Report on the outcomes of a Short-Term Scientific Mission[1]

Action number: CA22145

Grantee name: Dr. Mark Goadrich

Details of the STSM

Title: Comparison between CardStock and Belief-Stochastic Game Frameworks for Imperfect Information General Game Playing

Start and end date: 15/06/2025 to 21/06/2025

Description of the work carried out during the STSM

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

The goal of this STSM was twofold: first, to facilitate the exchange of theoretical and technical insights from CardStock, a general game playing engine specialized in card games; and second, to evaluate the practicality and performance of the Belief-Stochastic Game (BSG) framework by applying it to a diverse set of classic and modern games. We began with an in-depth discussion with Achille Morenville and Eric Piette on the CardStock engine and the ReCYCLE language for describing card game rules. One strength observed of the system is the way that ReCYCLE simplifies the development and iteration of game rules. The speed of simulations in CardStock is slow, but not infeasible to approach larger games and use more simulations to increase player performance. We also discussed how associating card visibility with the location of the card, and not with the card itself, creates complications for flipping a card in a game and retaining its location. Additionally, the knowledge representation of the AI players is extremely limited, such that if a card is given to another player, that card becomes untrackable.

We then moved to discussions and explanations of BSG from Morenville and Piette. Morenville provided detailed tutorials and examples on the Belief Propagation framework, how the knowledge of card identities is captured as a bipartite graph of variables and constraints in a Constraint Satisfaction Problem, and how when new information is revealed by players or the game in BSG, the variables are updated by the constraints using Constraint Propagation algorithms. Morenville also began showing how to implement a game in BSG. There were definite structural similarities to the rules and framework abstraction choices across both systems. I started to understand the difficulty of consistently updating variables for all players at all times.

[1] This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

Based on my initial understanding of the BSG system, we explored a few potential ideas for modification. These included subdividing the variable unit from cards into their component parts when representing with CSP, which could allow for more abstracted knowledge revision functions, a set approach to updating probabilities rather than an individual card approach, and the use of rollouts to estimate probabilities.

Throughout the week, I made multiple improvements to the CardStock system and ReCYCLE language. I reorganized the way output files from experiments are stored, simplified the way AI players are selected for experiments, and discovered and fixed multiple bugs in ReCYCLE. With these bugs fixed, I could make progress on fixing implementations for the games Crazy Eights and Golf.

I also implemented three new games, Kuhn's Poker, Leduc Poker, and Cuckoo, and began work on Scopa. Finally, I upgraded the Pure Monte Carlo AI implementation in CardStock to allow for more rollouts, and I standardized the approach using limited determinations to better align with the BSG framework. The upgrades allow for initial test runs to commence on the game Goofspiel as we begin calibrating our systems for the comparison experiments.

Our conversations led to creating a testbed of games for comparing our two systems, described below.

Description of the STSM main achievements and planned follow-up activities

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

Our main outcome was the creation of a traditional card games testbed, shown in Table 1, which will facilitate the comparison of General Imperfect Information Game Playing (GIIGP) systems.

Table 1: Traditional Card Games Testbed

Game	Genre	Origin	Deck Type	# of Play
BlackJack	Banking	France	52 French	2
Crazy Eights	Shedding	USA	52 French	3
Cribbage	Adding	England	52 French	2
Cuckoo	Exchange	England	52 French	6
Euchre	Euchre	Germany / USA	24 French (9, 10, J, Q, K, A)	4
Gin Rummy	Rummy	USA	52 French	2
Go Fish	Quartet	USA	52 French	4
Golf-6	Draw and Discard	USA??	52 French	4
Goofspiel	Collect	USA	52 French	2

Hearts	Avoidance	USA	52 French	4
Klaverjassen	Jack-Nine	Netherlands	32-card deck (7-10, J, Q, K, A)	4
Leduc Hold'em	Poker	Canada	6 cards (K, Q, J)	2
Pitch	High-Low-Jack	England	52 French	4
Schwimmen	Exchange	Austria	52 French	5
Scopa	Fishing	Italy	40 Italian (K, H, J, 7, 6, 5, 4, 3, 2, A)	2
Skittgube	Last In	Sweden	52 French	3
Solo Whist	Solo Trick	England	52 French	4
Sueca	Ace-Ten	Portugal	40 Spanish (K, H, J, 7, 6, 5, 4, 3, 2, A)	4
Zheng Shangyou	Climbing	China	52 French	5

These 19 games represent a variety of mechanical genres, number of players, and team vs individual play. Within each genre, we chose games that promote a diversity of cultural connections and experiences, utilizing card decks and games from different countries. We also chose games earlier in historical development of the game, with an eye toward a simpler version that retained the unique quality of the genre. This preference meant that we left out some classics that are complicated, such as Bridge, Skat, and Poker, but retained their essence through other games.

Implementing these games will challenge both of our systems; we expect each one to introduce new complications and edge cases that must be overcome to make progress. For CardStock and ReCYCLE in particular, this will likely result in new language keywords and parsing implementations, as well as a continued push toward efficiency of running rollout simulations. We will use the Card Game Rules website pagat.com into a standardized English description, making choices to resolve any remaining ambiguities. Many card games include rules to repeatedly play until one player or team has reached a certain number of points. However, for our experiments and implementations, we will limit each game to a single round, thus making the decision space more approachable for our AI players.

We also discussed various criteria we would use in this alignment comparison between our GIIGP systems. At a simple level, we can compare the memory and time usages for performing Monte Carlo rollouts across the full trajectory of the game state of each game. A second comparison will be understanding the differences in performance of AI players in each game. For consistency, AI players will compete against only random players.

A first attempt will be to align our approaches in a simple Monte Carlo player, so that each system is given the same budget of rollouts and uses them in an identical manner. For games with partners, we will have both players implement AI in games versus random players. The ease of coding up each game will be examined, relative to not just length of code but understandability. We plan to continue the conversations begun through this STSM throughout the fall, and we intend to publish our results in the 2026 IEEE Conference on Games proceedings.

This research will lay the foundations to formulate comparisons and understandings of the particular player experiences for each of the games in the testbed. Some games might be more amenable to play in one system versus the other, and we expect the explicit belief representation to contribute to higher-scoring AI players. When full games with all AI players are simulated, comparisons between

games will be attempted, which could shed light on the historical and cultural connections between the games in the testbed as well as modern games.