The Effects of Uncertainty on Behavioral Strategic-Policy Adaptation and the Neural Representations of Human Strategic Decision-making in Sequential Social Dilemmas

Sandy Tanwisuth

August 2020

1 Abstract

Successful navigation in the social world requires an agent to respond to social and environmental uncertainty adaptively. Though neural correlations of reinforcement learning (RL) were found in social strategic learning, the experimental set-ups are often in matrix games that do not capture the complexity of real-world social situations. Several questions remain unanswered: how do social uncertainty, environmental uncertainty, and their interaction affect the rate of change of adaptation in strategic-policy, how do associated brain networks relay when facing real-world social situations, and how does the brain reduce state-space information in the social world to compute a strategically-apt actionable value? We propose to study the behavioral and neural computational mechanisms of value-based strategic decision-making via online-behavioral and fMRI experiments. In these experiments, human participants will face *Sequential Social Dilemmas (SSD)* – multi-agent game-theoretical environments – against different types of pre-trained opponents. We hypothesize that the rate of strategic-policy-adaptation changes as a function of social-, environmental-reliability, and their interaction where higher social- and environmental-volatility result in a higher frequency of strategic-policy-adaptation. We assess variations of social- and environmental-uncertainty in online behavioral *SSD* tasks including *Cleanup*, and *Harvest*: a *public-goods* game, and a *tragedy-of-the-commons* game, respectively. Neurally, we expect the brain to track the reliability of the predictions within each strategic-policy in vlPFC, TPJ, and RCC. We also expect neural correlates of state-value in vmPFC, action-value in premotor regions, and mentalizing in mPFC, pSTS, and TPJ. We test the brain correlates against the family of models in *Partially Observable Markov Decision Process (POMDP)*. We also investigate how the brain performs a high-dimensionality reduction of the complex social world into key principal behavioral components such as state-value and action-value of self and others in regards to the observable environment through representation learning. We hope insights from these studies can elucidate neuropsychiatric social-function deficits, and inspired research in value alignment.

Keywords: reinforcement learning; social decision making; functional fMRI; intertemporal social dilemmas; partially observable Markov Decision Process

2 Specific Aims

Successful navigation requires an agent to respond to social and environmental uncertainty adaptively both in *Multi-agent Reinforcement Learning (MARL)* and human social interactions[19][14][16][26]. Several studies found the neural correlations of *Reinforcement Learning (RL)* algorithms in social strategic learning[10][33][13], where the experimental set-ups are often in matrix games: set-ups of payoff structure where two-players have opposing interests[30]. However, social situations are non-solipsistic, spatially-, and temporally-extended in the real world. Thereby, the neural dynamics in which the human brain represents the state of the social environment and makes strategic decisions to generate strategically-apt behavioral policies remain understudied. This proposal aims to investigate the unanswered questions: how do social uncertainty, environmental uncertainty, and their interaction affect the rate of change of behavioral adaptation in strategic-policy, how do associated brain networks relay when facing the complexity of the social world, and how does the brain reduce state-space information in the social world to compute a strategically-apt actionable value? Insights into these behavioral and neural computational mechanisms are crucial to understanding factors and interventions affecting behavioral and neuropsychiatric disorders related to social function deficits. The knowledge gained from these studies can also inform research in value-alignment[1][24] at the behavioral algorithmic levels and neural-systems representational levels.

This proposal leverages recent developments in game-theoretical *MARL*, social cognitive neuroscience, and biologically-inspired artificial intelligence to begin to bridge the fields using quantitative approaches. We hypothesize that the rate of strategic-policy-adaptation – switching between cooperative and defective policy – changes as a function of social-, environmental-reliability, and their interaction where higher social- and environmental-volatility result in less reliability and in a higher rate of strategic-policy-adaptation. We begin by assessing human behaviors in response to social and environmental uncertainty in online-tasks against different types of pre-trained agents. After the behavioral experiments, we will test cognitive neuroscience experiments using functional magnetic resonance imaging (fMRI) when human participants face the same tasks. We utilize model-based fMRI[23] and connectivity analysis to answer the question regarding the functionality and connectivity of brain regions associated with social functions. We also use the model-free analysis to assess how the brain performs a high-dimensionality reduction of the complex social world into key principal behavioral components. The details of these experiments are as follows:

Aim 1: Evaluating the influence of social uncertainty, environmental uncertainty, and their interaction on the rate of change of adaptation in strategic-policy. Prior works[26][3][7] led us to hypothesize that higher uncertainty in both social factors i.e. different types of opponents, environment factors i.e. availability of resources, and their interaction result in a higher rate of strategic-policy-adaptation. We administer varieties of environmental-uncertainty via *Sequential Social Dilemmas (SSD)* tasks while we assess the social uncertainty by having a human participant play the games against different types of opponents such as always-cooperated, always-cheated, random, and stable pre-trained RL agents. *SSD* is a series of social dilemmas, situations in which a conflict between personal and collective benefits is introduced[25][6][20]. Two categories of social dilemmas[17] are a *public-goods game (PGG)*, and a *tragedy-of-the-commons game (TCG)*. *PGG* is a game in which an individual is required to pay a cost to obtain shared resources[9], and *TCG* is a game in which there is an incentive to diminish the shared resources for personal gain[12]. We use *Cleanup*[14][16] and *Harvest*[19][14][16], which are *PGG* and *TCG* respectively, to evaluate the effects of environmental uncertainty in different incentive structures.

Aim 2: Understanding social function associated brain networks and their relationships. The brain regions of ventrolateral prefrontal cortex (vlPFC), temporoparietal junction(TPJ), and rostral cingulate cortex (RCC) have been shown to correlate with behavioral-policy control emulating others' goals in social learning[3]. Moreover, prior works[22][4][18][29] led us to expect that in social situations, the brain keeps track of the reliability of the predictions within each strategic-policy in the brain areas mentioned. Simultaneously, the neural correlations of state-value and action-value can be found in the ventromedial prefrontal cortex (vmPFC) in premotor regions[5]. We test brain correlates against the family of models in *Partially Observable Markov Decision Process (POMDP)* since this family of models has shown successful attempts in explaining social learning[26][2][8][31][11].

Aim 3: Investigating how the brain represents environment, other adaptive players, and the information integration mechanisms that produce strategically-apt action-value. Recent studies have shown similarities between biological and artificial neural networks, which can disentangle observational information through nonlinear-transformation into principal behaviorally-relevant components[5][32][15]. These findings led us to hypothesize that the neural computational framework of social decision making can be represented in similar manners. Perhaps, the brain needs to represent and performs a high-dimensionality reduction of the complex social world into key principal behaviorally-relevant components such as state-value and action-value of self and others in regards to the observable environment. Following successful accounts in finding signals of representation learning in the brains[21][28], we plan to utilize *Contrastive Unsupervised Representations for Reinforcement Learning (CURL)*[27] for mapping neural correlates of high-dimensionality reduction in social settings.

References

[1] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. Concrete Problems in AI Safety. *arXiv:1606.06565 [cs]*, July 2016. arXiv: 1606.06565.

[2] C. L. Baker, R. R. Saxe, and J. B. Tenenbaum. Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. page 6.

[3] C. J. Charpentier, K. Iigaya, and J. P. O'Doherty. A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. *Neuron*, 106(4):687–699.e7, May 2020.

[4] C. J. Charpentier and J. P. O'Doherty. The application of computational models to social neuroscience: promises and pitfalls. *Social Neuroscience*, 13(6):637–647, Nov. 2018. Publisher: Routledge _eprint: https://doi.org/10.1080/17470919.2018.1518834.

[5] L. Cross. Using Deep Reinforcement Learning to Uncover the Decision-Making Mechanisms - L. Cross - 10/25/2019.

[6] R. M. Dawes. Social Dilemmas. *Annual Review of Psychology*, 31(1):169–193, 1980. _eprint: https://doi.org/10.1146/annurev.ps.31.020180.001125.

[7] M. Devaine, G. Hollard, and J. Daunizeau. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLOS Computational Biology*, 10(12):e1003992, Dec. 2014. Publisher: Public Library of Science.

[8] P. Doshi and P. J. Gmytrasiewicz. A Framework for Sequential Planning in Multi-Agent Settings. *Journal of Artificial Intelligence Research*, 24:49–79, July 2005. arXiv: 1109.2135.

[9] E. Fehr and S. Gächter. Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4):980–994, Sept. 2000.

[10] A. N. Hampton, P. Bossaerts, and J. P. O'Doherty. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, 105(18):6741–6746, May 2008. Publisher: National Academy of Sciences Section: Biological Sciences.

[11] Y. Han and P. Gmytrasiewicz. IPOMDP-Net: A Deep Neural Network for Partially Observable Multi-Agent Planning Using Interactive POMDPs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6062–6069, July 2019. Number: 01.

[12] G. Hardin. The Tragedy of the Commons. *Journal of Natural Resources Policy Research*, 1(3):243–253, July 2009. Publisher: Routledge _eprint: https://doi.org/10.1080/19390450903037302.

[13] C. A. Hill, S. Suzuki, R. Polania, M. Moisa, J. P. O'Doherty, and C. C. Ruff. A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience*, 20(8):1142–1149, Aug. 2017.

[14] E. Hughes, J. Z. Leibo, M. Phillips, K. Tuyls, E. Dueñez-Guzman, A. G. Castañeda, I. Dunning, T. Zhu, K. McKee, R. Koster, H. Roff, and T. Graepel. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pages 3330–3340, Red Hook, NY, USA, Dec. 2018. Curran Associates Inc.

[15] K. Iigaya, S. Yi, I. Wahle, K. Tanwisuth, and J. P. O'Doherty. Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain | bioRxiv.

[16] N. Jaques, A. Lazaridou, E. Hughes, C. Gulcehre, P. Ortega, D. Strouse, J. Z. Leibo, and N. D. Freitas. Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. In *International Conference on Machine Learning*, pages 3040–3049, May 2019. ISSN: 2640-3498 Section: Machine Learning.

[17] P. Kollock. Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology*, 24(1):183–214, 1998. _eprint: https://doi.org/10.1146/annurev.soc.24.1.183.

[18] S. W. Lee, S. Shimojo, and J. P. O'Doherty. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3):687–699, Feb. 2014.

[19] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '17, pages 464–473, Richland, SC, May 2017. International Foundation for Autonomous Agents and Multiagent Systems.

[20] M. W. Macy and A. Flache. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7229–7236, May 2002. Publisher: National Academy of Sciences Section: Colloquium Paper.

[21] Y. Niv. Learning task-state representations. *Nature Neuroscience*, 22(10):1544–1553, Oct. 2019. Number: 10 Publisher: Nature Publishing Group.

[22] J. O'Doherty, S. Lee, R. Tadayonnejad, J. Cockburn, K. Iigaya, and C. J. Charpentier. Why and how the brain weights contributions from a mixture of experts. preprint, PsyArXiv, June 2020.

[23] J. P. O'Doherty, A. Hampton, and H. Kim. Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, 1104:35–53, May 2007.

[24] M. Peterson. The value alignment problem: a geometric approach. *Ethics and Information Technology*, 21(1):19–28, Mar. 2019.

[25] A. Rapoport. Prisoner's Dilemma — Recollections and Observations. In A. Rapoport, editor, *Game Theory as a Theory of a Conflict Resolution*, Theory and Decision Library, pages 17–34. Springer Netherlands, Dordrecht, 1974.

[26] T. Rusch, S. Steixner-Kumar, P. Doshi, M. Spezio, and J. Gläscher. Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia*, 146:107488, Sept. 2020.

[27] A. Srinivas, M. Laskin, and P. Abbeel. CURL: Contrastive Unsupervised Representations for Reinforcement Learning. *ArXiv*, 2020.

[28] K. L. Stachenfeld, M. M. Botvinick, and S. J. Gershman. The hippocampus as a predictive map. *Nature Neuroscience*, 20(11):1643–1653, Nov. 2017. Number: 11 Publisher: Nature Publishing Group.

[29] S. Suzuki and J. P. O'Doherty. Breaking human social decision making into multiple components and then putting them together again. *Cortex*, 127:221–230, June 2020.

[30] J. von Neumann, O. Morgenstern, and A. Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press, 1944.

[31] M. P. Woodward and R. J. Wood. Learning from Humans as an I-POMDP. page 5.

[32] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, June 2014.

[33] L. Zhu, K. E. Mathewson, and M. Hsu. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences*, Jan. 2012. Publisher: National Academy of Sciences Section: Social Sciences.