Abstract Models of Human-AI Teams

Pat Langley
Christopher J. MacLellan
Georgia Institute of Technology
Atlanta, Georgia

The research presented here was supported in part by Award No. W911NF2320203 from ARL's STRONG Program. Thanks to Erik Harpstead, Danny Wietekamp, Ying Wu, and others for useful discussions.

History of Human-AI Teaming

Research on collaboration between humans and AI systems has a long and successful history.

- Examples have ranged from simple recommender systems to sophisticated conversational agents.
- Recent computational advances have increased interest in the potential of human-AI teams.
- However, we have also seen cases in which AI technology has led to negative outcomes (e.g., Swartz, 2003).

The field would benefit from a cost-effective way to predict the behavior of human-AI collaborations.

Predicting Effectiveness

How might we predict the effectiveness of specific human-AI teams? Here are three options:

- Create AI agents and run experiments to measure performance
  - *This approach is both time consuming and expensive*
- Carry out Wizard of Oz studies with humans playing AI roles
  - *This is more reasonable but still relies on experiments*
- Develop and simulate *abstract models* of human-AI teams

The last option delays AI agent creation and experiments with subjects until we have plausible team designs.

What is an Abstract Model?

Classic computer models of human cognition (e.g., Newell & Simon, 1972) have doubled as AI systems.

- They encode mental states and generate detailed behavior.
- But they take time and effort to design, debug, and evaluate.

Instead, Ohlsson and Jewett (1995) proposed the development of abstract computational models that:

- Do not encode states / goals but still simulate steps to solutions.
- Predict speed and accuracy without building a full AI system.

Such abstract models offer a promising alternative to creating concrete models of cognition.

Abstract Models of Human-AI Teams

We can use abstract models to predict the effectiveness of human-AI teams. These would:

- Include ‘agents’ that play roles of human and AI members;
- Reflect an analysis of the steps required for a complex task;
- Simulate the time and error involved in executing the task.

They would not implement AI teammates that take the steps, but simulations would predict overall team behavior.

This would let us evaluate different team organizations and role assignments given members’ abilities.

Petri Net Models of Human-AI Teams

We have used Machinations (Dormans, 2012), an extension of Petri nets, to devise team models that specify:

- A network of tasks (nodes) the human-AI team must complete
  - Tasks are sequenced, but branches / parallel tracks allowed
- A set of agents (tokens) assigned to handle a subset of tasks
  - Each human or AI takes on a different role in the team

Each link between a pair of nodes has an associated time and probability of successful traversal.

During simulation, each token flows through the task network, possibly delayed by coordination bottlenecks.

Dice Adventure

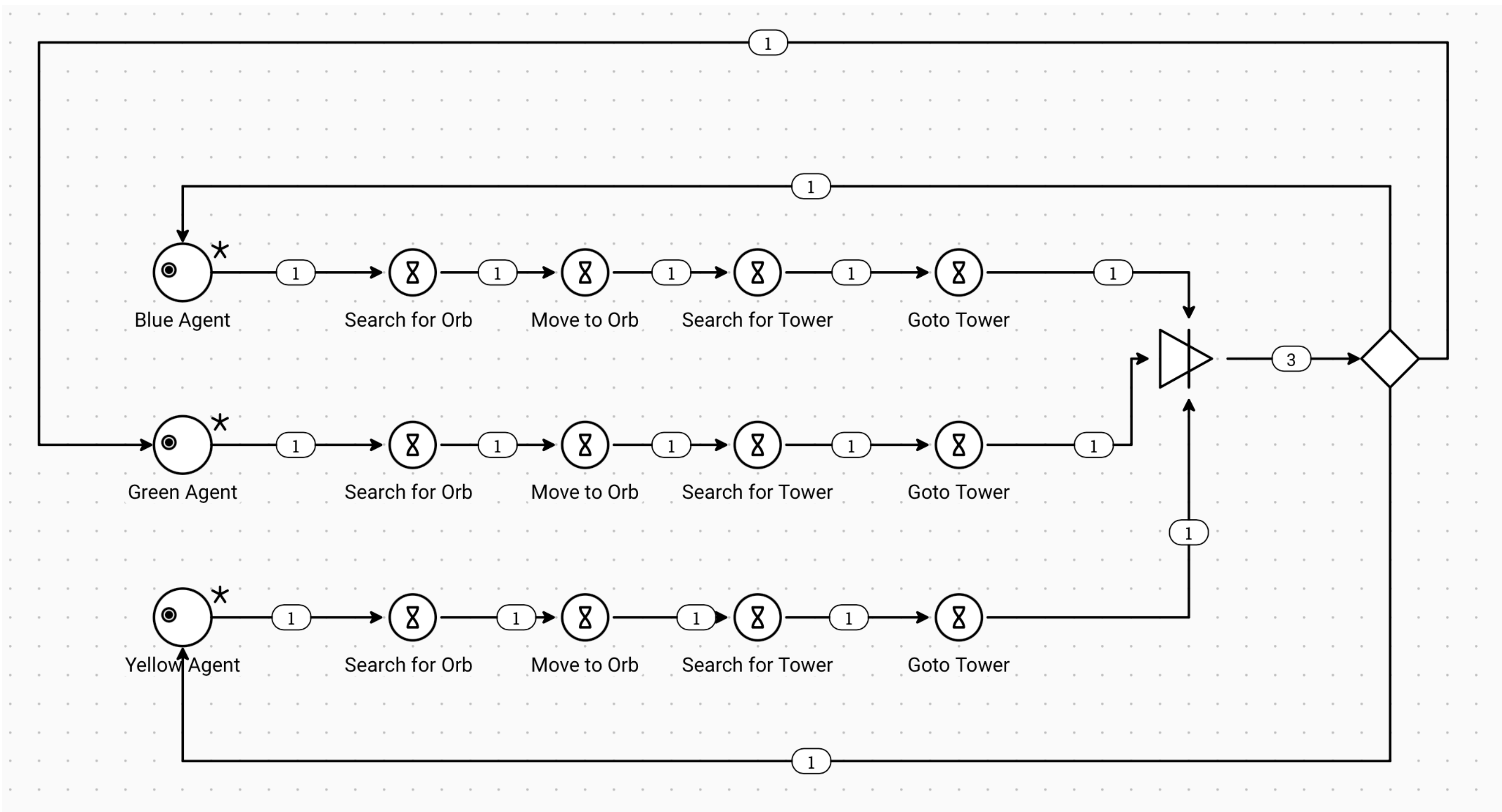
Harpstead et al.'s (2023) Dice Adventure is a cooperative video game that involves:

- A team of three agents – a giant, a human, and a dwarf
- Each searches a grid world to find an orb matching its color
- Once every agent has found its orb, they meet at a tower
- At this point, they advance jointly to the next level

Thus, the game requires both individual agent behavior and coordination among agents.

An Abstract Model for Dice Adventure

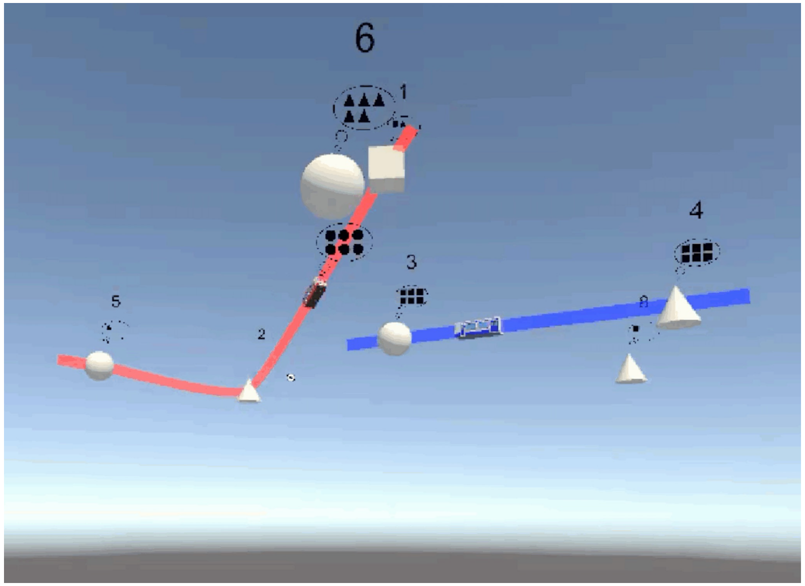
The model includes three sequential tracks, one for each agent.



These tracks converge, but only after each agent finds its orb.

Space Transit

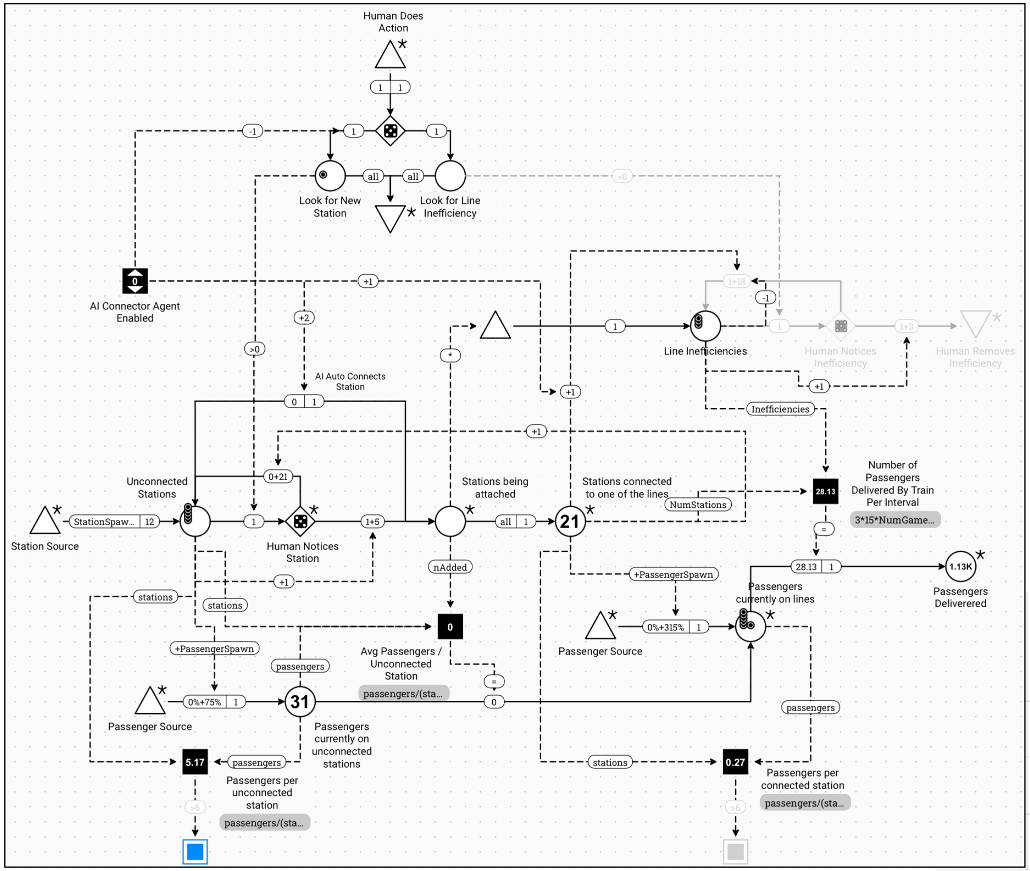
We also studied Space Transit, a logistics game where players connect stations to transit lines.



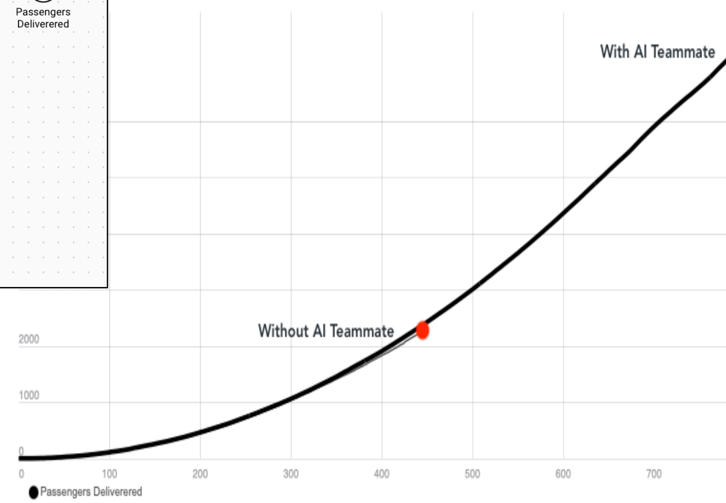
Useful lines let the passengers reach targets; players lose if the stations have too many people.

- Every day (19.2 sec) there is an 80% chance that a station will spawn.
- Every two hours (1.6 sec) there is a 15% chance a stations will spawn a passenger.
- Players try to either connect a station or optimize a line every two hours.
- Trains can deliver about 15 passengers every two hours, decreasing with line inefficiencies.
- The chance of successfully connecting a station or improving a line is proportional to the number of stations in the network.
- If the stations exceed six passengers on average, then the game is over.
- ***An AI agent connects stations to lines; this can create inefficiencies but lets humans focus on improving lines.***

An Abstract Model for Space Transit



The model for Space Transit is considerably more complex, with conditional branches and multiple pathways.



Simulation predicts the human-AI team will deliver more passengers and so keep playing twice as long.

Uses of Abstract Team Models

We can use abstract models of human-AI teams to predict how their behavior is influenced by:

- Overall design of the team
- Number / percentage of human vs. AI members
- Roles assigned to human and AI team members
- Inherent difficulty of each of the component tasks
- Speed and error rates of each member on components

Because simulation is inexpensive, we can search through a space of models to find effective team designs.

Related Research

We are not the first to propose abstract models of complex activity. Two closely related efforts are:

- The *GOMS framework* (John & Kieras, 1996), which bases its abstract models on cognitive task analyses.
  - This has been used widely to predict individual user behavior in human-computer interaction.
- *Joint activity graphs* (Vignati et al., 2022), which describe the interdependence in multi-agent teams.
  - They used this approach to inform agent behavior rather than to predict team performance.

Of course, we have also built on Ohlsson and Jewett’s (1995) ideas and on Dorman’s (2012) Machinations environment.

Closing Remarks

In summary, abstract models of human-AI collaboration hold great promise for predicting team performance.

- They can do this long before we make the effort to build AI teammates or run experiments with human subjects.
- Petri nets offer a good formalism for describing team activity and Machinations provides a useful software environment.
- We have used this framework to develop abstract models of human-AI teamwork and used them to predict effectiveness.

In future work, we will scale our models to many agents, add agent communication, and study tradeoffs among metrics.

References

Dormans, J. (2012). *Engineering emergence: Applied theory for game design*. Doctoral dissertation, University of Amsterdam, Amsterdam, Netherlands.

Harpstead, E., Stowers, K., Lawley, L., Zhang, Q., & Maclellan, C. (2023). Speculative game design of asymmetric cooperative games to study human-machine teaming. *Proceedings of the Eighteenth International Conference on the Foundations of Digital Games*, 1–4.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, *3*, 320–351.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S., & Jewett, J. J. (1995). Abstract computer models: Toward a new method for theorizing about adaptive agents. *Proceedings of the Eighth European Conference on Machine learning*, 33–52. Heraklion, Crete: Springer.

Swartz, L. (2003). *Why people hate the paperclip: Labels, appearance, behavior, and social responses to user interface agents*. Honors Thesis, Symbolic Systems Program, Stanford University, Stanford, CA.

Vignati, M., Johnson, M., Bunch, L., Carff, J., & Duran, D. (2022). Interdependence design principles in practice. *Frontiers in Physics*, *10*, 969544.