Data exploration, Playing styles, and Gameplay for Cooperative Partially Observable games: Pictionary as a case study

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering by Research

by

Kiruthika Kannan
2018702001
kiruthika.k@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
November 2023

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Data exploration, Playing styles, and Gameplay for Cooperative Partially Observable games: Pictionary as a case study" by KIRUTHIKA KANNAN, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. RAVI KIRAN SARVADEVABHATLA

To my grandma and grandpa...

Acknowledgments

I would like to express my sincere gratitude to my advisor, Prof Dr. Ravi Kiran, for his constant guidance and support. The skills and knowledge I gained under his supervision made my research thesis possible. I would like to express my sincere thanks to Prof Dr. Vinoo Alluri for her insightful guidance that helped me gain valuable knowledge. I would also like to thank the institute and CVIT for providing a research friendly environment that kept me motivated.

Secondly, I would like to express my thanks to my teammates who worked together on this project for their support and cooperation. I will treasure the fun days we had playing the game Pictionary game together. I want to extend my thanks to all the participants who enriched our Pictionary dataset with their creative gameplay.

To my dear friends Sowmya, Prathyusha, Mythri, and Tanu, I am grateful for the unforgettable memories we created during our time at IIITH. Lastly, I would like to convey my heartfelt appreciation to my family for their constant support and encouragement during the challenging phases of this long journey. Their unwavering belief in me has been a source of strength and motivation.

Abstract

Cooperative human-human communication becomes challenging when restrictions such as difference in communication modality and limited time are imposed. In this thesis, we present the popular cooperative social game Pictionary as an online multimodal test bed to explore the dynamics of human-human interactions in such settings. Pictionary is a multiplayer game where the players attempt to convey a word or phrase through drawing. The restriction imposed on the mode of communication gives rise to intriguing diversity and creativity in the players' responses.

To explore the player activity in Pictionary, an online browser-based Pictionary application is developed and utilized to collect a Pictionary dataset. We conduct an exploratory analysis of the dataset, examining the data across three domains: global session-related statistics, target word-related statistics, and user-related statistics. We also present our interactive dashboard to visualize the analysis results.

We identify attributes of player interactions that characterize cooperative gameplay. Using these attributes, we find stable role-specific playing style components independent of game difficulty. In terms of gameplay and the larger context of cooperative partially observable communication, our results suggest that too much interaction or unbalanced interaction negatively impacts game success. Additionally, the playing style components discovered via our analysis align with select player personality types proposed in existing frameworks for multiplayer games.

Furthermore, this thesis explores atypical sketch content within the Pictionary dataset. We present various baseline models for detecting such atypical content. We conduct a comparative analysis of three baseline models, namely BiLSTM+CRF, SketchsegNet+, and modified CRAFT. Results indicate that the image segmentation-based deep neural network outperforms recurrent models that rely on stroke features or stroke coordinates as input.

Contents

List of Figures

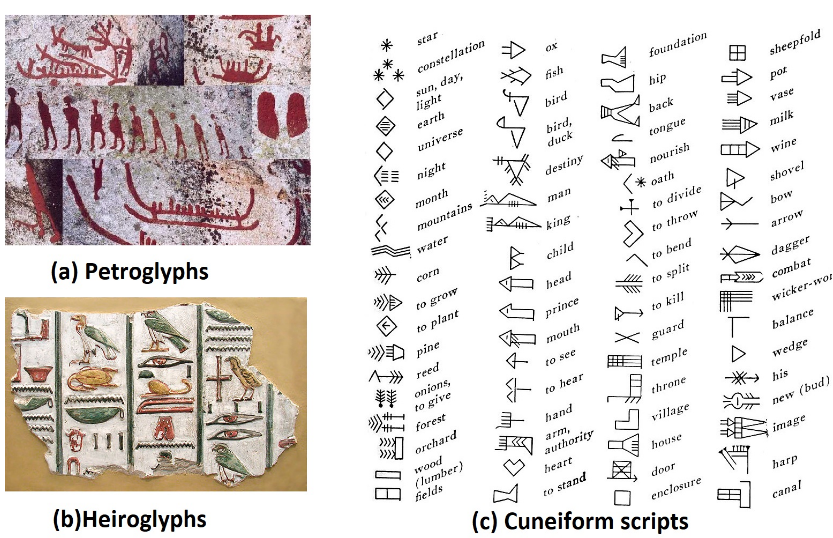
List of Tables

Chapter 1

Introduction

Cooperative human communication in a shared goal setting is ubiquitous. However, constraints often exist despite the cooperative setting. For instance, travelers throughout history have needed to communicate and convey intent despite constraints such as unknown language and cultural barriers. Gestures and pictorial representations are often utilized to overcome these barriers. Ancient pictorial communication involved pictograms such as petroglyphs (rock carvings), Egyptian hieroglyphs, and cuneiform writing systems (see Figure 1.1) [9, 81]. Shared understanding is achieved in such cases through iterative feedback. Studying these processes in a casual game setting allows us to tap into these fundamental human interactions occurring in constrained communication settings. In this thesis, Pictionary, a popular social casual game involving drawing and guessing, is used as a case study to understand the communication between players in a cooperative, constrained setting.

Figure 1.1: Forms of ancient communication.



1.1 Pictionary: A cooperative partially observable game

Pictionary is an interesting example of cooperative gameplay to achieve a shared goal in communication-restricted settings [29, 76]. The game of Pictionary consists of a time-limited episode involving two players - a *Drawer* and a *Guesser*. The *Drawer* is tasked with conveying a given target phrase to a counterpart *Guesser* by sketching on a whiteboard within a time limit [29]. Crucially, the rules of Pictionary forbid the *Drawer* from writing text on the whiteboard. Since the target phrase is not observable, this restriction leads to partial observability in an otherwise cooperative game. Coupled with time limit per game episode, the restriction also tends to unleash fascinating diversity and creativity in player responses and generated content. Figure 1.2 shows an illustrative description of the Pictionary game.

Figure 1.2: The Pictionary game.

The study of cooperative guessing games is not new [2, 33, 1]. However, existing works assume a *single* and *common* modality for communication (e.g., text). By contrast, players in Pictionary employ different modalities – the *Drawer* draws visual patterns on a 2-D canvas while the *Guesser* enters text guesses. In addition, the players employ out-of-modality 'iconic gestures' (e.g., the *Drawer* signaling 👎, 👍 to *Guesser* to convey the proximity of the guess phrase to the target phrase). Throughout the game, the players are forced to adopt strategies distinct from a single modality setting. This sets the stage for a multimodal theory of mind in which a player is required to 'imagine' other player's state of mind and that too, from responses expressed in a modality different from theirs. As we shall show, players exhibit different styles of interactions to achieve this multimodal communication. Examining these strategies and interaction styles could potentially be of relevance to game developers, researchers developing AI-based cooperative game agents, social scientists studying dynamics in social games, amongst other stakeholders.

1.2 Related work

Cooperative Partially observable (CPO) Games: For cooperative gameplay, the players require information on the state of each player as well as the game parameters. But in games, the players are not always allowed to fully observe their environments. The game design and game rules impose limitations on observability by hiding certain information from the players. Cooperative games with imperfect information are termed as Cooperative Partially Observable games (CPO) [2]. Bard et al. [8] introduced the card game Hanabi as a new challenge exhibiting a combination of cooperative gameplay and imperfect information. Eger et al. [23] demonstrated the use of intentionality and communication theory for designing Hanabi agents. Iconary [16] is a CPO game similar to Pictionary where the *Guesser* tries to guess a phrase instead of a word, and the *Drawer* repeatedly revises the icons on canvas to help the *Guesser* identify the phrase. Charades is a CPO game similar in spirit to Pictionary wherein the player has to act instead of drawing [32, 6, 66] and the role of *Guesser* remains similar to that in Pictionary. Ashktorab et al. [2, 1] used a word guessing game, *Wordgame* where a player, *'Giver'* is given a target word and asked to convey the word to the teammate, *'Guesser'* as in the case of Pictionary. Unlike Pictionary, this game uses a single modality of communication, viz., text.

1.3 Thesis Contribution

In this thesis, we make the following contributions

1. We introduce the game of Pictionary as a case study for multimodal Cooperative Partially Observable games.
2. We conduct an exploratory study on Pictionary telemetry data for understanding the dynamics of gameplay.
3. We characterize the player-player interactions in Pictionary and identify the playing styles exhibited by the players.
4. We examine the atypical activities in the Pictionary game data and compare the performance of baseline models for atypical sketch content detection.

1.4 Thesis Organization

The structure of this thesis is as follows. Chapter 2 discusses the related research work on different genres of games and introduces the game of Pictionary as an example of a CPO game. The rules of the game are explained, and the motivations for using Pictionary in the study are outlined. The browser based online Pictionary app used for data collection is also presented in Chapter 2. Chapter 3 summarizes the exploratory analysis of the collected Pictionary dataset. It presents a dashboard used for visualizing the statistics of each component of the game dataset. Chapter 4 delves deeper into characterizing the interactions in Pictionary gameplay and identifies Pictionary role-specific playing style components independent of game difficulty. It also examines the association of the game outcome with the playing style components and certain attributes of game interplay. Chapter 5 examines atypical activities observed in the Pictionary sketch data and compares different baseline models for detecting these atypical sketch content. Finally, the contributions, experiments, analysis, and observations are summarised in chapter 6.

Chapter 2

Pictionary: A Cooperative Partially Observable game

Games resemble a simplified model of reality [64]. They are designed to provide a controlled environment for variables to interact and influence each other. By considering game problems as cheap, formalized representatives of real-world tasks, we can learn how to solve problems in reality [69]. The study of games and artificial intelligence (AI) in games has branched into multiple research areas. These include behavioural learning, player modelling, games as AI benchmarks, believable agents, AI assisted game design, and general game AI [96, 94]. Another use case of Game AI in gaming industry is game analytics [69]. In game analytics, data about the player are analyzed to update game parameters such as difficulty levels and the user interface [25]. Analyzing when the players quit the game gives information on the game churn rate [34, 74]. A more extensive exploration of the game data includes player modelling that can benefit research focused on game design enhancement [21], personalized game content generation [55, 97], team composition [63], and adaptive gamification [71].

In the following sections, we first present how research on different genres of games has evolved from simple rule-based games to complex cooperative settings. Next, we introduce Pictionary as a case study of a Cooperative Partially Observable (CPO) game. We then explain the game rules of Pictionary and the unique characteristics that make it an interesting research challenge.

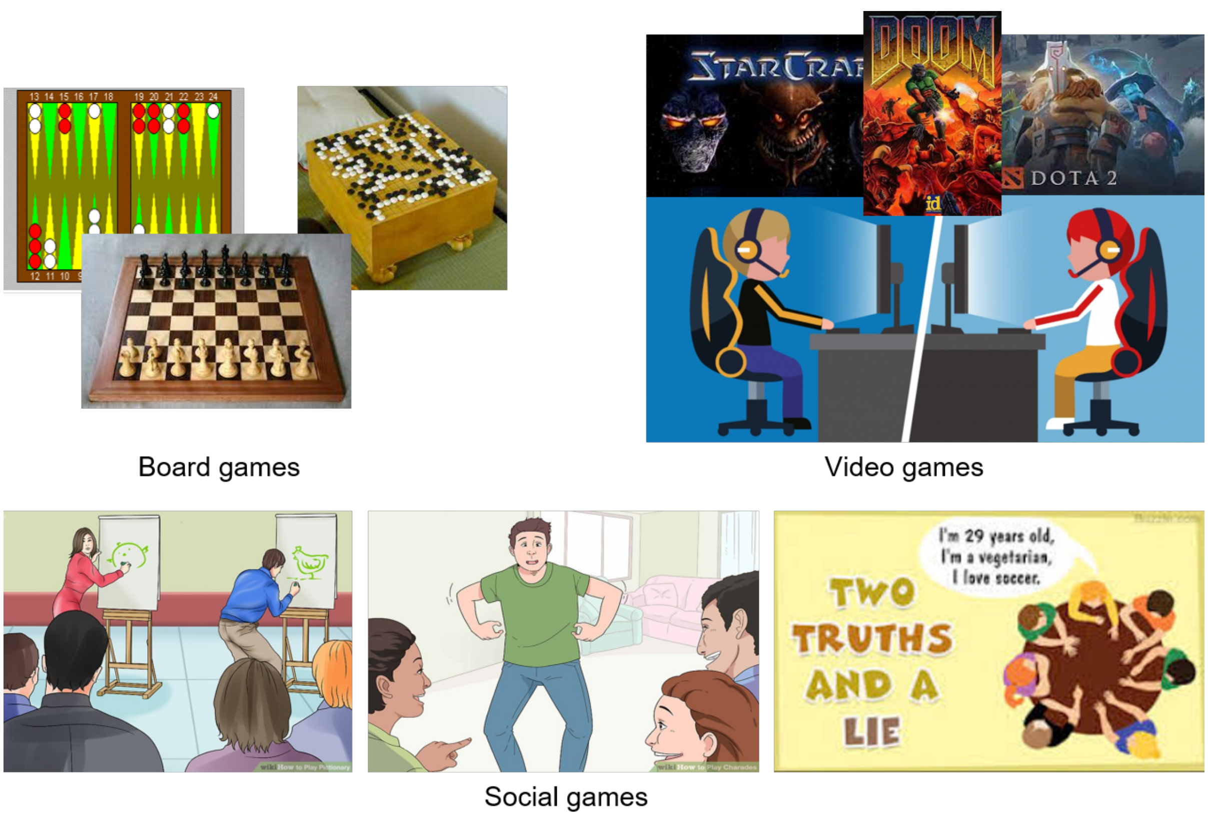
2.1 Diversity of Game Research

The research on games has explored different genres of games and has gone through a lot of breakthroughs [94, 98]. One of the earliest games played by AI, Pac-Man (1979) [61], relied on a simple state machine involving a random path-finding algorithm. Since then, artificial game playing agents have evolved to deal with far more complex tasks using highly sophisticated algorithms. A well-known example would be the chess playing system Deep Blue [15], which defeated the world chess champion in 1997. A more recent milestone was achieved by AlphaGo (2017) [80], a self-taught AI without learning from human games. One common aspect of these board games is that they have a finite number of states and actions. The deterministic nature of gameplay in these games is leveraged to engineer agents that excel, even surpassing human players.

Apart from traditional board games like Chess and Go, several genres of video games have also been of interest to AI researchers. Some commonly explored game genres are real-time strategy (RTS) games such as StarCraft [87], Multiplayer online battle arena (MOBA) games such as Dota2 [12], and first-person shooter (FPS) games such as Doom [44]. In contrast to board games, these video games have a larger number of variables acting independently in complex non-deterministic gameplay. As a result, the study of these games involves computationally expensive analysis of high-dimensional data.

Another characteristic of these games is the adversarial competitive nature of the gameplay. The player's motivation to play is predominantly the intent to win the game or defeat other players. Such games are usually zero-sum - i.e., one player loses, and the other player wins in the case of a two-player game. Contrary to this, cooperative social games involve multiple players with a shared goal. During cooperative gameplay, the players support each other to collectively achieve their targets. Needless to say, the gameplay in a competitive game is largely different from the gameplay in a cooperative game. Hence, most of the existing analyses on player behavior, playing styles, and gameplay analysis for competitive games are not extendable to cooperative games. We will explain more on this in Section 4.1

Figure 2.1: Types of games in research.



2.2 Pictionary as a case study

Pictionary is a popular casual social game played by people of all age groups. In Pictionary, a team of two or more players plays together towards a shared goal. One player of the team is chosen as the *Drawer* while the other player or players is the *Guesser*. The *Drawer* is given a randomly chosen *target word* which can be of any category, such as a name of an object, movie, action, or phrase. The other players do not know what this word is. The goal of the game is for the *Drawer* to convey the *target word* by sketching to the teammate *Guesser* within a time limit. The *Drawer* is provided with a canvas on which the *Drawer* can sketch the *target word* or hints related to the *target word*. However, the *Drawer* is not allowed to write any text on the canvas. Also, the *Drawer* is not allowed to communicate by speaking or acting. The *Guesser* looks at the canvas and provides guesses of the *target word*. The game ends successfully when the *Guesser* correctly guesses the *target word*, and the players earn points. The game ends in a failure if the time runs out without the *Guesser* guessing the *target word*. We consider a version of the game with a single *Drawer* and *Guesser*.

2.2.1 Characteristics of Pictionary

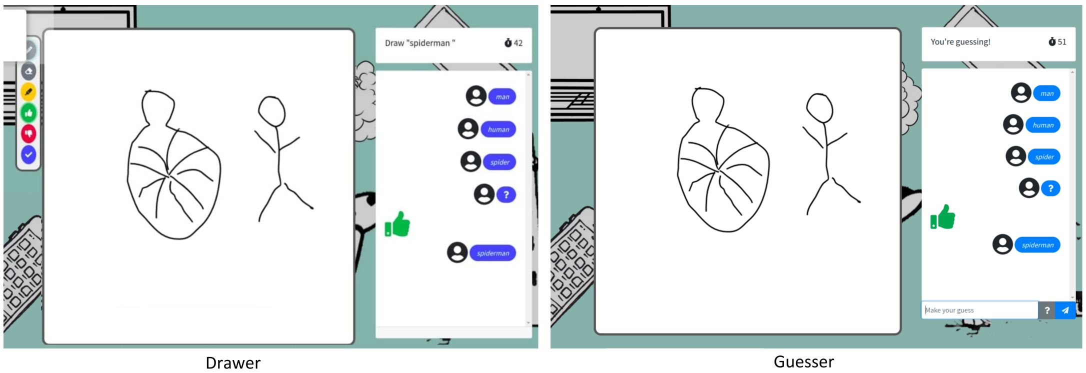
Casual social game: Pictionary is a social game where the players do not require any expertise to play the game. In casual social games, the motivation to play the game is not necessarily just to win in the game [47]. Players may participate to socialize with other players, as a recreation or more commonly for the gaming experience itself. Molinillo et al. [62] describe three types of motivations - hedonic, utilitarian, and relational motivations that players have while playing casual games. They show how hedonic motivation highly influences the players' attitude in a game. These types of games provide an opportunity to study aspects of gaming other than game achievement, such as sociability, immersion, creativity, and likability [2].

Cooperative game with Partial Observabililty: In CPO games, the players do not have perfect information of the game environment. [89, 93]. In Pictionary, the target word is unknown to the *Guesser*. Throughout the gameplay, the players continuously try to guess the other players' state of mind. The players achieve this by interacting with each other using the in-game communication channel. Thus, the interaction between the players constitutes the primary activity of Pictionary gameplay.

Communication in Pictionary: The objective of the game is quick communication under restriction. Since the goal of the game is to convey a message, the performance of the game is a strong indicator of how well players communicate with each other. An interesting feature in this game is that the mode of communication between players is dissimilar - while one player draws, the other player provides guesses. Players bring forth creative ways to express their information to other players. In addition to this, the players share other feedback to express their state of information. Thus, Pictionary provides an opportunity to study the diverse multimodal interaction between players.

2.3 Online Pictionary

Figure 2.2: Screenshot of our online Pictionary app used for Pictionary data collection.



To study the game of Pictionary, we first require a large, diverse dataset of Pictionary game sessions with actual human players. However, large-scale data collection for a physically embodied game entails several technological and logistical barriers. Therefore, we designed an online browser-based Pictionary game playing app. We designed the app to make the game-playing experience closely mimic the real-world counterpart. Our main focus was to facilitate and capture all the possible interactions between the *Drawer* and the *Guesser*. A screenshot of our web app is shown in Figure 2.2.

2.3.1 Data collection process

Our browser-based system is compatible with mouse and touch inputs, scalable and can handle up to 50 multiple concurrent Pictionary sessions. Informed consent is obtained and game instructions are provided when a player accesses the system for the first time. Players are assigned random names and paired randomly as *Drawers* and *Guessers*. Since the participants' information is anonymized, the players did not know who they were playing with. The targets provided to the *Drawers* are sampled from a dictionary of 200 target words. We emphasize that the target word dictionary consists of 138 nouns (e.g., airplane, bee, chair), 51 verbs (e.g., catch, call, hang), and 11 adjectives (e.g., happy, lazy, scary). To ensure uniform coverage across the dictionary, the probability of a target word being selected for a session is inversely proportional to the number of times it has been selected for elapsed sessions. The game has a time limit of 120 seconds. The game ends when the *Guesser* enters a word deemed correct by the *Drawer* or when the time limit is reached.

For the *Guesser*, a text box is provided for entering guess phrases. For the *Drawer*, the interface provides a canvas with tools to draw, erase and highlight locations (via a time-decaying spatial animation 'ping') for emphasis. In addition, thumbs up ( ) and thumbs down ( ) buttons enable *Drawer* to provide 'hot/cold' feedback on guesses. A question (?) button is provided to the *Guesser* for conveying that the canvas contents are not informative and confusing. The canvas strokes are timestamped and stored in Scalable Vector Graphic (SVG) format for efficiency. In addition to canvas strokes (drawing and erasure related), guesses and secondary feedback activities mentioned previously (👎 ,👍 , ❓, highlight) are also recorded with timestamps as part of the game session.

Participants:

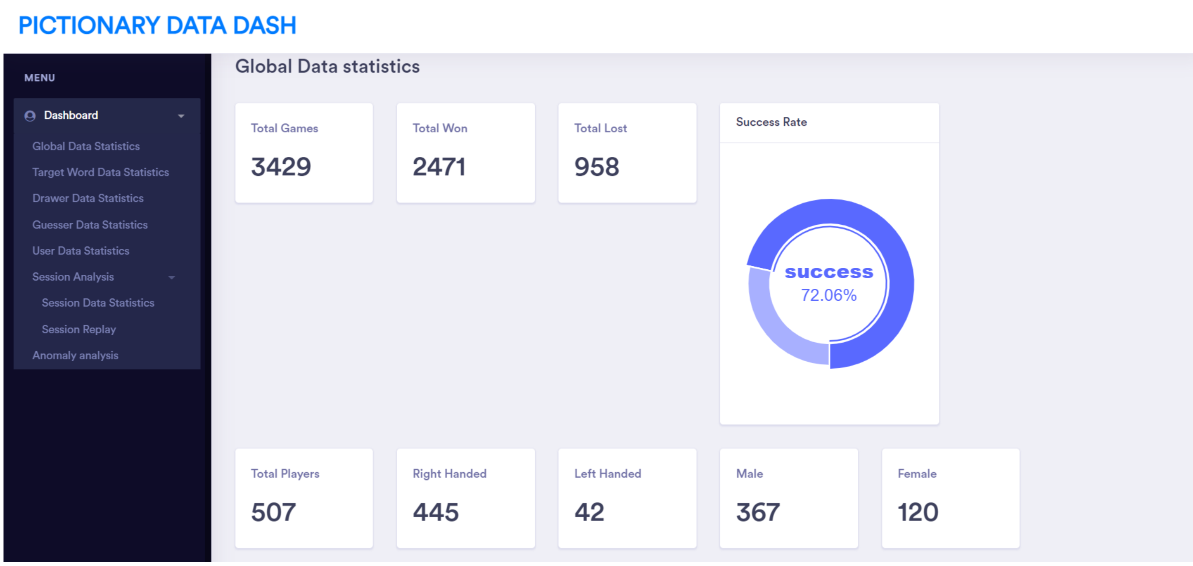
Using our online Pictionary game, we collected data of 3220 game sessions from 479 participants of diverse age groups (19 to 60 years, mean: 32.02, SD: 13.07), gender (328 male and 98 female), handedness (388 right-handed and 38 left-handed) and educational demographics (middle and high school students, graduate and undergraduate university students and working professionals). We chose participants with English as the primary language. Participants were not selected with any specific drawing skill.

Chapter 3

Exploration of Pictionary Game Data

3.1 Pictionary Dashboard

Figure 3.1: Screenshot of the Pictionary Dashboard.



To visualize and analyze the collected Pictionary Dataset, we built a browser based dashboard that presents different statistics of the collected Pictionary data. Figure 3.1 shows the screenshot of the Pictionary Dashboard. The dashboard visualizes the statistics of each component of the Pictionary dataset using interactive plots. The dashboard also has a feature to replay any specific game from the Pictionary dataset.

To thoroughly explore the Pictionary game data, we listed out the intuitive questions that can be answered regarding the game from the collected data. Though the list is not comprehensive, we analyzed the characteristics that are significant for understanding the Pictionary gameplay and the playing styles exhibited by the participants. The dashboard attempts to present the statistics for these questions in an organized, systematic way. We categorize these questions into

1. Global session related statistics
   - How much interaction takes place in a Pictionary game session?
   - Is the quantity of interaction similar for *Drawer* and *Guesser* roles?
   - How is the temporal distribution of the interaction in the game sessions?
   - How much time does it take for the *Drawer* and *Guesser* to start interacting?
2. Target word related statistics
   - Is the statistics of interactions in a game session correlated to the target word?
   - Does the target word influence the game outcome or game success rate?
3. User related statistics
   - What are the role-specific attributes of each player?
   - How frequently do players interact as *Drawer* or *Guesser*?

Figure 3.2: Histogram of the number of interactions in a game session.

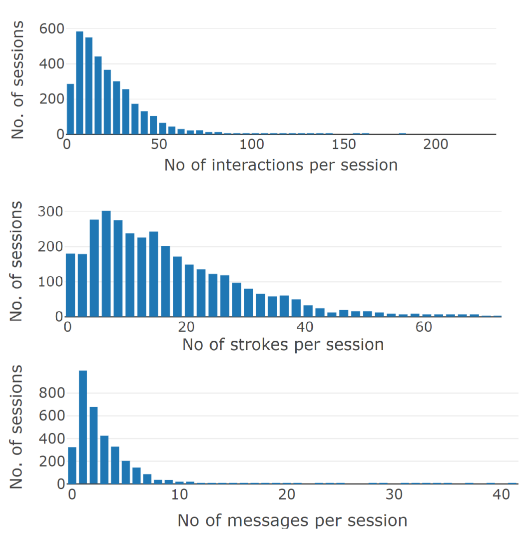


Figure 3.3: Histogram of time of interactions in a game session (in seconds).

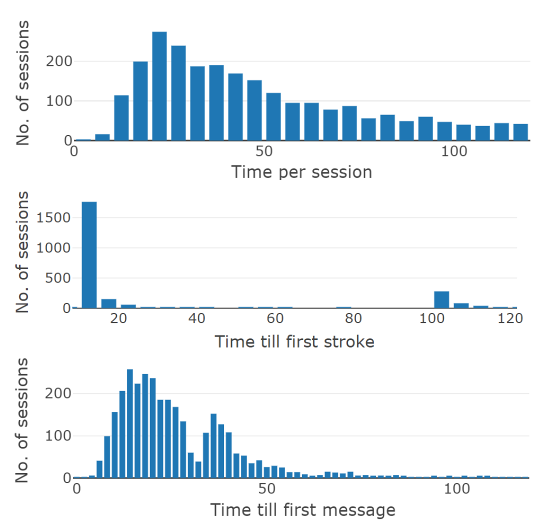
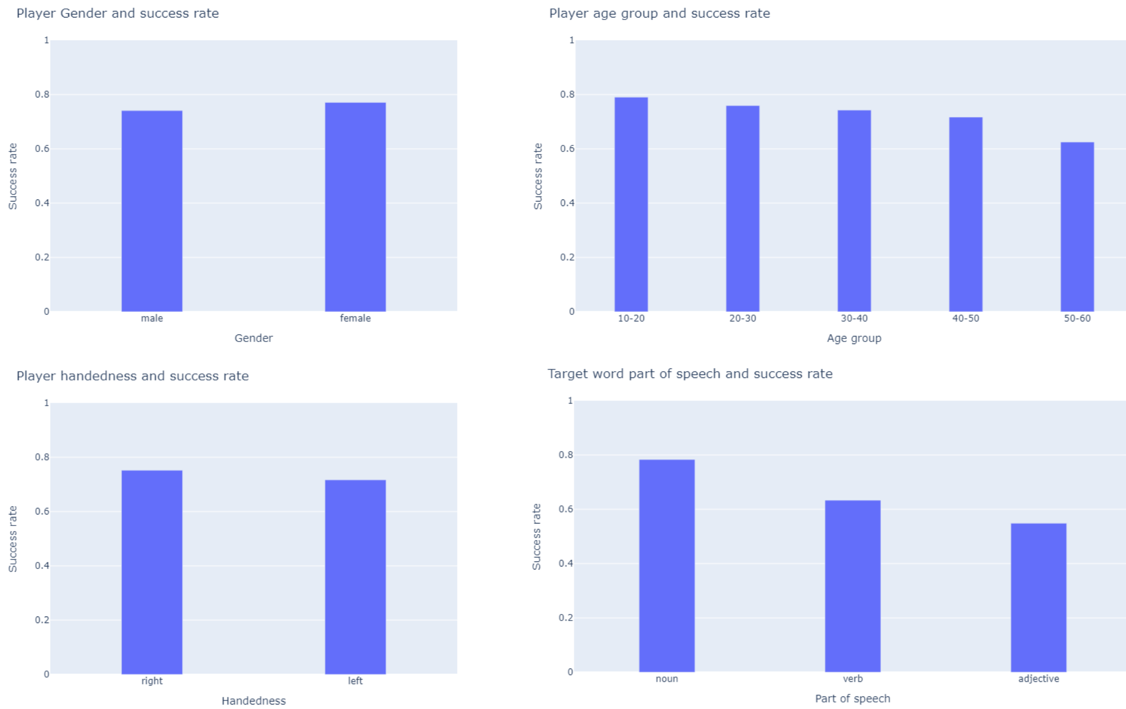


Figure 3.4: Game outcome statistics.



3.1.1 Global session related statistics

A game session in Pictionary consists of an asynchronous time series of interactions between *Drawer* and *Guesser*. The actions from *Drawer* and *Guesser* can be primary actions, such as sketching and guessing, respectively, or secondary actions, such as providing feedback using the icon menus in the game app. The distributions of these interactions in a session have a median value of 18 exchanges (both *Drawer* and *Guesser*) with a median absolute deviation of 10. (See Figure 3.2). Comparing the distribution of the strokes by the *Drawer* (median 14 ± 8) and the messages (median 2 ± 1) from the *Guesser*, we can see that the deviation is mainly contributed by the *Drawer* action and the number of guesses in most games remain small. This is an interesting observation of how the directionality of communications (i.e., from *Drawer* to *Guesser* and *Guesser* to *Drawer*) differs in intensity and hence require separate analysis [1].

In the temporal domain, the length of the game is defined by the duration of a game session in seconds. Figure 3.3 shows the histogram of game duration from the start of the game to the time of correct guess for successful games. We can see the duration is distributed over the entire time limit of 120 seconds. Another interesting attribute is the time of the first interaction for each player. For the *Drawer*, we can see the time of the first stroke is significantly biased to the beginning of the game while for the *Guesser* the distribution has a relatively higher deviation (See Figure 3.3).

Figure 3.5: Target word statistics.

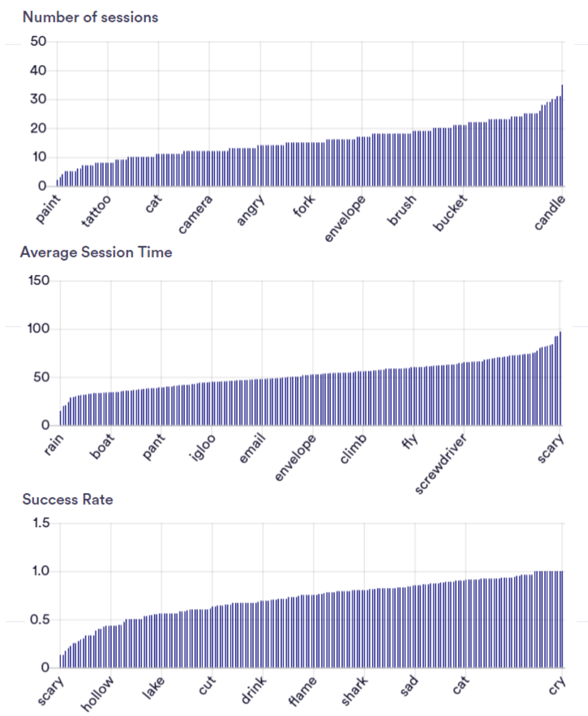
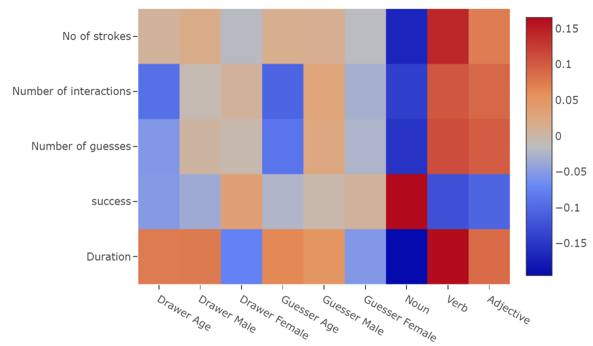


Figure 3.6: Correlation between various attributes of the dataset.



The outcome of any game session is either a success or failure. Figure 3.4 shows how the game outcome varies with the player age group, gender, handedness, and the parts of speech of the target word. It can be seen that the game outcome is biased by these attributes, specifically by the parts of speech of the target word.

3.1.2 Target word related statistics

In our Pictionary dataset, we manually selected a set of 200 target words of different difficulty levels. The set of target words includes 138 nouns, 51 verbs, and 11 adjectives. In the game, the target words are randomly sampled from this predefined set of words. However, the *Drawer* is given three chances to skip the target word. Hence a non-uniform distribution of target words over the dataset is seen (Figure 3.5). In addition, we can find a non-uniform distribution of game success rate and game duration which shows the influence of target word on the gameplay. Figure 3.6 shows a significant correlation between various attributes of target word part of speech and game attributes. We find the correlations for noun target words complementary to the correlation of both verb and adjective target words. This can be expected as the noun target word are easier to sketch compared to verb and adjective target words. To understand this relationship, we perform more experiments on target word in Section 4.2.3.

Figure 3.7: Histogram of Drawing time and Idle time of Drawers.

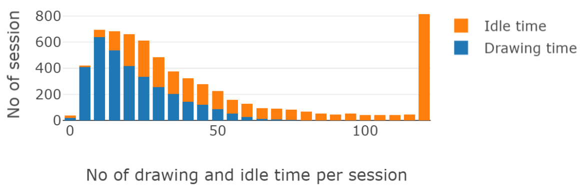
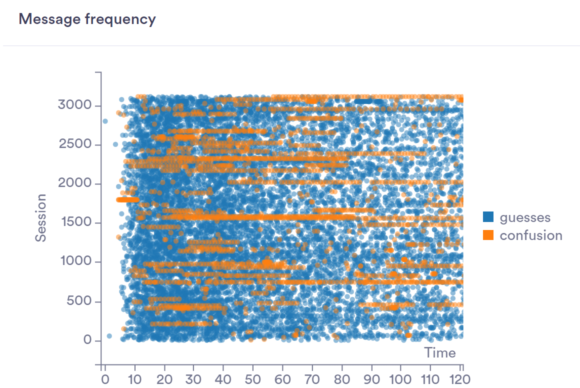


Figure 3.8: Distribution of the Guesser actions over time.



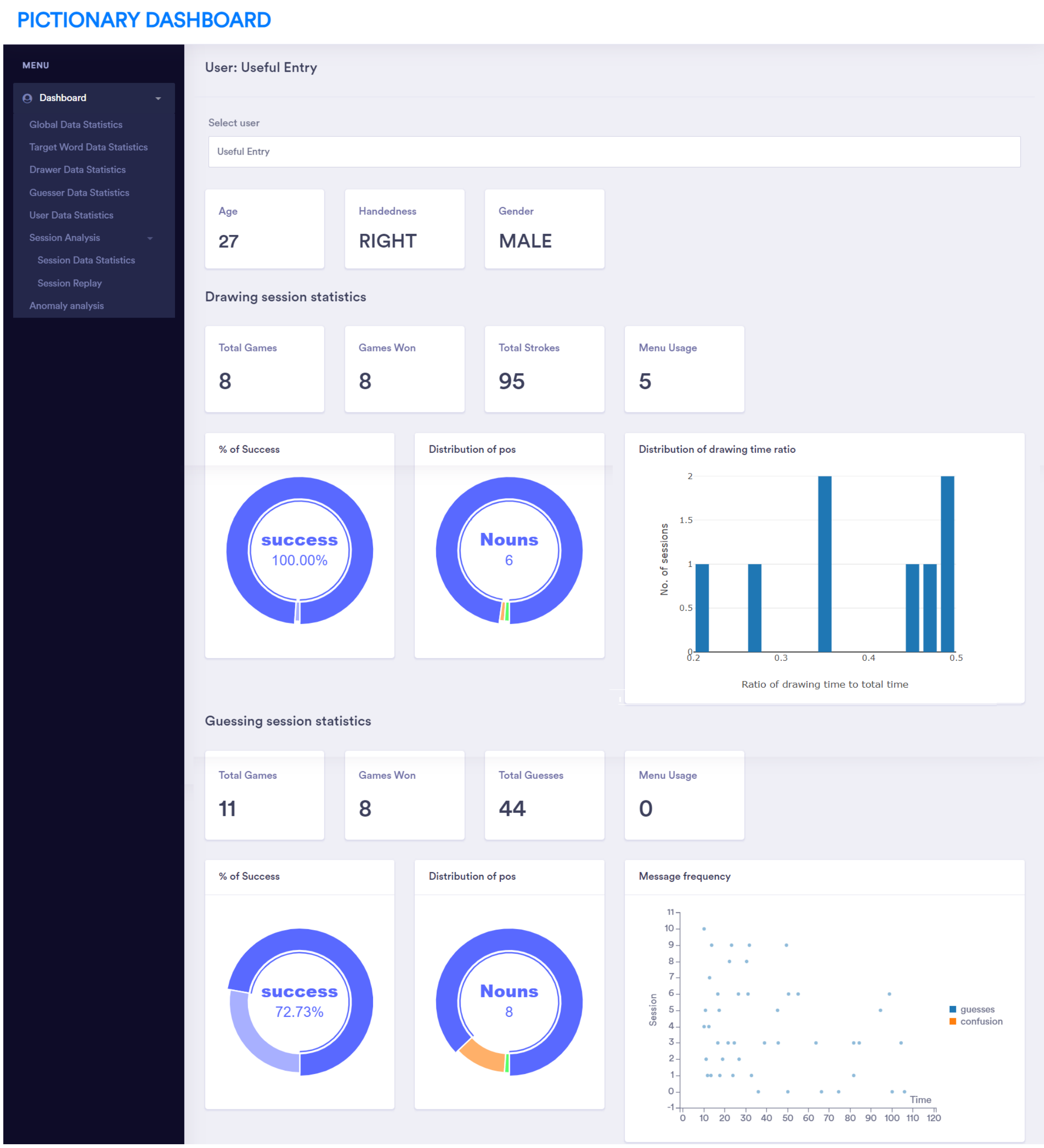
3.1.3 User related statistics

For each player, we collected data on their age (19 to 60 years mean: 32.02, SD: 13.07), gender (328 male and 98 female), and handedness (388 right-handed and 38 left-handed). Figure 3.6 shows the correlation between various attributes of player and game attributes. We find that only a small correlation exists between these attributes. Our dashboard also allows us to examine the activity of each player individually using their username. Since the actions of the players are role-specific, for each player, we need to examine their activity as a *Drawer* and a *Guesser* separately. Figure 3.9 shows a screenshot of the dashboard presenting the various statistics of the selected player. In addition to the above statistics, we also examine the relationship between the duration of drawing and the idle time of the *Drawer* during the game sessions (see Figure 3.7). The presence of a peak in the histogram with a large idle time may be indicative of a *Drawer* behavioural style. We also try to visualize the temporal distribution of the *Guesser* actions using Figure 3.8. We find that more guessing happens earlier in the game while the confusion action is distributed throughout the game duration.

3.2 Discussion

In this chapter, we provide a statistical analysis of the collected Pictionary data and present a customized dashboard to display the statistical information. However, it is important to note that this analysis is only preliminary in nature and lacks a thorough investigation of statistical rigor. To explore specific aspects of the game, we perform a more comprehensive analysis in the following chapter.

Figure 3.9: Screenshot of the Player Dashboard.



Chapter 4

Characterizing Interactions in Pictionary Game

In this chapter, we delve deeper by exploring the aspects of playing styles in an online version of Pictionary and associations between player actions and the game outcome using a data-driven approach. We use a large corpus of game telemetry data recorded from an online version of Pictionary game for our analysis. We identify attributes of player actions for each role and the interplay between them. We use Principal Component Analysis (PCA) to determine role-specific playing style components. We then investigate the impact of game difficulty on game attributes and playing style components. Finally, we discuss the relationship between player behaviour and game outcome. Although prior works have investigated player behavior and performance in cooperative gameplay [13, 3, 79, 72], these studies do not explore player interactions in a setting with restricted communication. We aim to address this research gap.

In this chapter, we investigate the following research questions:

RQ1: What are the different types of playing style components observed in a Cooperative Partially Observable (CPO) game such as Pictionary?

RQ2: Does the game's difficulty impact the playing style components and interplay attributes?

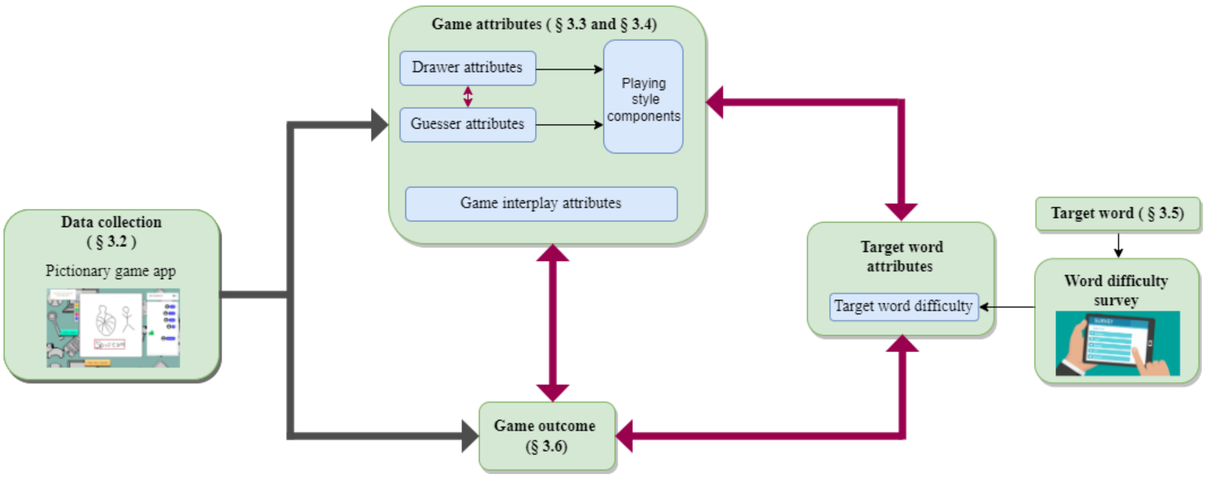
RQ3: Which playing style components and interplay attributes are associated with successful game outcomes?

A supplementary video summarizing the content of this chapter with animated examples of game sessions is provided here.

4.1 Related Work

Characterizing Communication in Games: Analysis of multiplayer games is a widely researched area involving study of communication between players with both cooperative [85, 1, 11] and competitive [46, 22, 40] goals. Emmerich et al. [28] studied the influence of various game design patterns on player interactions. They found that high player interdependence implies more communication. Toups

Figure 4.1: An illustration of Pictionary game analysis components. ↔ indicates that we perform statistical analysis between the linked components.



et al. [83] analyzed 40 digital games using a theoretical approach to develop a framework for cooperative communication game mechanics. They observed that players had lesser ability to request information from their partners and might develop emergent cooperative communication mechanics to compensate insufficiency in existing game design. Thus, cooperative communication mechanics is not generalizable to all game designs since it is dependent on the modality of communication, player motivation and player skills. Different modalities such as text messages, audio and video communication, and game specific activity such as sharing customized icons, cards, and location are commonly used in games. Spyridonis et al. [82] compared three communication modalities, namely, voice, textual chat, and pre-determined commands to propose an efficient communication model for games. They concluded that pre-existing commands lead to faster completion times compared to other modalities, such as textual chat. Communication in Pictionary happens through three modalities, namely, sketches on canvas, text for guesses, and pre-existing commands such as icons 👎 and 👍. The modalities are role-specific and hence, differ with direction of communication (i.e., *Drawer* interacts by sketching and icons and *Guesser* by text and icons). Ashktorab et al. [1] found that in *Wordgame*, the directionality of communication (as from *'Guesser'* or from *'Giver'*) caused significant differences in gameplay outcome and social perceptions of players. Our analysis also explores playing style components of each role (*Drawer* and *Guesser*) separately.

Playing Styles and Gameplay Analysis: Existing works have extensively researched on playing styles along numerous bases such as psychographic, behavioral, and in-game demographics for various games [35, 84]. One of the earliest works on playing styles proposed by Bartle et al. [10] classified players as *Killers* and *Achievers* or *Socializers* and *Explorers*. Though the study done by Bartle et al. [10] was based on Multi-User Dungeons games, this categorization was commonly adapted for many massively multiplayer online games [14, 54]. Loria et al. [56] classified player behavior into more elaborate types such as *competitive*, *striving*, *active*, *committed*, *purpose-driven*, amongst others. However, these player classifications do not examine the cooperative aspects of gameplay. Ferro et al. [30] classified pre-existing player models into five types, namely *Dominant*, *Objectivists*, *Humanists*, *Inquisitive*, and *Creative* based on personality traits and game mechanics. Although we focus only on cooperative game mechanics, our distinctive playing style components are comparable to only few of these player types. Kirman et al. [45] classified players based on social interactions in a game. Unlike Kirman et al. [45], our analysis takes into account the different styles of interaction between players.

Player types are identified either by user surveys in which the players report [2] or observational classification of the recorded gameplay [38]. Such analyses are not scalable and require manual intervention [35]. To overcome this, we use a data-driven approach on the game telemetry data to identify inherent laying style components in a cooperative game. The commonly used data-driven approaches for gameplay analysis include clustering algorithms [22, 40, 67], Hidden Markov Models [59], Archetypal Analysis [57] and Principal Component Analysis (PCA) [58, 48, 17, 70]. In our analysis, we use PCA to identify role-specific playing style components in Pictionary.

4.2 Methodology

The data collection for our game is done using the online browser based app described in Section 2.3. In Section 4.2.1, we introduce attributes of the game. We introduce the statistical analysis for identifying playing style components in Section 4.2.2 from game attribute data. In Section 4.2.3, we introduce attributes of game target phrase and methods to study target phrase's influence on game outcome. In Section 4.2.4, we describe our approach for analyzing the relationship of game outcome with the other attributes of our game. Figure 4.1 shows the outline of our methodology.

4.2.1 Game attributes

To characterize player behavior and interactive gameplay, we selected attributes that describe the player's actions and game interactions (see Table 4.1). The attributes quantify the degree of interaction and feedback provided. We broadly defined attributes based on player-role (i.e., *Drawer*, *Guesser*) and interactive aspects of the game (i.e., *Drawer-Guesser Interplay*).

We selected five *Drawer* attributes to describe the primary action (i.e., sketching) - see Table 4.1. Among the primary action attributes, `canvas area` is used as a measure of the canvas space utilized by the *Drawer*. The temporal aspects of the sketching activity are measured in terms of the time related attributes (e.g., instants of first and last sketch stroke, stroke frequency). We also selected attributes of secondary actions such as erase, highlight, thumbs up (👍), and thumbs down (👎). These actions are relatively rare. Hence, we record the percentage of games containing these actions for each *Drawer*.

Table 4.1: List of game attributes describing the player actions in Pictionary.

	Action	Attribute	Abbreviation
Drawer	Sketch	Area of canvas used	canvas area
		Time of first stroke	first stroketime
		Time of last stroke	last stroketime
		Time between first and last stroke	stroke duration
		Frequency of strokes	stroke freq
	Erase and highlight	Percentage of games with erase and highlight	erase highlight usage
	Thumbs up	Percentage of games with thumbs up	thumbs up usage
	Thumbs down	Percentage of games with thumbs down	thumbs down usage
Guesser	Guess	No of guesses	guess count
		Entropy of guesses	guess entropy
		Time of first guess	first guesstime
		Time of last guess	last guesstime
		Time between first and last guess	guess duration
	Confusion	Percentage of games with confusion	confusion usage
Drawer-Guesser Interplay	Interactions	No of UI interactions	feedback-laden
		Largest time of contiguous events of *Drawer*/*Guesser*	one-sided play

For the *Guesser*, we selected six *Guesser* attributes to describe the primary action (i.e., guessing) and secondary action (i.e., confusion ?) - see Table 4.1. We record the number of guesses in a session. Similar to the *Drawer* temporal attributes, we record the time of first guess, time of last guess, and duration of guessing. In addition to this, we calculate the entropy of the guess distribution by considering each guess as an impulse on the timeline of the game session. Since the secondary action (i.e., confusion ?) is present sparingly, the percentage of games where confusion is indicated by the *Guesser* is recorded.

For a player in *Drawer*'s role, we consider all sessions of the player in this role. For each *Drawer* attribute, the median attribute value across the sessions is considered the aggregated attribute score. A similar procedure is used to obtain the player's *Guesser* aggregated attribute scores by considering all the sessions with the player in *Guesser*'s role. The attribute scores for all the players are analyzed together to identify intra/inter-role associations and interplay dynamics between the *Drawer* and *Guesser*.

4.2.2 Characterization of Playing style components

To identify role-specific playing style components in a data-driven fashion, we perform Principal Component Analysis (PCA) on the corresponding *Drawer* and *Guesser* attributes separately. To estimate the number of relevant components, we use parallel analysis [31]. Subsequently, we examine the component loadings to label the playing style components. To examine the associations between the role-specific playing style components, we perform correlation between the *Drawer* and *Guesser* playing style scores obtained via the aforementioned PCA procedure.

4.2.3 Target word

Games that are either too difficult or too simple have a greater churn risk[49]. Hence, game difficulty is an important parameter in game studies[92]. The difficulty of a Pictionary game is primarily determined by attributes of target word, such as parts of speech. For instance, a simple noun such as 'tree' or 'boat' can be easily sketched. Verbs or adjective target words such as 'listen' or 'lazy' may require a more complex illustration. To quantify target word difficulty, we conduct a survey. Subsequently, we analyze the influence of target word difficulty on game attributes, playing style components and game outcome.

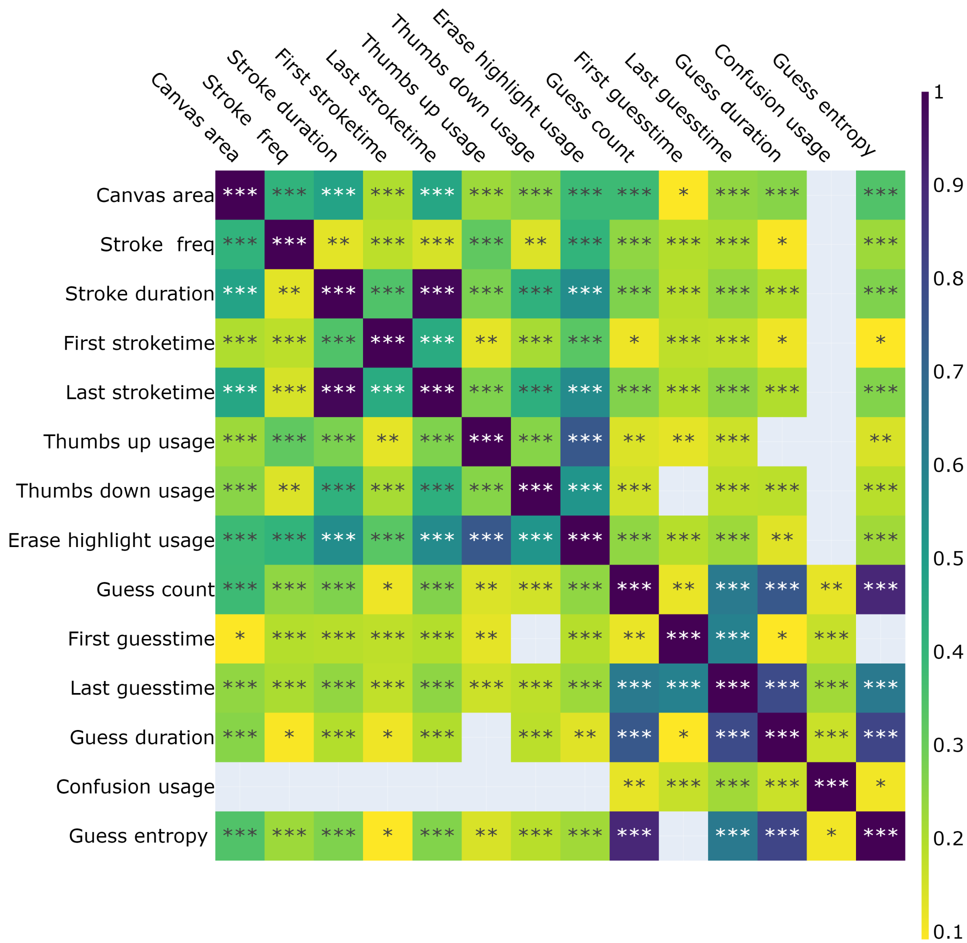
4.2.3.1 Target word difficulty:

40 participants (age: mean=21.83, SD=2.34, 17 females) rated the perceived sketching difficulty of the 200 target words. Each participant rated a subset (100 words) on a 5-point Likert scale ranging from 'very easy' to 'very difficult'. First, we assess the internal consistency of the acquired ratings by using Cronbach's alpha as a reliability measure. This measure was calculated on both subsets separately. Then, we use the median value of the resulting 20 ratings for each word as a measure of target word difficulty. All target words with a median rating of three and above were classified as 'difficult' while the rest were classified as 'easy'.

4.2.3.2 Influence of Target word:

We examine the influence of target word difficulty on game attributes and playing style components. To do this, we first divide the game sessions per player into two categories (i.e., 'easy' and 'difficult') based on target word difficulty. We then recalculate game attributes (Section 4.2.1) for each category separately. For difficult words, we expect an increase in amount of interaction and feedback, as well as a delayed response by both *Drawer* and *Guesser*, reflected by an increase in all the game attributes. We use Wilcoxon signed-rank test to examine the differences in game attributes based on target word difficulty level ('easy' and 'difficult'). Owing to multiple statistical testing, we perform Benjamini-Hochberg procedure to control for the false discovery rate.

Figure 4.2: Spearman's correlation between player attributes. Note: *p<.05, **p<.01, ***p<.001



Next, we obtain the playing style components for each category ('easy' and 'difficult') by performing category-wise PCA (as described in Section 4.2.2). We then examine the category-wise component loadings to determine if the playing style components are influenced by word difficulty. Finally, we perform paired t-test between the difficulty levels ('easy' and 'difficult') for all the scores of playing style components found with PCA.

4.2.4 Game outcome

The outcome of Pictionary is binary: a success or a failure. We investigate the relationship between game outcome and other aspects discussed above (i.e., playing style components, *Interplay* attributes, and target word difficulty) by performing point bi-serial correlation. Next, we group game sessions into two categories based on game outcome (success and failure). We recalculate *Interplay* attributes for each category separately. We then perform a Mann-Whitney U test to check for differences in *Interplay* attributes grouped by game outcome. Finally, to assess the relationship between the target word and game outcome, we compute the point bi-serial correlation between game outcome and word difficulty. We expect the evidence to support a negative association between the target word difficulty and game outcome.

4.3 Results

As a part of our overall analysis, we study the role of game difficulty, how it affects the game interplay, and the role-specific playing style components. In Section 4.3.1, we report the playing style components exhibited by the *Drawer* and *Guesser*. In Section 4.3.2, we demonstrate the effect of word difficulty on *Drawer*, *Guesser*, and *Interplay* attributes and playing style components. Finally, we demonstrate the association between the game outcome and aforementioned attributes in Section 4.3.3.

Table 4.2: PCA Component Loadings of *Drawer* attributes.

	PC1	PC2	PC3
Time between first and last stroke	**0.58**	0.01	−0.06
Time of first stroke	**0.36**	−0.13	0.15
Time of last stroke	**0.60**	−0.01	−0.04
% games with Thumbs up	−0.11	**0.70**	0.06
% games with Thumbs down	0.25	**0.37**	−0.22
% games with erase and highlight	0.14	**0.57**	0.08
Area of canvas used	0.26	−0.13	**0.57**
Frequency of strokes	−0.12	0.11	**0.77**
Explained Variance	46.30%	15.84%	11.95%
	Paced	Feedback	Intense

4.3.1 Playing style components

As can be seen in Figure 4.2, greater number of significant intra-role correlations are observed than inter-role correlations.

Table 4.3: PCA Component Loadings of *Guesser* attributes.

	PC1	PC2	PC3
No of guesses	**0.54**	-0.07	-0.02
Entropy of guesses	**0.56**	-0.12	-0.03
Time between first and last guess	**0.52**	-0.01	0.03
Time of last guess	**0.35**	**0.48**	0.02
Time of first guess	-0.08	**0.87**	-0.01
% games with confusion	0	0	**1**
Explained Variance	56.37%	20.90%	14.84%
	Intense	Delayed	Feedback

4.3.1.1 Drawer playing style components:

Parallel Analysis on *Drawer* attributes revealed an intrinsic dimensionality of three components. Subsequently, the three components obtained via PCA on the *Drawer* attributes explained a cumulative variance of 74.11%. As can be seen in Table 4.2, the first component ($PC1$), which we refer to as **Paced**, showed a high positive loading for attributes `stroke duration`, `first stroketime` and `last stroketime`. *Drawers* with higher values of **Paced** take longer to start sketching and spend a considerable amount of time drawing (see Figure 4.6(a) and Figure 4.6(b)). The second component ($PC2$), which we refer to as **Feedback**, showed high positive loadings for *Drawer* feedback attributes (`thumbs up usage`, `thumbs down usage`, and `erase highlight usage`). *Drawers* with higher values of **Feedback** use a significant amount of feedback elements on the GUI (see Figure 4.6(c) and Figure 4.6(d)). The third component ($PC3$), labelled as **Intense**, shows a high positive loading for `stroke freq` and a moderate positive loading for `canvas area`. *Drawers* with higher values of **Intense** draw fast and also use a large area to sketch (see Figure 4.6(e) and Figure 4.6(f)).

4.3.1.2 Guesser playing style components:

Parallel Analysis on *Guesser* attributes revealed three components. Subsequently, the three components obtained via PCA on the *Guesser* attributes explained a cumulative variance of 92.13% (Table 4.3). The first component ($PC1$), labeled as **Intense**, has high positive loadings for `guess count`, `guess entropy`, `guess duration`. *Guessers* with higher values of **Intense** make a lot of guesses over a significant time duration (see Figure 4.7(a) and Figure 4.7(b)). The second component ($PC2$), referred as **Delayed**, shows *Guesser* activity marked by high loading for the `first guesstime` and `last`

Table 4.4: Correlation between playing style components.

Playing style components		Guesser Intense	Guesser Delayed	Guesser Feedback
Drawer	Paced	0.325^{***}	0.252^{***}	-0.033
	Feedback	0.230^{***}	0.164^{***}	-0.034
	Intense	0.286^{***}	0.170^{***}	-0.059

Note: *p<.05, **p<.01, ***p<.001

`guesstime`. *Guessers* with higher values of **Delayed** wait for some time before guessing or guess only when they are confident (see Figure 4.7(c) and Figure 4.7(d)). The third component ($PC3$), labelled as **Feedback** has a high positive loading for `confusion usage`. *Guessers* with higher values of **Feedback** frequently use the confusion GUI button (see Figure 4.7(e) and Figure 4.7(f)).

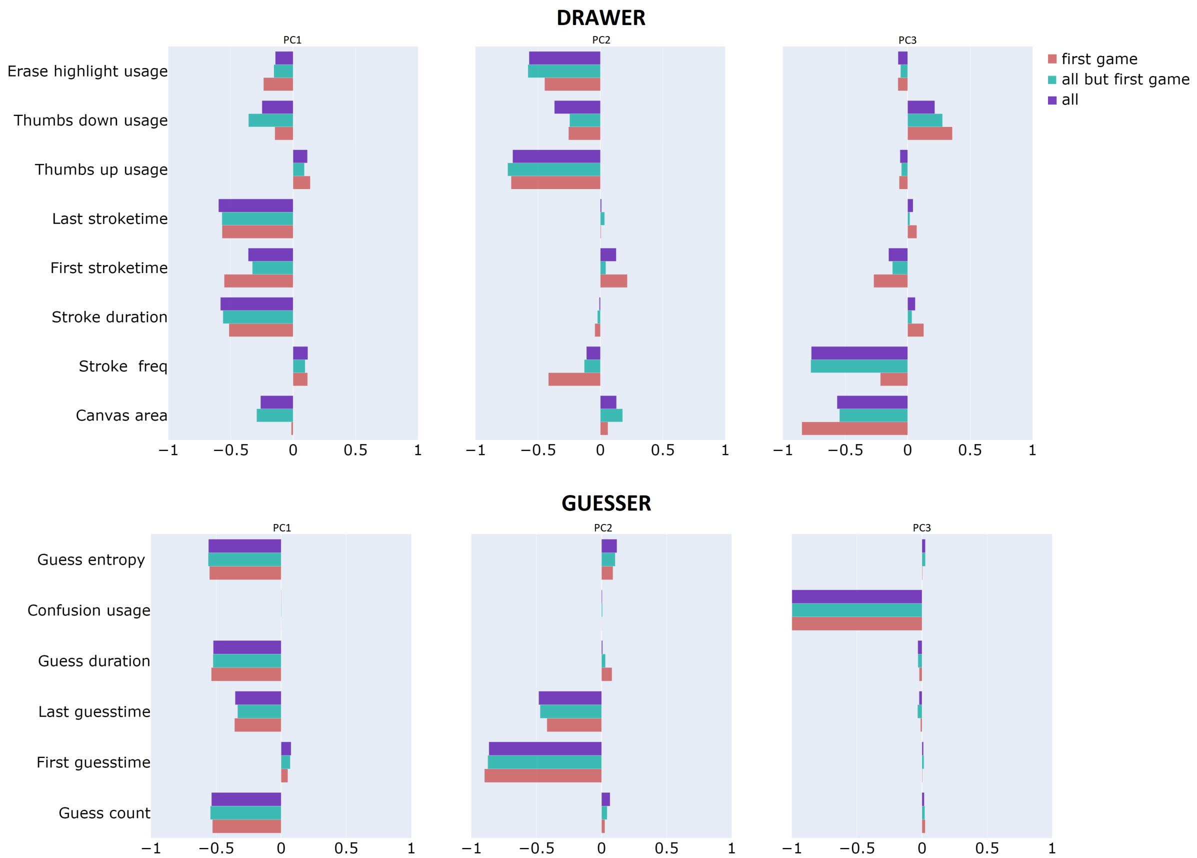
Spearman's correlation between *Drawer* and *Guesser* playing style components can be seen in Table 4.4. Significant positive correlations were observed between **Paced**, **Feedback** and **Intense** drawing style components and **Delayed** and **Intense** guessing style components. To ensure the playing style components are not affected by players playing for the first time, we repeated PCA analysis after removing the first game of each player (see Figure 4.3). We found that PCA loadings were near identical. This suggests that these playing style components are unaffected by the player's familiarity with the game app.

4.3.2 Target word difficulty

High reliability was observed for the ratings of both the subsets as evidenced by Cronbach's alpha (0.976 and 0.980, respectively). Out of 200 target words, 124 were rated as 'easy' and 76 as 'difficult.' 77% of *Drawers* and 73% of *Guessers* played games with both 'easy' and 'difficult' words. We used this subset of players ($n = 326$) for our analysis. As can be seen in Figure 4.4, for 'easy' and 'difficult' word categories, Wilcoxon signed rank test revealed significant differences for several *Drawer* attributes (as seen in Table 4.5). Games that had 'difficult' target words were characterized by higher values for all *Drawer* attributes except the frequency of strokes and usage of thumbs up. While no significant differences were observed for usage of thumbs up, reduced frequency of strokes was observed for 'difficult' target words. For the *Guesser*, Wilcoxon signed rank test revealed significant differences in all attributes (as seen in Figure 4.4 and Table 4.5). Games that had 'difficult' target words were characterized by higher values for all *Guesser* attributes.

PCA on *Drawer* and *Guesser* attributes separately for games with 'easy' and 'difficult' words revealed highly similar component loadings (Figure 4.5). This suggests no influence of target word difficulty on playing style components. Paired t-tests revealed no significant differences (refer Table 4.6)

Figure 4.3: PCA component loadings of *Drawer* and *Guesser* for first games of players and all but first game of players.



between the obtained playing style component scores at varying levels of word difficulty (i.e., 'easy' and 'difficult').

Spearman's correlation between target word difficulty and *Interplay* attributes revealed a positive correlation for both attributes, `feedback-laden` (r = 0.41, p<0.001) and `one-sided` (r = 0.47, p<0.001) interplay. Thus, we observed an increased usage of feedback and reduced activity from one of the players when faced with a 'difficult' target word.

4.3.3 Game Outcome

Point biserial correlation between drawing style components and game outcomes revealed a significant negative correlation for **Paced** drawing style component (r = -0.37, p<0.001), suggesting that slow and prolonged sketching is associated with greater incidence of unsuccessful games. On the other hand, **Intense** drawing style component was positively correlated (r = 0.09, p<0.001), albeit a small effect, with game outcome, suggesting that the games with higher frequency of sketching are mildly associated with success. Within guessing style components, **Intense** (r = -0.19, p<0.001) and **Delayed** (r =

Table 4.5: Wilcoxon signed rank test for each *Drawer* and *Guesser* attributes for 'easy' and 'difficult' word categories ($n = 326$).

Drawer Attributes	W	p
Canvas area	12954	8.82e-16
First stroketime	11948	5 .90e-18
Last stroketime	8066	1.00e-27
Stroke duration	9100	6.67e-25
Stroke frequency	21322	1.75e-03
Thumbs up usage	16265	4.99e-02
Thumbs down usage	9210	2.41e-06
Erase highlight usage	8487	4.37e-04

Guesser Attributes	W	p
Guess count	8328	2.58e-20
First guesstime	12554	1.26e-16
Last guesstime	7902	3.47e-28
Guess duration	7801	5.65e-25
Guess entropy	7868	3.38e-20
Confusion usage	1200	5.93e-2

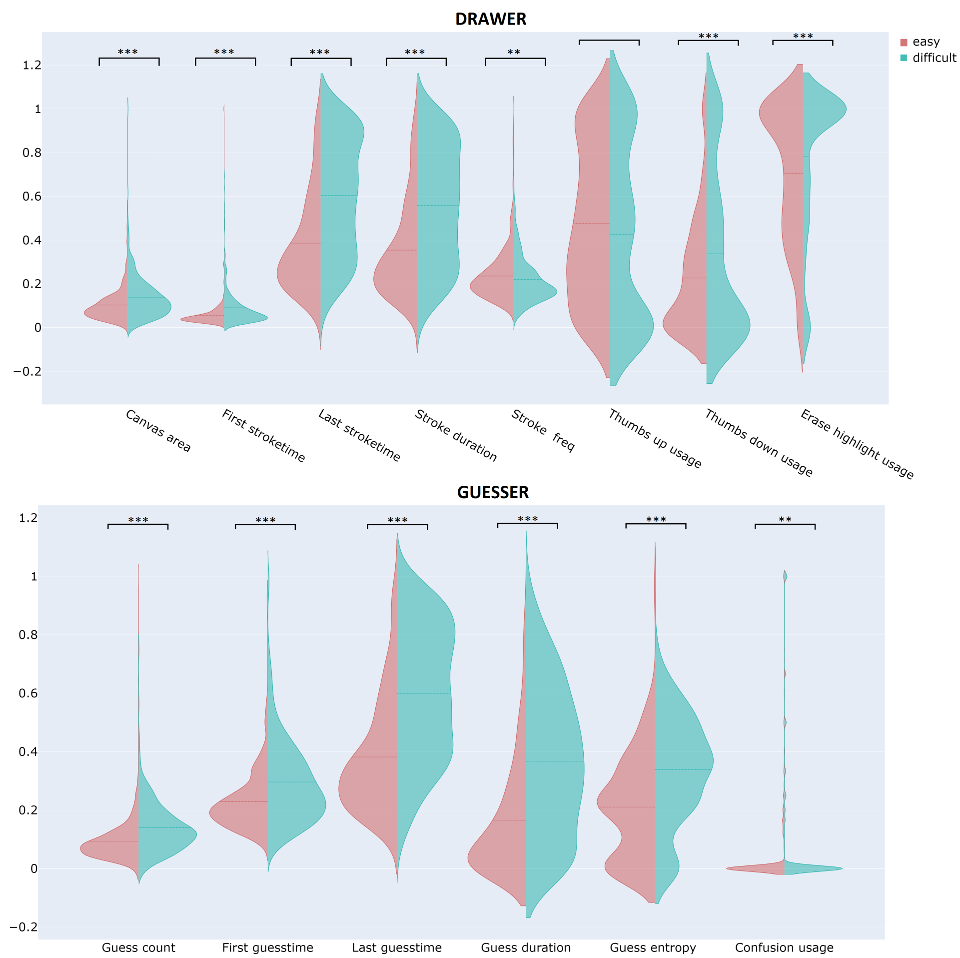
-0.19, p<0.001) were negatively correlated with game success. A negligible correlation was also found for **Feedback** guessing style component (*r* = -0.07, p<0.001). This suggests that excessive or belated guessing is associated with unsuccessful game outcome.

Mann-Whitney U test revealed significant differences for *Interplay* attributes grouped by game outcome (`feedback-laden`: $statistic = 990378.5$, $p = 1.14e - 97$ and `one-sided`: $statistic = 1073613.0$, $p = 3.62e{-}142$). Significant negative correlation was observed between *Interplay* attributes and game outcome (`feedback-laden`: *r* = -0.31, p<0.001 and `one-sided`: *r* = -0.46, p<0.001), suggesting that too little or too much of interaction is associated with unsuccessful game outcomes. Finally, as hypothesized, significant negative correlation was observed between word difficulty and game outcome (*r* = -0.59, p<0.001).

4.4 Discussion

Characterizing communication in a partially observable setting gives us interesting insights into how humans interact to convey information. In this thesis, we use Pictionary as a case study to understand communication styles in a shared goal setting. We identified three role-specific playing style components, namely, **Paced**, **Feedback**, **Intense** drawing style components and **Intense**, **Delayed**, **Feedback** guessing style components. Next, we analyzed the impact of target word difficulty on the game interactions. Finally, we explored the relationship between the game outcome and the playing style components and *Interplay* attributes.

Figure 4.4: Distribution of *Drawer* and *Guesser* Attributes for different levels of word difficulty. In the figure, ***p<0.001, **p<0.01, and *p<0.05 where p is the p-value of Wilcoxon signed rank test (Section 4.3.2).



4.4.1 Playing style components and characteristics of interactions in a collaborative game

Though player behavior has been extensively researched in the context of different kinds of games [35], to our best knowledge, we present the first attempt to analyze playing style components in a communication-restricted environment. Our analysis explores the dynamic nature of cooperative interaction. The attributes investigated in our study not only characterize the aggregated properties of the game but also includes the temporal aspects of gameplay. In line with Loria et al. [57], we find that both aggregated and temporal aspects are needed to understand the dynamics of the gameplay system.

Based on player actions obtained from Pictionary, we observe three types of playing style components for each role (*Drawer* and *Guesser*). The player attributes are classified by quantity and speed of interactions (**Intense** drawing and guessing style components), nature of temporal interactions (**Paced** drawing style component and **Late** guessing style component), and the quantity of feedback (**Feedback** drawing and guessing style components). We observe that the playing style components in Pictionary are comparable to the classification of player typology by Ferro et al. [30]. Ferro et al. [30] group the player types and personality traits from literature into five categories, namely, *Dominant*, *Objectivist*, *Humanist*, *Inquisitive*, *Creative*. The strong correlation we found between **Intense** drawing style component and **Intense** guessing style component appears to be suggestive of *Dominant* player type characterized by aggressiveness[30], exhibited in our study as intensive gameplay in terms of both speed and quantity. Furthermore, the playing style components, **Paced** drawing and **Delayed** guessing demonstrate traits of players who do not appear to be rushed despite the temporal constraint. Ferro et al. [30] describe *Objectivist* players as those who build upon their knowledge by focusing on self-directed tasks, which in our case is either to draw or to guess. The **Feedback** drawing and guessing style components involve responding with greater amount of icon-based feedback to the activity of the teammate. This style of playing is similar to *Humanist* player type, who involve themselves in tasks that rely on social engagement to solve problems.

Figure 4.5: PCA component loadings of *Drawer* and *Guesser* roles for different levels of word difficulty.

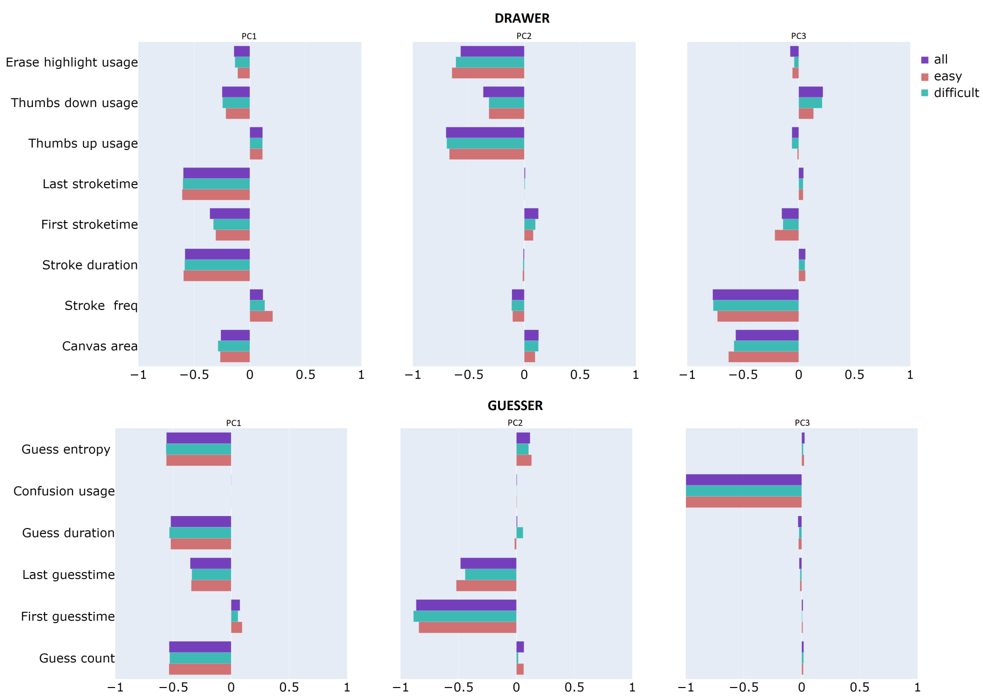


Table 4.6: Paired t-test for playing style component scores between the categories 'easy' and 'difficult'.

Drawer ($n = 326$):

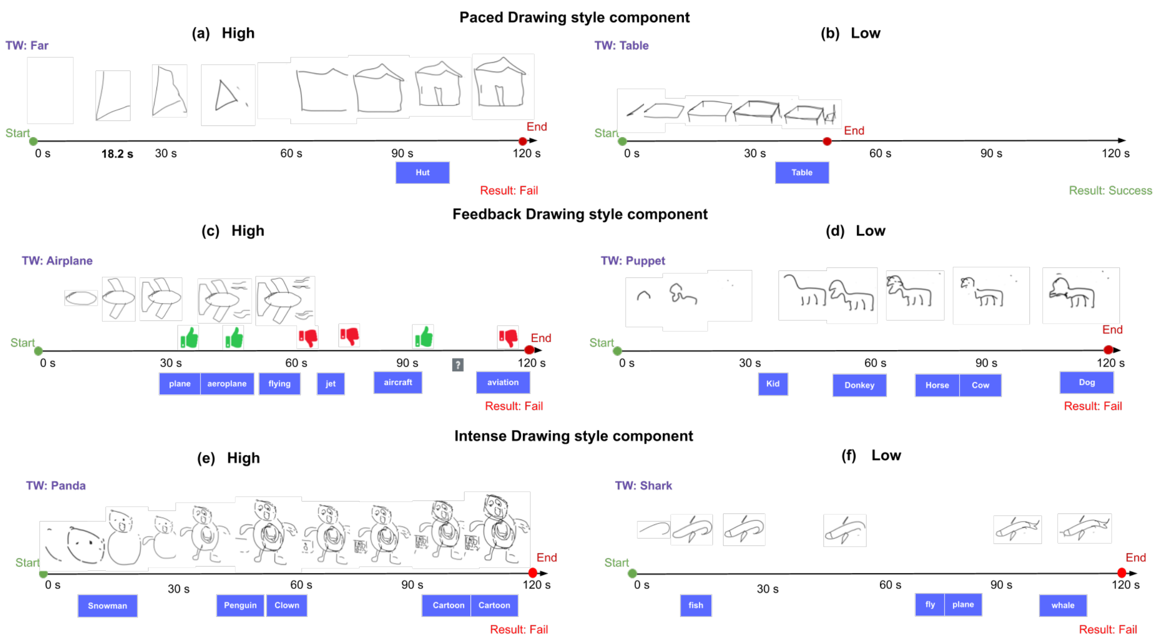
	T	p
Paced	-0.32	0.74
Feedback	-0.03	0.97
Intense	-0.53	0.59

Guesser ($n = 326$):

	T	p
Intense	-0.71	0.48
Delayed	-1.63	0.10
Feedback	-0.28	0.77

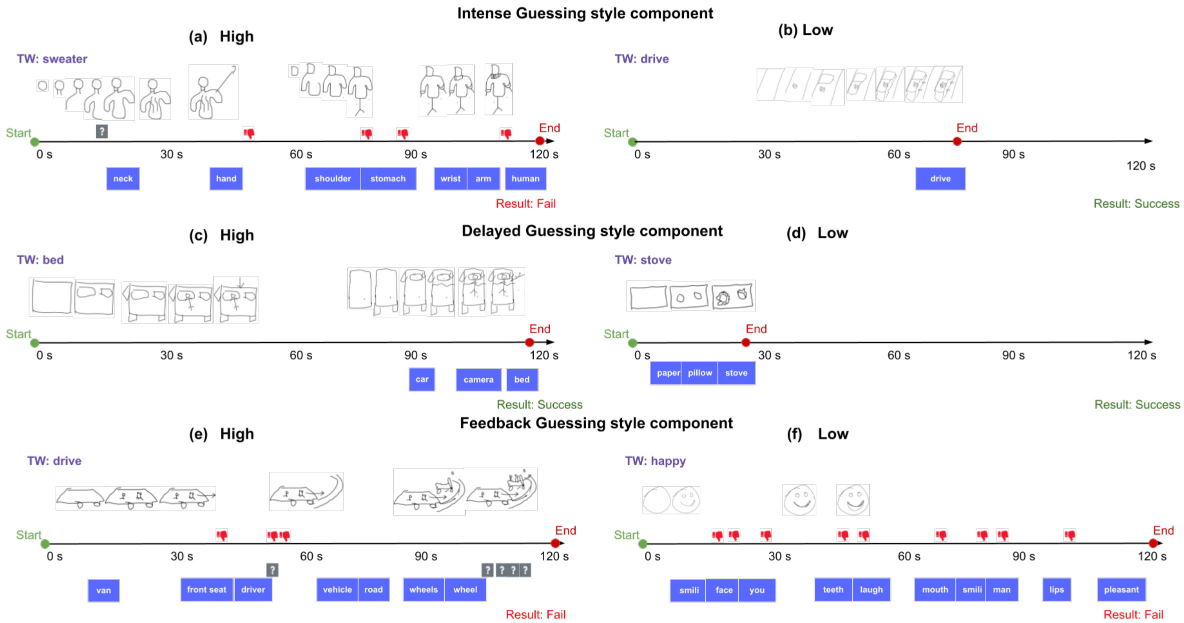
The identified playing style components from our analysis allow us to draw parallels to communication styles in general. Hwang et al. [37] classify communication styles of casino dealers and their influence on player satisfaction. They propose nine communication styles that affect player satisfaction.

Figure 4.6: Example of game with each *Drawer* playing style components.



We find that two of these playing style components, *Relaxed* and *Dominant*, are relevant to playing style components in Pictionary. *Relaxed* communicators tend to express themselves in an unperturbed and calm manner. Such a trait is seen in players with higher scores for **Paced** drawing and **Delayed** guessing style components. Hwang et al. [37] show that *Relaxed* communication style had a positive influence on player satisfaction. Although we do not have player satisfaction data, a potential proxy could be game outcome. We found that both higher **Paced** drawing and **Delayed** guessing style components were associated with more failed games. One possible explanation is the time limit imposed by the game. The *Dominant/Active* communication style that Hwang et al. [37] identify is reflective of players who tend to lead the interaction. This trait is often seen in players with higher scores for **Intense** drawing style component and **Intense** guessing style component. Although Hwang et al. [37] did not find a correlation between player satisfaction and *Dominant/Active* communication style, we find a negative correlation between **Intense** guessing style component and game success. **Intense** guessing style component is characterized by high volume and frequency of text messages leading to high cognitive load for *Drawer* and can be highly distracting. This is further supported by the negative correlation we found between **Feedback** guessing style component and game success. On the other hand, **Intense** drawing style component was positively associated with game success. These results suggest that aggressive drawing paired with early guessing in a controlled fashion may result in successful communication in Pictionary.

Figure 4.7: Example of game with each *Guesser* playing style components.



While analysis of individual playing style components outlines role-specific behavior, the *Interplay* attributes `feedback-laden` and `one-sided` revealed the overarching gameplay dynamics. The negative correlation of `feedback-laden` interplay with game outcome reinforces our earlier discussion that too much feedback can be distracting to players. Furthermore, we observe that `one-sided` interplay is significantly associated with failure in games. This suggests that balanced interactions between roles are vital for successful communication.

4.4.2 Impact of target word difficulty on gameplay

From our analysis, we found that the target word difficulty is associated with significant differences in most player actions. As expected, a difficult target word is associated with an increase in most of the *Drawer* and *Guesser* attributes. This indicates an increase in quantity of interactions and feedback and a delayed response from the players of both roles. Interestingly, we see a small decrease in the frequency of sketching for difficult words. This is in contrast to the expectation that a more difficult word requires more intense drawing.

While Figure 4.4 demonstrates differences in player attributes across word difficulty, the covariance between these attributes, as demonstrated in the loadings (Figure 4.5), remains intact. Furthermore, paired t-tests on playing style components scores between 'easy' and 'difficult' words (Table 4.6) revealed no significant differences ($p>0.05$). This implies that while there exist changes in player attributes, the overall playing approach seems to be constant. For instance, when a player is performing a task involving higher cognitive load (i.e., 'difficult' words) vs. lower cognitive load (i.e., 'easy' words), the activity of the players may differ based on the demands of the task, while the players approach to performing the task might remain stable. For example, a player playing a game with a 'difficult' word might draw for a longer period of time. However, the overall pace of the player (Paced playing style component) does not change. Thus, the playing style components identified in our analysis are inherent to the player and are not influenced by game difficulty.

4.4.3 Implications on Game and AI Agent design

Determining playing style components can help in designing games that appeal to a diverse range of players. Delayed start in communication as observed in playing style components **Paced** drawing, **Delayed** guessing, and intervals of inactivity as observed in `one-sided` interplay have higher churn risk. To address this, we can design a feature in the game which prompts a message to other players indicating that they are thinking or asking them to wait when they are inactive for a brief period. On the other hand, intensive interaction, as seen in playing style components **Intense** drawing and **Intense** guessing, might overwhelm the other player. To prevent this, the systems can incorporate suggestive mechanisms for players to match player speeds.

The analysis from this thesis can help design collaborative AI agents that mimic different styles of human communication[18]. Parameters of the agent's action, similar to our game attributes (refer Table 4.1), can be tuned for the agent to exhibit diverse playing styles. For example, an agent shouldn't start the game at the same time for every game. Furthermore, analysis of target word difficulty provides a guideline for designing agents that adapt their activity to different difficulty levels. For instance, the game logic for pairing players could rely on player types and word difficulty.

Chapter 5

Atypical activity in Pictionary

In Pictionary, the *Drawer* is assigned the task of conveying a given target phrase to a counterpart *Guesser* through the act of sketching on a canvas. The sketches created by the players typically serve as visual representations of the target word or provide hints that are related to the target word. When dealing with complex target words, such as adjectives and verbs, the sketches involve a composite mixture of individual components. For example, (see Figure 5.1) a simple illustration is used for the target word 'fish', whereas the addition of components such as a human or water is used for the target word 'fishing'. A more complex target word, such as 'dive', is represented by a set of components as well as arrows and symbols between these components. Through an exploratory study of our Pictionary sketch data, we find such instances of atypical activity exhibited by the players. This chapter presents the various types of atypical activity commonly observed in Pictionary. Additionally, different baseline models are compared to detect and identify this atypical content effectively.

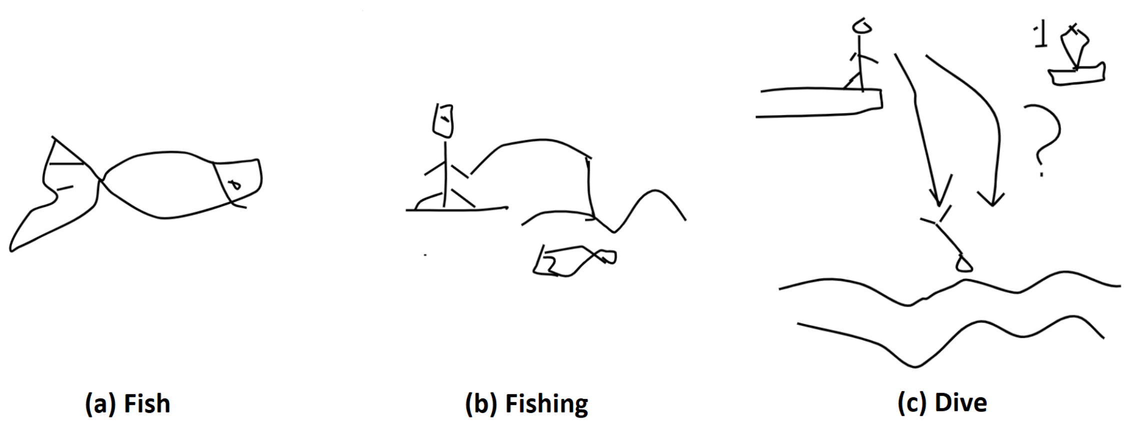


Figure 5.1: Examples of sketch data for target words of varying levels of difficulty

Figure 5.2: Some examples of atypical sketch content in Pictionary game sessions are shown as canvas screenshots. The content instances span text, numbers, question marks, arrows, circles and other icons (e.g. tick marks, addition symbol) categories - refer to Sec 5.1 for details.

5.1 Atypical activity in Pictionary Data

The rules of Pictionary forbid the *Drawer* from writing text on the whiteboard. This is usually not an issue when players are physically co-located. In the anonymized, web-based version of the game, however, the *Drawer* may cheat by writing text related to the target word on the digitally shared whiteboard, thus violating the rules. Apart from rule violation, atypical sketch content can also exist in non-malicious, benign scenarios. For instance, the *Drawer* may choose to draw arrows and other such icons to attract the *Guesser's* attention and provide indirect hints regarding the target word (see Figure 5.2).

An atypical sketch content instance can be thought of as a subsequence of sketch curves relative to the larger sequence of curves that comprise the game session. We first describe the categories of atypical content usually encountered in Pictionary sessions:

- *Text*: *Drawer* directly writes the target word or hints related to the target word on the canvas.
- *Numerical*: *Drawer* writes numbers on canvas.
- *Circles*: *Drawers* often circle a portion of the canvas to emphasize relevant or important content.
- *Iconic*: Other items used for emphasizing content and abstract compositional structures include drawing a question mark, arrow, and other miscellaneous structures (e.g. double-headed arrow, tick marks, addition symbol, cross) and striking out the sketch (which usually implies negation of the sketched item).

Each stroke in the dataset was manually annotated with labels of these four classes or a sketch class. The occurrence statistics of atypical sketch categories across game sessions can be viewed in Table 5.1.

Sketch Content Type class	Number of occurrences	Number of sessions containing	Number of target phrases containing
Text	**2419**	**478**	**180**
Individual letter	2244	460	178
Running hand	175	103	81
Numbers	**331**	**73**	**28**
Circles	**110**	**90**	**67**
Iconic	**750**	**377**	**147**
Arrow	497	292	129
Question mark	158	116	78
Miscellaneous	95	54	37

Table 5.1: Statistics of atypical sketch content categories in game sessions.

5.2 Atypical sketch content detection

Intervention is required to prevent rule violation in games. Manual intervention is impractical and not scalable to an online setting involving a large number of multiple concurrent game sessions. Providing user interface options for player-triggered flagging of rule violation is another possibility. But such mechanisms are not completely reliable since the *Guesser* benefits from the content written on the canvas and does not have real incentive to use the flagging mechanism. Also for the instances of benign activity, accurately localizing such activities can aid statistical learning approaches which associate sketch-based representations with corresponding target words [77]. Considering both malicious and benign scenarios, the broad requirement is for a framework which can respond to a variety of atypical whiteboard sketch content in a reliable, comprehensive, and timely manner. To this end, we attempt to design a baseline model to detect the presence of atypical sketch content in Pictionary.

5.3 Related work

Sketch datasets: Existing sketch datasets (e.g. TU-Berlin [24], Sketchy [75], QuickDraw [41]) have been created primarily in the context of sketch *object* recognition problem – assign a categorical label to a hand-drawn sketch. The category labels correspond to objects (nouns). Therefore, these datasets lack abstract sketches which tend to be drawn when words from other parts of speech (verbs, adjectives) are provided as targets. Existing datasets are also unnatural because they do not include canvas actions such as erase strokes or location emphasis. Also, no intermediate guess words are associated with sketched content. For a similar reason, these datasets do not contain atypical activities unlike the dataset we introduce. Sarvadevabhatla et al. [77] explore neural network based generation of human-like guesses, but for pre-drawn object sketches. However, they do not accommodate interactivity and non-sketch drawing canvas activities (e.g. erase, pointing emphasis). The Kondate dataset [60] contains on-line handwritten patterns of text, figures, tables, maps, diagrams etc. The OHFCD dataset [4] pertains to online handwritten flowcharts. Although challenging in their own way, these datasets are considerably more structured than our setting. Additionally, they share the sketch datasets' shortcoming of being too cleanly curated because actions such as erase are absent. As a unique aspect, our combination of a game setting and a time limit unleashes greater diversity and creativity, causing sketches in our dataset to be more spontaneous and less homogeneous compared to existing datasets.

Detecting canvas items: Recognizing atypical activities can be thought of as a stroke segmentation problem wherein each sketch stroke is labelled as either belonging to an atypical class or the default class (drawing). Stroke segmentation has been employed for labelling parts in object sketches either from stroke sequence information [95, 90, 43, 65] or within an image canvas [91, 50]. Recognizing atypical sketch content can also be posed as an object detection problem. In this case, the objective is to obtain 2-D spatial bounding boxes enclosing sketch strokes corresponding to the atypical content. We adopt this approach because it is faster and more amenable to near real-time operation compared to segmentation. Handwritten text is the most common atypical sketch content class in Pictionary. Hence, it is reasonable to consider approaches solely designed for text detection in domains such as outdoor scenes and documents [39, 20, 53, 52, 51, 100, 5]. Similarly, detection-based approaches have been proposed for mixed graphic structures [42, 78, 27]. However, graphic elements in these scenarios are more structured compared to our Pictionary setting.

Pictionary-like guessing games: Borrowing terminology from the seminal work of von Ahn and Dabbish [88], Pictionary can be considered an 'inversion game' with full transparency. Riberio and Igarashi [68] employ a sketching-based interactive guessing game to progressively learn visual models of objects. A review of Pictionary-like word guessing games involving drawing can be found in the work by Sarvadevabhatla et al. [77]. In general, most of the existing works are confined to idealized toy settings [36], with some not even containing any sketching aspect [16, 26].

5.4 Baseline Models for Atypical sketch content Detection

To our best knowledge, there are no existing models to detect and localize atypical events in a canvas. In this thesis, we present the first attempt to compare the performance of standard models for detecting atypical sketch content in our data. We customize models designed for tasks such as text detection, object detection, and sketch segmentation by making minor changes to improve performance on sketch data. We compare three models inspired by [19, 65, 5]. Table 5.3 shows the performance and detection time for each baseline model.

Table 5.2: Stroke features extracted for atypical activity detection.

Category	Feature	Description
Unary features	Length of stroke	Length of the stroke path in pixels
	Height of stroke	Projection of stroke length along the canvas height
	Width of stroke	Projection of stroke length along the canvas width
	Area of convex hull	Area of the convex hull enclosing the stroke points
	Length of major axis	Length of the longest straight line within the convex hull
	Length of minor axis	Length of the longest line perpendicular to the major axis
	Eccentricity	Deviation of the convex hull from a circle
	Rectangularity	Deviation of the convex hull from a rectangle
	Density	Ratio of the stroke length to the convex hull area
	Curvature	Curvature of the stroke path
	Closure	Ratio of distance between the endpoints of the stroke to the stroke length
	Time duration of stroke	Time in seconds between the first point and last point of the stroke
Pairwise features	Ratio of length	Ratio of stroke lengths of two consecutive strokes
	Ratio of width	Ratio of stroke widths of two consecutive strokes
	Ratio of height	Ratio of stroke heights of two consecutive strokes
	Ratio of area	Ratio of convex hull area of two consecutive strokes
	Intersection over union of area	Ratio of the overlapping area to the union of areas for two consecutive strokes
	Distance between centroid	Distance between the centroid of the strokes points of two consecutive strokes
	Distance between endpoints	Distance between last point of a stroke and the first point of next stroke
	Time Lapse between stroke	Time in seconds between last point of a stroke and the first point of next stroke

5.4.1 BiLSTM+CRF

A sketch is a series of time-stamped strokes. To leverage the sequential nature of the sketch, we first use a recurrent neural network (RNN) based model (refer Figure 5.3) inspired by [19]. We extract a set of unary and pairwise features (refer Table 5.2) for each stroke similar to [86]. The features are selected considering the characteristics of the atypical content in our dataset. The unary features are extracted from a single stroke, while the pairwise features are calculated for a pair of consecutive strokes. Since our data is sequential, we use a BiLSTM (BiDirectional Long Short Term Memory) layer to further encode the features. The output sequences of the BiLSTM are given to a Conditional Random Field (CRF) layer. Finally, the output classes are decoded using the Viterbi algorithm.

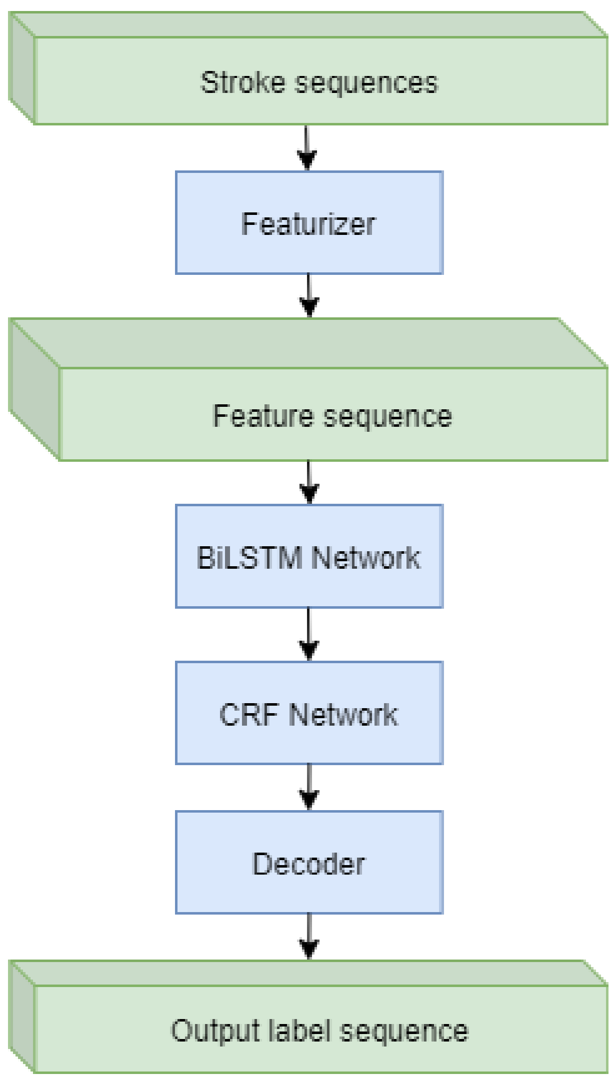


Figure 5.3: BiLSTM+CRF architecture.

5.4.2 SketchsegNet+

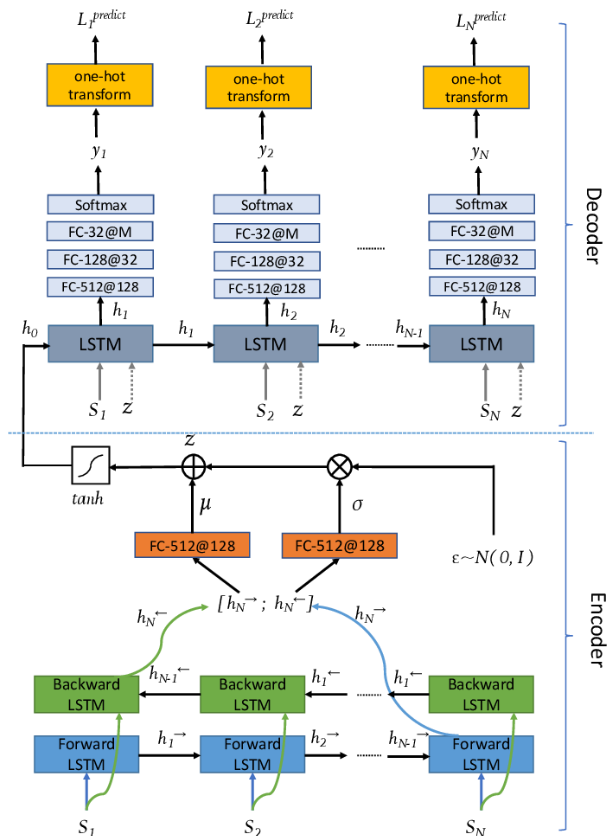


Figure 5.4: Sketchsegnet+ architecture.

SketchsegNet+ [65] is an RNN-based model designed for multi-class sketch semantic segmentation. Each point in the stroke is represented as a five-dimensional vector $S = [\Delta x, \Delta y, p_1, p_2, p_3]$ where $[\Delta x, \Delta y]$ is the differential offset of a point coordinates from the previous points and $[p_1, p_2, p_3]$ are binary flags to mark an ongoing stroke, last point of stroke and end of the drawing respectively. A sequence-to-sequence Variational Auto Encoder (VAE) is used to generate an output class for each point in a stroke (as shown in Figure 5.4). The input is first given to a BiLSTM encoder which predicts the mean and variance of a Gaussian latent vector variable. The sampled latent vector z is then given to an LSTM network along with the input S. Finally, a set of fully connected layers followed by a softmax layer is used to decode the output labels from the LSTM output sequence. The model is trained to optimize the mean square loss between the predicted labels and the ground truth labels.

5.4.3 Modified CRAFT

CRAFT (Character Region Awareness for Text Detection) [5] is a scene text detection method to effectively detect text area by exploring each character and affinity between characters. Unlike the RNN-

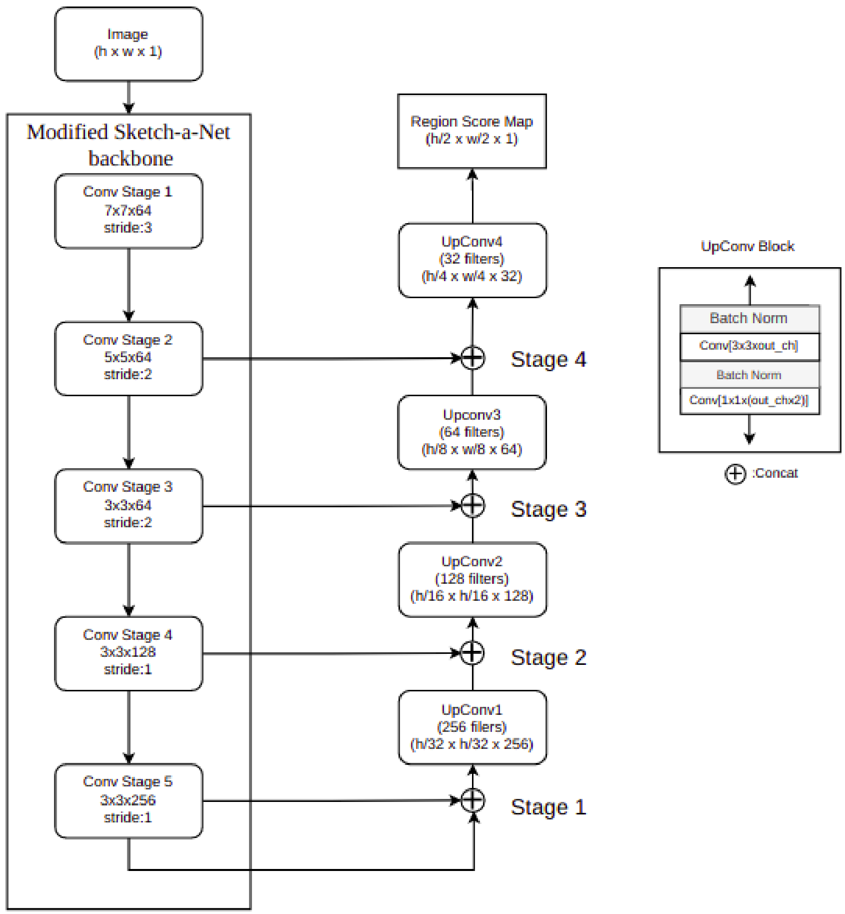


Figure 5.5: Architecture of modified CRAFT based on [5].

based models, CRAFT takes the rendered sketch image as input. The model adapts U-net [73] based architecture and learns a region segmentation score generated using Gaussian heatmaps at the center of every character in the input image. Image segmentation models such as CRAFT were designed to process information-rich natural scene images. Unlike text in natural settings, the text in sketch images is subjected to much more intrinsic variation but is relatively sparser in a blank canvas background. To incorporate the characteristics of sketch data, the VGG-16 backbone used in CRAFT [5] is replaced with a simpler Sketch-a-Net [99] backbone. To further improve the results, $15x15$ convolutions in the first layer of Sketch-a-Net [99] are reduced to $5x5$ kernel size. Additionally, the convolution stride is increased, and the max-pooling blocks are removed. This modified model was trained with mean squared error (MSE) loss to generate heatmaps scores. Finally, bounding boxes for each atypical object were generated from the heatmap using a modified watershed algorithm [5].

Table 5.3: Comparison of baseline model performance for atypical sketch content detection.

Method	Text only		Multiclass		# Parameters	ADT
	mAP	mAR	mAP	mAR	M=million	(m.sec)
BiLSTM+CRF [19]	0.06	0.04	0.02	0.03	0.01 M	85
SketchsegNet+[65]	0.56	0.32	0.04	0.11	3.90 M	21
Modified CRAFT [5]	0.47	0.69	0.17	0.30	1.18 M	34

5.5 Discussion

Detection of atypical content in sketches is a challenging task considering the diversity of sketch data. This thesis attempts to establish baselines for detecting the four classes of atypical content found in Pictionary sketch data. As seen in Table 5.1, these atypical content classes are highly imbalanced, with *Text* class having the maximum instances. Data augmentation was done to compensate for the class imbalance and increase the diversity of data. Three baseline models using different approaches were compared to detect the four atypical sketch classes. Additionally, the performance of these baselines on only the *Text* class is also reported. Table. 5.3 shows the performance (precision and recall) for these models for text-only and multi-class modes. The BiLSTM+CRF model (Section 5.4.1) used features hand-crafted for sketch strokes and a recurrent CRF model to classify these features. The poor performance of this model might signify the complex nature of the atypical classes that are not fully represented by the primitive features and simple model. The SketchsegNet+ model (Section 5.4.2) performed relatively better specifically for *Text* class. However, this model fails for other classes despite the data augmentation. The low performance for multi-class mode might be attributed to the varying length of atypical content in a sketch sequence. *Text* class usually occurs as a set of consecutive letters forming a word occupying a long subsequence of strokes in the sketch sequence. On the other hand, instances of other classes, such as arrows, numbers, or question marks, have relatively fewer strokes. The modified CRAFT (Sec 5.4.3) model processes the sketch content in the form of rendered images and is invariant to the number of strokes in a sequence. Although this model provides better performance, the scope for further improvement of atypical sketch content detection is proved necessary.

Based on the aforementioned comparison, it is evident that deep neural networks applied to process the rendered sketch image exhibit encouraging outcomes. Further research by Bansal et al. [7] has shown that a deep neural object detection network, CanvasNet, efficiently detects atypical content in Pictionary data. A summary of our related publication is provided here.

Chapter 6

Conclusion

6.1 Conclusion

In this thesis, we present the game of Pictionary as a case study for multimodal Cooperative Partially Observable games. The constrained mode of communication and the time-limited episodes of the Pictionary game leads to diverse interaction between the players. We present our browser-based Pictionary app, which is used for our data collection process. To analyze the large dataset collected, we developed an interactive dashboard.

We first present a preliminary statistical analysis of the Pictionary data in three domains: Global session related statistics, Target word related statistics, and User related statistics. The analysis shows that the distributions of interactions differ with directionality of communication. This indicates the need for role-specific analysis of the Pictionary interactions. We also find a significant bias caused by the target word part of speech on the game outcome. To examine the influence of the target word more accurately, we conducted a survey to quantify the difficulty of the target words used.

To further analyze the playing styles in Pictionary, we identify attributes of player interactions that characterize cooperative gameplay. We found stable role-based *Drawer* and *Guesser* playing style components unaltered by word difficulty. The playing style components reinforce the existence of certain personality types in games put forth by existing theoretical frameworks. In terms of gameplay and in the larger context of cooperative partially observation interaction, our results demonstrate that too much interaction or unbalanced interaction leads to unsuccessful communication. In future, our study can help in designing AI agents which mimic human behavior. In addition, our work also establishes a precedence for future studies on Cooperative Partially Observable games.

Finally, we explore atypical activity exhibited by the Pictionary *Drawers*. We identify four types of atypical sketch content, namely, Text, Numbers, Circles, and Iconic. We propose three baseline models to detect this atypical sketch content. A comparison of the baseline performances shows that deep neural models trained on the rendered sketch image performed better than recurrent models processing stroke sequence data.

6.2 Limitation and Future work

Pictionary as a case study poses certain limitations. The game has a fixed time limit of 120 seconds. Further analysis is required to validate the consistency in playing style components for variable time limits for the same levels of difficulty. Moreover, in our online version of Pictionary, the guesses are communicated as text messages. This may cause an increased latency in communication as compared to a real-world Pictionary game that uses voice modality for guessing [82].

To achieve a thorough analysis of multi-modal communication, further studies can be designed to use both text and voice modes of communication. Apart from Pictionary, similar games such as Wordgame [2] and Charades [66] are also interesting case studies that can be explored to better understand the differences caused by the modality of communication in partially observable environment.

Through our analysis, we explore only specific aspects of the game. A limitation of our analysis is that we do not take into account the drawing proficiency or prior experience of players. However, we use data from players of diverse age groups and demographics. Future work can also incorporate a post-game survey to rate social perceptions and player satisfaction to enhance communication in cooperative games.

The baseline models presented in this thesis require further improvements to reliably detect atypical sketch content. An efficient model for atypical activity detection can be used to create a monitoring system to warn the players of rule violations during the game. Moreover, a robust atypical sketch content detector can be used for other shared and interactive whiteboard scenarios.

Related Publications

1. **Kiruthika Kannan**, Anandhini Rajendran, Vinoo Alluri, Ravi Kiran Sarvadevabhatla, *"Draw Fast, Guess Slow: Characterizing Interactions in Cooperative Partially Observable Settings with Online Pictionary as a Case Study,"* In Human-Computer Interaction–INTERACT 2023: 19th IFIP TC 13 International Conference, York, UK, August 28 – September 1, 2023, Proceedings. Springer, 2023.

2. Nikhil Bansal, Kartik Gupta, **Kiruthika Kannan**, Sivani Pentapati, Ravi Kiran Sarvadevabhatla, *"Drawmon: A distributed system for detection of atypical sketch content in concurrent pictionary games,"* In Proceedings of the 30th ACM International Conference on Multimedia, pages 2852–2861, 2022.

Bibliography

[1] Z. Ashktorab, C. Dugan, J. Johnson, Q. Pan, W. Zhang, S. Kumaravel, and M. Campbell. Effects of communication directionality and ai agent differences in human-ai interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[2] Z. Ashktorab, Q. V. Liao, C. Dugan, J. Johnson, Q. Pan, W. Zhang, S. Kumaravel, and M. Campbell. Human-ai collaboration in a cooperative game setting: Measuring social perception and outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–20, 2020.

[3] M. Ashton and C. Verbrugge. Measuring cooperative gameplay pacing in world of warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 77–83, 2011.

[4] A. Awal, G. Feng, H. Mouchère, and C. Viard-Gaudin. First experiments on a new online handwritten flowchart database. In *Electronic Imaging*, 2011.

[5] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee. Character region awareness for text detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[6] M. F. Bafadal and H. Humaira. The use of charades games in teaching vocabulary to the junior high school students. *Linguistics and ELT Journal*, 6(2):14–21, 2019.

[7] N. Bansal, K. Gupta, K. Kannan, S. Pentapati, and R. K. Sarvadevabhatla. Drawmon: A distributed system for detection of atypical sketch content in concurrent pictionary games. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2852–2861, 2022.

[8] N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, et al. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280:103216, 2020.

[9] D. Barry. Picturing the first writing. 2013.

[10] R. Bartle. Players who suit muds. In *Journal of MUD Research, vol. 1, no. 1*, 1996.

[11] A. Bauer and Z. Popović. Collaborative problem solving in an open-ended scientific discovery game. *Proceedings of the ACM on human-computer interaction*, 1(CSCW):1–21, 2017.

[12] C. Berner, G. Brockman, B. Chan, V. Cheung, P. Debiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

[13] A. Beznosyk, P. Quax, K. Coninx, and W. Lamotte. The influence of cooperative game design patterns for remote play on player experience. In *Proceedings of the 10th asia pacific conference on Computer human interaction*, pages 11–20, 2012.

[14] S. S. Borges, R. Mizoguchi, V. H. Durelli, I. I. Bittencourt, and S. Isotani. A link between worlds: Towards a conceptual framework for bridging player and learner roles in gamified collaborative learning contexts. In *Advances in social computing and digital education*, pages 19–34. Springer, 2016.

[15] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.

[16] C. Clark, J. Salvador, D. Schwenk, D. Bonafilia, M. Yatskar, E. Kolve, A. Herrasti, J. Choi, S. Mehta, S. Skjonsberg, et al. Iconary: A pictionary-based game for testing multimodal communication with drawings and text. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1864–1886, 2021.

[17] J. Colwell and M. Kato. Investigation of the relationship between social isolation, self-esteem, aggression and computer game play in japanese adolescents. *Asian Journal of Social Psychology*, 6(2):149–158, 2003.

[18] D. Conroy, P. Wyeth, and D. Johnson. Modeling player-like behavior for game ai design. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, pages 1–8, 2011.

[19] A. Darvishzadeh, T. Stahovich, A. Feghahati, N. Entezari, S. Gharghabi, R. Kanemaru, and C. Shelton. Cnn-blstm-crf network for semantic labeling of students' online handwritten assignments. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1035–1040. IEEE, 2019.

[20] L. Deng, Y. Gong, Y. Lin, J. Shuai, X. Tu, Y. Zhang, Z. Ma, and M. Xie. Detecting multi-oriented text with corner-based region proposals. *Neurocomputing*, 334:134–142, Mar 2019.

[21] H. Desurvire and M. S. El-Nasr. Methods for game user research: studying player behavior to enhance game design. *IEEE computer graphics and applications*, 33(4):82–87, 2013.

[22] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau. Guns, swords and data: Clustering of player behavior in computer games in the wild. In *2012 IEEE conference on Computational Intelligence and Games (CIG)*, pages 163–170. IEEE, 2012.

[23] M. Eger, C. Martens, and M. A. Córdoba. An intentional ai for hanabi. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 68–75. IEEE, 2017.

[24] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10, 2012.

[25] M. S. El-Nasr, A. Drachen, and A. Canossa. *Game analytics*. Springer, 2016.

[26] A. El-Nouby, S. Sharma, H. Schulz, D. Hjelm, L. El Asri, S. Ebrahimi Kahou, Y. Bengio, and G. W. Taylor. Tell, draw, and repeat: Generating and modifying images based on continual linguistic instruction. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019.

[27] E. Elyan, L. Jamieson, and A. Ali-Gombe. Deep learning for symbols detection and classification in engineering drawings. *Neural networks : the official journal of the International Neural Network Society*, 129:91–102, 2020.

[28] K. Emmerich and M. Masuch. The impact of game patterns on player experience and social interaction in co-located multiplayer games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 411–422, 2017.

[29] N. Fay, B. Walker, and N. Swoboda. Deconstructing social interaction: The complimentary roles of behaviour alignment and partner feedback to the creation of shared symbols. In *CogSci*, 2017.

[30] L. S. Ferro, S. P. Walz, and S. Greuter. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, pages 1–6, 2013.

[31] S. B. Franklin, D. J. Gibson, P. A. Robertson, J. T. Pohlmann, and J. S. Fralish. Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science*, 6(1):99–106, 1995.

[32] K. Frans. Ai charades: Language models as interactive game environments. In *2021 IEEE Conference on Games (CoG)*, pages 1–2. IEEE, 2021.

[33] K. I. Gero, Z. Ashktorab, C. Dugan, Q. Pan, J. Johnson, W. Geyer, M. Ruiz, S. Miller, D. R. Millen, M. Campbell, et al. Mental models of ai agents in a cooperative game setting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[34] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage. Predicting player churn in the wild. In *2014 IEEE Conference on Computational Intelligence and Games*, pages 1–8. Ieee, 2014.

[35] D. Hooshyar, M. Yousefi, and H. Lim. Data-driven approaches to game player modeling: a systematic literature review. *ACM Computing Surveys (CSUR)*, 50(6):1–19, 2018.

[36] F. Huang, E. Schoop, D. Ha, and J. Canny. Scones: towards conversational authoring of sketches. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 313–323, 2020.

[37] J. Hwang and S. Park. An exploratory study of how casino dealer communication styles lead to player satisfaction. *Journal of Travel & Tourism Marketing*, 35(9):1246–1260, 2018.

[38] I. Iacovides, A. L. Cox, A. Avakian, and T. Knoll. Player strategies: Achieving breakthroughs and progressing in single-player and cooperative games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pages 131–140, 2014.

[39] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading text in the wild with convolutional neural networks, 2014.

[40] J. Jiang, D. Maldeniya, K. Lerman, and E. Ferrara. The wide, the deep, and the maverick: Types of players in team-based online games. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–26, 2021.

[41] J. Jongejan, H. Rowley, T. Kawashima, J. Kim, and N. Fox-Gieg. The quick, draw! ai experiment. *Mountain View, CA*, 2016.

[42] F. D. Julca-Aguilar and N. S. T. Hirata. Symbol detection in online handwritten graphics using faster r-cnn. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 151–156, 2018.

[43] K. Kaiyrbekov and M. Sezgin. Stroke-based sketched symbol reconstruction and segmentation, 2019.

[44] M. Kempka, M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. Vizdoom: A doom-based ai research platform for visual reinforcement learning. In *2016 IEEE conference on computational intelligence and games (CIG)*, pages 1–8. IEEE, 2016.

[45] B. Kirman and S. Lawson. Hardcore classification: Identifying play styles in social games using network analysis. In *International Conference on Entertainment Computing*, pages 246–251. Springer, 2009.

[46] Y. Kou. Toxic behaviors in team-based competitive gaming: The case of league of legends. In *Proceedings of the annual symposium on computer-human interaction in play*, pages 81–92, 2020.

[47] J. Kuittinen, A. Kultima, J. Niemelä, and J. Paavilainen. Casual games discussion. In *Proceedings of the 2007 conference on Future Play*, pages 105–112, 2007.

[48] A. Lange, A. Somov, A. Stepanov, and E. Burnaev. Building a behavioral profile and assessing the skill of video game players. *IEEE Sensors Journal*, 22(1):481–488, 2021.

[49] J. Li, H. Lu, C. Wang, W. Ma, M. Zhang, X. Zhao, W. Qi, Y. Liu, and S. Ma. A difficulty-aware framework for churn prediction and intervention in games. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 943–952, 2021.

[50] L. Li, H. Fu, and C.-L. Tai. Fast sketch segmentation and labeling with deep learning, 2018.

[51] M. Liao, B. Shi, and X. Bai. Textboxes++: A single-shot oriented scene text detector. *IEEE transactions on image processing*, 27(8):3676–3690, 2018.

[52] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. *Lecture Notes in Computer Science*, page 21–37, 2016.

[53] Y. Liu, S. Zhang, L. Jin, L. Xie, Y. Wu, and Z. Wang. Omnidirectional scene text detection with sequential-free box discretization, 2019.

[54] C. E. Lopez and C. S. Tucker. The effects of player type on performance: A gamification case study. *Computers in Human Behavior*, 91:333–345, 2019.

[55] E. Loria. A framework to infer player experience and inform customized content generation in gameful systems. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pages 21–28, 2019.

[56] E. Loria and A. Marconi. Player types and player behaviors: Analyzing correlations in an on-the-field gamified system. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pages 531–538, 2018.

[57] E. Loria, J. Rivera, and A. Marconi. Do they play as intended?-comparing aggregated and temporal behavioral analysis in a persuasive gamified system. In *HICSS*, pages 1–10, 2021.

[58] S. Mader and E. Tassin. Playstyles in tetris: Beyond player skill, score, and competition. In *International Conference on Entertainment Computing*, pages 162–170. Springer, 2022.

[59] Y. Matsumoto and R. Thawonmas. Mmog player classification using hidden markov models. In *International Conference on Entertainment Computing*, pages 429–434. Springer, 2004.

[60] T. Matsushita and M. Nakagawa. A database of on-line handwritten mixed objects named "kondate". In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 369–374, 2014.

[61] I. Millington and J. Funge. *Artificial intelligence for games*. CRC Press, 2018.

[62] S. Molinillo, F. Muñoz-Leiva, and F. Pérez-García. The effects of human-game interaction, network externalities, and motivations on players' use of mobile casual games. *Industrial management & data systems*, 2018.

[63] H. Y. Ong, S. Deolalikar, and M. Peng. Player behavior and optimal team composition for online multiplayer games. *arXiv preprint arXiv:1503.02230*, 2015.

[64] V. Peters, G. Vissers, and G. Heijne. The validity of games. *Simulation & Gaming*, 29(1):20–30, 1998.

[65] Y. Qi and Z.-H. Tan. Sketchsegnet+: An end-to-end learning of rnn for multi-class sketch semantic segmentation. *IEEE Access*, 7:102717–102726, 2019.

[66] A. Rahmah and Y. Astutik. Charades game: Does it affect students' learning on english vocabulary. *EnJourMe (English Journal of Merdeka)*, 5(1):75–89, 2020.

[67] D. Ramirez-Cano, S. Colton, and R. Baumgarten. Player classification using a meta-clustering approach. In *Proceedings of the 3rd Annual International Conference Computer Games, Multimedia & Allied Technology*, pages 297–304, 2010.

[68] A. Ribeiro and T. Igarashi. Sketch-editing games: Human-machine communication, game theory and applications. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, page 287–298, New York, NY, USA, 2012. Association for Computing Machinery.

[69] S. Risi and M. Preuss. From chess and atari to starcraft and beyond: How game ai is driving the world of ai. *KI-Künstliche Intelligenz*, 34(1):7–17, 2020.

[70] L. A. Rodrigues and J. D. Brancher. Improving players' profiles clustering from game data through feature extraction. In *2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 177–17709. IEEE, 2018.

[71] I. Rodríguez, A. Puig, and À. Rodríguez. Towards adaptive gamification: A method using dynamic player profile and a case study. *Applied Sciences*, 12(1):486, 2022.

[72] M. J. Rogerson, M. R. Gibbs, and W. Smith. Cooperating to compete: the mutuality of cooperation and competition in boardgame play. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

[73] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[74] J. Runge, P. Gao, F. Garcin, and B. Faltings. Churn prediction for high-value players in casual social games. In *2014 IEEE conference on Computational Intelligence and Games*, pages 1–8. IEEE, 2014.

[75] P. Sangkloy, N. Burnell, C. Ham, and J. Hays. The sketchy database: Learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, 2016.

[76] R. K. Sarvadevabhatla, S. Surya, T. Mittal, and R. V. Babu. Pictionary-style word guessing on hand-drawn object sketches: Dataset, analysis and deep network models. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):221–231, 2018.

[77] R. K. Sarvadevabhatla, S. Surya, T. Mittal, and R. V. Babu. Pictionary-style word guessing on hand-drawn object sketches: Dataset, analysis and deep network models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):221–231, 2020.

[78] B. Schäfer and H. Stuckenschmidt. Arrow r-cnn for flowchart recognition. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, volume 1, pages 7–13. IEEE, 2019.

[79] M. Seif El-Nasr, B. Aghabeigi, D. Milam, M. Erfani, B. Lameman, H. Maygoli, and S. Mah. Understanding and evaluating cooperative games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 253–262, 2010.

[80] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[81] A. B. Siyanbola, S. F. Omolola, and A. O. Adeyemi. Evolution of graphic design from the paleolithic era. *KIU Journal of Humanities*, 6(1):219–227, 2021.

[82] F. Spyridonis, D. Daylamani-Zad, and M. P. O'Brien. Efficient in-game communication in collaborative online multiplayer games. In *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pages 1–4. IEEE, 2018.

[83] Z. O. Toups, J. Hammer, W. A. Hamilton, A. Jarrah, W. Graves, and O. Garretson. A framework for cooperative communication game mechanics from grounded theory. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pages 257–266, 2014.

[84] J. Tuunanen and J. Hamari. Meta-synthesis of player typologies. In *Proceedings of Nordic Digra 2012 Conference: Games in Culture and Society, Tampere, Finland*, 2012.

[85] P. Tylkin, G. Radanovic, and D. C. Parkes. Learning robust helpful behaviors in two-player cooperative atari environments. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1686–1688, 2021.

[86] T. Van Phan and M. Nakagawa. Text/non-text classification in online handwritten documents with recurrent neural networks. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 23–28. IEEE, 2014.

[87] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, et al. Starcraft ii: A new challenge for reinforcement learning. *arXiv preprint arXiv:1708.04782*, 2017.

[88] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, Aug. 2008.

[89] J. Walton-Rivers, P. R. Williams, and R. Bartle. The 2018 hanabi competition. In *2019 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2019.

[90] F. Wang, S. Lin, H. Wu, H. Li, R. Wang, X. Luo, and X. He. Spfusionnet: Sketch segmentation using multi-modal data fusion. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1654–1659, 2019.

[91] W. Wang, E. Xie, X. Song, Y. Zang, W. Wang, T. Lu, G. Yu, and C. Shen. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network, 2020.

[92] D. Wheat, M. Masek, C. P. Lam, and P. Hingston. Modeling perceived difficulty in game levels. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–8, 2016.

[93] P. Williams. *Artificial intelligence in co-operative games with partial observability*. PhD thesis, University of Essex, 2019.

[94] B. Xia, X. Ye, and A. O. Abuassba. Recent research on ai in games. *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 505–510, 2020.

[95] L. Yang, J. Zhuang, H. Fu, K. Zhou, and Y. Zheng. Sketchgcn: Semantic sketch segmentation with graph convolutional networks, 2020.

[96] G. N. Yannakakis and J. Togelius. A panorama of artificial and computational intelligence in games. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(4):317–335, 2014.

[97] G. N. Yannakakis and J. Togelius. Experience-driven procedural content generation. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 519–525. IEEE, 2015.

[98] Q. Yin, J. Yang, W. Ni, B. Liang, and K. Huang. Ai in games: Techniques, challenges and opportunities. *arXiv preprint arXiv:2111.07631*, 2021.

[99] Q. Yu, Y. Yang, Y.-Z. Song, T. Xiang, and T. Hospedales. Sketch-a-net that beats humans. In M. W. J. Xianghua Xie and G. K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 7.1–7.12. BMVA Press, September 2015.

[100] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang. East: an efficient and accurate scene text detector. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 5551–5560, 2017.