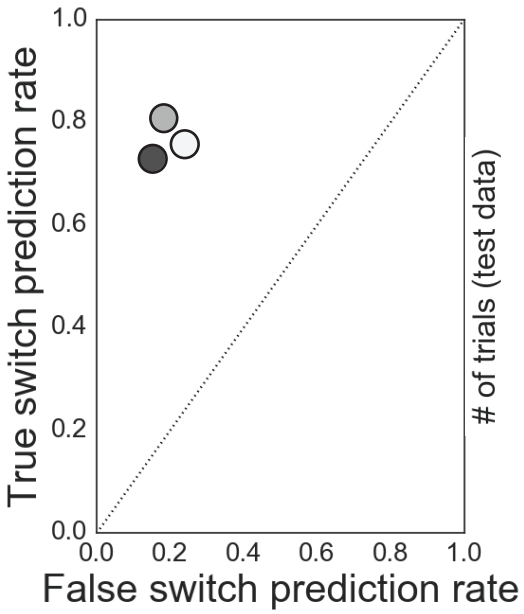
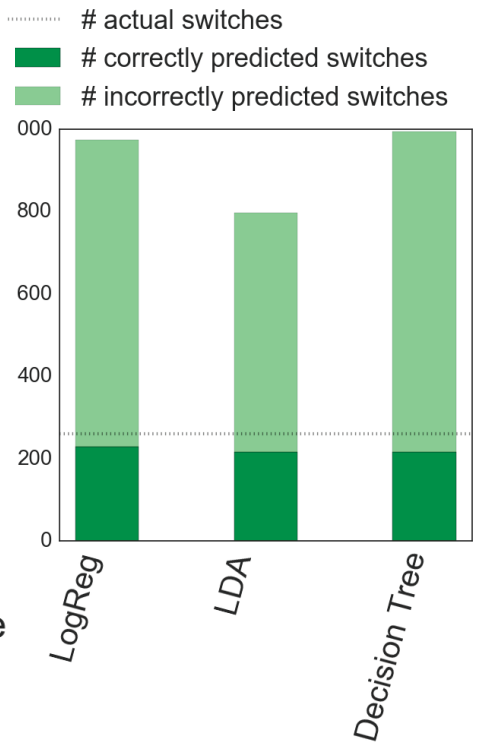
Classifying ‘switch’ & ‘stay’ decisions

Logistic Regression
Regularization : 10e-4
penalty : l2
Class Weight : balanced

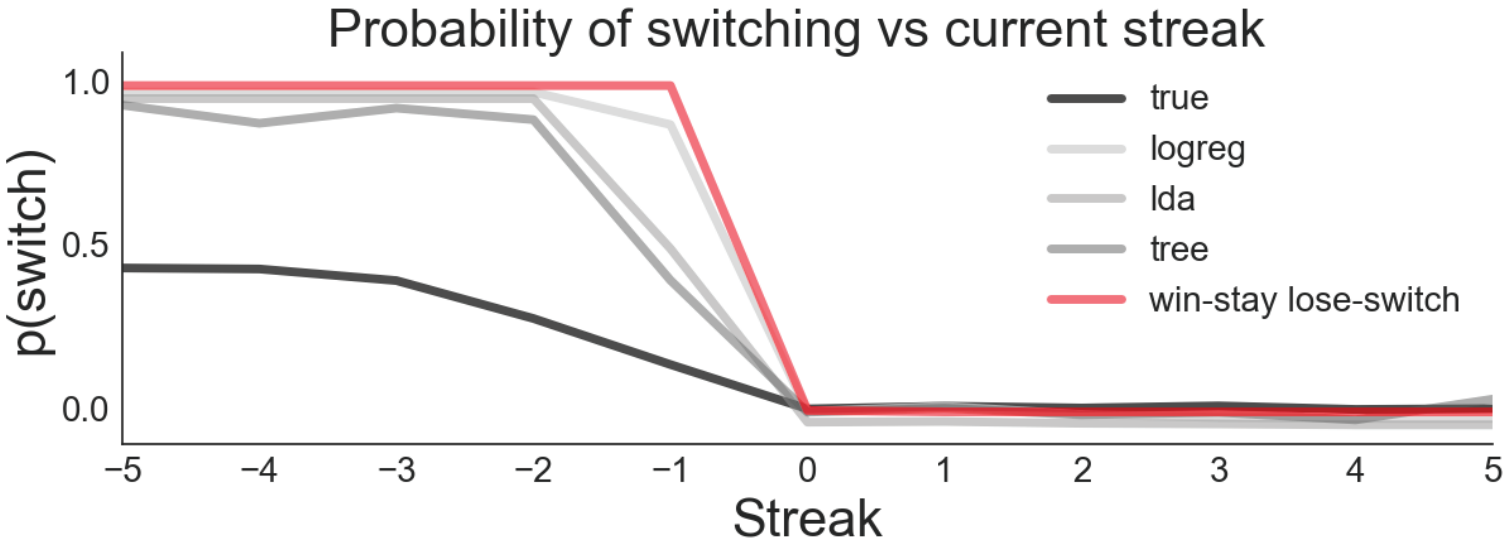
Linear Discriminant Analysis
Priors : {Stay 0.6, Switch 04}
penalty : l2

Decision Tree Classifier
Depth : 5
Class Weight : balanced
criterion : gini





Although these models are ~80% accurate at predicting both switches and stays, the switching accuracy is achieved by *drastically over-predicting the number of switches.* This presents a severe limitation to using model to interpret the behavior (e.g. by studying the beta coefficients). What is causing the models to predict switch when the mice decide to stay?



Above we can see that our models are essentially employing ‘win-stay, lose-switch’ strategy. However, in the actual data, mice only switch ~20% of the time following a single ‘no-reward’ outcome. Nonetheless, this result inspired us to take a closer look at the experimental data: