Deliverable-2 Report

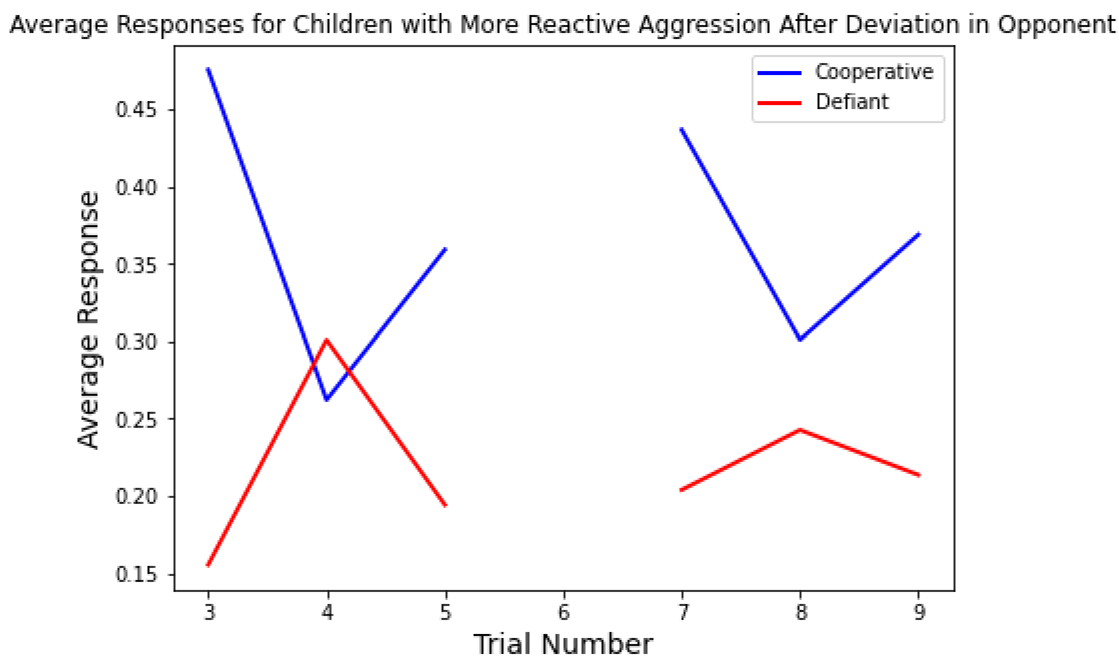
Dataset - Non-Twin Dataset

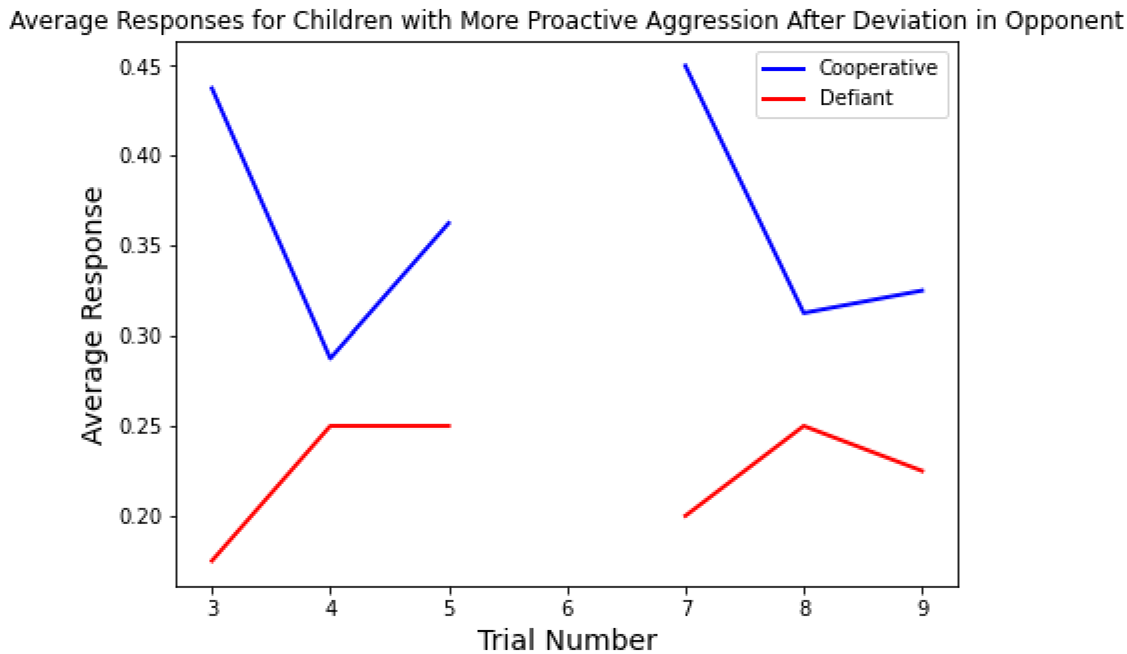
Refined Preliminary Analysis

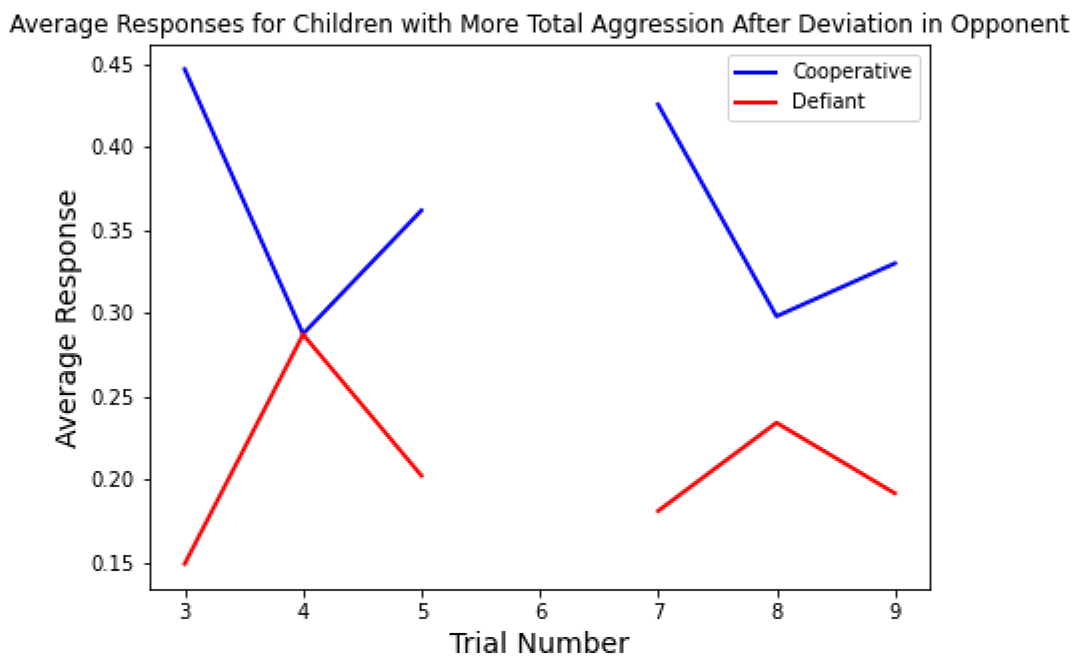
We are looking to answer the following question:
Does the partner's deviation from their initial pattern (as pre-programmed) cause a change in state (e.g preference) in the child?

Preliminary analysis included looking at how the average responses changed in rounds 3-5 and round 7-9 as the partner deviated from the initial pattern of cooperation and defection in the rounds 3 and 7. Previously we considered the subjects if their reactive, proactive and total aggression score was more than mean of each aggression score. But now after discussion with the client, we realized that considering the subjects whose aggression scores are more than equal to the median is more suitable for our data.
Hence, we refined our preliminary analysis to be based on median of aggression scores. The average responses vs round number plots for each of reactive aggression, proactive aggression and total aggression considering subjects whose aggression score was more than median of each aggression score are below :

Figure-1







These plots can be compared with previous deliverable plots which considered mean aggression scores instead.

Question Being Answered:
Can we predict aggression based on actions and reaction times?

We answered this question by training various regression models using the kids' decision data, and then predicting the kids' aggression scores using these models. In general, we built two kinds of models: one for discrete aggression levels prediction and one for continuous aggression score prediction.

1. Logistic Regression Classifier for Discrete Variable Prediction

Logistic regression classifier was trained using the sklearn library. Since this classifier can only predict discrete categorical variables, and some aggression score group has too few members, we first classified the children into above-the-median and below-the-median in respective to each aggression score (active, proactive, and total); samples with missing aggression scores are dropped. We will train a classifier for each kind of aggression score and each kind of opponent and report their accuracies by conducting 5-fold cross validation experiments.

We first trained the classifiers using all ten decisions of the children against each type of opponent. The classifiers for predicting all three kinds of aggression scores when trained using the decisions against the TFT opponent had much higher accuracies than the ones trained using decisions against COOP or DEF opponents. However, all classifiers had accuracies below 0.6. We hypothesized that this low accuracy is probably because the kids' aggression level is only reflected through their decisions on particular rounds instead of all ten rounds together.

In an attempt to improve the accuracies of our classifiers, we trained classifiers using only the later 7 decisions, since the opponents are programmed to make an unexpected move on round 3 (i.e. COOP opponent will defect, and DEF opponent will cooperate). This time, almost all classifiers had an accuracy of around 0.56. The accuracies of COOP and DEF classifiers noticeably increased, whereas the accuracies of TFT classifiers decreased.

Lastly, we trained classifiers using only the kids' decisions on round 4 and 8, the two rounds immediately after the opponents make unexpected moves. No significant improvements are seen, but the accuracies of COOP classifiers when predicting the reactive and total aggression scores decreased to only 0.44, indicating that just the two immediate rounds following the unexpected moves of the opponents are not sufficient to predict the kids' reactive or total aggression scores.

2. Continuous Variable Prediction

Continuous models for the prediction of aggression based on round choices and reaction times were implemented with linear regression, random forest regression, and support vector regression. Root mean squared error and R^2 were calculated for every model. The difference in models trained and tested on the choices and reaction times for different rounds of each game type were evaluated to determine which rounds were most predictive.

As a general rule, proactive regression was the easiest type of aggression to predict. Predictions of reactive aggression and total aggression were significantly less accurate for all partner types, models, and round ranges. For the most part, all three regression models performed comparably, with no clear winner for any situation. Models predicting aggression based off of reaction time often resulted in different predictions than round choice models. For the most part, round choice data predicted proactive aggression somewhat better than reaction time models, but reaction time data occasionally provided better results elsewhere.

Overall, the RMSEs of the predictions were far less than ideal for reactive aggression, and the R^2 scores seemingly showed nearly no correlation between the model and the data. A more in depth analysis of the model for reactive aggression will need to be carried out in the future. An integrated model using both reaction times and round choices for prediction will be constructed as well.

Refined Project Scope

Through literature reading and online research, we have decided to move away from the Drift Diffusion Model because 1. We are dealing with reaction times that are much longer than 1000-1500 ms, which are the ideal reaction time length for DDM, and 2. The decisions made in this experiment are not one-step intuitive decisions but rather a process that requires reasoning and strategy planning. We have reached accordance with the client that we will be providing an equivalent amount of insight into the data using alternative methods, such as training regression models and Bayesian Models.

Note

The code files 'data_processing.py' (Python file) and 'data_processing.ipynb' (Jupyter Notebook) are placed in the 'Deliverable_2' folder which is inside the 'Deliverables' folder.
The code has to be further optimized and structured for better understanding and presentation.