Meta-reasoning: Deciding which game to play, which problem to solve, and when to quit

Lionel Wong,[1] Tracey Mills,[2] Ionatan Kuperwajs,[3] Katherine Collins,[4] and Thomas Griffiths[3]
[1]Stanford University, [2]Massachusetts Institute of Technology, [3]Princeton University, [4]University of Cambridge

Overview and motivation

People are general purpose problem solvers. We obtain food and shelter, manage companies, solve moral dilemmas, spend years toiling away at thorny math problems, and even adopt arbitrary problems through puzzles and games. The cognitive flexibility which allows us to represent and reason about such a wide range of problems, often referenced as a distinguishing feature of human intelligence (Tomasello, 2022), presents us with an especially ubiquitous one: *deciding which problem to solve.* The meta-level problem of *what problem to choose* exists, in part, because people have limited problem solving resources (Griffiths, 2020). While this challenge has been examined through various lenses across cognitive science, implicit in many of these perspectives is the notion of bounded rationality. Given our limited time and energy, how do we decide which problems are worthwhile and when we should quit to pursue something new?

Resource-rational analysis has thus been especially fruitful in establishing normative and computational frameworks for understanding fundamental aspects of problem selection (Lieder & Griffiths, 2020; Gershman, Horvitz, & Tenenbaum, 2015; Icard, 2023). In particular, cognitive scientists have found that the way people define or construe problems (Ho et al., 2022), select problem-solving approaches and strategies (Lieder & Griffiths, 2017; Binz, Gershman, Schulz, & Endres, 2022), and decompose goals into subproblems (Binder, Mattar, Kirsh, & Fan, 2023) can be understood as attempts to maximize expected rewards and minimize costs based on the broader reward structure of the environment as well as their own cognitive limitations. This broad framework has also been productive in explaining how people solve a different but related problem: that of deciding how much time to spend on a particular problem or subproblem before focusing on something new (Vul, Goodman, Griffiths, & Tenenbaum, 2014; Callaway, Rangel, & Griffiths, 2021; Callaway et al., 2022; Kuperwajs, Ho, & Ma, 2024).

At the same time, it is not straightforward to translate these approaches – based on notions of rationality as expected reward maximization – to many of the more naturalistic versions of these problems that people continually face. The reward structure of human experience is opaque and highly complex, and the space of possible problems is infinitely rich and expansive. Indeed, under more descriptive analyses, many problems that people pursue in practice seem to be idiosyncratic or impractical (Chu, Tenenbaum, & Schulz, 2023). This is perhaps especially true early in life, with children's play often revolving around constructing and assigning utility to seemingly arbitrary and costly problems ("the floor is lava!") – a tendency which Chu and Schulz (2023) argue supports the process of generating new ideas. By studying what kinds of problems or games people adopt in settings lacking clear extrinsic reward structure, other work has begun to outline and formalize the abstract qualities of an intrinsically worthwhile problem, such as creativity, interestingness, novelty, and fun (Chu, Hu, & Ullman, 2024; Davidson, Todd, Togelius, Gureckis, & Lake, 2024; Zhang, Collins, Wong, Weller, & Tenenbaum, 2024; Brändle, Wu, & Schulz, 2024).

Much of this work focuses on problem selection as an individual enterprise. The ability to continuously construct and revise what problems we choose to pursue is profoundly personal, having been linked to our sense of agency and self (Paul, Ullman, De Freitas, & Tenenbaum, 2023). However, problem solving and selection is also undoubtedly socially and culturally embedded. Collective problem solving was likely key to our evolutionary history, introducing novel coordination problems between agents with competing goals which shaped modern perspective taking abilities (Tomasello, 2022). In turn, opportunities to collaborate with others and build upon cultural innovation shape the landscape of problems one might pursue (Vélez, Wu, Gershman, & Schulz, 2024; Colas, Karch, Moulin-Frier, & Oudeyer, 2022).

Understanding how people engage in meta-reasoning for problem selection can have far-reaching applications. For instance, a better understanding of how people select, or struggle to select, problems could be used to guide the development of algorithms to support more well-motivated project selection (Heindrich & Lieder, 2024), inform the design of principled curricula in education (Corbett, Koedinger, & Hadley, 2001), and even help scientists understand how to pick impactful research problems (Fischbach, 2024). With this workshop, we hope to make progress towards this goal by drawing connections between diverse perspectives on problem selection to identify common frameworks and concepts as well as contrasting assumptions.

Approach and workshop structure

We will offer a forum with leading researchers in computational cognitive science, psychology, and philosophy to engage the broader cognitive science community with an interdisciplinary discussion on problem selection grounded in recent empirical advances. Our speakers span a range of career stages. The program will include several sessions, each beginning with an invited talk. During the talk component, speakers will be encouraged to engage with one of the following guiding questions:

- What role should the concept of rationality play in the study of meta-reasoning for problem selection?

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

- What is the relationship between individual and collective problem selection? How and to what extent does social or cultural context shape problem selection?
- How is meta-reasoning for problem selection shaped by learning and development?
- How does our sense of agency or identity influence the problems we consider and choose to solve, especially over long timescales where we may not know how our choices will ultimately change us?

Following each talk, we will moderate an interactive discussion in which both speakers *and audience members* engage with the central question relevant to the session. Our goal is to create a stimulating environment that is principally a workshop - where researchers of all career stages can engage in the questions and early answers at the cutting-edge of research on the topic of meta-reasoning for problem selection. Each session will be approximately 40 minutes, with time equally split between the talk and discussion portions. There will be a coffee break for informal conversation among attendees between the second and third sessions.

Organizers and speakers

Organizers will not be giving full presentations.

Lionel Wong (organizer) is a Postdoctoral Scholar at Stanford, studying how people (and computational models) reason about open-world situations described in language, and integrate information in language with their background knowledge and cognitive capacities from other domains.

Tracey Mills (organizer) is a PhD student in computational cognitive science at MIT. She studies how people approach open-ended inference and reasoning problems.

Ionatan Kuperwajs (organizer) is a Postdoctoral Research Associate at Princeton University. He studies how people make decisions and plan sequences of actions in complex environments.

Katherine Collins (organizer) is a PhD student at the University of Cambridge. Her research centers around applied computational cognitive science and human-AI interaction.

Thomas Griffiths (organizer) is a Professor of Psychology and Computer Science at Princeton University. His group aims to understand the computational and statistical foundations of human inductive inference.

Mark Ho is an Assistant Professor of Psychology at New York University. He works on planning and social interaction between humans and AI systems.

Natalia Vélez is an Assistant Professor of Psychology at Princeton University. She works on group-level innovation and coordination.

Junyi Chu is a Postdoctoral Fellow at Stanford University. She studies children's play and creativity.

Laurie Paul is a Professor of Philosophy and Cognitive Science at Yale University. Her research interests are in metaphysics, decision theory, and philosophy of mind.

Program
Introduction
Session #1: resource rationality Invited talk: Mark Ho Discussion
Session #2: social context Invited talk: Natalia Vélez Discussion
Coffee break and informal discussion
Session #3: learning and development Invited talk: Junyi Chu Discussion
Session #4: agency and identity Invited talk: Laurie Paul Discussion
Closing

References

Binder, F. J., Mattar, M. G., Kirsh, D., & Fan, J. E. (2023). Humans choose visual subgoals to reduce cognitive cost. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*, *129*(5), 1042.

Brändle, F., Wu, C. M., & Schulz, E. (2024). Leveling up fun: Learning progress, achievement, and expectations influence enjoyment in video games. *PsyArXiv*.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Computational Biology*, *17*(3), e1008863.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, *6*(8), 1112–1125.

Chu, J., Hu, J., & Ullman, T. D. (2024). The task task: Creative problem generation in humans and language models. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Chu, J., & Schulz, L. E. (2023). Not playing by the rules: exploratory play, rational action, and efficient search. *Open Mind*, *7*, 294–317.

Chu, J., Tenenbaum, J. B., & Schulz, L. E. (2023). In praise of folly: flexible goals and human cognition. *Trends in Cognitive Sciences*.

Colas, C., Karch, T., Moulin-Frier, C., & Oudeyer, P.-Y. (2022). Language and culture internalization for human-like autotelic ai. *Nature Machine Intelligence*, *4*(12), 1068–1076.

Corbett, A. T., Koedinger, K., & Hadley, W. S. (2001). Cognitive tutors: From the research classroom to all classrooms. In *Technology enhanced learning* (pp. 215–240). Routledge.

Davidson, G., Todd, G., Togelius, J., Gureckis, T. M., & Lake, B. M. (2024). Goals as reward-producing programs. *arXiv preprint arXiv:2405.13242*.

Fischbach, M. A. (2024). Problem choice and decision trees in science and engineering. *Cell*, *187*(8), 1828–1833.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–883.

Heindrich, L., & Lieder, F. (2024). Leveraging automatic strategy discovery to teach people how to select better projects. *arXiv preprint arXiv:2406.04082*.

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature*, *606*(7912), 129–136.

Icard, T. (2023). Resource rationality. *Book manuscript*.

Kuperwajs, I., Ho, M. K., & Ma, W. J. (2024). Heuristics for meta-planning from a normative model of information search. *PsyArXiv*.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1.

Paul, L., Ullman, T., De Freitas, J., & Tenenbaum, J. (2023). Reverse-engineering the self. *PsyArXiv*.

Tomasello, M. (2022). *The evolution of agency: Behavioral organization from lizards to humans*.

Vélez, N., Wu, C. M., Gershman, S. J., & Schulz, E. (2024). The rise and fall of technological development in virtual communities. *PsyArXiv*.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Zhang, C. E., Collins, K. M., Wong, L., Weller, A., & Tenenbaum, J. B. (2024). People use fast, goal-directed simulation to reason about novel games. *arXiv preprint arXiv:2407.14095*.