

Cognitive Analysis of Recursive Strategies in “Odd Man Out”

Weston Gaylord (wag@stanford.edu)

Stanford University, 531 Lasuen Mall, PO Box 15390
Stanford, CA 94309 USA

Abstract

I explore the effectiveness of different recursive strategies of a game called “Odd Man Out”: Rock Paper Scissors with two options and three players. Using a recursive cognitive model based on Schelling Coordination games, I compare Static-Prior, Dynamic-Prior and 2-Back Models. I conclude that the 2-Back Model with recursive depth 2, in which players make assumptions about opponent’s behavior based on their last two moves, is the best predictor of empirical data.

Keywords: RPS; Cognitive Modeling; Recursion; Schelling Coordination; K-Level Game; Iterative Strategy

Introduction

Social behavior is based in part on *theory of mind*, the assumption that others have the same subjective consciousness--thoughts, desires, beliefs, etc.--that we do. Ascribing mental states to others is the basis of empathy, and is an important factor in the evolution of social groups. It can also be used to strategize in games and other logical tasks.

For example, in the popular game “Rock Paper Scissors,” each player chooses a move based on guesses about what the other player will choose. Such a game can be described as a “multi-agent learning problem,” in which players optimize their strategies based on their mental models of each other’s strategies.

But this type of thinking can recurse indefinitely: after all, my strategies depend on my mental model of you, which depends on your mental model of me, and so forth. In reality, humans can only strategize to a certain recursive depth before they must simply make a decision. This paper examines the change over time in recursive theory-of-mind strategies in a multi-agent game similar to Rock Paper Scissors.

Background

When I was a child, I invented a simple psychological game called “Odd Man Out” that my friends and I played whenever we swam in a lake or a pool. The game went like this: three players stand at the edge of the water. Each makes a silent decision to either jump in or stay on the shore; on the count of three, all three players act on their decision. You win if you are the odd man out: *either* the only player who jumped or the only player who stayed.

The game can be seen as a variant on Rock Paper Scissors: there are only two options (jump or stay) but each of three players must make a decision simultaneously based on the decisions of both the other players. It’s easy to see why the game encourages players to use a recursive theory-

of-mind strategy, trying to imagine what the other two players will do and choosing a move accordingly.

The research question I investigate is: “How do people develop strategies in ‘Odd Man Out’?” Specifically, what cognitive model best predicts the behavior of human contestants in the game? In this paper, I compare three related cognitive models, which vary in the way players determine priors of whether their opponents are more likely to jump or stay.

In the Static-Prior model, the contestants begin the game with a static belief for their opponents and do not vary it over many rounds. In the Dynamic Prior model, the contestants update their priors each round based on the complete history of their opponents’ moves. In the 2-back model, contestants update their priors each round using only the previous two moves of their opponents.

Related Research

Stuhlmüller et al. describe a class of games called Schelling Coordination games, which use recursive mental models to represent coordination between human agents (Stuhlmüller 2013).

In particular, they describe a game first postulated by Schelling in which two agents, Alice and Bob, want to meet at one of two bars. Each has a prior over which bar they think the other wants to go to, and then use recursive reasoning to make decisions in order to arrive at the same bar as the other. Crucially, these Schelling models use a “level-k” approach (Camerer 2004). If Bob is ambivalent about the two bars, at level 0 he is equally likely to go to either. At level 1 he makes a decision based on his belief about Alice’s prior -- if he knows she likes Bar A, he will be more likely to go there.

At level 2, he might realize that Alice *knows* he’s ambivalent and is therefore somewhat more likely to go to Bar B. The algorithm described in this paper formed the basis for my Church model -- with the difference that the players are trying to be the “odd man out” rather than coordinating with the others.

In order to apply this type of k-level coordination game, De Weerd et al. suggests that the success of a theory-of-mind strategy is dependent on the existence of an optimal response, explaining that “a slight asymmetry, such that one option is preferable over the other, may therefore benefit agents making use of higher-order theory of mind. Such asymmetries may create a focal point for agents with a lower-order theory of mind, which may result in more predictable behavior” (De Weerd 2012).

RPS and “Odd Man Out” are both games *without* such an optimal strategy, unless it is found in the predictable

behavior of one or more of the players. De Weerd et al. show diminishing returns on higher-order theories of mind in Rock-Paper-Scissors. This may be a result of a “flip-flopping” effect at high orders: because of the symmetrical nature of RPS, an n-level strategy might be the same as level $n+3$.

Additionally, humans run out of the cognitive power (and time) to determine such high levels; in my modeling I set a k-level of 2 for each experiment.

Stottinger et al. describe an experiment in which people played RPS against computers that used either a frequency bias (choosing each option with a static frequency) or player-dependent bias (choosing based on the opponent’s previous choice).

The computers then switched either to a modified version of the same bias or to the other type; the researchers found that people were more efficient at updating their strategy for a within-strategy shift; finding it difficult to switch from playing against a frequency bias to playing against a player-dependent bias.

This research sheds light on the flexibility of agents in a multi-player game to change their strategies based on the changing strategies of the other players (Stottinger 2014).

Modeling Strategy

I chose a level-k recursive modeling strategy based on the Schelling Coordination game (Stuhlmüller 2013). Rather than modeling all three players simultaneously, I model one player and his or her beliefs about the other two players.

The player cognitive model (and the player’s models of each of the opponents) takes in three parameters: a cognitive depth, and the move history of the two opponents. At level 0, the player simply chooses to jump or stay based on his or her prior.

At level 1, the player models each opponent at level 0 -- that is, each opponent chooses a move based on *their* prior. Armed with a guess about each opponent, the player then makes a decision: if he/she guesses that both opponents will make the same move, he/she will chose the opposite; otherwise, he/she chooses based on his/her own prior.

At level 2, the player models each opponent at level 1 -- that is, each opponent models the other two players at level 0 and then makes a decision. As before, the player then makes a decision based on the guesses about his/her opponents’ moves.

This leads to a recursive strategy that is similar to the human thought process when playing “Odd Man Out.” As mentioned above, I chose a depth of $k=2$ for my experiments, representative of the assumption that players primarily act based on guesses of what their opponents believe they will do, but not based on guesses of what their opponents believe they believe their opponents will do.

The model uses an enumeration-query on the player’s choice, conditioned on the fact that the player’s choice is different than the choices of both opponents -- and is therefore the odd man out. This is similar to the Schelling Coordination game, which uses a rejection query to query

on the actors’ choices with the condition that their choice matches that of the other actor. I use an enumeration query to work with the Data Analysis Model, which uses Bayesian inference.

The basic pseudo-code for this recursive model can be found below:

```
(define (player depth opponent1-history opponent2-history)

  if depth is 0:
    choose a move based on your own prior

  else:

    choose a prior for opponent 1
    make a depth-1 guess for opponent 1’s move
    choose a prior for opponent 2
    make a depth-1 guess for opponent 2’s move

    choose a move based on the two guesses
)
```

The key difference is in how the priors are chosen. In the Static-Prior model, the priors are chosen beforehand and hard-coded into the model; alternatively, they are set as variables and we use a Data Analysis model to run Bayesian Inference on the data to determine what the variables must have been given the empirical data observed.

In this model, the opponents’ history is not used. In the Dynamic-Prior model, the priors are recalculated every round based on the ratio of jumping to staying that each opponent embodied.

In the 2-back model, the priors are recalculated every round based on the previous two moves of the player in question: (jump stay) and (jump jump) increase the likelihood of jumping, whereas (stay jump) and (stay stay) increase the likelihood of staying.

The degree to which these likelihoods increase were set based on intuition; other methods of determining the weights can be found in the discussion section.

Methods

To gather empirical data to test against my model, I tested three groups of three players, taken from the Stanford undergraduate population.

Each participant, rather than standing by the side of a pool, raised a hand to indicate a “jump” and remained still to indicate a “stay.” On the experimenter’s count of three, all three participants acted on their choice; the three moves were recorded and a new round began.

Each group of three contestants participated in 50 consecutive rounds, for a total of 150 rounds across all three groups.

Results

Static-Prior Model

I began by separating the data into two sections: a training set (the actions of two participants in separate experiments) and a testing set (the rest of the data). I proceeded to run

Bayesian inference on the basic model, querying for the values of the jump/stay base rate priors of each of the two opponents given the training data.

The results of this inference are shown below:

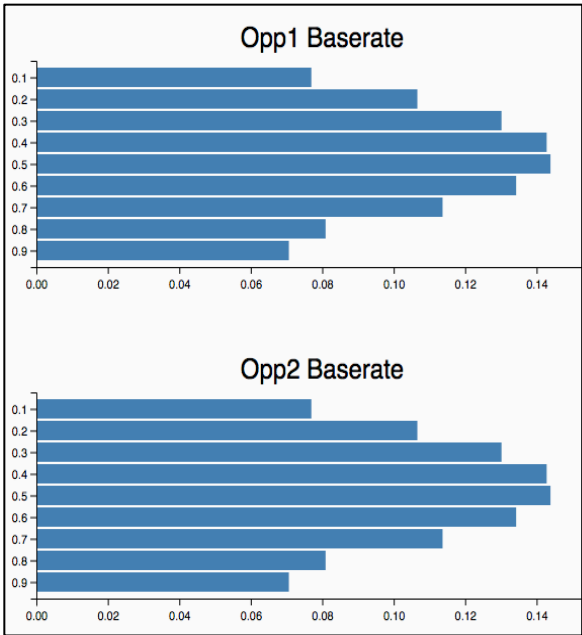


Fig 1. Inferred Baserate Parameters for Training Data

Since 0.4 and 0.5 are nearly equally weighted, I decided to take 0.45 as the inferred parameter for each opponent. This shows that the average player displays the belief that his/her opponents have a very slight preference for staying over jumping. This makes sense because we'd expect people to lazily prefer stillness to moving.

I then took these inferred parameters and hard-coded them into the cognitive model, and sampled data from this trained model. The resulting model predicts that the player would jump 65% of the time and stay 35% of the time.

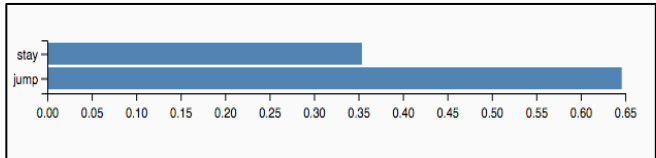


Fig 2: Sampled data from the Static-Prior trained model

Does this match the empirical data? In fact, the testing data showed a percentage of 56% jumping and 44% staying:

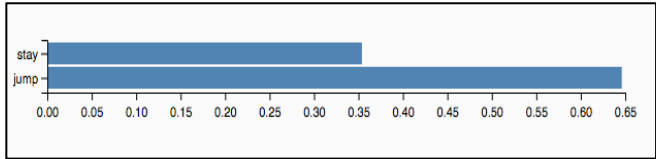


Fig 3: Testing set of the empirical data

Dynamic-Prior Model

For the Dynamic-Prior model, rather than inferring the priors, we update them every round. To see how this model behaves, I sampled the change over time of jump/stay percentages after three, seven and ten rounds of one of the experiments.

For example, see the following table of Opponent 1's first 10 moves.

Table 1: Ten Moves of a Sample Opponent.

ROUND	MOVE
1	stay
2	jump
3	stay
4	stay
5	jump
6	jump
7	jumpj
8	jump
9	stay
10	stay

After three rounds, the model predicts a ratio of 64% jump to 36% stay based on the opponent's 2/3 staying history. After seven rounds, the opponent has jumped 4/7 times, and the strategy shifts back to staying 55%. After ten rounds, the opponent is balanced with 5 jumps and 5 stays, and so the model predicts equal

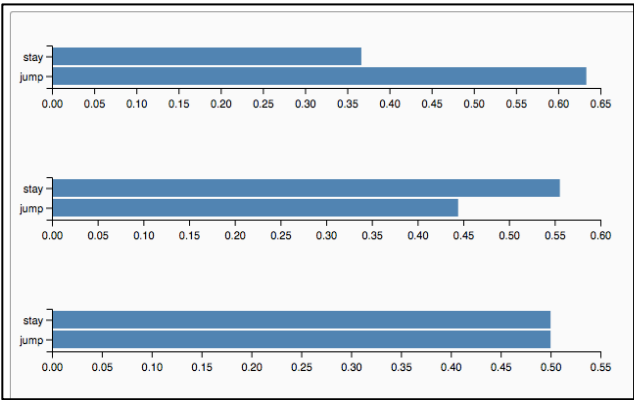


Fig 4: Sampled data from the Dynamic-Prior model

The problem with this model is that after 40-50 rounds, players cannot be expected to make minute adjustments to strategies or remember complex strings of move histories for multiple opponents.

Thus, to predict behavior over many rounds, I decided to try a 2-back model, which uses similar updating based on only the last two moves of each player.

2-Back Model

This model is based on the observation that in the empirical data, each player's moves tended to be either *alternating* (flipping from jump to stay in each consecutive round) or *streaky* (continuing on a run of the same choice for many consecutive rounds). These alternating or streaky strategies usually lasted for 4-5 rounds.

My intuition was that a 2-back method could determine a prior based on the last two moves. An opponent who had jumped the last two moves was 60% to jump again; an opponent who had jumped and then stayed was 55% to jump. An opponent who had stayed the last two moves was 60% likely to jump again; an opponent who had stayed and then jumped was 55% likely to stay.

This model ended up being the best predictor of empirical data. Sampling from the 2-back model produces the following jump/stay ratio:

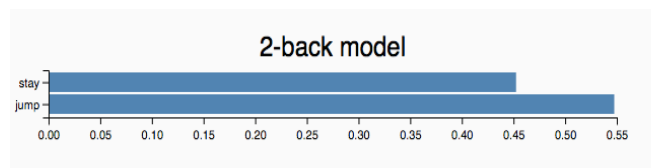


Fig 5: Sampled data from the 2-back model

Discussion

Analyzing the Results

Recall that the empirical results have the following distribution:

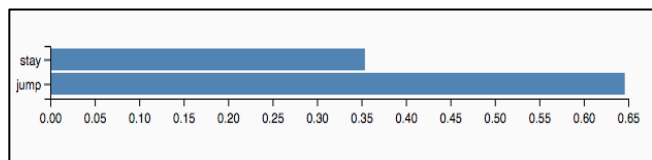


Fig 6: Jump/stay distribution of the empirical data

With a sampling prediction of 55% jump and 45% stay, the 2-back model is both an intuitively good model for how humans play the “Odd Man Out” game (due to its realistic depiction of memory) and the objectively closest predictor of the observed data.

Methodological Concerns

The upside to this method was that it allowed for many more consecutive rounds more quickly than jumping into a pool would afford.

However, there are a number of ways in which the hand-raising variant of the game is different than the pool-

jumping variant. First, the stakes of getting wet may be somewhat higher than the stakes of raising one's hand, so an initial aversion to jumping might exist in the pool variant but not in the hand-raising variant. Similarly, an increased willingness to jump once one is already wet would be present in the pool variant.

Additionally, the satisfaction of winning is somewhat higher in the pool variant, having made one's companions jump while you remain on shore, or being the only one to enjoy the cool water. This higher satisfaction might lead to more competitive strategizing in the pool variant.

Finally, some people are more inherently risk-taking or risk-averse and so might have a more unbalanced prior on whether they would rather jump or stay at the Depth 0 level. This effect is considerably lessened in the hand-raising variant.

However, all of these considerations likely have a small effect on the overall strategy, and the recursive thinking is clearly present in both variants of the game. Therefore, the hand-raising variant was an acceptable replacement.

Further Research

One of the limiting factors of the study was the comparatively small amount of data I was able to collect in person over six weeks. In the future, this research would benefit from an online survey that could automatically combine groups of three over many rounds and obtain thousands of data points.

The other missing piece seems to be inference to determine the priors on different combinations of previous-two-moves in the 2-back model. The 60%/55% parameters were set by me as an estimation of my own strategizing, but with more data I could run Bayesian inference to determine what they should be.

An interesting question to consider in the future could be: what matters more in choosing a strategy and setting priors, one's *own* last move or the opponent's?

It is possible that we are more interested in producing a seemingly random series of moves than we are in reacting to an opponent's series of moves, and therefore would make choices based primarily on our own history. Alternatively, it may be that we are competitively trying to outsmart our opponents and consider their countermoves very heavily.

Additionally, it's possible that we have a longer memory for our *own* past moves than for our opponents' past moves. In this case, rather than a generalized 2-back model, it might be worthwhile to test a model that considers 3 or 4 moves back for ourselves and only 1 or 2 for our opponent.

I'm curious about how rounds where no one wins affect strategy. In empirical observation, all three experiments had streaks of five or six rounds where no one won, leading to increased competitive spirit and frustration. Do we change our strategies during such a dry spell?

It will also be important to run more sophisticated analysis on the data itself, perhaps looking at small sequences of moves rather than the overall ration of jumping to staying. How do strategies change after a win, or a series

of wins, or a series of losses? How long do players typically spend on a micro-strategy such as streaking or alternating, and what kinds of factors might be involved in switching from one to another? Further analysis of the data could bring some of these answers to light.

Finally, I am curious about recursive depth itself: how does recursive depth change over the course of many experiments? I have an intuition that we begin with deep recursion and back off to lower-level thinking (perhaps until prompted to be more competitive again).

It also may be that "thinking time" between tests (as one might have when climbing out of the pool) allows for greater depth in strategizing, as opposed to quick consecutive trials. Further research could bear this out.

Conclusion

After modeling human behavior in the "Odd Man Out" game with Static-Prior, Dynamic-Prior and 2-Back Models, we conclude that the 2-Back Model with k-level 2 is the best predictor of empirical data.

Acknowledgments

Thanks to Noah Goodman, Desmond Ong and Michael-Henry Tessler.

References

- Bowling, M., & M. Veloso (2001). *Rational and Convergent Learning in Stochastic Games*. IJCAI'01 Proceedings of the 17th international joint conference on Artificial intelligence (2): 1021-1026.
- Camerer, C., T. Ho, & J. Chong (2004). *A cognitive hierarchy model of games*. Q J. Econ., 119(3):861-898.
- De Weerd, H., R. Verbrugge, & B. Verheij (2012). *Higher-order social cognition in rock-paper-scissors: A simulation study (extended abstract)*. International Conference on Autonomous Agents and Multiagent Systems: 1195-1196.
- Stottinger, E., A. Filipowicz, J. Danckert, & B. Anderson (2014). *The Effects of Prior Learned Strategies on Updating an Opponent's Strategy in the Rock, Paper, Scissors Game*. Cognitive Science, 38(7): 1482-1492.
- Stuhlmüller, A. & Noah D. Goodman (2013). *Reasoning about Reasoning by Nested Conditioning: Modeling Theory of Mind with Probabilistic Programs*. Cognitive Systems Research.