**usr\_start0-4** : the users first 5 moves when joining the game (before really adapting to the game). user\_start0 being the very first move against a real opponent.

**user\_overall\_mem0-9**: the user's memory of the past 10 moves overall. (payoff) index 0 means the oldest, 9 means the most recent.

**direct\_history0-4:** the user's memory of the past 5 games against the **same opponent** (payoff)

Also note that the 0.5 is just a placeholder for "no\_data".

Let me know if that is unclear. There is no indication in the data I gave you that the user has played the same player exactly last round. This can be extracted from raw data, but it is significant work, if you absolutely need it, I can see if I can do it (or give you data that helps you extract it).

==================================================================

This repository contains some simple analysis and raw data for the prisoner's dilemma game we ran on Facebook a few years back. Please don't share the data with anyone, as in it's current state it might still be possible to recover identities.

The two pdf files show some summary statistics from the analysis of the data back then (some of that data is not available to you due to privacy concerns, I can anonymize and give it to you if you need it).

Key notes: 1- For this data set, the "usr" means the responder (moves second), and the "opp" means the requester (moves first).

2- Different stages of the game have different payoffs (they are described as the payoff to the user under CC, SC, CS, SS). CC is for both cooperating, SS for both defecting, and CS for user cooperating and opponent defecting.

3- Each row is a completed game.

4- History/memory data are numbered as this: memory\_0 means oldest memory, and memory\_9 means the most recent memory. History is about user moves overall, memory is about direct interactions with the same opponent.

5-user history was shown to the opponent (0 means they defected 0 percent of the last 5 games, 100 means they defected 100 percent of last five games overall).

==================================================================

Anywhere that the values are 0,0.5,1 these are actually strategies, not payoffs (with 0.5 being imputed (lack of) data).

There is no equivalent to opponent start data, I simply haven't compiled that into this dataset.

In order to make things a bit easier to trace back (because I don't currently have the bandwidth to debug this) and give you the opportunity of using the raw data more directly (so you can mine any sort of information you want), I added two additional csvs to the folder.

"games\_raw\_anon" and "users\_raw\_anon"

The users csv includes the IDs, gender, year of birth and score (you should ignore the score, it was degenerate).

(I have done a simple transform on the IDs to anonymize the data. )

The games csv is basically a history of all the games in the system.

In there whenever you have "t" for a strategy, it means the user -defected-.

The seen\_bit says whether the responder has seen the request and maybe chose not to respond.

You should ignore same\_parity at least for now, because I have to go back and find in my notes how exactly we implemented that. If I remember correctly it meant that for same parity of fb id's they would see each other's history while playing each other. But I have to check this.

note that "stage 1" was just the entry point in the game and the user would just play a robot. In these games we pretended the user played themselves.

There are also 3 "anomalous" users. IDs for them are:

7970- This is a bot that played tit-for-tat

9970- This is a very active user (skews your data)

9512- This is a bot that almost always cooperated (a user wrote this, but he may have played some games too).

So you may want to filter them in certain parts of your analysis or reweight them.

I will take a look at the processed data that I gave you originally and see if I can find the correct explanations, but meanwhile, you can use the source data to produce any analysis you want.