

---

# Formatting Instructions for NIPS 2015

---

**David S. Hippocampus\***  
Department of Computer Science  
Cranberry-Lemon University  
Pittsburgh, PA 15213  
hippo@cs.cranberry-lemon.edu

**Coauthor**  
Affiliation  
Address  
email

**Coauthor**  
Affiliation  
Address  
email

**Coauthor**  
Affiliation  
Address  
email

**Coauthor**  
Affiliation  
Address  
email  
(if needed)

## Abstract

Abstract here.

## 1 Introduction

Describe hanabi here + cooperative AI in general

Hanabi [1, 2] is a cooperative card game in which two or more players act on partial information towards a common goal. Each player can see the cards of all other players, but not their own. On each turn, a player can play a card (without looking at it), discard a card, or give information to another player. The goal is to play as many cards as possible in a specified order.

Using machine learning to play adversarial games is a well-studied topic. We believe that using machine learning to play a cooperative game will bring about new questions. In Hanabi, for example, the emphasis on cooperation means that its advantageous for all players to agree on certain protocols beforehand. One question that we'd like to answer in this project is whether, given a human player that is already following a certain protocol, an AI can be automatically taught to follow the same protocol.

Another question is whether it's possible to simultaneously train multiple models. The most obvious way to train a model may be to simulate games played with an initial, hardcoded policy, which has deterministic behavior and can therefore act as part of the environment. We believe this will work well for a two-player game, but is unlikely to scale to more players, when more coordination is needed. As part of this project, we would also like to explore the effect of simultaneously training two or more models, possibly with one hard-coded player.

---

\*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledged funding agencies.

## 2 Prior Work

## 3 Approach

### 3.1 Overview

### 3.2 Algorithms

#### 3.2.1 TRPO

[3]

#### 3.2.2 VPG (if space)

#### 3.2.3 CEM (if space)

#### 3.2.4 CMA-ES (if space)

### 3.3 Heuristics

#### 3.3.1 Heuristic

#### 3.3.2 SimpleHeuristic

## 4 Evaluation

## 5 Conclusion

### References

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to ‘small’ (9-point) when listing the references. **Remember that this year you can use a ninth page as long as it contains *only* cited references.**

### References

- [1] Hanabi — board game — boardgamegeek. <https://boardgamegeek.com/boardgame/98778/hanabi>. Accessed: 2017-4-30.
- [2] Hanabi (card game). [https://en.wikipedia.org/wiki/Hanabi\\_\(card\\_game\)](https://en.wikipedia.org/wiki/Hanabi_(card_game)). Accessed: 2017-4-30.
- [3] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel. Trust region policy optimization. *CoRR*, abs/1502.05477, 2015.