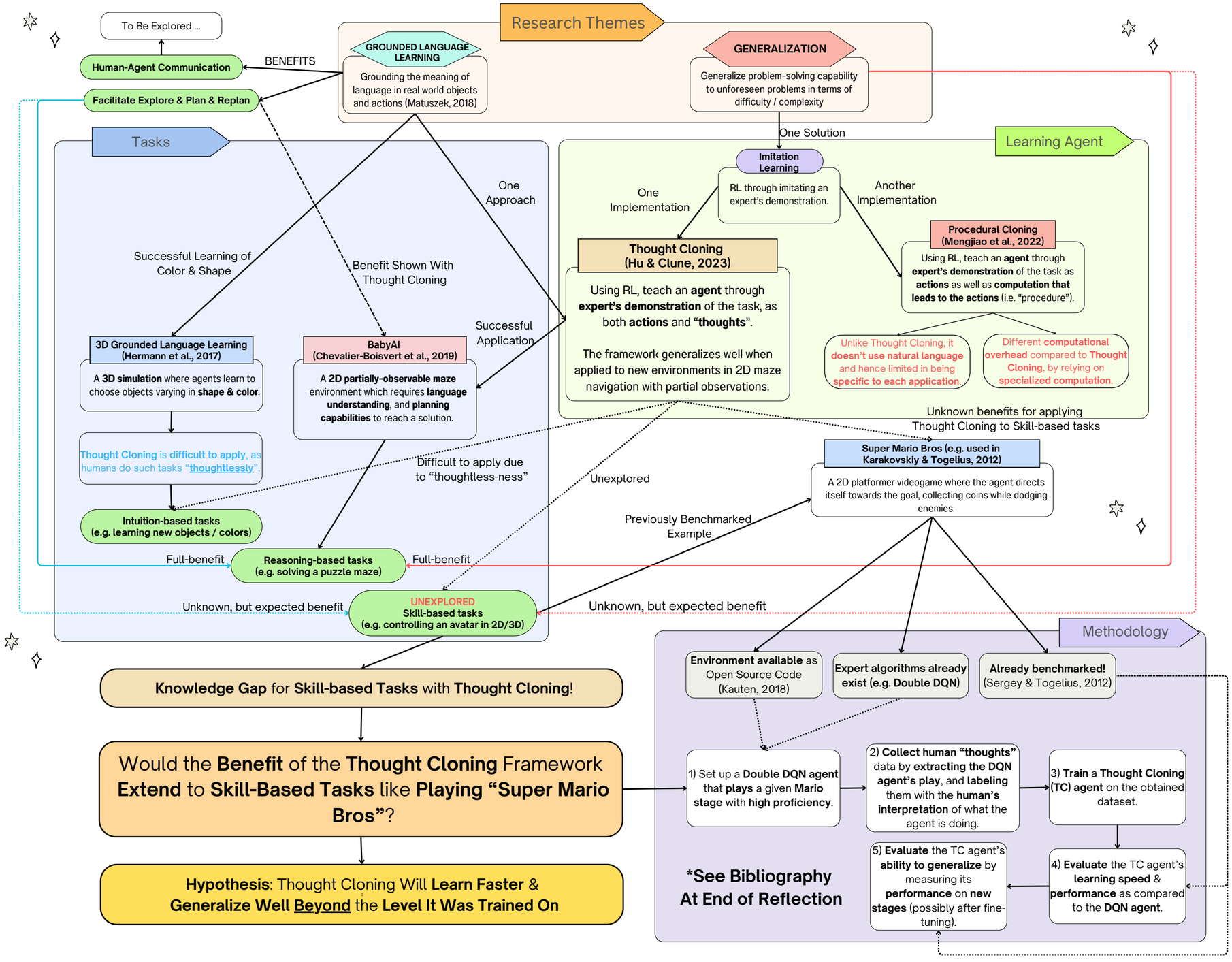
To Be Explored ...
Human-Agent Communication
BENEFITS
Facilitate Explore & Plan & Replan
Research Themes
GROUNDED LANGUAGE LEARNING
Grounding the meaning of language in real world objects and actions (Matuszek, 2018)
GENERALIZATION
Generalize problem-solving capability to unforeseen problems in terms of difficulty / complexity
One Solution
Tasks
Learning Agent
One Approach
Successful Learning of Color & Shape
Benefit Shown With Thought Cloning
Successful Application
Imitation Learning
RL through imitating an expert's demonstration.
One Implementation
Another Implementation
Thought Cloning (Hu & Clune, 2023)
Using RL, teach an agent through expert's demonstration of the task, as both actions and "thoughts".
The framework generalizes well when applied to new environments in 2D maze navigation with partial observations.
Procedural Cloning (Mengjiao et al., 2022)
Using RL, teach an agent through expert's demonstration of the task as actions as well as computation that leads to the actions (i.e. "procedure").
Unlike Thought Cloning, it doesn't use natural language and hence limited in being specific to each application.
Different computational overhead compared to Thought Cloning, by relying on specialized computation.
3D Grounded Language Learning (Hermann et al., 2017)
A 3D simulation where agents learn to choose objects varying in shape & color.
BabyAI (Chevalier-Boisvert et al., 2019)
A 2D partially-observable maze environment which requires language understanding, and planning capabilities to reach a solution.
Thought Cloning is difficult to apply, as humans do such tasks "thoughtlessly".
Difficult to apply due to "thoughtless-ness"
Unknown benefits for applying Thought Cloning to Skill-based tasks
Super Mario Bros (e.g. used in Karakovskiy & Togelius, 2012)
A 2D platformer videogame where the agent directs itself towards the goal, collecting coins while dodging enemies.
Unexplored
Previously Benchmarked Example
Intuition-based tasks (e.g. learning new objects / colors)
Full-benefit
Reasoning-based tasks (e.g. solving a puzzle maze)
Full-benefit
Unknown, but expected benefit
UNEXPLORED Skill-based tasks (e.g. controlling an avatar in 2D/3D)
Unknown, but expected benefit
Knowledge Gap for Skill-based Tasks with Thought Cloning!
Would the Benefit of the Thought Cloning Framework Extend to Skill-Based Tasks like Playing "Super Mario Bros"?
Hypothesis: Thought Cloning Will Learn Faster & Generalize Well Beyond the Level It Was Trained On
Methodology
Environment available as Open Source Code (Kauten, 2018)
Expert algorithms already exist (e.g. Double DQN)
Already benchmarked! (Sergey & Togelius, 2012)
1) Set up a Double DQN agent that plays a given Mario stage with high proficiency.
2) Collect human "thoughts" data by extracting the DQN agent's play, and labeling them with the human's interpretation of what the agent is doing.
3) Train a Thought Cloning (TC) agent on the obtained dataset.
4) Evaluate the TC agent's learning speed & performance as compared to the DQN agent.
5) Evaluate the TC agent's ability to generalize by measuring its performance on new stages (possibly after fine-tuning).
*See Bibliography At End of Reflection

COGS401 Midterm Reflection

Potential Knowledge Gap That We Fill

Thought Cloning (TC) is an Imitation Learning (IL) framework, which extends traditional IL approaches (such as behavioral cloning) by providing natural language “thoughts” to be imitated by the Reinforcement Learning (RL) agent. The authors successfully applied TC to a 2D maze navigation task, known as BabyAI (Chevalier-Boisvert et al., 2019) to show speed up in training. The obtained agent also was able to adapt better to environments that are novel both in terms of difficulty and complexity.

Because it uses natural language prompts, TC can be applied to a variety of tasks where “expert” thoughts can be collected efficiently. I argue that BabyAI is but one example of such tasks, and is especially characterized by a smaller and deterministic action space, relying heavily on planning to reach a solution. Many RL environments found in the literature are in contrast characterized by high-dimensional action spaces, and require precise frame-by-frame control of the agent (e.g. MineRL, an RL environment based on the videogame “Minecraft” by Guss et al., 2019).

It is plausible that such “skill-based” task environments would still benefit from TC: for example, people routinely communicate strategies for such dynamic environments in natural language (e.g. a friend could explain how to best fight a hostile NPC in Minecraft). Furthermore, the true benefit of TC might lie in the agent’s ability to generalize their learning to novel environments.

By applying TC to the well-established and benchmarked (Sergey & Togelius, 2012) “Super Mario Bros” environment (Kauten, 2018), this project could then fill the research gap of whether “Thought Cloning is beneficial to skill-based RL tasks”, evaluating it both in terms of learnings speed gain and generalization to novel environments.

Potential Next Steps

The next steps in this project will be to consolidate my experimental plan.

First, the current taxonomy of reinforcement learning (RL) tasks I have discussed (planning-based, skill-based, intuition-based) largely builds on my experience working with RL. Time did not allow me to fully research a formal taxonomy of reinforcement learning tasks and appropriate terminologies. Completing such research should inform me to reconsider my choice of the best skill-based task to apply Thought Cloning (TC) (Hu & Clune, 2023).

Second, further research would be needed in order to establish a method for collecting “thoughts” data for Super Mario Bros efficiently. Hu & Clune (2023) cleverly avoided this problem by using synthetic data. My current experimental idea relies on the manual labor of the researcher, which could potentially be reduced by using previously established frameworks (e.g. leverage prior RL agents, such as obtained by Sergey & Togelius (2012).

Bibliography

Please note that some papers below are drawn from “arXiv”, a Computer Science pre-print platform that does not require peer-reviewing so as to accelerate the development of machine learning. That said, all papers included here are reputable, having either 1) been accepted to peer-reviewed conferences (such as ICLR 2019), or 2) been published by a reputable organization, such as Google DeepMind.

Chevalier-Boisvert, M., Bahdanau, D., Lahlou, S., Willems, L., Saharia, C., Nguyen, T. H., & Bengio, Y. (2019). BabyAI: A Platform to Study the Sample Efficiency of Grounded Language Learning. *arXiv:1810.08272v4 [cs.AI].* https://doi.org/10.48550/arXiv.1810.08272

Guss, W. H., Houghton, B., Topin, N., Wang, P., Codel, C., Veloso, M., & Salakhutdinov, R. (2019). MineRL: A Large-Scale Dataset of Minecraft Demonstrations. *arXiv:1907.13440 [cs.LG]*. https://doi.org/10.48550/arXiv.1907.13440

Hermann, K. M., Hill, F., Green, S., Wang, F., Faulkner, R., Soyer, H., Szepesvari, D., Czarnecki, W. M., Jaderberg, M., Teplyashin, D., Wainwright, M., Apps, C., Hassabis, D., & Blunsom, P. (2017). Grounded Language Learning in a Simulated 3D World. *arXiv:1706.06551v2 [cs.CL]*. https://doi.org/10.48550/arXiv.1706.06551.

Hu, S., & Clune, J. (2023). Thought Cloning: Learning to Think while Acting by Imitating Human Thinking. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt & S. Levine (Eds.), *Advances in Neural Information Processing Systems* (pp. 44451-44469). Curran Associates, Inc.

Kauten, C. (2018). gym-super-mario-bros: Super Mario Bros for OpenAI Gym. https://github.com/Kautenja/gym-super-mario-bros (visited on 2024/10/15).

Matuszek, C. (2018). Grounded Language Learning: Where Robotics and NLP Meet. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Early Career*, pp 5687-5691. https://doi.org/10.24963/ijcai.2018/810

Mengjiao, Y., Abbeel, P., Schuurmans, D., & Nachum, O. (2022). Chain of Thought Imitation with Procedure Cloning. *arXiv:2205.10816 [cs.LG]*. https://arxiv.org/abs/2205.10816

Sergey, K., & Togelius. J. (2012). The Mario AI Benchmark and Competitions. *IEEE Transactions on Computational Intelligence and AI in Games*, *4*(1). pp55–67. https://doi.org/10.1109/TCIAIG.2012.2188528.

Supplementary Bibliography

“Learning Agent” Category

Torabi, F., Warnell, G., Stone, P. (2018). Behavioral Cloning from Observation. *arXiv:1805.01954v2 [cs.AI]*. https://doi.org/10.48550/arXiv.1805.01954

> Discusses Behavioral Cloning, which informs both Thought Cloning (Hu & Clune, 2023) and Procedural Cloning (Mengjiao et al., 2022).

“Tasks” Category

Baker B., Akkaya, I., Zhokhov, P., Huizinga, J., Tang, J., Ecoffet, A., Houghton, B., Sampedro, R., & Clune, J. (2022). Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos. *arXiv:2206.11795 [cs.LG]*. https://doi.org/10.48550/arXiv.2206.11795

> Discusses an approach for collecting video data from the online website YouTube, utilizing its data to train agents. This has very high synergy with Thought Cloning, especially for tasks where much online data is available.

“Methodology” Category

Lee, G., Luo, M., Zambetta, F., & Li, X. (2014). Learning a Super Mario Controller from Examples of Human Play. *2014 IEEE Congress on Evolutionary Computation.* https://doi.org/10.1109/CEC.2014.6900246

> Discusses an approach to training an AI agent for playing Super Mario, utilizing Inverse Reinforcement Learning as the core mechanism. Successfully trains agents at different levels of proficiency.