Drift Diffusion Models of children's interactions in a Repeated Prisoner's Dilemma Game (Team 2)

Dakota Clark-Singh
Audrey Tucker
Ryan Shnitman

- At the moment Ryan is having some medical problems so he has not been able to contribute directly to this deliverable. Given that he created the regressions, we were not able to explain those on our own, so we've highlighted and noted where we would need his input to know what exactly he tried and what his results were. So far Cody and I have done as much as we can on our own.

Deliverable 3

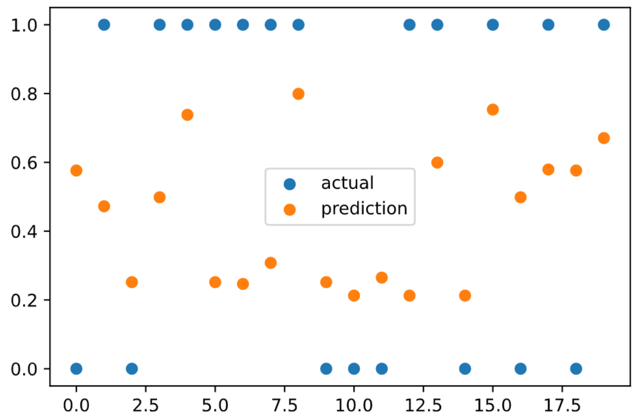
Approaches Taken to answer project questions:

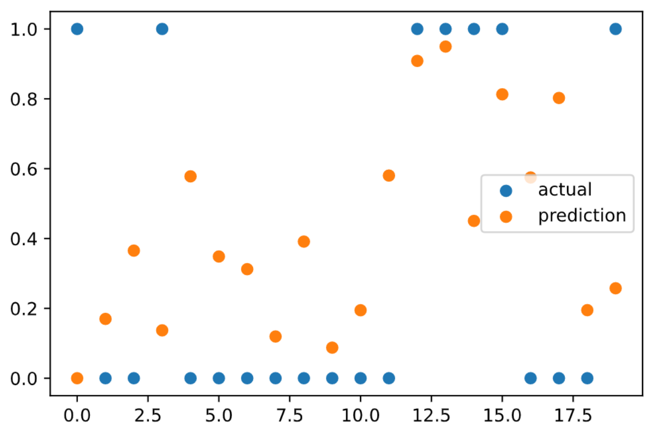
1. Decision Tree Model

Following advice from Steve and Gowtham, we decided to move away from the LSTM model and instead create a decision tree model. Since our dataset is so small, the more complex LSTM model would be susceptible to overfitting by memorizing our limited samples and not generalize well to other data. A decision tree model could be more robust to overfitting given less training data. Our goal for the decision tree model was the same as the LSTM: predict whether a child would cooperate or defect in a given round using information about the child's attributes, specifically aggression, and the decisions of the child's partner in the proceeding rounds. Such a model could provide greater insights into the patterns in the decision making process of the children rather than looking at the total number of times the child cooperated and defected as an aggregate.

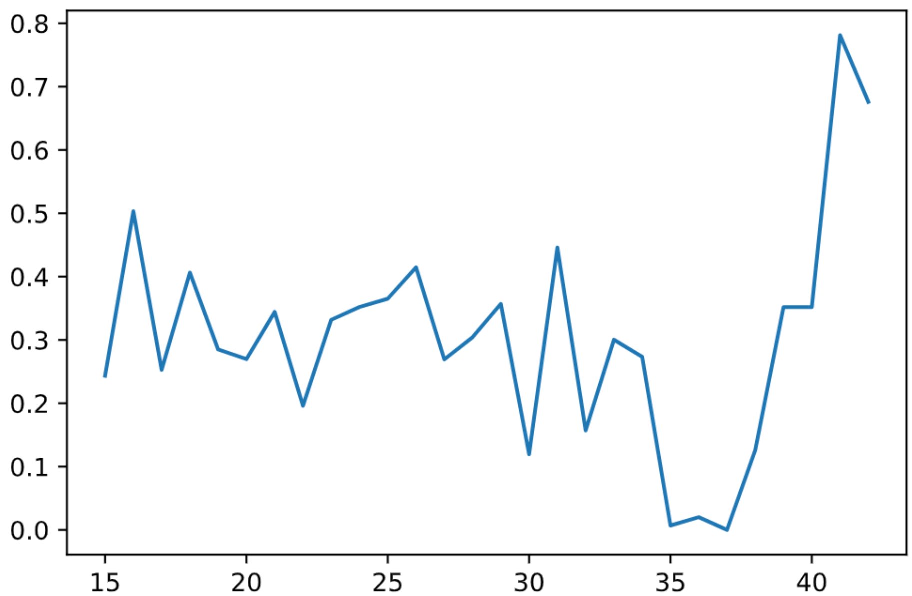
To combat overfitting in our decision tree, we decided to use a random forest model. A random forest is a model made of many decision trees and takes a random sampling of the training data when training each tree, so each tree is exposed to a slightly different subset of the data. Also, a random forest can use a random subset of features when splitting nodes. The prediction of the random forest is the average of each tree's prediction. These characteristics make the random forest more robust to overfitting. We constructed two simple random forest models. The first used the child's self reported aggressiveness and their partner's last decision to predict the child's next decision. The second model used the partner's last three decisions rather than just one. Each model predicted a number between 0 and 1, closer to 0 being more likely to defect and closer to 1 being more likely to cooperate.

Both models could make generally accurate predictions, with the second model consistently making predictions closer to 0 or 1. These figures show the first 20 predictions with orange dots and the actual decisions with blue dots, for our first and second random forest models respectively. The closer the orange and blue dots are for a given index on the x-axis, the more accurate the prediction.





We were hoping this model would allow us to answer the questions “Are children with higher aggressive traits more less likely to return to cooperation over the three rounds after partner defection (less likely to forgive)?” and “Does the partner's deviation from their initial pattern (as pre-programmed) cause a change in state (e.g., preference) in the child?”. Using the model trained on the three previous partner decisions, we constructed various sequences of three moves and tested each across a range of aggression values to see if the predicted behavior was different. Unfortunately, there was no clear linear trend between aggressiveness and the likelihood of cooperation given a specific sequence. For example, these are the results for the sequence “defect”, “defect”, “cooperate” over a range of aggressiveness values:



We chose this sequence to test a child's tendency to forgive and cooperate if their partner defects twice then cooperates. If this tendency could be linked to aggression, this would answer one of our key questions. However, the results do not demonstrate any clear relationship here. We hypothesize that this is again due to our lack of data. The random variation in the sample of children that participated in the experiment could be hiding a meaningful relationship.

We could try to improve this model by also including the response time as a feature.

2. Classification Approach

In addition to predicting the next move based on trends in aggression and responses to different sequences of moves, we explored models based on assigning each child a strategy based on their decisions when playing against the cooperative simulated partner. We thought this would be a good baseline to see how children behave when their opponent is mostly cooperative. The strategies we defined were:

- draw blood: defects in the first three rounds
- Full grudge: cooperates in first three rounds and defects for all of the last seven rounds
- Partial grudge: cooperates in the first seven rounds (allowing one retaliatory defection in round 4) and defects in the last three rounds
- Tit-for-tat: same as the simulated tit-for-tat, mimics partner's last move
- Full cooperation: cooperates in every rounds
- Other: any other sequence of moves

We modified the dataset by adding a column with the strategy that each child falls into based on their decisions with the cooperative partner.

Next, we designed a kNN and a logistic regression model to predict a child's strategy given their age, parent's education and income level, gender, and their various behavioral indexes. However, the results of these models did not show a strong relationship between these features and the strategy. Again, we think that this is mostly due to our lack of data. The mean squared error on the testing set against the kNN predictions was 3.26 and its mean accuracy was 0.46. For the logistic regression predictions, the mean squared error was 3.25 and the mean accuracy was 0.34.

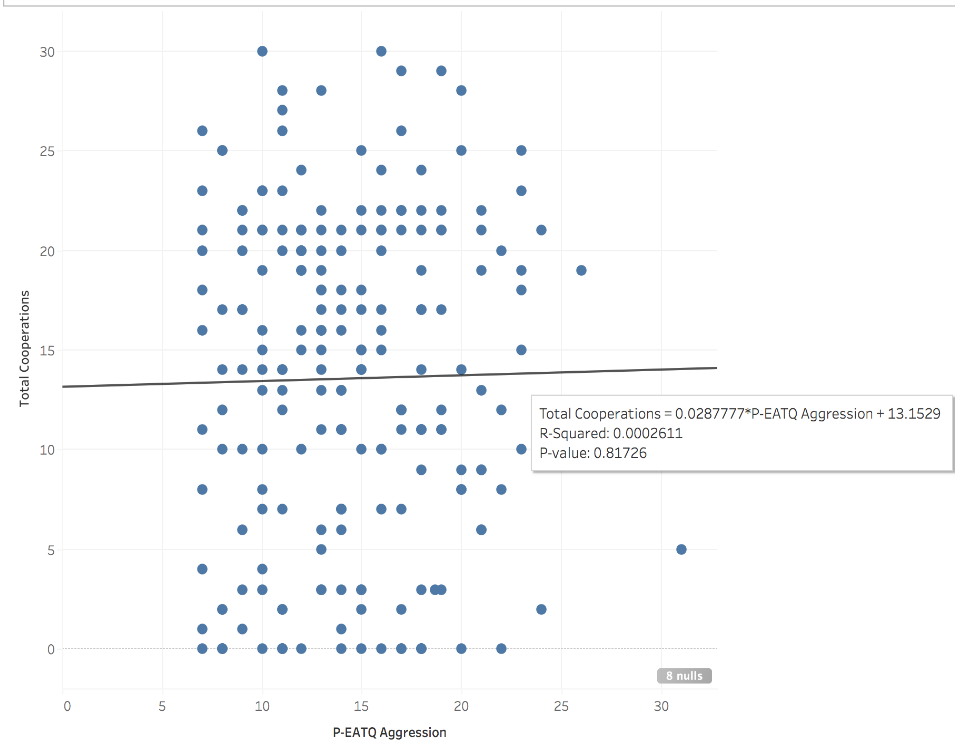
With more data, we think that this classification based approach could prove useful to understanding and quantifying the relationship between characteristics of children and their parents and their behavior and decision making. Our pre-defined strategies could give predictions more real world relevance than only predicting a child's next decision. We intend to explore this approach further and combine it with our time series decision tree model.

Our regression model - and why we suspect there aren't any clear patterns:

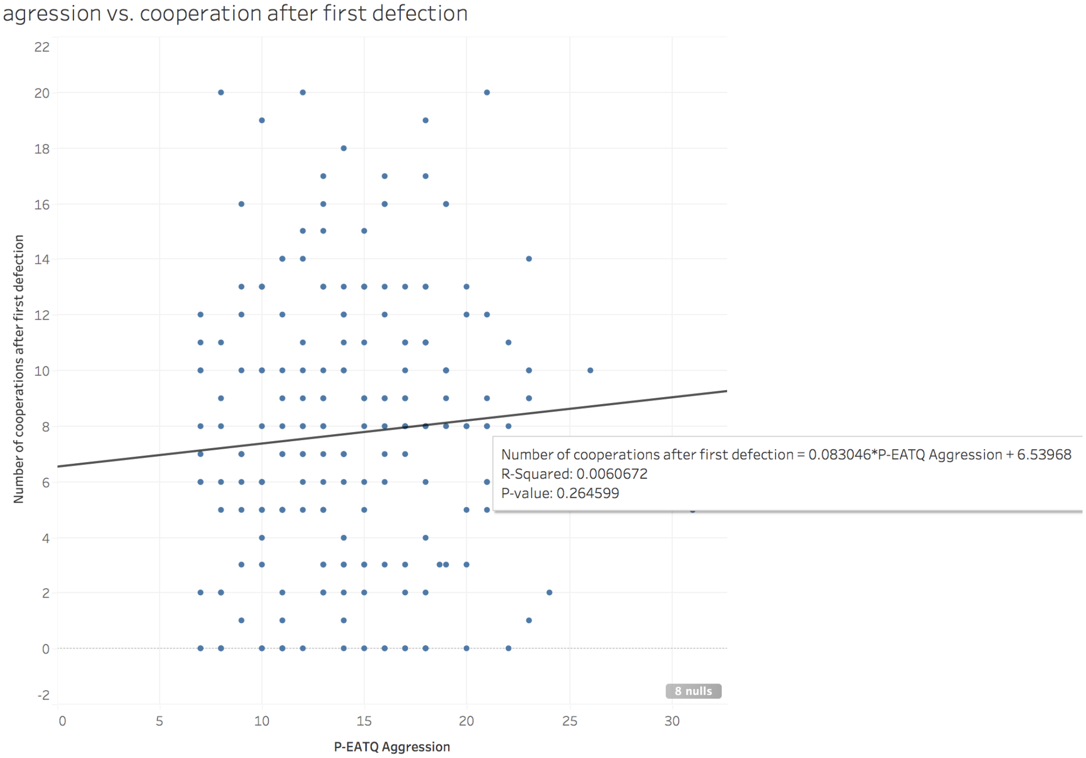
Consistent with what we found for our last deliverable, our regression model did not show any clear patterns between aggression and cooperation/defection (consult Ryan about what regressions were tried in particular). In addition to using Python's tools, we also created several scatterplots using Tableau to try and better understand the relationship between aggression and defection. As shown below, the high p-values and low R-squared values show little to no patterns in the data with regard to cooperation and aggression scores. P-values higher than 0.05 are often used as a benchmark for trend lines showing statistically significant correlations or not, so these very high p-values indicate little to no correlation. To the same effect, R-squared values whose absolute value is close to 1 indicate a perfect correlation between two variables, so low R-squared values indicate low correlation between variables.

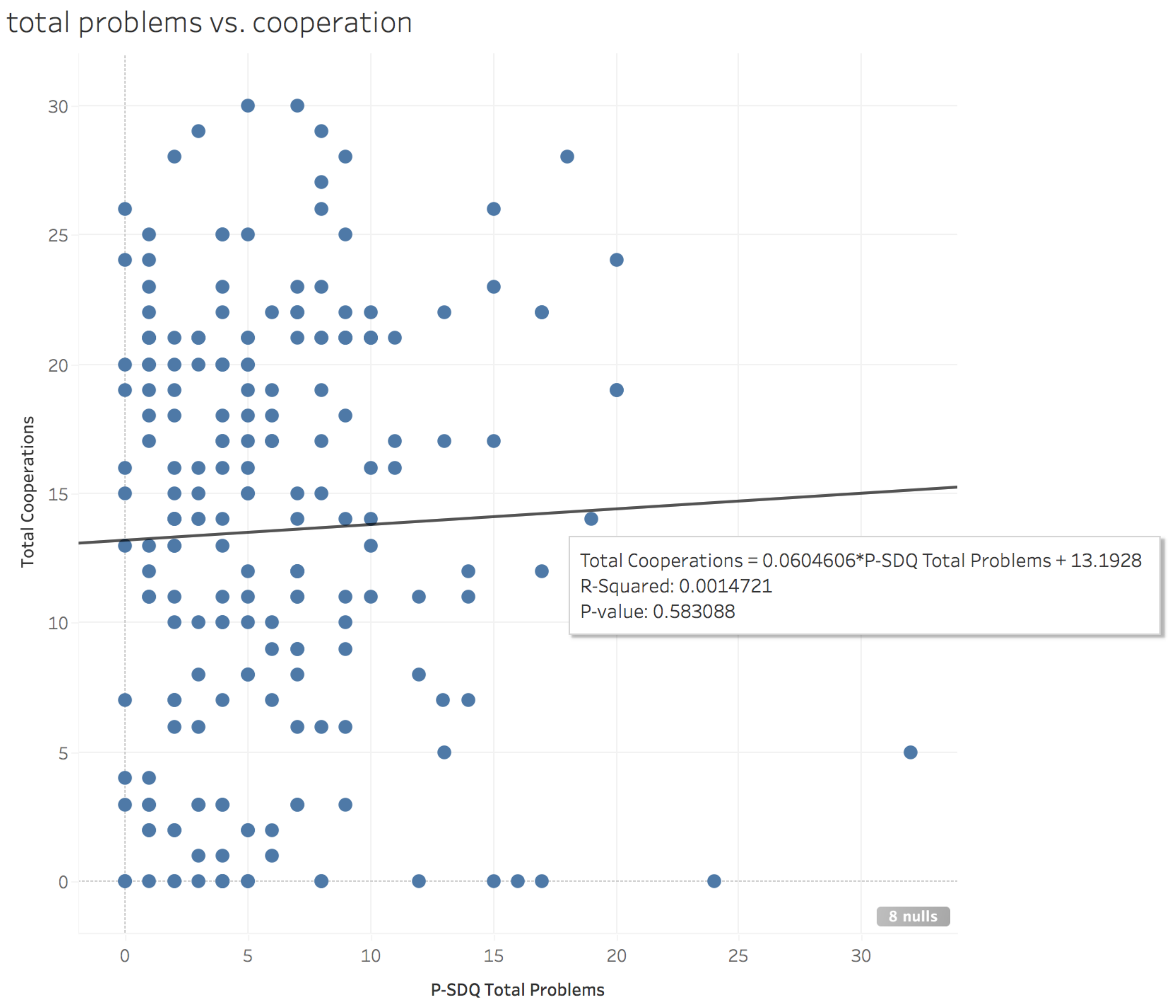
To answer one of our original questions, we looked at scatterplots of cooperation after the first defection, and in total across all rounds. Similar to the results of our regression analysis (consult Ryan), the trends shown below were not statistically significant. So far our results indicate that there may be no connection between high aggressive scores and decisions to defect or cooperate.

agression vs. total cooperations



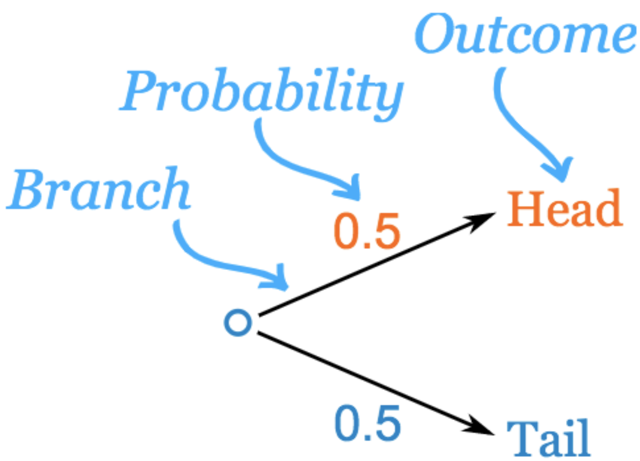
agression vs. cooperation after first defection





As we begin to more clearly see that there was no overwhelming pattern in our data (at least using a regression model), our focus is shifting from trying to find a trend, to trying to explore reasons and more concretely prove that there may not be a trend in our data.

After speaking with one of the project managers Gowtham, we learned that a regression may be a bad fit for our data, because the decisions in each round are connected - so the child's decision to defect or cooperate is related to what they, as well as their partner did for the last round. It is for this reason that he recommended looking into the decision tree model - given that this model predicts future decisions based directly on past events or decisions. An example of a simple decision tree is shown below. This is a decision tree for a coin flip, where 0.5 is the likelihood that it will land on heads or tails for each flip (represented by the node labeled zero).



[1]

Potential reasons why we are not seeing strong correlations in our regression models:

- **Violated assumption of Independence:**

Our data does not follow the assumption for linear regressions - that the different variables (so the different coefficients) must not be correlated (meaning all our variables should be independent).[2] This assumption also must be met in logistic regressions.[3]

---

1 https://www.mathsisfun.com/data/probability-tree-diagrams.html

2 https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R5_Correlation-Regression/R5_Correlation-Regression4.html

3 http://www.statisticssolutions.com/wp-content/uploads/wp-post-to-pdf-enhanced-cache/1/assumptions-of-logistic-regression.pdf

(consult with Ryan to see if we tried any other types of Regressions). However, as shown in our correlation matrix, many of the variables are strongly correlated. Especially given that the experiment was a prisoner's dilemma game, where each decision to defect or cooperate is dependent on what the partener did the last time.

This lack of independence between rounds may explain why we are not seeing patterns when using the regression approach. When speaking to one of the spark project managers (Gowtham), he suggested using a decision tree model. This model predicts the probability of each possible series of decisions, where each decision or event is represented by a node in the tree. Therefore, this model accounts for the dependence of each decision on what happened last.

- **Small data set:**

As we have already discussed in previous deliverables, our data set is quite small so even if there are patterns, they do not appear clearly when analysing this data set. When speaking with Gowtham he confirmed that our data set is very small to perform machine learning, and still pretty small to run regressions on. Unfortunately there is nothing we can do about this.

- Include correlation matrix - use this to justify/prove collinearity and correlation between variables in our regression model (what were the actual variables used in the regression model? - consult Ryan)
- Include new regressions Ryan tried
- Mention feature selection using Lasso model - I want to talk about this but I don't know what Ryan ended up using it for

---

Limitations with data/risks of achieving project goal:

As we run more analysis, it seems likely that there is no clear pattern in the data (which is possible with a dataset so small). If this is the case, we still plan on giving our client a detailed report of what we did find, and some possible explanations (along with supporting graphs) detailing why there might not be a clear pattern in the data. In our meetings with the other team, they seem to be running into the same issue of not finding clear patterns in the data, so at the moment this data set/experiment may be inconclusive.