CS711: Introduction to Game Theory and Mechanism Design **Sep-Nov 2020**

Exploration-Exploitation Behaviour of Humans in Foraging Tasks

Group Members:
Pratham Shukla (170497), Kisha Singh (180357), Prakansh Mishra (180521), Priyanka Talwar (19211268)

Abstract

Foraging is an important evolutionary behavior responsible for the development of many cognitive features unique to humans. However, only a little is known about foraging behavior in humans. Like any other decision-making problem, any agent involved in foraging also has to balance exploration and exploitation. For our study, we use a 2X2 game scenario (2 agents on 2 foraging patches) to investigate the exploration-exploitation strategy of the agents involved. We propose a model where each agent tracks the changes in the patches through a Kalman Filter and relies on random exploration to maximize rewards. The model fits human data well and suggests that humans deploy a highly explorative mixed strategy between the greedy and the non-greedy choice. We also see that increasing exploration beyond a point does not affect the expected cumulative reward much. Therefore humans behave suboptimally not in the sense of procuring less reward but expending more effort to collect the same amount of reward. Our study opens windows to evaluate human behavior in naturalistic games from an algorithmic perspective.

1 Introduction

The fundamental assumptions of game theory: rationality and intelligence are central to most developments in game theory. But the validity of these assumptions in real life is certainly different from their literary definitions [M92]. It is well established that human cognitive resources are bounded, and perfect rationality and intelligence are exceedingly complex for humans to achieve. Humans evolve in a manner that optimizes resource utilization and reward acquisition simultaneously ([SS77], [PY94]). Moreover, unlike the conventional economic definitions of rationality, humans value prosocial behavior, pay heed to affective states and also consider behavioral dynamics to make decisions. Behavioral game theory [C97] is the branch that explores game theory with a humanized perspective.

Decision-making in new and stochastic environments is characterized by a need to balance the desire for information with the desire for reward. This is termed as the exploration-exploitation dilemma. Exploration-exploitation tradeoff is "ubiquitous in nature" and needs to be constantly moderated for optimization [WBCB21]. For example, sitting for a math test where one has limited time to maximize one's grades, a balance has to be struck between solving the known problems to the end or wandering off to unknown problems hoping to land an easier one that can increase the reward rate. Exploration is often characterized by risk-taking [LCPR05] and can be random and directed in nature. Random exploration [T33] is symbolic of variability in behavior when the agent randomizes its actions over the available set of choices. Directed exploration [B56] serves the purpose of seeking information. An agent tends to directly explore options it knows less about. Humans usually use a mix of these strategies adapting them to the needs of the environment and the task at hand.

Foraging behavior carries indispensable importance in the development of the features that make humans, human. Sociocultural components that are necessary and unique to human existence could have well been developed as a consequence of foraging [JE00]. The evolution of bipedalism, language, and socializing has been the key to making a place for humans at the top of the animal kingdom. Examining foraging behavior hence opens a window to understanding the developmental aspects of human cognition.

1.1 Related work

Most research combining game-theory and foraging focuses on animal behavior, largely the predator-prey interaction ([BLG99], [KAKRAV13]). Very little literature exists that builds on human foraging and its evaluation. Our research is one such step, and simply, is a mix-and-match of semantically similar subproblems. We focus ourselves on achieving a well-defined model for human foraging behavior and comparing it to the optimal explore-exploit strategy given the game state transformations and their perception.

As seen in [PP77], according to optimal foraging theory, the rule is, "People move on when the latest intake drops below the average rate". This is known as the Marginal Value Theorem [C76]. Previous studies of foraging often assume this average is fixed and do not take into account how the behavior of competitive agents can alter the decisions of agents in such tasks. But in a recreation of the foraging task in a laboratory [ZGFW15], with degraded input, agents tended to stick to the patches that yielded better previously disregarding exploration. THis hampers the average rate value assumption. Non-explorative strategies also do not guarantee convergence to Nash Equilibria.

In most similar studies, agents are exposed to unknown environments with a set of options each of which changes noisily. This is commonly modeled as the multi-armed bandit problem [KP14], where an agent is supposed to face the challenge of finding the optimal action from a set of actions with unknown expected rewards. Such problems inherently suffer from the exploration-exploitation tradeoff. We have discussed already that agents resort to random and/or directed exploration strategies to tap their environment. Our case is a multi-agent multi-bandit analog of the problem. In [SAJ10], 2 player 2×2 games are used to study explore-exploit in repeated games from an algorithmic perspective, revolving around convergence, and space and time complexity, but no prominent literature on simulating human fraging behavior artificially are available.

Through our project, we go on to establish a novel approach combining algorithmic and behavioral game theoretic aspects.

1.2 Brief Overview of the Report

In Section 2, we formally introduce the game. In Section 3, we explain the experiment that we made our volunteering human agents perform as well as the simulations we did using a reinforcement learning model to predict optimal behaviour in foraging tasks.

2 Foraging Task Game

We set up a foraging task and create an environment in which there are two agents (P, Q), who have been given two choices (A, B) each (two bushes of berries). They do not have prior knowledge of the rewards of the choices (r_A, r_B). They are required to make this choice repeatedly for a set number of trials (N), after each of which they receive a certain reward from their choice on the basis of two rules - if they make different choices, they end up with the complete reward that is available but if they make the same choice, they end up sharing the reward equally among themselves. Their goal is to maximise their own cumulative reward after N trials.

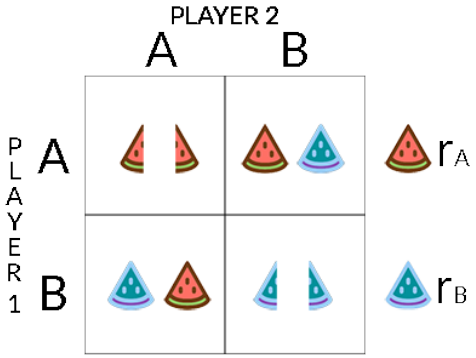
To simulate a realistic game, the rewards that can be obtained are updated after each trial. The rewards are updated by a noisy process. The expected value of the process is updated as follows -

1. If a choice is made by only one agent, the expected value of the reward gets reduced by a factor (f_D), just like the berries available would be lower the next time the bush is chosen.

2. If a choice is made by both the players together, the expected value of its reward gets reduced by a factor ($f_D{}^2$).
3. If a choice is left unused in a trial, the expected value of its reward gets replenished by a factor (f_R).

For our experiment, both with human agents and artificial agents, we use the following values for the parameters defined above.

- $r_A = 20$
- $r_B = 12$
- $N = 30$
- $\sigma = 1$
- $f_D = 0.95$
- $f_R = 1.05$



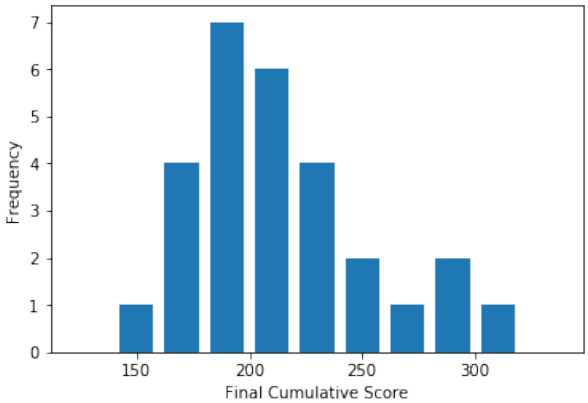
3 Methods

We perform the experiment part of the project in two parts - the experimental human trials and the simulation & data analysis.

3.1 Experimental Human Trials

We had 28 volunteers play our game in pairs. The players were not allowed to communicate with each other mid-game and they did not know what choices the other player was making or the rewards the other player was getting. They were required to figure out in the 30 trials what the game parameters were and how they could maximise their cumulative reward at the end.

The distribution obtained from the cumulative reward data of the human trials has a mean of 212.42 and a standard deviation of 37.94 with the following histogram -

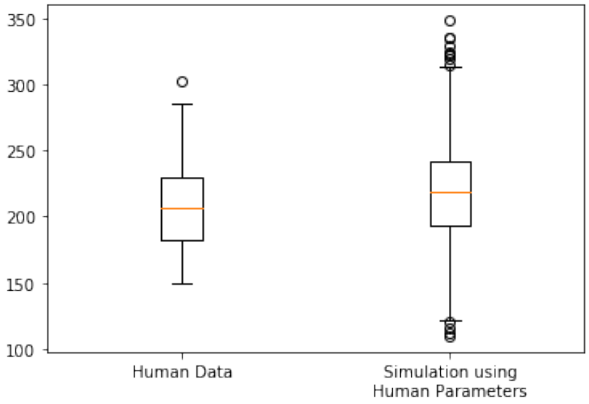


3.2 Simulations & Data Analysis

We investigate the human decision-making process in our game through a modeling process very similar to the restless bandit setting in Speekenbrink and Konstantinidis (2015) and Daw et al. (2006). Both of these papers implement Kalman Filters to mimic human learning dynamics in Gaussian-noisy bandit settings. These filters estimate the joint distribution of the variables under observation for every timestep of incoming measurements. Kalman Filters are preferred over simpler learning models like Rescorla-Wagner (Siegel & Allan, 1996), because of their ability to modulate the learning rates as a function of the tracked uncertainty in the system. Given the volatile nature of our problem, we use Kalman Filter for the learning process, where every individual estimates the dynamics of every patch using a different filter. For choice, we used the ϵ-greedy model (Wunder et al., 2010). The ϵ parameter is responsible for forcing random exploration in the choice. Otherwise, it relies on the greedy choice, which is the higher estimate of the two values being maintained.

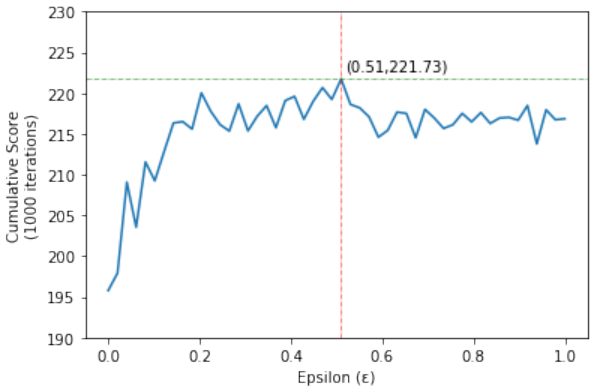
For the first part of the analysis, we use our Kalman Filter, ϵ-greedy model to fit the data obtained from 28 participants who played the game. The fitting was left to be done by the pykalman 0.9.5 package (Duckworth, 2013) for python, which uses the Expectation-Maximization algorithm (Do & Batzoglu, 2008) for the process. In conjunction with the filter fitting, the ϵ parameter was evaluated by a Bayesian update process. The ϵ thus obtained ($\mu = 0.85$, $\sigma = 0.4$) suggests high exploration, which is inevitable given the restless nature of the reward structures. Another interpretation could be a mixed-strategy between the greedy and the non-greedy strategy. The expected probabilities for using the greedy and the non-greedy strategies, hence, are 1 - $(\epsilon/2) = 0.575$ and $(\epsilon/2) = 0.425$ respectively.

To evaluate the fit of the Kalman filter over the human data, the parameters (transition_matrices, transition_covariance, obseravtion_matrices, observation_covarince) retrieved from the data were used to play 1000 instantiations of the game. The cumulative rewards obtained for these simulations produced distribution parameters ($\mu = 218.47$, $\sigma = 6.038$) very similar to the human data ($\mu = 212.42$, $\sigma = 37.94$). This implies that the parameters obtained from the data very well capture the dynamics of human behavior (t = 0.84, p = 0.79). Our model closely emulates human choice, learning, and uncertainty estimation processes.



Given that our model imitates human learning dynamics, and after having established the choice strategies that humans use, we now find the optimal strategy for the game. We use the same filter set to play the game, and perform grid search over the domain of epsilon ($0 < \epsilon < 1$). 1000 simulations were performed over each value to determine the value with the highest expected reward. The model initially benefits from exploration and the cumulative expected rewards climb quickly between $\epsilon = 0$ and $\epsilon = 0.2$. Beyond $\epsilon = 0.2$, the cumulative rewards stay nearly the same, and rather show a very slight decrease as ϵ grows. The decrease was obtained for all the instances of the simulation exercise. Interestingly, even though the values

stay nearly the same, the maxima is obtained only between the interval $0.510 < \epsilon < 0.612$. Hence, there does exist a very weak maximum, but the agent can choose to explore a lot less and maintain a similar payoff.



4 Summary & Discussions

We see that the agents involved in our game are involved in a highly explorative mixed strategy, which lies very close to the optimal cumulative reward but has led them to explore more. This in itself is an interesting behavioral adaptation. We see that our human subjects failed to successfully track the dynamics of the environment and hence could not foresee that increased exploration does not gather higher rewards. Given the parameters of our task, this becomes a benign strategy in terms of the rewards collected, but naturalistic settings are usually not the same. In a realistic scenario, a high amount of exploration often does not come at zero cost. Walking from one patch to another in itself has an energy cost [BTK01] which would increase the demand for food procured from foraging [PD17]. The increased exploration could be a result of a stress response elicited during uncertainty tracking. Each patch in our case is intrinsically a binormal distribution which is being tracked by a single normal distribution. Given that the agent has to inflate its uncertainty estimates, exploration forced by these inflated estimates seems conducive [BRM16]. Our model tracks the binormal distribution as a single one; a more sensitive learning model can assume distributed bandits which shall be updated probabilistically. These probabilistic updates can be decided by an added layer in the model hierarchy [PD20]. Hierarchical Gaussian Filters are picking up popularity in the neuroscience literature, given their ability to track multiple levels of uncertainty [MLDIBFS14].

Our study is one of the few that have tried studying competition in evolutionarily important behaviors using game-theoretic settings (BLG99 and CKBG14). Similar research can build upon uncovering the intricacies of human computations in natural environments. With the ongoing surge in computational capabilities, the resources available now make it more possible than ever for us to understand the computations in the neural circuitry, and their interaction with the environment as a function of their boundedness.

References

[M92] MYERSON RB. On the Value of Game Theory in Social Science. Rationality and Society. 1992;4(1):62-73. doi:10.1177/1043463192004001008

[SS77] Shaw ML, Shaw P. Optimal allocation of cognitive resources to spatial locations. J Exp Psychol Hum Percept Perform. 1977 May;3(2):201-11. doi: 10.1037//0096-1523.3.2.201. PMID: 864393.

[PY94] C.H. Papadimitriou and M. Yannakakis, On complexity as bounded rationality, in: Proceedings Symposium on Theory of Computation (STOC-94) (1994)

[C97] Camerer, C. F. (1997). "Progress in behavioral game theory. Journal of economic perspectives", 11(4), 167-188.

[WBCB21] Robert C Wilson, Elizabeth Bonawitz, Vincent D Costa, R Becket, "Balancing exploration and exploitation with information and randomization, Current Opinion in Behavioral Sciences" Volume 38,Pages 49-56, April 2021. (https://www.sciencedirect.com/science/article/pii/S2352154620301467)

[LCPR05] Law, E. L., Coggan, M., Precup, D., & Ratitch, B. (2005). "Risk-directed exploration in reinforcement learning. Planning and Learning in A Priori Unknown or Dynamic Domains, 97.

[T33] Thompson, W. (1933). "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. Biometrika, 25(3/4), 285-294. doi:10.2307/2332286

[B56] Bellman, R. (1956). "A Problem in the Sequential Design of Experiments". Sankhyā: The Indian Journal of Statistics (1933-1960), 16(3/4), 221-229. Retrieved December 5, 2020, from http://www.jstor.org/stable/25048278

[JE00] Johnson, A. W., & Earle, T. K. (2000). "The evolution of human societies: from foraging group to agrarian state". Stanford University Press.

[BLG99] Brown, J., Laundré, J., & Gurung, M. "The Ecology of Fear: Optimal Foraging, Game Theory, and Trophic Interactions". Journal of Mammalogy, 80(2), 385-399, 1999

[KAKRAV13] Katz MW, Abramsky Z, Kotler BP, Rosenzweig ML, Alteshtein O, Vasserman G. "Optimal foraging of little egrets and their prey in a foraging game in a patchy environment". Am Nat. 2013 Mar;181(3):381-95. doi: 10.1086/669156. Epub 2013 Jan 29. PMID: 23448887.

[PP77] Graham H. Pyke, H. Ronald Pulliam, "Optimal Foraging: A Selective Review of Theory and Tests",The Quarterly Review of Biology 52(2), 1977.

[C76] Eric L. Charnov, "Optimal Foraging, the Marginal Value Theorem", Theoretical Population Biology 9, 129-136, 1976 https://doi.org/10.1016/0040-5809(76)90040-X

[KP14] Kuleshov, V., & Precup, D. (2014). "Algorithms for multi-armed bandit problems". arXiv preprint arXiv:1402.6028.

[ZGFW15] Jinxia Zhang, Xue Gong, Daryl Fougnie, Jeremy M. Wolfe, "Using the past to anticipate the future in human foraging behavior". Vision Research, Volume 111, Part A, 2015. https://doi.org/10.1016/j.visres.2015.04.003 (http://www.sciencedirect.com/science/article/pii/S0042698915001236)

[SAJ10] Adam M. Sykulski, Niall M. Adams, Nicholas R. Jennings, "Exploitation by Exploration: 2-player Repeated 2×2 Games with Unknown Rewards", 2010 (https://www.researchgate.net/publication/45086291_Exploitation_by_Exploration_2-player_Repeated_22_Games_with_Unknown_Rewards)

[DOD06] Daw, N., O'Doherty, J., Dayan, P. et al. "Cortical substrates for exploratory decisions in humans". Nature 441, 876–879, 2006. (https://doi.org/10.1038/nature04766)

[SB98] Sutton, R. S. & Barto, A. G. "Reinforcement Learning: An Introduction" MIT Press, Cambridge, Massachusetts, 1998.

[CB98] C. Claus and C. Boutilier, "The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems". Proceedings of the National Conference on Artificial Intelligence, pages 746–752, 1998.

[W13] Jeremy M Wolfe, "When is it time to move to the next raspberry bush? Foraging rules in human visual search", Journal of Vision, 2013.

[BTK01] Luis M. Bautista, Joost Tinbergen, Alejandro Kacelnik. "Proceedings of the National Academy of Sciences" Jan 2001, 98 (3) 1089-1094; DOI: 10.1073/pnas.98.3.1089

[PD17] Anthony J Porcelli, Mauricio R Delgado, "Stress and decision making: effects on valuation, learning, and risk-taking", Current Opinion in Behavioral Sciences Volume 14, , Pages 33-39, April 2017, doi: 10.1016/j.cobeha.2016.11.015. PMID: 28044144; PMCID: PMC5201132.

[BRM16] de Berker, A., Rutledge, R., Mathys, C. et al. "Computations of uncertainty mediate acute stress responses in humans". Nat Commun 7, 10996, (2016).

[PD20] Piray P, Daw ND. "A simple model for learning in volatile environments". PLoS Comput Biol. 2020 Jul 1;16(7):e1007963. doi: 10.1371/journal.pcbi.1007963. PMID: 32609755; PMCID: PMC7329063.

[MLDIBFS14] Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ and Stephan KE "Uncertainty in perception and the Hierarchical Gaussian Filter". Front. Hum. Neurosci., 2014 (`https://doi.org/10.3389/fnhum.2014.00825`)

[CKBG14] Cressman R, Křivan V, Brown JS, Garay J. Game-theoretic methods for functional response and optimal foraging behavior. PLoS One. 2014 Feb 28;9(2):e88773. doi: 10.1371/journal.pone.0088773. PMID: 24586390; PMCID: PMC3938838.

GitHub Repository - `https://github.com/Pratham-04/CS711A-Fall-2020/tree/main/Project`