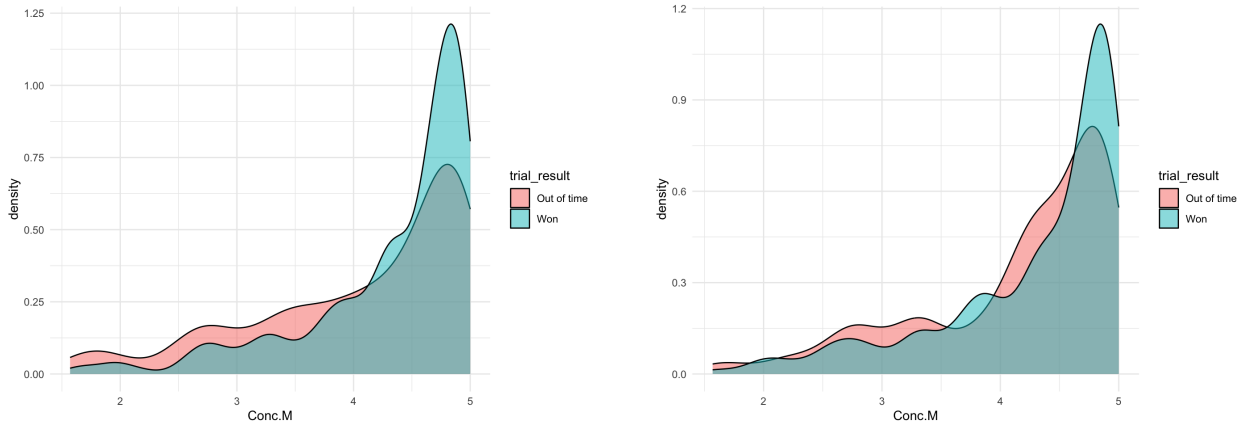
Next, for further analysis, I looked at the density of concreteness for adjectives, comparing Won vs Lost outcomes, and similarly looked at the density of concreteness for nouns in Figure 3. I also looked at the density of concreteness for friends, comparing Won vs Out of time, and the density for strangers in Figure 4.

The analyses across Concreteness for Adjectives vs Nouns with Won vs Lost outcomes, and Concreteness for Friends vs Strangers with Won or Out of time outcomes produced significant effects in terms of interactions. The visualizations in Figure 3 and Figure 4 help show how higher concreteness helps with winning for nouns compared to adjectives, and how higher concreteness helps with winning for friends more than strangers respectively.



(a) Figure A (b) Figure B

Figure 3: This figure focuses on the Won vs Lost outcomes. Fig. A looks at the density of Concreteness across either outcome for Adjectives. Fig B. shows the same distribution for Nouns.



(a) Figure A (b) Figure B

Figure 4: This figure focuses on the Won vs Out of time outcomes. Fig. A looks at the density of Concreteness across either outcome for Friends. Fig B. shows the same distribution for Strangers.

4 Analyses

4.1 Linguistic Analyses Replicating Original Findings

My initial analysis closely follows the analysis in the publication, looking at models predicting outcomes (Won vs Out of Time, Won vs Lost) based on linguistic factors (Age of Acquisition, Concreteness, Frequency), and produces essentially identical results (Trott et al., 2023). The models include random intercepts for Speaker, Guesser, Order, Session, and Deck, and fixed effects of Trial with Partner, Part of Speech, Concreteness, Log Frequency, and Age of Acquisition. Following the original methodology, a full model was created with all of the linguistic variables, and then models were created by dropping each of the linguistic variables one by one. The independent effects of Age of Acquisition, Concreteness, and Frequency on the model fit are evaluated by using ANOVA chi-square tests to compare the full model with each model with a dropped variable.

I only compared the full model with the models for no Concreteness, no Age of Acquisition, and no Frequency, since these were the factors I was most interested in, and some of the other variables in the original analyses, such as average semantic distance, were found to have no independent effects on outcomes in the paper (Trott et al., 2023). They were kept as random effects in the models to replicate the original

analysis accounting for these variables as having possible effects.

When running these analyses on Won vs Out of Time predictions in R, I had to make 2 changes: I had to code Won or Out of time as a binary (1,0) variable for glmer to work correctly, and I removed the random effect of order since it had a variance of zero, leading to a singular fit warning. After making these changes, the analysis corresponded to the findings in the original paper (Trott et al., 2023).

For the Won vs Out of time predictions, the following results were obtained, corresponding to the relevant results in the original paper (Trott et al., 2023):

- Concreteness improves model fit ($\chi^2(1) = 21.81$, $p < .001$).
- Age of Acquisition improves model fit ($\chi^2(1) = 19.36$, $p < .001$).
- Log Frequency improves model fit ($\chi^2(1) = 30.37$, $p < .001$).

Analyzing Won vs Lost predictions on Concreteness, Age of Acquisition, and Frequency, I had to make a few minor edits to the R code. For this dataset, both Order and ppt_id were found to have zero variance, so I dropped them as random effects. In the model without concreteness, color also had zero variance and had to be dropped. After these corrections, the analysis once again corresponded to the original analyses in the paper (Trott et al., 2023). The following results were found, comparable to the original results (Trott et al., 2023):

- Concreteness improves model fit ($\chi^2(1) = 20.67$, $p < .001$).
- Age of Acquisition improves model fit ($\chi^2(1) = 13.7$, $p < .001$).
- Log Frequency improves model fit ($\chi^2(1) = 61.85$, $p < .001$).

4.2 Further Analyses on Interactions with Concreteness

I decided to further analyze the full noun phrases dataset in the original SCARFS database. I wanted to observe the interaction effects between Concreteness of target words across a few variables, and see whether the interactions had significant effects on the outcomes (Won vs Out of time and Won vs Lost). For all of the models, random effects of participant ID and card were added, similar to the procedure followed in the analyses of the linguistic factors in the original paper (Trott et al., 2023).

First, I looked at the outcomes Won vs Out of Time, comparing a model with an interaction term between Concreteness and Part of Speech to a model without one. I focused on Adjectives and Nouns for this analysis, since I was interested in the effects of concreteness for adjectives compared to generally more concrete nouns. I hypothesized that concreteness would affect adjectives more than they did nouns, since nouns have a higher concreteness on average. This interaction between Concreteness and Adjective or Noun did not improve model fit, since the chi-square test comparing the two models was not statistically significant.

However, when following a similar procedure comparing models with an interaction between Concreteness and Part of Speech to a model without an interaction, I found a significant effect across Won vs Lost outcomes. For the Won vs Lost comparison, the interaction effect model improved predictions ($\chi^2(1) = 4.60$, $p < .05$). The effect of concreteness is stronger for nouns than for adjectives for these outcomes, since the interaction term is positive. This could be due to the higher concreteness of nouns specifically helping in winning when compared to losing by saying a taboo word, compared to higher concreteness for adjectives. More concrete nouns might be easier to communicate than more concrete adjectives.

Next, I analyzed the outcomes Won vs Out of Time for a model with an interaction term between Concreteness and Pair Type (Friends or Strangers) to a model without this interaction term. I hypothesized that higher concreteness would have a more positive effect on winning for strangers than for friends. The interaction improved model fit based on the ANOVA chi-square test ($\chi^2(1) = 6.18$, $p < .05$), which shows that concreteness does predict winning or running out of time differently for friends and strangers. Interestingly, the coefficient on the interaction term between concreteness and pair_type being stranger was negative (-0.37), indicating that higher concreteness actually helps strangers less than it helps friends in winning vs running out of time. This could be due to concreteness, which is independently effective in improving winning

outcomes, being even more valuable when communication is between a pair of friends who already share common ground and have a communication pattern or style, than between a pair of strangers.

Analyzing the same interaction term but for the outcomes Won vs Lost, the interaction term did not significantly improve model fit. This could be due to higher concreteness between pairs of friends and strangers helping winning outcomes to the same extent compared to saying a taboo word, perhaps because friend and stranger pairs have more similar effects of concreteness for wins and losing by saying a taboo word.

These analyses produce interesting interaction effects for linguistic properties of target words (Adjectives vs Nouns) and relationships between participants (Friend vs Stranger) with Concreteness on Won vs Lost and Won vs Out of Time outcomes. Since each of these 'losing' outcomes happens due to a different failure in communication, it makes sense to analyze these outcomes separately. It would be helpful to further analyze these interaction effects, especially why their significance differs across outcome comparisons. The direction of the interaction is also interesting in either case, since my hypotheses expected opposite directions of these effects.

5 Limitations

My analyses reproducing results from the original publication closely followed original results, so there were no discrepancies in these results (Trott et al., 2023). However, my independent analyses on the full noun phrases dataset may have limitations, since there are multiple possible hypotheses that can be tested using similar interaction effects. While I followed the approach using a comparison of models with relevant random effects and chi-square tests from the original paper, I could have tested more interactions to find out which ones, if any, were significant across both sets of outcomes (Trott et al., 2023). An unbalanced dataset for Adjectives vs Nouns may have affected the effects of the interaction between concreteness and part of speech. Additionally, more theoretical domain knowledge could have led to the construction of other alternative hypotheses for these analyses. The correlation matrices for the fixed effects could be inspected further to understand any lack of independence between variables in the dataset as well.

6 Conclusion

Overall, my analyses corresponded to the findings in the original publication, and produced interesting results about the interaction of concreteness with other factors in the full noun phrases dataset (Trott et al., 2023).

Specifically, I reproduced the original results finding that 3 linguistic factors (Age of Acquisition, Concreteness, Frequency) of the target words independently improve predictions of both Won vs Out of time and Won vs Lost outcomes (Trott et al., 2023). Lower age of acquisition, higher concreteness, and higher frequency of the target word positively predict wins.

Additionally, from further independent analysis, I found that concreteness affects winning differently for adjectives than nouns, compared to losing by saying a taboo word. The effect of concreteness is stronger on nouns than on adjectives. Also, the effect of concreteness is stronger for friends than it is for strangers, when comparing winning to running out of time.

References

Marc Brysbaert and Boris New. Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, 44(4):978–990, 2012.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46(3):904–911, 2014.

Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods*, 44(4):978–990, 2012.

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

Sean Trott, Benjamin Bergen, and Eva Wittenberg. Spontaneous, controlled acts of reference between friends and strangers. *Language Resources and Evaluation*, 57:1081–1105, 2023. doi: 10.1007/s10579-022-09619-y. URL `https://doi.org/10.1007/s10579-022-09619-y`.