Comparing Human Behavior to an Optimal Policy for Innovation

Bonan Zhao[1], **Natalia Vélez**[2], **Thomas L. Griffiths**[1,2]
[1]Department of Computer Science, Princeton University
[2]Department of Psychology, Princeton University
{bnz, nvelez, tomg}@princeton.edu

Introduction

AI models have achieved great successes when the learning target is well-defined and training data are abundant, but human intelligence faces a fundamentally different challenge: human learning does not stop at solving a single problem; instead, we seek new challenges, define new goals, and come up with new ideas (Chu and Schulz 2020). Pursuing these different goals results in the body of knowledge that gets transmitted, and potentially leads to the distinctive human ability of accumulating beneficial cultural traits (e.g., knowledge, skills, and tools) over time (Tennie, Call, and Tomasello 2009). Unlike the classic explore-exploit trade-off between known and unknown options, making new tools or generating new ideas is not about collecting data from existing unknown options, but rather about *create* new options out of what is currently available. Sometimes this involves discovering novel ways of representing old information (Kuhn 1970), as in how heliocentric models superseded the geocentric ones. Such ability to choose new learning targets and set new goals is key to explaining and understanding our current conceptual constructs (Bramley et al. 2023).

The Discovery Game

We introduce a discovery game designed to study how people make decisions about pursuing innovations, viewing discovering new ideas as a process of combining existing ideas. Sometimes the combination itself becomes a stand-alone idea, potentially more powerful and rewarding than its subparts (Basalla 1988; Youn et al. 2015). In a discovery game, players can collect rewards from the available items, and they may discover novel items by combining existing items (similar to ascending the tech tree in a crafting game). Since not all combination leads to successful discoveries, players need to make decisions between gathering rewards from what they have, or attempting to create new items.

We formalize this decision problem as a Markov Decision Process and present analytical solutions for the optimal policy in finite horizon discovery games. In particular, we examine two key factors that drive innovation-seeking behaviors: the success rate of discovery and how much more rewarding a discovery is (the incentive for discovery). We show that both higher success rate and greater incentive encourage innovation-seeking and that the optimal level of innovation-seeking is independent from how many opportunities there are in total.

We report an online behavioral experiment ($n = 210$) that tested these predictions. In the experiment, we manipulated the success rate and incentive to be high or low. While we find that the majority of people's decisions align with the theoretical predictions, there are interesting phenomena such that people seem to assign unequal weights to success rates and incentives.r We also analyze the rich body of strategies people use in different conditions, collecting insights about when people go further, or stop, seeking to enrich their currently available toolkits.

Implications

Our task offers a rich space in which to experiment with various assumptions. For example, instead of using constant parameters, one may manipulate the success rate and incentives to grow, decrease, or randomly change over time. Recent advances in generative AI have shown that self-goal generation may be achieved with the help of semantic domain-specific knowledge (Wang et al. 2023). We could enrich the feature space of this simple discovery game to reflect such intuitions, and study how people grow domain-specific expectations of whether pursuing innovation under certain directions is worthwhile. When multiple domains are at play, we also expect to observe an intellectual division-of-labour phenomenon, where the increasing returns of being an expert in a particular domain could lead to garden-pathing effects in technology development (Arthur 1994). These dynamics may contribute to design artificial learning systems that benefit from parallel, distributed computation in human-like ways, and therefore discover human-like knowledge that can be better understood and used by people.

Acknowledgments

This work was supported by a grant from the Templeton World Charity Foundation (TWCF 20648) to TLG and NV

References

Arthur, W. B. 1994. *Increasing returns and path dependence in the economy*. University of michigan Press.

Basalla, G. 1988. *The evolution of technology*. Cambridge University Press.

Bramley, N. R.; Zhao, B.; Quillien, T.; and Lucas, C. G. 2023. Local search and the evolution of world models. *Topics in Cognitive Science*.

Chu, J.; and Schulz, L. E. 2020. Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, 2: 317–343.

Kuhn, T. S. 1970. *The structure of scientific revolutions*, volume 111. Chicago University of Chicago Press.

Tennie, C.; Call, J.; and Tomasello, M. 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528): 2405–2415.

Wang, G.; Xie, Y.; Jiang, Y.; Mandlekar, A.; Xiao, C.; Zhu, Y.; Fan, L.; and Anandkumar, A. 2023. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Youn, H.; Strumsky, D.; Bettencourt, L. M.; and Lobo, J. 2015. Invention as a combinatorial process: evidence from US patents. *Journal of the Royal Society interface*, 12(106): 20150272.