Tabletop games as Multimodal Datasets for Social AI

JŪRA MINIOTAITĖ, KTH Royal Institute of Technology, Sweden
ANDRÉ PEREIRA, KTH Royal Institute of Technology, Sweden

Progressing the work to create believable, data-driven, social AI starts where most difficult AI challenges start, with the dataset. Tabletop games provide an excellent opportunity to get a dataset with complex social behavior in a structured context. In this work, the tabletop games; Pandemic Hot Zone - Europe, Hanabi, Poker, and a custom designed game are analyzed to assess their quality as tasks for data collection. We will analyze possible datasets collected in the abovementioned games in respect to: (1) Dialog overlapping and lexical diversity; (2) Cooperation, Competitiveness and Social Context; (3) Game components and non-verbal behavior; (4) Research Opportunities.

CCS Concepts: • **Human-centered computing → User studies**.

Additional Key Words and Phrases: datasets, games, multimodal interaction, social ai

ACM Reference Format:
Jūra Miniotaitė and André Pereira. 2022. Tabletop games as Multimodal Datasets for Social AI. In . MMAI, Amsterdam, The Netherlands, 5 pages.

1 INTRODUCTION

Current multimodal datasets are often not very social and sometimes feature acted rather than natural interactions. There are some exceptions to this [3, 5, 7], however, there is still a great need for more multimodal, high-quality datasets. Interaction is difficult to capture because it is a connection that is happening in a physical and social space on many levels at the same time. Naturalistic social interaction is chaotic and, by definition, impossible to force. To catch natural interaction without trying to force it, we propose employing tabletop games. Tabletop games provide social structure that can enable data collection of target interaction patterns.

In this paper, we address how to exploit tabletop games as the context and instrument for collecting rich multimodal social data. We present an initial framework for analyzing tabletop games for data collection and give an overview on the type of research opportunities that this type of tasks enable. Four different tabletop game candidates are explored to establish how different game mechanics facilitate different types of interactions and hence different chances for collecting multimodal datasets.

2 BACKGROUND

To analyze what game mechanics are suitable for data collection three popular games were chosen, and one game was custom made. Games that promote social interaction via cooperation or asymmetric information were chosen for analysis. A play time of 30 minutes was targeted so that the interaction is long enough to facilitate complex social interactions, but not so long that would reduce the capacity of collecting data from a varied pool of participants.

2.1 Hanabi

Hanabi is a cooperative game where the players play their cards in the correct order to build fireworks. The cards are numbered and have one of five different colors. The players cannot see their own cards, they can only see each other's cards. The players take turns performing one of the following actions: give one piece of information, discard a card, or play a card. Giving information to other players is limited in such a way so that it is impossible for the players to

have all information about the cards they play. Instead, the hints they give each other must be informed by the context in which they are given. Several authors, including Google's DeepMind and Google Brain, have proposed the Hanabi challenge as a new frontier for AI research as this game deals with imperfect information. The game forces the players to inform their play by theory of mind reasoning [1].

2.2 Pandemic Hot Zone - Europe

Pandemic was released in 2008 and it is considered one of the most successful cooperative board games and can be played in a limited time by a wide range of players. The game is collaborative where the players 'play against the game'. The goal is to stop the spread of viruses in the big cities in the world by travelling to cities with player markers and curing diseases. There is a short version of Pandemic that was released in 2021 called 'Pandemic hot zone' that contains the main game elements of the popular game and can be played in 30 minutes. Pandemic is also a boardgame with lots of physical game components that makes it easy to keep track of the state of the game and provides opportunities for collecting rich non-verbal behavior data such as joint attention and communication via gestures.

2.3 Poker

Poker is a competitive card game with complex social interaction such as bluffing, betting and trash talking. The game gained a lot of popularity in the beginning of this millennia and continues to attract players at tournaments and viewers on Youtube [2, 8]. Texas Hold'em was the specific variation of poker chosen for analysis because it is a popular version of poker that promotes intense social interactions with its several betting stages. It is also a game that has been used for AI research in the past [9, 10]. The goal is to make the best five-card hand from two personal cards that are dealt in the beginning and three communal cards that are revealed during the game. The players bet after they get their personal cards as well as after each communal card is revealed.

2.4 PeekerPicker (custom game design)

Some research projects require specific interaction patterns. For our own research goals, we looked into the possibility of using one of the three games proposed above and explored designing a game that would be tailored to our needs. One of the risks with creating a custom game is that it takes a long time to determine if a game is fun to play long-term.

A game design workshop [11] explored a few different types of game mechanics and themes as guidance to develop a tabletop game. The result was the game Peeker&Picker where similarly to [6], a player collaborates with a teammate while at the same time engaged in a competitive interaction with another team of two players. The game is based of memory games where the goal is to collect pairs of cards. The cards are laid out on a table and each team has a 'picker' and a 'peeker', where the picker can pick a card and then hold it and the peeker can look at cards. The picker cannot hold more than 4 cards at a time. If the picker picks up the fifth card, they discard a card. The peeker can either look at two cards that are on the table or see their teammate's hand. Both teams take turns playing one action at a time.

The idea is that the players convey enough information to their teammate to describe a card but keep it fragmented and incomplete to not give away information to the other players. This is combining describing cards like in the game 'Taboo'[1] with theory of mind thinking. The cards also provide opportunities for joint attention and gesturing to discuss the physical location of the cards the players should pick up or look at.

[1] https://boardgamegeek.com/boardgame/1111/taboo

3 ANALYSIS OF GAMES

For analyzing the games, we performed 4 gameplay sessions where they were informally playtested and recorded like in Figure 1. Three of the games (Hanabi, Pandemic, Peeker&Picker) were tested by the first author and three people aged between 26 and 36 (2 males, 1 female). Poker was tested with a partly different group including the first author and one of the participants in the previous group. In addition, three more participants aged between 25 and 33 (2 males, 1 female) were recruited for playtesting Poker. In all game tests most of the players had not played the game before, therefore the games had to be introduced to the players before the start of the game.

Fig. 1. Playtesting poker

3.1 Dialog overlapping and lexical diversity

For describing the dialog the word 'chaotic' will be used. It means there is a lot of overlapping dialogs, uncertainty in who the speaker is addressing and very rapid turn taking between multiple participants. Dialog described as chaotic is extremely challenging to be recognized by speech recognition algorithms or annotated. The dialog in Hanabi is not very chaotic, turn-taking is strictly controlled by the rules of the game. Since it is easy to slip up and cheat accidentally, players are encouraged to keep silent. In practice, when playing a game of Hanabi, the players tend to cheat and talk about tactics anyway. The cards in the game are spoken about in terms of color and number, rather than being engaged in the fantasy of building fireworks. In Pandemic, the theme of the game shines through in the dialog where the players talk about curing and travelling to different cities. The dialog is chaotic because the group discusses and take decisions together about each players actions during the game. In poker, the topics that are discussed are keeping track of possible bluffing and the turns, while in Hanabi it is about tactics. Peeker&Picker is designed to facilitate dialog with lexical diversity using memory cards that promote it. The lexical diversity can also be considered versatile as by changing the theme of the game, that is the cards, you can get the players to change the content of the dialog. Regarding dialog, there tends to be an overlap between the turns as players speak to their teammate when it is not their turn to pick a card which makes for quite a chaotic dialog situation with two pairs talking to each other and only occasionally addressing the other players.

3.2 Cooperation, Competitiveness and Social Context

The two purely collaborative games in this analysis are Hanabi and Pandemic. They are collaborative in very different ways. Playing Pandemic promotes more of a team feeling, because the setbacks are caused by the game rather than by a player. The players commonly make a group decision about the moves that they make, so if there is a setback the whole group is to blame. In contrast, Hanabi has a competitive feel to it because the players can often play in a way that the others disagree with. When one player is using the common resources in the game in way that the others feel is suboptimal that causes social tension in the group. In a similar way, Peeker&Picker is challenging the cooperation between teammates by providing the risk of making the wrong move by disclosing too much information to the opponents openly.

Social adaptability is important in tabletop game design [4] for promoting dynamic social environments that can adapt to different playing styles or personalities. Pandemic promotes several collaborative roles such as mediator and helper, while Poker promotes competitive social roles such as dominator and violator. Peeker&Picker has both competitive and collaborative roles, however, interaction between teams is limited which can make the game less socially adaptable.With respect to our observations, Hanabi appears to be the least socially adaptable game because there is no room for changing one's social behavior in any other way than being less engaged, that is taking the role of recluse.

3.3 Game components and non-verbal behavior

For multimodal data collections, game components matter. They provide opportunities for non-verbal behavior as well as grounding the conversations by referring to the objects that are in play. The game components also help with determining the state of the game. In a multimodal dataset this can give insights into the context of the human behavior.

Hanabi has moments of communication with pointing when the players give hints about each other's cards. Other non-verbal behavior that could be found in the game is much more subtle and could be interpreted as cheating. One example would be looking for reactions in facial expressions when talking about playing a card. In Pandemic, gesturing about what moves to make on the board is the most prominent non-verbal behavior. There is also gaze interaction in joint attention when discussing tactics referring to the pieces on the board.

In Poker, gaze and interpreting facial expressions play a big role in trying and determine the other players state of mind, but there is no pointing or referring expressions that are essential to play the game. There is also indirect communication in how confident you are in your hand by betting, which is done by moving the chips. In Peeker&Picker, players use pointing and looking at cards to communicate with their team-mate. Because the players want to sneak information past the other team there is also communication via gaze and facial expressions like in Poker and Hanabi.

3.4 Research opportunities

The games presented in this work can be used to provide data for different types of research. The Hanabi challenge is described as the new frontier for AI research as it challenges the AI to adapt to a player's social behavior [1]. Collecting multimodal data within this game provides an opportunity to study, for example, how other players' non-verbal communication can help an AI in playing the game. Collecting a multimodal Poker dataset could also be used to explore game theory with social cues and challenges in emotion recognition. Since pandemic and Peeker&Picker do not put restrictions on speaking to the teammates it offers the opportunity to study group decision making and turn-taking in 'chaotic' dialog environments. Most games with game components can also be used to study gestures in dialog and

mutually grounded dialog. This study is particularly motivated by the opportunity to use multimodal datasets to train social robots to play tabletop games. Rich multimodal data with gaze, natural language and social behavior is needed for training robots that can mimic the interaction patterns found in humans.

4 CONCLUSION

Tabletop games are a good source for rich multimodal social data that can be used for various research endeavors. When collecting multimodal data using a tabletop game it is important to consider the style of the dialog. It can be very chaotic or very structured, and the two extremes offer different challenges and opportunities. Using themes in games is an efficient way to change the content of the speech in the dataset. In addition, game components in the game should be considered since they offer an opportunity to study mutual attention and physically grounded dialog. Different games offer opportunities to study various research areas and can be employed to push several boundaries in research such as creating embodied robots that can navigate social environments.

REFERENCES

[1] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H. Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibl Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. 2020. The Hanabi Challenge: A New Frontier for AI Research. *Artif. Intell.* 280, C (mar 2020), 19 pages. https://doi.org/10.1016/j.artint.2019.103216

[2] Rachel Croson, Peter Fishman, and Devin G Pope. 2008. Poker superstars: Skill or luck? Similarities between golf—thought to be a game of skill—and poker. *Chance* 21, 4 (2008), 25–28.

[3] Metehan Doyran, Arjan Schimmel, Pınar Baki, Kübra Ergin, Batıkan Türkmen, Almıla Akdağ Salah, Sander CJ Bakkes, Heysem Kaya, Ronald Poppe, and Albert Ali Salah. 2021. MUMBAI: multi-person, multimodal board game affect and interaction analysis dataset. *Journal on Multimodal User Interfaces* 15, 4 (2021), 373–391.

[4] Daniel Eriksson. 2005. Socially adaptable games. (2005).

[5] Lucien Maman, Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Laurence Likforman-Sulem, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. 2020. Game-on: A multimodal dataset for cohesion and group analysis. *IEEE Access* 8 (2020), 124185–124203.

[6] Raquel Oliveira, Patrícia Arriaga, Patrícia Alves-Oliveira, Filipa Correia, Sofia Petisca, and Ana Paiva. 2018. Friends or foes? Socioemotional support and gaze behaviors in mixed groups of humans and robots. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 279–288.

[7] Cheul Young Park, Narae Cha, Soowon Kang, Auk Kim, Ahsan Habib Khandoker, Leontios Hadjileontiadis, Alice Oh, Yong Jeong, and Uichin Lee. 2020. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* 7, 1 (2020), 1–16.

[8] PokerStars. 2022. Online Qualifier DESTROYS Poker Pros PokerStars. https://www.youtube.com/watch?v4Uo1PxVZlsI

[9] Martin Schmid, Matej Moravcik, Neil Burch, Rudolf Kadlec, Josh Davidson, Kevin Waugh, Nolan Bard, Finbarr Timbers, Marc Lanctot, Zach Holland, et al. 2021. Player of games. *arXiv preprint arXiv:2112.03178* (2021).

[10] Finnegan Southey, Michael P Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. 2012. Bayes' bluff: Opponent modelling in poker. *arXiv preprint arXiv:1207.1411* (2012).

[11] Jonathan Thwaites. 2020. Board Game Design Workshop: Okanagan Regional Library (ORL). https://www.orl.bc.ca/read-listen-watch/programs-for-adults/ongoing-activities/board-game-design-workshop