What Should I Do Now?
Goal-Centric Outlooks on Learning, Exploration, and Communication

Cédric Colas (ccolas@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139, USA

Junyi Chu (junyichu@fas.harvard.edu)
Department of Psychology, Harvard University
Cambridge, MA 02138, USA

Gaia Molinaro (gaiamolinaro@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94704, USA

Robert Hawkins (rdhawkins@wisc.edu)
Department of Psychology, UW-Madison
Madison, WI 53715, USA

Keywords: goals; decision-making; language; reinforcement learning; developmental psychology; cultural evolution

Overview

Goals are a central pillar of everyday mental activity. From finding your way home to solving a puzzle or ordering food delivery, much of human action and cognition is goal-directed. Perhaps unsurprisingly, theories of goals are a central focus in the psychology of motivation (Elliott & Dweck, 1988), social and personality psychology (Fishbach & Ferguson, 2007), as well as research aimed at understanding factors contributing to task achievement in educational and industrial settings (Ames & Ames, 1984; Locke & Latham, 2002).

In the cognitive and computational sciences, however, goals have mostly been taken for granted. In the vast majority of studies, participants are presented with experimenter-defined task objectives, whereas their intrinsic motives are typically viewed as a nuisance variable at most (Karayanni & Nelken, 2022). While some research has considered how information processing may differ as a function of the task assigned to participants (Salverda, Brown, & Tanenhaus, 2011), this approach still neglects a fundamental aspect of human psychology: that human learners and decision-makers are *autotelic* agents who represent, generate, and flexibly pursue their own goals. This realization raises important open questions: What are goals useful for? Why is there so much diversity and complexity in human goals? How do we represent and prioritize goals?

The design and study of artificial agents may help shed some light on the above questions. Autotelic machine learning, for instance, endows agents an intrinsic motivation to pursue their own goals and studies how different representation learning processes or intrinsic utility functions can influence the agent's ability to explore or grow repertoires of skills (Colas, Karch, Sigaud, & Oudeyer, 2022).

In this symposium, we highlight recent work emphasizing a goal-centric outlook on learning, exploration, and communication. We bring together researchers from a range of fields (psychology, neuroscience, artificial intelligence), perspectives (computational, evolutionary, developmental, social, linguistic), and career stages to examine **how and why we represent, select, and pursue our own goals**. This symposium will initiate an interdisciplinary conversation aimed at bridging computational accounts of cognition with theories of motivation.

Gaia Molinaro, a PhD candidate at the University of California, Berkeley, will discuss the importance of addressing goals from a cognitive science perspective. She will emphasize the role of goals in shaping key components of human learning and decision-making, and thus the need for a better understanding of goal selection in its own right.

Junyi Chu, a postdoctoral fellow at Harvard, will present empirical studies examining how children and adults choose actions and goals in playful and non-play contexts. She will use these findings to motivate novel proposals about the development of flexible goal-oriented reasoning, the value of play for cognitive development, and the kinds of intrinsic motivations learners are sensitive to.

Robert Hawkins, an assistant professor at UW-Madison, will present a computational framework for goal-relevant communication among social agents. He will discuss the role of questions, how listeners make context-sensitive inferences about questioner goals, and the tractability of theory-of-mind reasoning over the space of all possible goals.

Cédric Colas, a postdoctoral scholar at MIT, will offer a computational perspective informed by the design of intrinsically motivated artificial agents. He will detail how goal generation can influence state abstractions, increase adaptiveness, and allow cumulative learning of behavioral repertoires in artificial agents.

Gaia Molinaro: Goal-Centric Learning

Psychology, neuroscience, and artificial intelligence research communities have vastly benefited from reinforcement learning as a model of interactions between agents and their environments. Key components of learning highlighted by this framework include states, actions, and rewards. I will argue that goals play a pivotal role in animal learning by affecting state representation, action selection, and reward delivery (Molinaro & Collins, 2023a). To exemplify, I will illustrate how invoking the notion of goals helps formally account for people's rational and irrational choices in decision-making tasks (Molinaro & Collins, 2023b). Finally, I will propose that, due to the crucial role of goals in learning, both experimental cognitive scientists and artificial intelligence researchers would benefit from a more precise understanding of goal selection, management, and pursuit.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Junyi Chu: Play for Problems and Proposals

Play is one of the most recognizable and universal forms of behavior. But what is the value of play? Despite a long-standing scientific and practical interest in play, there is relatively little agreement on what constitutes play and how play supports development. In this talk, I propose that play in humans reflects a novel kind of exploration (Chu & Schulz, 2020; Chu, Tenenbaum, & Schulz, 2024), in which players are trying to figure out what problems they can pose and solve. I will provide a number of empirical studies illustrating how children and adults choose goals and actions when trying to have fun compared with under other objectives. Together, this work suggests that inventing and pursuing novel goals is an intrinsically rewarding activity, and raises new questions about the nature and development of action, motivation, and innovation.

Robert Hawkins: Communicating about Goals

Effective social collaboration depends on the ability to represent others' goals alongside our own. While we can learn a lot about another agent's goals by simply observing their behavior, humans also explicitly communicate about their goals. In this talk, I will discuss two lines of research investigating *questions* as one of the primary vehicles people use to navigate the space of goals through natural language. First, I will discuss a study showing that people spontaneously go beyond the literal information requested in a yes/no question based on considerations of relevance to the inferred questioner goal. Comparing human performance to large language models like GPT-3 suggests that models often fail to provide the relevant information people prefer to provide (Tsvilodub et al., 2023). Second, I will describe a computational framework formalizing goal-relevance as the result of recursive social reasoning about decision problems (Sumers et al., 2023), which provides insight into classic effects of goal-sensitivity in dialogue.

Cédric Colas: Autotelic Machine Learning

Nature cannot evolve biological functions to face every possible environmental situation. Instead, it evolved goal-directed organisms able to represent, prioritize and flexibly pursue goals (Tomasello, 2022). Artificial agents also benefit from goal-directedness: they can execute more complex tasks, learn what to attend to, explore better and grow behavioral repertoires. But where do goals come from? Although artificial agents usually pursue goals provided by their designer, humans seem to be autotelic: they come up with their own! How do we know which goals to pursue? I will review different intrinsic objectives from the recent AI literature and discuss how socio-cultural aspects may influence the representation and selection of goals (Colas et al., 2022). I will finally argue that the joint generation and pursuit of goals drives a developmental process towards more autonomous, flexible and capable agents.

Acknowledgments

Cédric Colas received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101065949. Gaia Molinaro's contribution was supported by the Chateaubriand Fellowship of the Office for Science & Technology of the Embassy of France in the United States. Junyi Chu is grateful for support from the MIT Quest for Intelligence, the Center for Brains, Minds and Machines (NSF STC CCF-1231216), and the Ditmars foundation.

References

Ames, C., & Ames, R. (1984). Goal structures and motivation. *The Elementary School Journal*, *85*(1), 39–52.

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

Chu, J., Tenenbaum, J. B., & Schulz, L. E. (2024). In praise of folly: flexible goals and human cognition. *Trends in Cognitive Sciences*.

Colas, C., Karch, T., Sigaud, O., & Oudeyer, P.-Y. (2022). Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, *74*, 1159–1199.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of personality and social psychology*, *54*(1), 5.

Fishbach, A., & Ferguson, M. J. (2007). The goal construct in social psychology. *Social psychology: Handbook of basic principles*, *2*, 490–515.

Karayanni, M., & Nelken, I. (2022). Extrinsic rewards, intrinsic rewards, and non-optimal behavior. *Journal of Computational Neuroscience*, *50*(2), 139–143.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

Molinaro, G., & Collins, A. G. (2023a). A goal-centric outlook on learning. *Trends in Cognitive Sciences*.

Molinaro, G., & Collins, A. G. (2023b). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLoS Biology*, *21*(7), e3002201.

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta psychologica*, *137*(2), 172–180.

Sumers, T. R., Ho, M. K., Griffiths, T. L., & Hawkins, R. D. (2023). Reconciling truthfulness and relevance as epistemic and decision-theoretic utility. *Psychological Review*.

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

Tsvilodub, P., Franke, M., Hawkins, R. D., & Goodman, N. D. (2023). Overinformative question answering by humans and machines. *arXiv preprint arXiv:2305.07151*.