There’s no such thing as cheap talk:
A machine learning analysis of pre-play communication on Golden Balls

Ethan Bevington, Class of 2019

Motivation

In game theory, “cheap talk” refers to communication between players that does not directly affect the payoffs of the game, no matter what the players actually know. That is, a message in a game theory model is cheap talk if the player can freely lie. Game theorists have shown that even cheap talk can be informative if players’ interests are aligned (Crawford and Sobel, 1982) but cheap talk should not contain any information, and should therefore not be correlated with actions, if the players’ interests are directly opposed. However, modern behavioral economists have questioned this prediction. One reason that it may not be true is that unconscious biases may affect the content of messages, even when players do not intend this.

Testing the information content of cheap talk in real-world situations is difficult. Our project tests this hypothesis by analyzing transcripts from the final round of the British game show “Golden Balls.” On this show, contestants play a game similar to the classic Prisoner’s Dilemma, in which the players’ material interests are directly opposed. The game consists of three rounds, the first two of which are not important to our project. The final round, however, requires players to pick between a split and steal ball, which determines the outcome regarding the jackpot they have obtained. If they both split, they each get half the jackpot. If one splits and one steals, the one who steals takes home the entire jackpot. If both steal, both players receive nothing. Prior to the final decision, players are given time to discuss what they are going to do, which creates the speech we have used to build our models.

Methods

Transcripts for 132 players were created, allowing us to generate tables of the particular unigram (single words), bigram (pairs of words), and trigram (groups of three words) usage for each player. We also consider interactions between usage of each unigram by the player and by their opponent, and the observed characteristics pot size, race, gender, and age. The goal was to see which text terms, if any, were predictive of the player’s action (choice of split or steal).

Because there were many possible predictor variables (hundreds), it was appropriate to use a machine learning statistical method that selects which variables are the best predictors. We focused on LASSO regression and compared many sets of variables, which helped us understand which variables are the best predictors. We used a model with no text terms as a baseline (race, gender, age, and pot size). We then considered variants with different combinations and sets of unigrams, bigrams and trigrams, and interactions.

Results

Players chose ‘split’ 55 % of the time. Thus, if one simply predicted split, then the prediction error rate would be 45%. When observed non-text variables are included, the error rate decreases to 0.44. When just the top 30 unigram terms are included, the error rate goes to 0.4. Thus, cheap talk matters. The error drops to 0.35 as models increase in complexity. When we consider more text, our models suggest the number of times an opponent says ‘money’ predicts a player will split. ‘The money’ predicts splitting in top 50 models, indicating people trying to maximize their own money by stealing avoid the word. Additionally, the phrases ‘want to split,’ ‘lot of,’ and ‘go home’ predict splitting while ‘through,’ ‘change,’ ‘son,’ ‘we,’ or ‘team’ predict stealing. In the top 40 and top 50 models, individual usage of ‘change,’ ‘son,’ ‘we,’ and ‘team’ remains predictive of stealing. This suggests people who reference past cooperation tend to be lying about their intention to split.

Faculty Mentor: Professor Daniel Stone
Funded by the Craig A. McEwen Student Research Fellowship in the Social Sciences