APPENDIX: LISTS OF MEASURES FROM THE TTT AND CFG TASKS

April 5, 2020

Measures from the TTT task

- **Shutter size:** The "spatial shutter" size of each participant is the mean distance between the chosen moves and the potential paths for winning of their previous move. Specifically, we used the following procedure: If the chosen square belongs to one of the potential paths stemming from the previous 'X' move, the distance is zero. Otherwise, the distance is the minimal Manhattan distance between the chosen square and the potential paths stemming from the previous 'X' move. A potential path is a valid path (horizontal, vertical or diagonal) of 5 squares, which is not already blocked by the 'O' player.
- **X misses [total and percent]:** Number of times that a participant could make a winning move for the 'X' player, but made another move instead.
- **O misses [total and percent]:** Number of times that a participant could make a winning move for the 'O' player, but made another move instead.
- **Vector of probabilities to use each scoring heuristic:** We assess by the log likelihoods of the player's moves the probability she uses a specific scoring heuristic. We define scoring heuristics in increasing levels of complexity:
  (1) Density: A square's score is the number of its neighboring squares marked by X.
  (2) Linear: A square's score is the sum of scores of its potential winning paths, where each path score is the number of squares with an 'X' in it.
  (3) Non-linear: Same as Linear, but each path score increases substantially the closer it gets to a win: $\frac{1}{n-n_i}$ (where n is the number of Xs needed to win, and n_i is the number of X in path i).
  (4) Interaction: Same as the Non-linear scoring, augmented by an interaction term between the shared paths. The added score is given by $\frac{n_i \cdot n_j}{(n-1)^2-(n_i \cdot n_j)}$.
  (5) Forcing: Same as the Interaction scoring, but includes an additional constant bonus score for forcing player O to move to a specific square to block an immediate threat by X.

  For each heuristic, we compute the likelihood of each move by normalizing the scores into probabilities. We then compute log-likelihood fit between the participant's moves and each heuristic.
- **Time of play:** Total time to complete the game for each player.
- **Solved [yes/no]:** Whether the participant solved the problem correctly.
- **Entropy:** The entropy of all moves chosen by the participant throughout the game.
- **Correlation of log likelihood vs. predicted log likelihood:** For each scoring heuristic, we compute the correlation between the log-likelihood predicted by the scoring heuristic and the log-likelihood based on participants' moves.

- **Total # of search moves:** The number of actions taken by a participant in the game.
- **Total # of resets:** The number of times the participant resets the board.
- **Mean and Median duration length of search until reset/done (time and steps).**
- **Mean and Median duration length of search in exploration (time).**

Measures from the CFG task

- **Total play time:** The total time players play the game, measured from after they finish the tutorial session until the game ends (12 min) or they push the "end game" button.
- **Total # moves:** The total number of square moves players did during the entire game.
- **Average Speed:** Total number of moves divided by total play time.
- **# galleries:** Number of shapes saved to the "gallery".
- **Self avoidance:** Number of repeated shapes in the entire game trajectories (how many times a player produced the same shape, in the graph of shapes this is pronounced as a crossing of its own path and we count the number of total crossings).
- **Uniqueness of rated shapes:** Uniqueness score is defined as the minus log of the frequency of the shape being created by all players in this experiment (Uniqueness $= -\log[\text{frequency}]$). The rated shapes are the 5 shapes players choose at the end of the game as most creative.
- **Uniqueness of rated shapes (united):** Uniqueness score is defined as the minus log of the frequency of the shape being created by all players in all experiments thus far (~ 1200 players) (Uniqueness $= -\log[\text{frequency}]$). The rated shapes are the 5 shapes players choose at the end of the game as most creative.
- **# clusters:** The number of exploitation bouts a player had in the game.
- **% shapes in exploration:** The number of shapes created in the exploration phase. The benefit of this measure is that it scales out the total number of shapes created and the speed of the player.
- **% time in exploration:** The percent of time spent in exploration phase. This measure accounts for the speed of play in exploration and exploitation.
- **Exploration optimality:** The median ratio between the minimal number of moves between two consecutive chosen shapes (in the exploration phase) and the number of moves the player actually took in the game. The closer this number is to 1 the more "optimal" is the search. In the original study this number for the exploration phase was around 0.5 indicating meandering exploration process.
- **Scavenging (Exploitation) optimality:** The median ratio between the minimal number of moves between two consecutive chosen shapes (in the exploitation phase) and the number of moves the player actually took in the game. The closer this number is to 1 the more "optimal" is the search. In the original study this number had a mean of 0.8 and a median of 1 for the exploitation phase, indicating a directed search in the exploitation phase.
- **Optimality:** The ratio between exploration and exploitation optimalities, the higher is the difference the more different the two phases are.
- **Median # steps between shapes in exploration:** The median number of steps (moves) between chosen shapes in exploration.
- **Median # steps between shapes in exploitation:** The median number of steps (moves) between the first and last chosen shapes in each cluster (entire exploitation phase).

- **Speed at exploration:** The mean ratio between the number of moves divided by the time of each exploration phase.
- **Speed at scavenging (exploitation):** The mean ratio between the number of moves divided by the time of each exploitation phase.
- U**:** The mean uniqueness score over all chosen shapes of the player. The Uniqueness score for each shape is defined as the minus log of the frequency of the shape being created by all players in this experiment's dataset (Uniqueness $= -\log[\text{frequency}]$).
- U**(united):** The mean uniqueness score over all chosen shapes of the player. The Uniqueness score for each shape is defined as the minus log of the frequency of the shape being created by all players over all experiments (~ 1200 players) (Uniqueness $= -\log[\text{frequency}]$).
- U **exp:** The mean uniqueness score over all chosen shapes in the exploration phase.
- U **exp (united):** The mean uniqueness score over all chosen shapes in the exploration phase, shape's frequency is assessed over all datasets.
- U **scav:** The mean uniqueness score over all chosen shapes in the exploitation phase, shape's frequency is assessed over this experiment's dataset.
- U **scav (united):** The mean uniqueness score over all chosen shapes in the exploitation phase, shape's frequency is assessed over all datasets.
- U **trans:** The mean uniqueness score over all transition shapes (the first chosen shape in an exploitation phase that signifies the transition between exploration and exploitation), shape's frequency is assessed over this experiment's dataset.
- U **trans (united):** The mean uniqueness score over all transition shapes (the first chosen shape in an exploitation phase that signifies the transition between exploration and exploitation), shape's frequency is assessed over all datasets.
- **Unique Shapes:** The number of shapes only the player discovered.
- **Trans rated:** The number of rated shapes that are transitional (transition the search from exploration to exploitation).
- **# clusters in GC:** The number of clusters (created in exploitation bouts) that are in the giant component (for that specific player). How we create the giant component: We take all clusters created by all players (this experiment's dataset) and connect any two clusters if they share 2 shapes. This forms a network of clusters. We then use the Girvan-Newman algorithm to find modules in the network. The giant component of the network is all the connected modules (and there are other modules that do not connect to the giant component). In the original study we had 13 modules that belonged to the giant component.
- **% clusters in GC:** The percent of clusters that are in the giant component. This measure scales out the total number of clusters the player created (i.e., the number of exploitation bouts they had). This is a measure of "out-of-the-box" thinking, since the more clusters you have in the GC the more "common" your clusters are and they share a thematic thread with other players.