Hiding Behind Cards: Identifying Bots and Humans in Online Poker

by

Benjamin Altman

A Thesis
Presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Master of Science
in
Applied Computer Science

Guelph, Ontario, Canada

© Benjamin Altman, May, 2013
ABSTRACT

HIDING BEHIND CARDS: IDENTIFYING BOTS AND HUMANS IN ONLINE POKER

Benjamin Altman
University of Guelph, 2013

Advisor: Dr. Blair Nonnecke

As online gaming becomes more popular, it has also become increasingly important to identify and remove those who leverage automated player systems (bots). Manual bot detection depends on the ability of game administrators to differentiate between bots and normal players. The objective of this thesis was to determine whether expert poker players can differentiate between bot and human players in Texas Hold ‘Em Poker. Participants were deceived into thinking a number of bots and humans were playing in gameplay videos and asked to rate player botness and skill. Results showed that participants made similar observations about player behaviour, yet used these observations to reach differing conclusions about whether a given player was a bot or a human. These results cast doubt on the reliability of manual bot detection systems for online poker, yet also show that experts agree on what constitutes skilled play within such an environment.
Acknowledgements

This thesis has been a long and arduous journey, but it was not one that I could complete on my own. These are the ones who made this possible:

- Dr. Blair Nonnecke: my advisor, for his immense patience, and his trust and support throughout my graduate school career.
- Miriam Benarroch-Altman: my wife, for her love, motivation and intellectual support, even through the rough patches.
- Mark Altman: my father, for that final motivational push fueled by the power of financial logic.

I also wish to thank the rest of my family for sticking by my decisions, and trusting me to follow it through to the end. Lastly, I want to thank the rest of the fellow students from my lab – it was always reassuring to know that we were all in this together.
# Table of Contents

**ACKNOWLEDGEMENTS** .................................................................................................................. III

**TABLE OF CONTENTS** ................................................................................................................... IV

**LIST OF TABLES** .............................................................................................................................. VI

**LIST OF FIGURES** ............................................................................................................................ VII

1 **INTRODUCTION** ......................................................................................................................... 1

1.1 **RESEARCH FOCUS** ................................................................................................................... 2

1.2 **RESEARCH STATEMENT** ........................................................................................................... 3

1.3 **DOCUMENT SUMMARY** ............................................................................................................ 3

2 **LITERATURE REVIEW** .................................................................................................................. 4

2.1 **THE RISE OF ONLINE GAMING** ............................................................................................... 4

2.2 **ONLINE GAME GENRES** ........................................................................................................... 5

2.3 **THE GAME OF POKER** .............................................................................................................. 8

2.3.1 **Texas Hold’Em Poker** ............................................................................................................ 10

2.3.2 **Online Poker** ......................................................................................................................... 12

2.4 **GAMING BOTS** ........................................................................................................................... 14

2.4.1 **Overview of Gaming Bots** ..................................................................................................... 15

2.4.2 **Poker Bots** ............................................................................................................................. 16

2.4.3 **CPRG Poker Bots** ................................................................................................................. 17

2.5 **BOT DETECTION METHODS** .................................................................................................... 19

2.5.1 **Behaviour-Based Bot Detection** .......................................................................................... 20

2.5.2 **Traffic-Based Analysis** .......................................................................................................... 21

2.5.3 **Manual Bot Detection** .......................................................................................................... 22

2.6 **ACCURACY OF MANUAL BOT DETECTION** ............................................................................ 24

2.6.1 **Experimental Environments** ............................................................................................... 24

2.6.2 **How Players Detect Bots** ..................................................................................................... 26

2.6.3 **Ambiguous Interpretations of Player Behaviour** .................................................................. 28

3 **METHODOLOGY** ............................................................................................................................ 30

3.1 **OVERVIEW** ............................................................................................................................... 30

3.2 **RATIONALE** ................................................................................................................................ 31

3.2.1 **Use of a Poker Environment** ............................................................................................... 31

3.2.2 **Control Over Bot Behaviour** ............................................................................................... 32

3.3 **PARTICIPANTS** .......................................................................................................................... 35

3.4 **MATERIALS** .............................................................................................................................. 36

3.5 **PROCEDURE** .............................................................................................................................. 41

4 **RESULTS & DISCUSSION** ............................................................................................................ 44

4.1 **IDENTIFYING A DIFFERENCE BETWEEN BOTS AND HUMANS** ........................................... 44

4.1.1 **Experimental Results** ........................................................................................................... 44

4.1.2 **Quality of Poker Bots** ........................................................................................................... 45

4.1.3 **Identifiers used in Differentiating Humans and Bots** ............................................................. 46

4.1.4 **Terms Used to Describe Humans and Bots** ......................................................................... 47

4.1.5 **Confidence in Identification of Bots and Banning of Offenders** ......................................... 48
List of Tables

Table 2.1 Description of Common Online Game Genres ................................................................. 8
Table 3.1 Participant Age and Play Location .............................................................................. 35
Table 3.2 Gameplay Experience and Frequency of Observing Poker ............................................. 36
Table 3.3 Differences Between Bot Configurations ....................................................................... 39
Table 4.1 Bot Opinions and Believability ......................................................................................... 46
Table 4.2 Classification Identifiers ................................................................................................. 47
Table 4.3 Gameplay Descriptors ...................................................................................................... 47
Table 4.4 Percentage of Bots Identified Successfully ......................................................................... 48
Table 4.5 Confidence and Banning .................................................................................................. 49
Table 4.6 Player Traits .................................................................................................................... 55
Table 4.7 Participants’ Opinions on Best Traits vs Identifiers and Descriptors ................................. 57
Table D.1 Botness Rating per Seat per Participant ......................................................................... 83
Table D.2 Skill Ranking per Seat per Participant .......................................................................... 83
List of Figures

Figure 3.1 Screenshot of Hand 1, Video 2.................................................................37
Figure 4.1 Botness Ratings vs Seat.................................................................45
Figure 4.2 Participants’ Botness Ratings for Player in Seat 3 ...........................................49
Figure 4.3 Participants’ Botness Ratings for Player in Seat 3 and Seat 7..............................50
Figure 4.4 Disagreement on Botness as Expressed by 3 Participants.................................51
Figure 4.5 Heterogeneity of Botness Ratings by all Participants.........................................51
Figure 4.6 Frequency of Each Botness Rating..........................................................52
Figure 4.7 Heterogeneity of Skill Rankings by all Participants..........................................54
Figure 4.8 Skill Ranking by Participant for Seat 9.......................................................54
Figure 4.9 Average Botness Rating and Skill Ranking (inverted) by Seat.............................56
It has been decades since the Turing Test proposed that it might be possible for humans to mistake a computer for another human when communicating remotely (Turing, A., 1950). Since then, we have experienced the growth and spread of the Internet, and as a consequence, increased remote interactions. Multiplayer gaming is an example of such an interaction, and one that has become increasingly common. Both video game consoles and computers now provide the means for users to engage in gameplay with other players online, rather than being confined to the same room. With increased remote interactions has come the very situation proposed in the Turing Test: human players engage in remote interactions with other players with no way to verify whether other players are actually human or not. As a result, individuals are able to exploit the online system by using well-designed computer players, known as bots, that can participate in online gaming in their stead.

Bots are capable of playing without experiencing fatigue or frustration, as well as analyzing and reacting to situations far more quickly than a human in some scenarios. In solitary play, a bot may be considered a benign gaming aid, but the situation changes when other human players are involved. Depending on the game, a human controller can potentially leverage a bot’s superior abilities in a number of ways: it can upset a game’s economy through endless task repetition (Cornelissen & Grootjen, 2008; van Kesteren, Langevoort & Grootjen, 2009); advance quickly in the game through endless play; and outperform human opponents through its superhuman response rate (Laird & Duchi, 2001). If such advantages are leveraged in a situation involving financial gain, such as online gambling, the impact on the game and community are more substantial. Consequently, improving bot detection systems has become increasingly important in online gaming.
Although recent research has focused on automated bot detection solutions (Chen, Liao, Pao & Chu, 2008; Cornelissen & Grootjen, 2008; Chen et al., 2009), manual bot detection by humans remains a relevant research area (Chen et al., 2009). One possible reason for this is that automated solutions are not yet perfect – due to false positives and limited sample sizes, such systems cannot yet be trusted to remove players from games automatically (Cornelissen & Grootjen, 2008). For instance, many online gaming communities still allow members to report other players as bots, which leads to their account being flagged for manual review by an administrator, and possible disciplinary action (Aion Community Team, 2009; Botwatch, n.d.). Consequently, it becomes important to consider whether humans can accurately differentiate bots from humans in online play. Without such accuracy, wrongful banning can easily occur, or bots can continue operating within a supposedly secure system.

1.1 Research Focus

Prevalence of bots within online gaming communities is a persistent issue, and has resulted in the creation of a number of detection systems. While there are many automated systems that may examine player behavior, data traffic, or require the player to answer simple questions to prove their humanity, manual intervention and review is still typical in many communities and may complement automated systems as well. Previous research on the ability of humans to detect bots has focused on highly graphical games with player avatars, or simplistic games with little player choice (see Chapter 2). Consequently, this study sought to extend the current literature by examining manual bot detection in an environment that falls between those two extremes: Texas Hold ’Em Poker.

While poker has a limited set of actions available to players (Appendix C), it is well-known as a game with deep player-versus-player strategy. As a result, observing and
understanding opponent behavior is a skill that must be developed by expert players, and they would be expected to be able to differentiate “human” gameplay from that of bots.

1.2 Research Statement

This thesis demonstrates that expert poker players will use similar observations to reach differing conclusions when asked to differentiate bots from human players in a game of poker. In other words, the same behavior is considered to be indicative of both a bot and a human, depending on the observer’s inclination.

1.3 Document summary

The body of this thesis has been split into four distinct sections. Chapter 2 contains a literature review comprising online gaming and game genres, the game of Texas Hold ‘Em Poker, gaming bots, and both manual and automated bot detection in various game environments. This review forms the basis for the hypotheses and the rationale for the methodology. Chapter 3 follows with the study’s methodology, including rationale, participant information, materials and procedure. Chapter 4 contains the quantitative and qualitative results of the study, as well as the analysis and discussion of their significance, organized by hypothesis. The final chapter, Chapter 5, contains the conclusions, limitations and future extensions of the study. An appendix provides copies of the questionnaires, raw data, and Texas Hold ‘Em terminology.
2 Literature Review

This chapter discusses the background literature relevant to the hypotheses and thesis statement. It begins by providing a brief background to the rise of online gaming, and the common varieties of games found online. Next, it delves into one genre to examine one specific game, Texas Hold ‘Em Poker, and explains rules, strategies and online gameplay. A description of gaming bots follows, covering how bots may differ depending on game genre, how they can impact a game, and a summary of the evolution of poker bots. Next, the multitude of methods for detecting bots are examined, outlining both automated and manual methods and their respective benefits and drawbacks. Finally, the accuracy of manual bot detection is considered in a variety of game environments, and key factors influencing manual bot detection are identified.

2.1 The Rise of Online Gaming

With the increased availability of the Internet, video gaming has steadily moved from the living room to the online world. The revenue of online video games has risen drastically in the past decade, from an approximate valuation of $1 billion in 2002, to over $8 billion in recent years (Meloni, 2008). The popularity of online gaming has also expanded the idea of “online societies”, giving users a new outlet for socializing and entertainment, and becoming increasingly important to them as a result (Yee, 2006). While the idea of online gaming for entertainment is hardly uncommon, the sheer profitability of online gaming ventures may be less commonly known.

Some professional gamers have moved from face-to-face conventions to online competitive circuits – a move that has proven quite lucrative. One strategy game, known as StarCraft II, has a number of annual competitions hosted by companies and enthusiasts. The top
prize for one of such competitions, run by Major League Gaming, is currently set at $25,000 (Major League Gaming, n.d). Some of these competitions are even televised, and have exceeded 3 million viewers over the course of the competition broadcast (Stark, 2012). Clearly, even games that do not involve money directly can be turned into professional competitions for large cash prizes – the stakes are raised, however, in online gambling.

Unlike most video games, gambling almost always involves actual money – whether online or offline, competitive or casual. With the large growth in usage and access to the Internet, online gambling has continued to grow into a multi-billion dollar industry (Research and Markets, 2013). With online tournament prizes exceeding $10 million, online gambling can become a very lucrative opportunity for a lucky few (Willis, 2011).

Depending on the audience, the world of online gaming can represent a new society, an entertaining break from everyday life, a competitive sport, or a lucrative career. As access to the Internet continues to grow, and more activities move to online environments, its value to users of all ages is likely to continue to increase year after year.

2.2 Online Game Genres

There are a wide variety of game genres that have become popular online, however, for the sake of brevity, this section will focus on those that are the most common (see Table 2.1 for an overview).

To begin, the **First-Person Shooter (FPS)** is one of the more common game genres found online. These games are typically characterized by a player taking the viewpoint of a character, and engaging in violent, fast-paced gameplay with a number of opponents. Goals in this genre include capture-the-flag, in which players seek to invade an enemy base and safely
return to their own, king-of-the-hill, in which players seek to defend a specific position from the enemy, and death matches, in which players seek to kill as many of the opposition as possible.

Other games in the FPS genre have sought to take a different approach by creating smaller player match-ups, and in some cases even encouraging co-operation. For example, Borderlands by 2K Games allows players to team up in order to conquer difficult fights. Although player-versus-player combat is available, co-operative gameplay is one of the promoted features of Borderlands multiplayer. Regardless of whether a FPS encourages co-operation or competition, the underlying mechanics typically involves killing one’s enemies.

A new genre made possible through online gaming is the **Massively Multiplayer Online Role-Playing Game (MMORPG)**. Games in this genre are characterized by enormous worlds, a large number of non-player characters, and an underlying storyline. Players tend to either co-operate or compete in order to complete quests, gain levels, find equipment, and establish themselves within the game world. Additionally, games in this genre tend to follow one of three pricing models: pay-to-play, in which the consumer purchases the game and can play online for free; subscription, in which the consumer purchases the game and pays a monthly fee for online access; and free-to-play, in which the game is free, and the consumer only pays for optional benefits and advantages in the game itself.

This genre is dependent on the Internet due to its need to connect a large number of players to fill enormous game worlds (Bartle, 2010). This is best demonstrated by the number of MMORPGs, e.g., The Secret World and Age of Conan, that “failed” over the past decade, as they were unable to maintain a large enough subscriber base to attract and retain new players (Onyett, 2012). These were typically subscription-based games forced to either shut down or move to a “free-to-play” or “buy-to-play” model in which they attempted to reinvigorate player
population by removing the subscription fee. This inability to reliably hold onto a player population has been the downfall of many MMORPGs seeking to supplant the leading game, Blizzard Entertainment’s World of Warcraft (WoW). Although it was released in 2004, WoW continues to dominate the MMORPG genre with over 11 million subscribers.

MMORPGs are also known for establishing entire communities dedicated to deep dissection of the gameplay. Whether it is analysis of in-game actions to establish the most effective strategies, or elaborate discussion of the in-game economy, the MMORPG genre represents a whole other world to many of its players (Whynot, 2011). This has a downside in that many individuals become highly addicted to the game world, and are unable to play in moderation (Ng & Wiemer-Hastings, 2005).

The Real-Time Strategy (RTS) game genre has been a staple of the game industry for years. Games in this genre are characterized by an overhead view of a map, micromanagement of a large number of units to achieve strategic objectives, and a focus on skill and strategy over luck. This genre is known for its “pro gaming”, particularly with older games such as StarCraft and its expansions. Matches tend to have fewer than 8 participants, and social structures, such as alliances, can easily emerge in a competitive game.

Lastly, there are classic card games that have moved online. Online gaming communities exist for a number of classic games, including Hearts, Euchre, Bridge, and Poker. Games involving wagers, such as Texas Hold ‘Em Poker, have contributed to the creation of the multi-billion dollar online gambling industry. There are many Hold ‘Em Web sites, and some, such as PartyPoker.com, nearly rank among the top 1000 visited Web sites on the Internet (Alexa, n.d.). The game of poker is deceptively simple on the surface, but its complexities lie in the strategies players must apply with limited information, e.g., cards visible and the behaviour
of other players. The terms, gameplay and strategies involved in Texas Hold ‘Em Poker are outlined in the following section, and will demonstrate the rationale for using it as the focus of this study.

<table>
<thead>
<tr>
<th>Genre</th>
<th>Viewpoint</th>
<th>Notable Points</th>
<th>Game Size (Approx. Average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person Shooter (FPS)</td>
<td>First-Person (see through avatar’s eyes)</td>
<td>Tend to revolve around military objectives – killing and capturing.</td>
<td>Large (15+)</td>
</tr>
<tr>
<td>Massively Multiplayer Online Role Playing Game (MMORPG)</td>
<td>Third-Person (behind avatar)</td>
<td>Completion of quests, in-game economies, social gameplay.</td>
<td>Enormous (500+)</td>
</tr>
<tr>
<td>Real-Time Strategy (RTS)</td>
<td>Overhead / Bird’s Eye</td>
<td>Micromanagement of units to capture military objectives and outmaneuver opponents.</td>
<td>Small (6+)</td>
</tr>
<tr>
<td>Card</td>
<td>Overhead / Bird’s Eye, no avatar</td>
<td>Usage of betting and bluffing in order to beat opponent cards.</td>
<td>Medium (10+)</td>
</tr>
</tbody>
</table>

Table 2.1 Description of Common Online Game Genres

2.3 The Game of Poker

Poker’s popularity has surged in the past 10 years or so, as more online gambling sites open, and the game is highlighted in TV shows and other media. The variant that receives the most exposure at this time is Texas Hold’Em. However, prior to the shift online, Seven Card Stud was the most well-known poker variant. Prior to describing the specifics of Texas Hold’Em Poker, the following briefly describes the basics of poker. Additional information on playing poker, terms, and hand values can be found in Appendix C.

Poker has a small set of rules and actions making it appear quite simple. The game is played with a standard deck of 52 cards, and has a defined hierarchy of card values. There are only four actions that a player can take, and each player takes their turn one at a time. Many poker variants begin by dealing two cards face-down to all players. Game rounds typically alternate between dealing additional cards to players and having a round of betting, in which
players take one of the four actions until all remaining players have wagered the same amount of money.

When taken as a set of five, a player’s cards are considered their hand, and are more valuable if specific criteria are met. An example of this would be a “full house”, in which the player holds three of one card value, and two of another. If a player’s hand contains no special combination, the player may be left with the worst hand, known as “high card”, in which they will only win if their opponents do not possess a single card of a higher value. Due to the variety of these combinations, the exact strength of a player’s hand cannot be determined by their opponents – there is no way to know what cards they possess, and consequently, their highest valued combination.

During normal gameplay, there are four possible actions: players can pass, known as a check; match the current bet, known as a call; increase the bet, known as a raise; or refuse to match the highest bet and sit out the hand, known as a fold. It is important to note that a player cannot perform a check if any bets have been placed, as they must, at a minimum, call the highest bet made before their turn. Once a player folds, they cannot take an action until the next hand has begun.

It is during the betting phase that players can employ strategy to deceive their opponents. This is known as a bluff, and involves taking actions that do not reflect the true strength of your hand. An example might be an enormous wager with a poor quality hand, hoping that other players will fold, leaving the bettor with all bets made to that point, known as the pot. Conversely, a player with a high quality hand may choose to make small bets, or even check, in order to deceive others into betting aggressively with what they believe to be superior hands.
Bluffing leverages the lack of information available to opponents in order to maximize profits or turn a losing hand into a winner.

Even if not bluffing, placing large wagers can still be an effective tactic to maximize the chances of taking the pot. In a **no-limit** game, a player can wager all the money they have at the table, known as “going all-in”, forcing their opponents to either make the same substantial investment or fold. Betting can play a significant role in no-limit games, as it allows players to determine how risk adverse their opponents are. This information is vital, as it allows a player to tailor their strategy to best exploit a given opponent’s weakness. In a **limit** game, players are only able to bet a maximum amount in each round, eliminating the viability of an “all-in” strategy, and also increasing the likelihood that the other players will call bets with mediocre hands.

Ultimately, poker is a deceptively simple game. The limited number of actions and alternating round structure seem quite straightforward, but it is difficult for a player to win poker simply by having good hands. Effective betting is necessary to best profit from opponents, and effective betting can only be done by learning opponent weaknesses. By observing the betting habits and body language of opponents, a poker player can factor in additional information when choosing their actions, and consequently, play more effectively. This is typically summed up as “playing the player, not the cards”, and is a vital point to keep in mind when examining poker as a game environment.

### 2.3.1 Texas Hold’Em Poker

At first glance, Texas Hold’Em Poker seems like a straightforward game of standard poker. All hand values are consistent with standard poker, betting follows similar patterns and rules, and player roles, such as dealer and blinds, remain present. However, there is a crucial difference –
there are five visible communal cards, meaning that the majority of each player’s hand is shared with everyone at the table.

Players are dealt two personal cards at the start of the game, and forced to immediately decide to either raise, call or fold. This decision point occurs as a result of the forced starting bet made by the large blind, which is the player two seats to the left of the dealer. At this point, only 40% of a player’s potential hand is known to them, and there is no additional information available on the table. The only way to make this decision is by evaluating one’s own cards, position at the table, and the betting behavior and body language of others.

For this reason, position is an absolutely critical piece of information when playing Texas Hold’Em Poker. Being in early position means that a player is seated to the left of the large blind, and is thus forced to act before seeing the actions of others at the table. The consequences of this fact are significant, as additional wagers may be placed as each player takes their turn, forcing the player in early position to increase their investment simply to continue playing the hand. Conversely, being in late position is beneficial since nearly all players have taken action, and consequently, there is additional information available beyond the two cards held by a player. The dealer is the last player to take an action, and thus has the most information available.

The flow of the game after the deal is quite straightforward, and alternates between revealing cards, and a round of betting. Initially, three communal cards, known as the flop, are revealed. At this point, players have sufficient cards to form a full hand, and additional betting occurs. Next, an additional communal card, known as the turn, is revealed. This is followed by further betting. Then, the last communal card, known as the river, is revealed. After a final round of betting, player hands are shown, and the winner takes the pot. It is entirely possible that
the strongest hand in a round is formed entirely of communal cards, in which case the pot is split between all players still in the hand.

This gradual reveal of information reinforces the point that information is paramount in Texas Hold’Em Poker. Information about one’s own cards, communal cards, and opponent behaviour are the only inputs available for making decisions. The third item, however, becomes significantly limited when the environment shifts online. Suddenly, players at the table cannot read opponent body language and conversation is likely reduced – both these changes result in a reduction in information.

2.3.2 Online Poker

The loss of information is not the only factor that changes when poker moves online. The pool of potential participants increases dramatically, as players can now join from anywhere in the world to play at the same table. A wide variety in entry fees and betting limits ensures that there is a table suitable for whatever a given player wishes to wager.

Beyond joining international tables, there is nothing limiting players to a single table. Although face-to-face playing of multiple games simultaneously may be feasible on a small scale, it becomes considerably simpler online. Consequently, players are now able to multiply their potential gains or losses without needing to play games with high stakes. This is known as multi-tabling, and can be taken to extremes - demonstrated by the current world record in which a player achieved a profit after playing 62 tables simultaneously for an hour (Connors, 2009).

Finally, the lack of both physical cards and a human dealer can lead to quicker gameplay. Cards are dealt instantly, players can pre-select their actions to have their turn occur immediately when available, and bets are placed without any moving, counting or verification of chips. Site
operators no longer have to pay staff to act as dealers or security, and players are able to both win and lose money more quickly.

The combination of international access, multi-tabling capabilities, quicker gameplay, and low costs has led to the growth of an enormous industry. One of the more popular poker sites, PokerStars, has an average of over 28,000 players wagering real money at any given time. When including those playing for solely entertainment, that number rises to over 100,000 (PokerScout, n.d).

While PokerStars does account for over 50% of online poker gameplay, it is estimated that over 60,000 players are gambling in online poker with real money at a given moment (PokerScout, n.d.). Taken with the fact that there are over 600 known poker sites at this time, and PokerStars has an estimated annual revenue of $1.4 billion (Vardi, 2010), the size and scope of the online poker industry cannot be disputed.

Given the gameplay changes, multi-tabling capabilities and amount of money changing hands, it’s no surprise that some have sought to enhance their playing capabilities with new tools. One example is SharkScope, a database that stores gameplay information on a large number of poker players. This information could be as simple as a win/loss ratio, or as detailed as their performance and profits in specific tournaments. Using this knowledge, players can attempt to supplement the limited information available in online poker, which results from the lack of body language and face-to-face interactions. Some poker sites have explicitly banned many such tools from usage as violation of their Terms of Service (TOS) (PokerStars, n.d.).

The PokerStars TOS identifies five categories of tools that are prohibited: two are directly related to illegal guidance, such as software that would recommend a specific action or share information privately with other players; two are related to mass gathering of information,
such as software that tracks hand history or player activity; and the last one relates to software that can play the game without requiring human input – this software is typically known as a **bot**.

Bots are capable of integrating the capabilities of the other tool types. They can gather information passively while playing or spectating, without compromising gameplay. They are able to actively analyze a situation to determine the most appropriate action. Ultimately, they can integrate a large amount of additional information and act on it, without requiring direct human intervention. By having a bot programmed to play in their place, someone can enhance their multiplaying capabilities and maximize potential profits by leveraging the bot’s advanced computing power and infallible memory.

Beyond the TOS limitations, there are other reasons why poker bots are not used by more players. In addition, there are many bots that are either far too detectable, through manual or automated means, and there are many bots whose skill level is too low to be effective. The creation of effective poker bots, and determining how they can be configured to maximize their skill, is no small feat as will be outlined in the next section.

### 2.4 Gaming Bots

Gaming bots are not a new concept – they have long been a focus of Artificial Intelligence (AI) research (Turing, 1953), and have been investigated in a large number of game genres. The work in the AI field has led to simple games like Checkers being “solved”, i.e., these game bots are capable of playing perfectly, every single time (Schaeffer *et al.*, 2007). Consequently, a draw is the best result that a human opponent can hope to achieve, if avoiding a loss is even possible. However, not all bots are quite so successful.
2.4.1 Overview of Gaming Bots

Games that have been solved range from the exceedingly simple, such as Tic-Tac-Toe (Pilgrim, 1995), to more complex games, such as Checkers (Schaeffer et al., 2007). The most capable bots are able to push opponents to a draw in every game, while others simply compete at the same level. There are other games that have been solved to lesser degrees, such as Chess (Hsu, 2002), and Othello (Buro, 1997), in which bots have been known to defeat the best players. In these cases, the game is not fully solved, but rather solved for simpler variants, such as a reduced board size or for a reduced portion of the game.

A trait common to the aforementioned games is that they are deterministic, meaning that the immediate result of a given action will always be the same and can thus be reliably predicted. Additionally, all players are aware of the same amount of information at any given time – there are no hidden pieces or unknowns to factor into a decision-making process. When a given action does not have a consistent outcome, the game is, by definition, no longer deterministic.

This can be demonstrated with an example scenario from one-on-one poker, in which a single bot is playing against a single player. The bot’s private cards are an ace and a king, while the opponent has a seven and a two. The bot raises due to the quality of its hand, and its opponent folds. Two rounds later, the bot’s private cards are, once again, an ace and a king. The opponent’s private cards are now two aces. The bot raises again, yet this time the opponent also raises. To the bot, the situation was identical, as all known information was identical. However, the missing information, the opponent’s cards, led to a completely different outcome. This example serves to not only demonstrate the fact that poker is non-deterministic, but also shows exactly how lacking information can have a significant impact on poker gameplay.

From an AI point-of-view, poker has value as a research environment due to the uncertainty of the information a bot needs to use to make decisions. As the visible information is
insufficient, bots are forced to track and interpret opponent behavior to determine the best path to victory – this is typically described as opponent modeling.

2.4.2 Poker Bots

Gaming bots have demonstrated considerable success against human players, and continue to be investigated within the field of AI. For example, poker has become the focus of a number of research groups, such as the Computer Poker Research Group (CPRG) at the University of Alberta.

A number of reasons for this focus have been identified by Papp (1999). To begin with, poker games have multiple players, each competing to maximize their own profit. This increases the amount of information that must be factored into every decision. In addition, that information is both unreliable, as opponents may bluff, and incomplete, since opponent cards are hidden. A successful bot must use this incomplete, unreliable information in order to predict opponent behaviour. These predictions then cause the bot to adapt its behavior in order to employ betting and bluffing strategies to exploit opponents’ weaknesses, and maximize its profits. Due to the absolute necessity of modeling and adapting behavior for success, poker is a fruitful environment for AI researchers to investigate.

One of the most important points worth noting is that attempting to examine poker in simplistic environments, such as one-on-one games, has a risk of changing the problem being examined. A considerable amount of complexity is lost when the game is limited to two players, and this may only be exacerbated by instituting betting limits (Papp, 1999). Consequently, bots are typically tailored towards a specific type of play: ring, in which many players participate, and heads-up, in which there is only one opponent.
Poker strategy can be viewed as a combination of two factors known as aggression and tightness. Aggression relates to how much money a player is willing to risk, while tightness refers to the likelihood of playing a given hand. The relationship between the two factors is then modified according to opponent behavior (Chaddock, Pickett, Armstrong & Oates, 2007). Due to the wide variety in gameplay strategies employed by poker players, any given combination of these two factors may result in an approach that will exploit a given player’s weaknesses (Johanson, 2007).

Though there are many poker bots in existence today, those from the CPRG are some of the best known. This group has been dedicated to poker bot research since 1995 (University of Alberta Computer Poker Research Group. n.d.), and their bot won in three of six events at the most recent Computer Poker Competition (Jackson, 2013). Consequently, their research forms a solid base for both the variety and evolution of Poker Bots.

### 2.4.3 CPRG Poker Bots

The Loki bot was the first bot proposed by a member of the CPRG – it was not sufficiently skilled to compete with humans, but served as a proof-of-concept; a bot can play poker effectively, and model its behavior to take advantage of opponent weaknesses. Its weakness likely stemmed from a rigid betting strategy that did not take opponent behavior into account (Papp, 1999).

The next attempt, Poki, was far more successful. It was crafted in 1999 and was still winning competitions as recently as 2008 (Abou Risk & Szafron, 2010). It determines an overall betting strategy at the start of every hand, and then adjusts its strategy based on hand strength, and potential value of available actions. Lastly, and possibly most importantly, it adapts its
behavior based on observed opponent actions to determine which action has the best possible outcome (Johanson, 2007).

According to Johanson, Poki excels in ring play but tends to fall short in heads-up play. Consequently, it can be beaten by intermediate human players in many scenarios. Its heads-up weakness may be a result of the different approach required for one-on-one play – specifically, the factor of honesty. When only two players remain in a game, the success of a bluffing strategy can increase dramatically.

Poki was further refined in a later bot, called Polaris (Johanson, 2007). Polaris took the lessons learned from previous bots, like Poki and one called Hyperborean, to create a flexible strategy that takes a different approach for different situations. It is capable of reading opponent behavior to determine which tactic is most appropriate, essentially acting as a combination of the previous CPRG bots. The resulting bot played competitively against professional poker players and at all-bot competitions.

The most recent bot exploration from CPRG involves expanding the Hyperborean bot to both no-limit (Schnizlein & Bowling, 2009) and ring (Abou Risk & Szafron, 2010) games. Both involve further exploiting the lack of information in a Hold ‘Em Poker game, and tailoring the strategy towards complete betting freedom or additional opponents. Hyperborean’s Ring version is currently capable of soundly defeating the previous champion, Poki.

Overall, as poker bots continue to evolve, they are very likely to be able to defeat even the most skilled professional players in a ring game. Heads-up bots are already capable of defeating world-class players (Abou Risk & Szafron, 2010), so it is just a matter of determining the appropriate approach. As bots begin to handle more and more diverse situations, such as ring
poker or no-limit betting, it is very likely that regulation and restriction of bots in online gambling will become a pressing problem to solve.

2.5 Bot Detection Methods

Due to the refinement of gaming bots, poker and otherwise, it has become more and more important to be able to detect and remove them from games. As stated earlier, bots are capable of upsetting game balance, including both internal and external economies. For online gambling, the issue is even more urgent, as real money is being lost to bots with distinct advantages. As bots become more and more profitable, their ability to play without experiencing fatigue allows them to play far more hands than any human.

In recent years, there has been a significant research focus upon automated bot detection solutions, particularly for the MMO RPG genre. This is likely due to the rapidly rising popularity of that game genre and the valuation of some virtual items in real dollars (Woo, Kwon, Kim, Kim & Kim, 2011; Platzer, 2011). Most automated bot detection solutions are passive, and involve analyzing player behavior during actual gameplay. The most common detection solution, the CAPTCHA (von Ahn, Blum, Hopper, & Langford, 2003; Golle & Ducheneault, 2005), is an active one in which users must answer a question to prove that they are human. The CAPTCHA is still used in some online games today, including many discussion board systems and internet communication sites. CAPTCHAs have been criticized, however, for ruining the immersive nature of a game and interrupting the gameplay of players who are obeying the rules (Gianvecchio, Wu, Xie & Wang, 2009).

Automated systems tend to focus on a particular aspect of gameplay in order to identify bots – this could be on their social gameplay, movement patterns, communication with the game server, input patterns, etc. Detecting each of these aspects requires customization to a particular
game environment, and may prove to have only short-term success if the detection mechanism becomes known to users (Kang, Woo, Park & Kim, in press). Each automated detection method has benefits and drawbacks – none have been demonstrated to be infallible.

2.5.1 Behaviour-Based Bot Detection

Behaviour-based bot detection systems are likely the most diverse – depending on the system, any given factor may be considered indicative of “botness”. Behaviours examined may include movement patterns, frequency of actions, and group play dynamics.

Laurens, Paige, Brooke and Chivers (2007) opted to examine line-of-sight tracing and movement patterns in FPS games. The focus on this environment means that the bots being examined are supportive rather than fully autonomous – they improve aiming and allow human players to see through opaque obstacles. By examining anomalous line-of-sight tracing and navigation around obstacles, Laurens et al. were able to achieve a successful detection rate of 70%.

By contrast, Chen and Hong (2007) focused on the lack of action, or idle time, as their key identifier. This resulted in a successful detection rate of 90%. In a similar effort, Thawonmas, Kashifuji and Chen (2008) examined a MMORPG using server logs of user actions. By examining the actions that users took, the variety of actions, and the frequency in which they were performed, they were able to identify autonomous bots with a successful detection rate of 95%.

While both these solutions are generalizable to other environments where logs exist, they are also dependent upon secrecy of the detection mechanism – once bot creators are aware of the detection method, bot behavior can be adjusted to reduce the chance of being detected.
Kang et al. (in press) investigated the same type of environment, a MMORPG, but looked specifically at logs of players who “team up”, or “group”, to complete objectives. By considering the duration and size of groups, as well as the types of actions they performed, Kang et al. were able to achieve a successful detection rate of 95.92%. Although slightly superior to the detection rate of Thawonmas et al. (2008), this solution suffers from the same drawback – once the detection method is known, bot creators can adjust bot behavior to avoid detection by engaging in a larger variety of actions or reducing the duration and frequency of their group play.

Altogether, utilizing an automated bot detection system based on behavior analysis is a valid approach with high accuracy – however, it must be continually maintained to counter changing bot behavior, and in many cases, must be tailored towards the game environment in which it is being used. Additionally, it requires a data set to analyze and is dependent upon the presence of differences in bot and human behavior.

2.5.2 Traffic-Based Analysis
Traffic-based analysis is a lower-level approach to bot detection – instead of relying upon observable behavior, it relies upon server-side monitoring of packets from the client. Chen et al. (2009) propose that examining frequency of packets, as well as how this rate is impacted by network latency and bulk sends, will provide an accurate indicator of whether a player is a bot. In the limited sample, the proposed method achieved an accuracy rate of 90%, while minor adjustments improved this rate to 95% at a cost of 1% false negatives. Due to simply analyzing traffic, this approach is far more generalizable than those bound to behavior-based analysis.
However, a sophisticated solution by a bot creator to make the bot better emulate human player traffic, rather than simply attempting to randomize packet send intervals, could render this system ineffective as well. Chen et al. conclude by stating that manual detection coupled with
automated detection would be the most effective solution. By collating user reports, administrator review and automated detection, bots can be reliably identified and removed from game environments.

2.5.3 Manual Bot Detection

The time and effort involved in manual detection on a large scale can be extremely difficult to accommodate (Wendel, 2012) and many automated solutions depend upon a final step of manual analysis by an administrator prior to taking action against a given player. This is done to reduce the chance of false positives, which may have significant backlash from a game community (Wendel, 2012; Chen et al., 2009). Manual methods also lack the need to be customized to a given community, as investigations are conducted by administrative staff familiar with the game environment.

Since a manual detection system would be impossible to scale to examine the behaviour of every single player, a flagging method is typically established. Many online games, regardless of genre, support the notion of community reporting. The other primary method is the aforementioned flagging by automated systems. Both of these flagging methods are subject to a number of concerning flaws. While community reporting can technically increase the number of investigators who can help flag bot usage, it can also expose the system to a number of sociological risks (Chen et al., 2009). The findings of Dellarocas (2003), who examined community reporting systems in relation to reviews, can also be applied to gaming communities due to the similarities.

One common sociological risk is ballot stuffing – bots and collaborators can simply increase the number of reports and report randomly in an effort to overwhelm the system and
hide their own actions. While simplistic, this method can still effectively inundate reviewer queues, making them much more difficult to process.

Second, there is the simple issue of false accusations – players or bots can collaborate in order to cast a target in a negative light, wasting administrative resources for reviewing and risking the ban of an innocent. This is exacerbated in games where players can quickly create new accounts and identities, or amass them over time, to collaborate for malicious purposes. For this reason, systems often take account age and reputation into account when considering community reports.

Lastly, there is the simple matters of numbers – a single report against a single player is likely insufficient to warrant assessment by an administrator. Acting on every single report would just re-introduce the scalability problem, so each administrative team needs to define a threshold for investigation. This method may also increase vulnerability to false accusations as colluding accounts can increase reporting rate fairly easily.

Although these flaws can be significant, community reporting will typically reduce the overall scope required for manual bot detection, making it more manageable for administrators to review flagged accounts. However, even if a flagging method were to work flawlessly, a manual detection system may break down at the last step – administrator review. If an administrator is unable to reliably differentiate a bot from a human, then detection systems, whether manual or automatic, may fail. This would result in false positives, banning those who are not using bots, and false negatives, permitting those who are using bots. For this reason, it is of the utmost importance that an administrator can differentiate between bots and humans based on observing behavior.
2.6 Accuracy of Manual Bot Detection

Due to its subjective nature, the success of manual bot detection efforts will depend upon the methodology used by the detecting agents. This variance has led to studies exploring how successful observers are in differentiating bots from humans. These studies examined manual bot detection in a small variety of game environments, and determined a number of factors that are leveraged as rationale for detection. The following subsections describe these studies in detail, including their environments, methodologies, and findings.

2.6.1 Experimental Environments

Though there have been a number of studies exploring bot detection by human participants, there tend to be many similarities in the experimental environments utilized by the researchers. A number of researchers have focused on bot identification in First-Person Shooter (FPS) games (Laird & Duch, 2001; Clarke & Duimering, 2006; Gorman, Thurau, Bauckhage & Humphrys, 2006; Ryan, 2007; Chen et al., 2008; Hingston, 2009; Conroy, Wyeth & Johnson, 2011) while others have focused on the environment of a Massively Multiplayer Online Role-Playing Game (MMORPG) (Cornelissen & Grootjen, 2008; Chen et al., 2009). Within these environments, player behavior and interactions are both easily observed. This large amount of available information may lead to potentially higher success in differentiating bot players from human players (Clarke & Duimering, 2006). For example, a player in a FPS may be stuck behind a relatively minor obstacle, which could be indicative of a poor quality bot, or be able to score a headshot against a target from an enormous distance, which could be indicative of a high-quality bot. With MMORPGs, players tend to use bots to avoid the tedium of repeating actions, known as “grinding”, and this pattern of repetition is apparent when observing the character over a lengthy period of time (Cornelissen & Grootjen, 2008).
Conversely, McGlinchey and Livingstone (2004) examined bot detection within the simple game environment of Pong. Pong presents an extremely limited set of actions for a player to take: moving the bat up, and moving the bat down. Participant accuracy in bot detection varied significantly, resulting in either being correct in most cases, or incorrect in most cases. It was concluded that humans were able to perceive a difference in the two player types, but unable to reliably assign the correct player type once the difference was identified. Participants reported that stuttering observed in player bat movement was a large factor in their decision – however, some felt this was indicative of a bot, while others felt it was indicative of a human.

The environments described so far mostly represent either extremely simple games, wherein there is a limited element of strategy and available actions, or reflex-based games where player behavior provides a number of visible yet ambiguous cues for bot detection such as: accurately aiming a gun, poor path finding, lightning-fast reflexes, lack of hesitation, etc. (Clarke & Duimering, 2006; Laird & Duchi, 2001). A recent study by Hagelbäck and Johansson (2010) provides a comfortable medium between these two extremes by examining bot detection within a Real-Time Strategy (RTS) game. Within RTSs, there are typically an enormous number of actions available to players, but the decision time observed is slower than in a FPS. The large number of possible actions, taken in combination with the precise unit micromanagement required, causes it to be a more complex environment than Pong, but lacking the reflex-based gameplay of a FPS.

Hagelbäck and Johansson’s 2010 examination of bot detection in a RTS game also gathered a large amount of qualitative data in the form of comments from judges. Three different bots were used in a number of head-to-head games, and judges observing the games were asked
to determine whether each player was a bot or a human. It was found that factor categories such as multitasking, overall strategy, base planning, skill, decision speed, game knowledge, intelligence, and ability to adapt were all used by judges to detect bots. Out of these factors, only multitasking, which involved splitting attention between a number of individual unit movements, and decision speed reliably led to a correct response. The remaining factors were not found to be reliable for bot detection, as they could conceivably be indicative of both bots and humans, whether skilled or not, and were thus ambiguous.

Conroy, Wyeth & Johnson (2011) demonstrated that multitasking can be considered a differentiating factor in a FPS. In such an environment, a player will frequently combine multiple actions simultaneously – scanning for threats while running, tracking an enemy and firing. In the 2011 study, however, it was found that the lack of multitasking by a given player in this environment that led judges to correctly assert that the player was a bot. The bots would perform convincingly well at a single given task, but tended to avoid combining actions and tasks, which judges found suspicious. Thus, multitasking can be a valuable indicator for bot detection in both a RTS and FPS – however, its influence is different in each, and thus dependent on the environment and actions available to players.

2.6.2 How Players Detect Bots

Irrespective of game environment, there are a number of factors that have been established as significant influences upon bot detection. Decision speed has been identified as one of these factors (Laird & Duchi, 2001; Chen et al., 2009; Cornelissen & Grootjen, 2008; Hagelbäck & Johansson, 2010), and can be interpreted in two ways: decision intervals and decision time. In order to detect bots, automated detection systems tend to utilize regularity of decision intervals (Cornelissen & Grootjen, 2008; Chen et al., 2009), which are the time periods between each
decision made, while human observers tend to rely on the time a player takes to make a given decision, which could be considered a player “thinking”. Laird and Duchi (2001) also found that both faster and slower decision times were considered increasingly botlike as they deviated from the expected human time of 0.1 seconds. Evidently, the fact that a decision time of 0.1 seconds is indicative of a human player applies primarily to games in which reflexive actions are expected; in less reflex-based games, significantly longer decision times could be expected. The evaluation matrix employed by Hinkannen et al. (2008) suggests a similar level of importance as outlined in Laird and Duchi (2001), with “human-like reaction times” being a heavily-weighted measure of believable AI. Thus, the current literature suggests that decision time has a very strong influence on bot detection.

A second factor, skill, has also been identified as an influence on botness (Laird & Duchi, 2001; Gorman et al., 2006; Ryan, 2007; Hinkkanen, Kurhila & Pasanen, 2008; Hingston, 2009; Hagelbäck & Johansson, 2010), and can be interpreted differently for every game. In a FPS, skill have been measured by the ratio of kills to deaths and aiming precision (Laird & Duchi, 2001). In a RTS, skill has been measured as an efficient use of resources and simultaneous unit movements (Hagelbäck & Johansson, 2010). In a MMORPG, skill has been measured by completing multiple quests simultaneously, strategic use of combat abilities and leading a large group of players effectively (Taylor, de Castell, Jenson & Humphrey, 2011). In Pong, skill is communicated by the difference in player scores, providing an approximation of player reflexes and effectiveness in the game. In a FPS, Laird and Duchi (2001) found that aiming precision was a significant influence upon player type classification. It was explicitly noted by all participants that the best shooters could not possibly be human due to their high accuracy levels. Conversely, Hingston (2009) has also investigated the relationship between skill and botness in a
FPS environment, and found that there was no significant correlation between the two factors – a skilled or unskilled player could not be reliably identified as a bot.

Similar ideas are included in the evaluation methodology by Hinkannen et al. (2008). “Humanness” points are given for “bad aim upon seeing player for the first time”, indicating that good reflexes and aiming precision are associated with bot players. Lastly, Gorman et al. (2006) describe player skill as the possession of extensive knowledge of the game mechanics. If a player demonstrated knowledge of how to effectively utilize the “rocket jump” maneuver, an advanced movement technique in the game “Quake”, participants were influenced to classify them as human. Thus, a given player may be identified as a bot or a human, depending on how skill can be demonstrated within the game environment.

The last factor, error rate, has been identified as a significant influence on bot detection (Gorman et al., 2006; Ryan, 2007; Hinkannen et al., 2008; Hingston, 2009) and includes two elements: intentional mistakes and frivolous actions. Both of these factors have been shown to increase the humanness of a given player in most cases (Gorman et al., 2006; Hinkannen et al., 2008), though they are highly dependent upon the game environment. Hingston (2009) has also identified mistakes considered “missing behavior”, such as a bot not reacting to enemies or failing to execute basic tasks, can be a reliable indicator of botness for observers.

### 2.6.3 Ambiguous Interpretations of Player Behaviour

One of the more significant influences discovered in prior studies is the fact that a given behavior can be considered botlike or humanlike, depending on the observer (McGlinchey & Livingstone, 2004; Ryan, 2007; Hagelbäck & Johansson, 2010). Since participants identify bots based on observed behaviour, two participants may view the exact same behavior and come to different conclusions as to whether the player is a bot or a human – the behavior is considered to be
ambiguous, as it can be attributed to either player type. For example, in Ryan (2007), a participant judged a bot to be human because there “seemed to be fewer ‘good’ items … [and the other player was] obtaining them before I was”, whereas a participant judged a human to be a bot because it was “finding the food quicker”. Additional examples of ambiguity can be found in the study examining Pong (McGlinchey & Livingstone, 2004). Participants observed that “the bots moved with more jerky and sudden movements”, and some concluded that this indicated a bot player while others believed the same behaviour indicated a human player. Lastly, many of the factors identified in Hagelbäck and Johansson’s study (2010), such as intelligence, overall strategy, ability to adapt, and game knowledge, can be classified as ambiguous since they were attributed to both bots and humans, depending on participant opinions.

Thus, it has been demonstrated that manual detection can be both ineffective and effective at differentiating bot players from human players, depending on the environment studied. The environment is the key differentiator – it determined whether a given factor reliably identified bots, did not influence bot identification, or confounded successful bot identification by being attributed to humans. This thesis extended the study of manual bot detection into an entirely new type of environment, Texas Hold ‘Em Poker.
3 Methodology

This chapter describes the study hypotheses, rationale, participants, materials and procedure. The participant recruitment strategy is outlined, as well as some key demographics and questionnaire responses. Study materials are described, including the experimental software, the bot type, and the particular trait configuration for each bot used in the study videos. Lastly, the full procedure of the study is described in order to facilitate any efforts to replicate the study.

3.1 Overview

A single study was conducted in order to explore three hypotheses:

1. Expert poker players will be unable to identify a difference between human and bot players when observing a game of online poker.

2. Expert poker players will come to similar conclusions when asked to identify the best and worst players in a game of online poker.

3. Skill will not be a reliable indicator as to whether a given player is a bot or a human.

Expert poker players were recruited as participants for the study, and they observed pre-recorded footage of online poker gameplay. Nine of ten players in the recorded game were bots, configured to emphasize different preset gameplay traits, but participants were deceived into believing that a larger number of humans could be playing. One player was human, which was needed due to a requirement of the software since bots could only be used in a mode made for practicing poker. Participants were then asked to rate how botlike each player was on a scale of 1 to 5, where 5 equated “very botlike”, and 1 equated “very human”. Additionally, participants ranked all players in terms of their playing skill, regardless of whether they considered the player to be a bot or a human. Consequently, this study gathered expert observations, ratings and rationale concerning how botlike and how skilled each player was perceived to be. As each
player was configured to emphasize different traits, the goal was to identify which traits correlated with botlike ratings.

### 3.2 Rationale

This study design and methodology was primarily driven by modification of methods used by prior studies examining other game environments (Laird & Duchi, 2001; Ryan, 2007; Hagelbäck & Johansson, 2010; McGlinchey & Livingstone, 2004; Gorman et al., 2006). Major decisions, such as the use of an online poker environment and fixed decision times, were guided by the findings and methods of these previous studies.

#### 3.2.1 Use of a Poker Environment

A wide variety of genres for online gaming were candidates for use in this study, but a decision was made to use online poker as the experimental environment. As described in Section 2.3, online poker is a complex strategy game that has a limited number of actions, an emphasis on opponent observation, and can be quite lucrative for skilled players. It can be considered more similar to Pong since it has a limited set of actions and less observable information when compared to a FPS, RTS or MMORPG. While a player in a FPS, RTS or MMORPG has visible movement patterns and takes precise actions within the environment, a player in online poker only demonstrates betting and folding behavior by way of the same number of cards available to all players.

Online poker is a turn-based, strategic game with a large number of ways in which player behaviour can vary within the limited set of actions. This produces a large quantity of behavioural information for participants to use in identifying bots, and which become more detailed as a game is played. The following is an example of this: poker players may temporarily
forego their primary goal, gaining the most chips, in order to deceive opponents and encourage them to bet and thus have a chance of achieving a larger payoff overall. Poker is an unstudied research environment for this line of research, a game in which there is more strategy than Pong, but less behavioural information than previously used experimental environments. Player strategies and evaluations tend to evolve throughout a game and the game environment itself does not contain the same quantity of visible behavior as found in FPS, RTS or MMORPGs.

Another reason for using poker as the experimental environment was due to its players. Poker is a game in which players often study their opponents in great detail in order to understand their strengths and weaknesses. This information is then used to determine how other players can be exploited for further benefit. As some online poker games may not have an upper limit on decision times, observers are not forced to rush their assessments of opponents. When compared to a FPS game, where a “human” decision time is a mere 0.10 seconds (Laird & Duchi, 2001), it is clear that poker is a more suitable environment for encouraging opponent observation. Consequently, poker affords the opportunity for greater strategy development due to additional time available for analysis. Understanding player strategy and choices, and how their play patterns can be leveraged to benefit the observer is a fundamental aspect of expert poker. As a result, it was anticipated that using a poker environment would mean that the study tasks would come more naturally to the participants recruited. This was the same rationale for recruiting expert participants, as it was assumed they would be more experienced with opponent observation.

3.2.2 Control Over Bot Behaviour

Previous research has defined a number of factors that influence success of bot detection in specific environments (see Section 2.6.2), some of which were isolated and excluded from this
study. Additionally, in order to better study the influence of skill, the experimental environment was leveraged to manipulate poker gameplay traits, and consequently bot skill, on a per-bot basis.

Since the benefit of a short decision time is likely to be less significant in a non-reflex-based game, decision time was expected to be a less influential factor in a poker environment. However, because decision time has been recognized as such a strong influence on bot detection in multiple studies, the current study eliminated this factor altogether by fixing the decision time for all bot players to 1 second. This allowed additional hands of gameplay to be shown in the limited session time and eliminated response time as a factor to be dealt with during analysis, i.e., it reduced the need for more participants as there was one less variable. Lastly, the experimental software used for this study did not permit decision time modification on a player-by-player basis – since this could only be modified for all bots, it was deemed more efficient to negate the factor rather than develop a custom game environment enabling individual/random bot timing.

Although participants were asked to rate the skill of each player in the game, it can be difficult to precisely define skill in poker due to the random element of the cards dealt. As a result, participants were tasked with forming their opinions of player skill after observing a number of hands of gameplay. Depending on the observer, skill can be evaluated by examining chip count, hand win/loss ratio, overall play style, betting habits, actions taken with both good and bad hands, and so on. However, in order to leverage the attributes built into the Pokibot, a bot developed for play with multiple participants (Davidson, 2002), the concept of skill was separated into the four configurable behavioural factors: **risk aversion**, which determines the likelihood to fold a hand that is unlikely to win; **honesty**, which determines the likelihood to
bluff, playing as if different cards are held; **aggression**, which determines the likelihood to bet or re-raise an opponent; and **adaptiveness**, which determines whether the player adjusts strategy to exploit opponent weaknesses.

The aforementioned four factors of skill represent the behaviour scales found in the bot customization for the software that was used in this experiment. This particular bot was developed by the CPRG in Alberta, which is known for its poker bots submitted to international competitions.

There is little data available concerning which aspects of poker skill are likely to be most influential, and this was not the primary reason for leveraging this environment – therefore, this segmentation of skill was primarily exploratory. However, it had been hypothesized that due to the complexity of the concept of skill within this environment, and the fact that humans can compete with bots due to the lack of reflex-based gameplay, it would not have a strong influence upon success of detecting bots.

Although identified as a potential indicator of bots in prior studies, it was not believed that error rates would have any impact on humanness within a poker environment, as the strategy is more complex than in other games. For example, many actions that appear to be mistakes may actually be done intentionally: used to deceive opponents who are attempting to identify a given player’s weaknesses and patterns. When other environments like MMORPGs and FPSs were considered, “unnecessary jumping” and “fires gun for no reason” (Gorman *et al.*, 2006; Hingston, 2009) have been given as examples of errors. While players in a poker environment may continue betting with a poor hand, it is not clear whether the choice was intentional, and does not provide additional hints about whether the player is a bot.
3.3 Participants

Participants recruited for this study were required to be 18 years of age, and established by a reputable source to be highly skilled at Texas Hold ‘Em Poker. The participants were seven members of an online Texas Hold ‘Em Poker tour and two members of a well-respected online poker forum. Details about participant demographics can be found in Table 3.1 and Table 3.2 below. The first set was recruited through an administrator of the Red Hot Poker Tour (RHPT), an organization that hosts a number of poker tournaments in the provinces of Ontario and British Columbia in Canada, and consisted of 7 members who were among the Tour’s top 50 players. The second set was recruited through a well-respected member of the Two Plus Two Poker Forum (2p2), a poker strategy forum consisting of over 250000 members, and both were considered to be highly skilled by their peers in the community. Ratings from the third RHPT participant (P3) were dropped due to a refusal to respond to any questions concerning the classification of player type or skill. All participants had over four years of experience playing Texas Hold ‘Em Poker, and the majority of their playtime was completed online rather than in-person. One participant had briefly used the trial of the experimental software, Poker Academy, but had stopped using it long before the study took place. The other eight participants had no previous experience with the software.

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>43.6</td>
<td>22-57</td>
</tr>
<tr>
<td>% played Online vs Offline</td>
<td>85.4%</td>
<td>50%-99%</td>
</tr>
</tbody>
</table>

Table 3.1 Participant Age and Play Location
<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of experience (offline)</td>
<td>5+ years (6)</td>
</tr>
<tr>
<td></td>
<td>4 years (3)</td>
</tr>
<tr>
<td>Years of experience (online)</td>
<td>5+ years (3)</td>
</tr>
<tr>
<td></td>
<td>4 years (3)</td>
</tr>
<tr>
<td></td>
<td>3 years (3)</td>
</tr>
<tr>
<td>Frequency of watching poker (TV/Online/etc.)</td>
<td>Monthly (2)</td>
</tr>
<tr>
<td></td>
<td>Weekly (1)</td>
</tr>
<tr>
<td></td>
<td>Every few days (4)</td>
</tr>
<tr>
<td></td>
<td>Daily (2)</td>
</tr>
</tbody>
</table>

Table 3.2 Gameplay Experience and Frequency of Observing Poker

3.4 Materials

Stimuli for the study consisted of 141 brief videos, each of which contained between 1-20 seconds of recorded gameplay from the Texas Hold ‘Em Poker software called Poker Academy (see Figure 3.1 for a still from a video). A free version of BSR Screen Recorder was used to capture the videos, and ffmpeg was used to convert them to a flash video format. For local sessions, the facilitator’s laptop was used to display the videos. For remote sessions, the videos were hosted on spotthebot.boxcarhosting.com, and shown using a free video plugin called FlowPlayer.
The software used for constructing gameplay videos was Poker Academy Pro 2 (Poker Academy Pro, n.d.) which was a Texas Hold ‘Em Poker application that allowed players to either compete locally against artificial intelligence players (bots) for the sake of training, or compete online against human players. Poker Academy’s bots were based upon research by the Computer Poker Research Group (CPRG) from the University of Alberta (UACPRG, n.d.), and the Pokibot AI, a bot specializing in poker play with multiple participants (Davidson, 2002), was used for the players in the study videos. The videos showed nine Pokibots and a single human, the facilitator, playing 30 hands of Texas Hold ‘Em Poker. Due to the limitations of the application, a single human player, represented by Seat 1, was necessary to create the gameplay videos and played an arbitrary action each round. All player names were modified to simply display as their seat number to avoid any non-gameplay-based indicators of humanity.
Each of the bots represented a specific configuration of Pokibot, in order to identify whether certain traits are more likely to be viewed as “botlike”. The default configurations for Pokibot were randomized when the bot player was added to the game, likely to ensure varied gameplay for users of the software. The software permitted a maximum of 10 players per game, of which only 9 could be bots, so all options were able to be singularly emphasized within a single bot. However, a number of configuration options were available for further customization:

- The first, pre-flop looseness, dictated how likely a bot was to stay in the game based on its initial hand, before any other information was available.
- The second, a “Loose/Tight” slider, dictated how likely a bot was to stay in the game based on its hand, including consideration of all other information currently visible on the table. The tighter a bot was set to play, the more likely it was to fold if its chances of winning were low.
- The third configuration is a “Passive/Aggressive” slider, and dictated how likely a bot was to bet or raise, in which additional wagers are made. The more aggressive a bot is set to play, the more likely it was to bet or raise.
- The fourth, a “Honest/Tricky” slider, determined how likely the bot was to bluff, in which it bet and behaved as if it possessed cards it did not.
- The final configuration, a “Math/Model” slider, determined how much the bot selected its actions based on the behavior of other players rather than the mathematical odds of success.

The configurations used for each bot are shown in Table 3.3 below. Cells highlighted in white indicate deviations from the standard bot settings, while cells in grey indicate parity. Due to the lack of a program-defined default bot configuration, the base configuration used for this
The experiment was for the trait sliders to be set directly in the center, at position 5, to result in balanced behaviour (50%). With the exception of the bots in seats 9 and 10, each bot configuration involved a single trait slider being increased or decreased by 3 positions (30%).

<table>
<thead>
<tr>
<th>Seat#</th>
<th>Risk Aversion (Tightness)</th>
<th>Honesty (Bluffing)</th>
<th>Aggression (Betting)</th>
<th>Adaptiveness (Modeling)</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HUMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>80%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3 Differences Between Bot Configurations. Grey cells indicate equality, slider set at position 5 (50%)  

- Seat 2 had the “Loose/Tight” trait slider set to the 80% tight position, and was set to be a tight player in the Pre-Flop as well. This led to fewer hands being played, as the bot was less likely to continue playing unless the odds were in its favour.
- Seat 3 had the “Loose/Tight” trait slider set to the 20% tight position, and was set to be a loose player in the Pre-Flop as well. This led to more hands being played, as the bot was more likely to continue playing even when it had a slim calculated chance of winning.
- Seat 4 had the “Honest/Tricky” trait slider set to the 20% tricky position, and was set to be a 20% tricky player in the Pre-Flop as well. This led the bot to bluff rarely, only betting when it had a strong hand, and folding or calling when it had a poor or mediocre hand.
● Seat 5 had the “Honest/Tricky” trait slider set to the 80% tricky position, and was set to be an 80% tricky player in the Pre-Flop as well. This led the bot to bluff often, bet more with poor hands, and make small bets to conceal good hands and maximize payoff.

● Seat 6 had the “Passive/Aggressive” trait slider set to the 80% aggressive position. This led the bot to bet and raise more often than it called or checked.

● Seat 7 had the “Passive/Aggressive” trait slider set to the 20% aggressive position. This led the bot to check and call more often than it bet and raised.

● Seat 8 had the “Math/Model” trait slider set to the 20% modeling position. This led the bot to rely on the absolute chance of its own hand winning based on the information known to it at the time the decision was made. There was no converse of this bot, as the 30 hands were not a sufficient period of time for behaviour modeling to impact the bot’s actions.

● Seats 9 and 10 were initially set to play faster or slower than the other bots. However, as the decision times were set to one second for each player, this difference was dropped. Therefore, there were no differences between players 9 and 10, and both represented a default, balanced bot, by having each configuration set to 50%.

● Seat 1 was the only human, and simply selected an action quickly in order to avoid taking an amount of time inconsistent with the decision time of the bot players.

All bot player actions were set to take exactly one second, under the guise of ensuring the videos were shown quickly enough to enable participants to view a large number of hands in a short period of time. Due to software limitations, this configuration could not be customized for each bot by the application, and therefore the response time was fixed for all bots in the game. Although this eliminated the influence of decision time, a primary identifier of botlike behavior,
it also enabled investigation of other factors in isolation and reduced number of participants required.

3.5 Procedure

After a brief introduction, participants were asked to fill out the first of two consent forms, both of which were approved by the Research Ethics Board of the University of Guelph. This consent form reiterated much of the overview information provided at the time of recruitment for the study (Appendix A.1). When conducted remotely, this consent form was emailed to participants, and they were required to reply with a message stating that they consent to participate.

Next, participants completed the pre-session questionnaire (Appendix B.1). This questionnaire included demographic information such as age and gender, but focused on Texas Hold ‘Em experience in both offline and online scenarios. Additionally, the questionnaire contained an optional section concerning experience with playing against computer opponents. When conducted remotely, this questionnaire was emailed to participants, who filled it out and emailed a copy back to the facilitator.

At the start of the study itself, participants were told to assume the role of administrators tasked with identifying computer players in online poker games – this was the reason that all cards were shown face-up. Additionally, participants were assured that any data they provided would not be used directly in the development of improved bots for online gameplay, as it was assumed that some participants may have negative feelings towards bots and those who use them. They observed a minimum of 93 videos, which comprised a total of 20 hands of gameplay. Each video contained one of the five portions of a hand of poker: the Deal, the Pre-Flop, the Flop, the Turn, or the River (see Appendix C for description). In some cases, a hand ended early due to players folding, with the result that fewer than 5 videos were produced.
for that particular hand. Each video ranged from 1 to 20 seconds, depending on the number of players who had not folded at the time. Participants were told that the videos showed 10 players engaged in a game of online poker, and that player response times had been fixed to one second in order to guarantee that a sufficient number of hands could be viewed in the time allotted for the experiment. The 10 players were described as a mixture of humans and bots, and participants were told that there was “at least one” of each type in the game. Participants were free to pause, rewind and replay clips as desired, since it was difficult for some to adjust to the brisk pace set by the one second response time of players. Lastly, they were asked to take any notes on behaviour that they felt would be relevant to classifying players as humans or bots.

After the first 20 hands of gameplay had been viewed, participants were asked if they wished to stop and identify players as bots or humans. If they wished to continue to gather additional information, an additional 10 hands of gameplay were available for viewing. All participants stopped after viewing 20-23 hands of gameplay, with an average of 21 hands viewed across all participants.

The questionnaires presented after each video contained both quantitative and qualitative elements (Appendix B.2, Appendix B.3). Participants were asked to identify each player’s “botness” on a Likert scale (1 = not very botlike, 5 = very botlike) – this formed the quantitative portion of the results when paired with the bot settings. The qualitative portion of the questionnaire consisted of the notes taken throughout the study, as well as the written justification for the “botness” rating. By using similar metrics to evaluate elements of humanness and requiring participants to form a scale judgment concerning player types, a direct comparison was possible between the findings from the strategy-based experimental environment and the findings of simple or reflex-based environments such as Pong and Quake.
Additionally, although the environments in the aforementioned studies differed significantly from the experimental environment, the factor of skill was consistently investigated as an influence upon bot detection.

Lastly, the experimenter posed some brief interview questions (Appendix B.3) to obtain additional details concerning the ratings and justifications, as well as to gather thoughts concerning the videos and bots in general. These questionnaires used questions comparable to the scales used in Laird and Duchi (2001), McGlinchey and Livingstone (2004), and Gorman et al. (2006), due to the similarity in experiment methodology.
4 Results & Discussion

This study gathered both quantitative and qualitative data concerning expert classification of poker players as bots or humans. The quantitative results focused on ranking player skill and player botness, and the raw data can be found in Appendix D. The qualitative data focused on general user opinion on bot players, factors that influenced skill and botness decisions, confidence concerning classification, and what could make bots appear more human. The three hypotheses put forth at the end of Chapter 2 are used to structure the discussions in Sections 4.1-4.3. Within each of these section, the analysis of quantitative data is examined first, followed a look at the qualitative findings.

4.1 Identifying a Difference between Bots and Humans

The first hypothesis stated that expert players would be unable to identify a difference between human and bot players when observing a game of online poker in which players are unable to communicate. Raw data relating to this hypothesis can be found in Table D.1.

4.1.1 Experimental Results

The hypothesis was evaluated using an Independent Samples Kruskal-Wallis Test to determine whether bot characteristics (as defined by seat number in Table 3.3) had a significant impact on botness rating. The null hypothesis, $H_0$, was that there would be no difference between the botness ratings across all players. The alternate hypothesis, $H_a$, was that there would be a significant difference between the botness ratings across all players. An alpha value of 0.05 was used for this test.

The results of the test were not significant [$H(9)= 12.140, p = 0.206, ns$], indicating that no seat’s botness rating differed significantly from that of the others (see Figure 4.1).
rejects the alternate hypothesis and supports the first hypothesis, as participants were deceived into believing both humans and bots were playing the game, yet no player was significantly more or less botlike than the others. This finding stands in contrast to the higher bot detection ratings seen in simple games like Pong (McGlinchey & Livingstone, 2004) or those with an abundance of observable behavior and reflex-based gameplay like Quake (Laird & Duchi, 2001). This analysis also shows that the various bot configurations did not have a significant effect on the botness of a given player.

![Botness Ratings vs Seat](image)

**Figure 4.1 Botness Ratings vs Seat**

As a follow-up, participants were asked a series of questions designed to understand their opinions of bots in general, their rationale for player classification, and their confidence in that classification (Appendix B.3). This provided additional insight into key factors influencing botness ratings. Their responses were categorized, and overlapping comments were noted. In some cases, there are fewer comments than there were participants – this is due to a lack of opinion on the part of the participant, or for not responding to a question.

### 4.1.2 Quality of Poker Bots

Questions 6, 8 and 9 in Appendix B.3 focused on exploring participant opinion of the quality of poker bots, and how they could be improved. Participant responses and comments are shown in
Table 4.1 below and grouped under three categories 1) Opinions on Bots, 2) How to Improve Bots and 3) How to Make Bots Believable.

Responses to Q6 (question 6) demonstrate that participants were divided on the quality of poker bots in general: four felt that bots cannot play poker well, three felt that they cannot adapt and only play based on calculations, and two felt that they play perfectly. Few suggestions were provided to improve bots in response to Q8, short of general gameplay hints like adapting to opponents. Q9 explored how bots could be made more believable and evoked a larger number of responses from the participants. The most frequent comment was that intentional errors would make a bot believable, as would taking additional risks and avoiding consistent gameplay. Essentially, participants felt that avoiding “mathematical play”, in which the bot constantly tailors its behavior according to its precise chance of success, was the best way to make a bot believable.

<table>
<thead>
<tr>
<th>Opinions on Bots (Q6)</th>
<th>How to Improve Bots (Q8)</th>
<th>How to Make Bots Believable (Q9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Bots cannot play poker well (4)</td>
<td>• Adapt to other players (2)</td>
<td>• Intentional errors (4)</td>
</tr>
<tr>
<td>• Bots operate purely on probability/math, do not adapt (3)</td>
<td>• Better position playing (1)</td>
<td>• Take risks / bluff (4)</td>
</tr>
<tr>
<td>• Bots play perfectly (2)</td>
<td>• Play aggressively (1)</td>
<td>• Avoid consistent gameplay (3)</td>
</tr>
</tbody>
</table>

Table 4.1 Bot Opinions and Believability, (#) = number of participants who made this comment, n=9

4.1.3 Identifiers used in Differentiating Humans and Bots

Participants were asked about what they watch for when asked to differentiate bots from humans (Table 4.2). This was explored in Q2 from Appendix B.3, as well as in the spoken rationale provided by participants during the experiment.

Again, players identified as bots were seen to suffer from a failure to adapt to the current table situation, but were also seen as maximizing value at all times. Players identified as
humans, on the other hand, showed less aggression, knew the correct time to fold, and adjusted their behavior based on their position at the table. Interestingly enough, all the top human identifiers can be considered aspects of “profitability”, as they all involved minimizing risk and avoiding bad situations.

<table>
<thead>
<tr>
<th>Identifying a Player as a Bot (Q2)</th>
<th>Identifying a Player as a Human (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Profitability (4)</td>
<td>• Hesitant, bet/raise less (2)</td>
</tr>
<tr>
<td>• Does not adapt to situation (2)</td>
<td>• Play position better (2)</td>
</tr>
<tr>
<td>• Overvalue pot odds (2)</td>
<td>• Knowing when to fold (2)</td>
</tr>
<tr>
<td>• Play position better (1)</td>
<td>• Prone to mistakes (1)</td>
</tr>
<tr>
<td>• Consistent playstyle (1)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2 Classification Identifiers, (#) = count of participants who made this comment, n=8

4.1.4 Terms Used to Describe Humans and Bots

Table 4.3 contains the terms used to describe both bot and human gameplay. This data is also sourced from Q2 in Appendix B.3 and from spoken rationale during the experiment.

In this case, humans are seen to be inconsistent and driven by emotions. Conversely, bots were thought to play consistent, mathematical poker and play competently. Only 16 comments concerning gameplay style were provided, but the themes are consistent – humans make mistakes, and bots do not adapt to other players.

<table>
<thead>
<tr>
<th>Bot Gameplay (Q2)</th>
<th>Human Gameplay (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Plays decently (2)</td>
<td>• Anomalous moves (3)</td>
</tr>
<tr>
<td>• Plays consistently, mathematical (2)</td>
<td>• Emotion-driven (2)</td>
</tr>
<tr>
<td>• Bizarre moves (1)</td>
<td>• Plays too many hands (1)</td>
</tr>
<tr>
<td>• Value-betting (1)</td>
<td>• Predictable (1)</td>
</tr>
<tr>
<td>• Not pot-invested (1)</td>
<td>• Good (1)</td>
</tr>
<tr>
<td></td>
<td>• Advanced strategy (1)</td>
</tr>
</tbody>
</table>

Table 4.3 Gameplay Descriptors, (#) = count of participants who made this comment, n=8
4.1.5 Confidence in Identification of Bots and Banning of Offenders

Next, Table 4.5 contains the responses to Q3-Q5 from Appendix B.3, which concerned confidence levels and opinions on banning players identified as bots. Participant responses and are grouped under three categories 1) Confidence in Classifications, 2) Banning Bots Using Manual Detection and 3) Banning Bots Using Automated Detection.

Only two participants claimed to be very confident in their ability to detect bots, while three participants could not qualify their confidence levels. Most participants seemed to feel uncomfortable banning players detected as bots through this manual banning procedure – three explicitly said they would want a large number of additional hands to observe first, while another three would expect bot accounts to be flagged for further review as well. Three participants felt they would be comfortable banning those identified as bots after viewing the videos – this is surprising, considering the fact that no participant successfully identified more than 50% of the bots (Table 4.4).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Bot (Botness rating&gt;3)</th>
<th>Human (Botness rating&lt;3)</th>
<th>Correct (% of players identified as bots)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.4 Percentage of Bots Identified Successfully

Many of the same limitations would apply to an automated banning system – most participants would want extremely high levels of accuracy or a warning system with an appeal
path for impacted players. Two participants felt that automatic banning had no place in online poker, and would never be an acceptable solution.

<table>
<thead>
<tr>
<th>Confidence in Classifications (Q3)</th>
<th>Banning Bots Using Manual Detection (Q4)</th>
<th>Banning Bots Using Automated Detection (Q5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Not confident (&lt;20%) (2)</td>
<td>• Needs many more (1000+) hands (3)</td>
<td>• Warning/appeal system (4)</td>
</tr>
<tr>
<td>• Unsure of confidence (3)</td>
<td>• Need to flag accounts for further monitoring (3)</td>
<td>• Very high confidence required (90-100%) (3)</td>
</tr>
<tr>
<td>• Very confident (&gt;70%) (2)</td>
<td>• Would ban after viewing videos (3)</td>
<td>• High confidence required (75-80%) (2)</td>
</tr>
<tr>
<td></td>
<td>• Look for “perfect play” (2)</td>
<td>• No automatic banning (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• CAPTCHA to ensure humanity (1)</td>
</tr>
</tbody>
</table>

Table 4.5 Confidence and Banning, (#) = count of participants who made this comment, n=8 (Q3), n=9 (Q4, Q5)

4.1.6 Individual Differences Amongst Participants

In many cases, participants were not in agreement concerning which players were bots, and a player’s botness. An example of botness rating variation within a single player can be seen below, in Figure 4.2.

![Figure 4.2 Participants’ Botness Ratings for Player in Seat 3](image-url)
A second example, Figure 4.3, shows a comparison of botness ratings for a subset of the players. It is clear that participants were in stronger agreement in some cases (S7), yet disagreed in others (S3).

![Bar chart showing botness ratings for Player in Seat 3 and Seat 7]

**Figure 4.3 Participants’ Botness Ratings for Player in Seat 3 and Seat 7**

When examining botness ratings for all players from a subset of the participants, it is seen that some players, like Seat 8 and Seat 9, received highly polarized ratings from these three participants (Figure 4.4). Conversely, Seat 5 received very similar ratings from each of the three participants in this case.
Finally, Figure 4.5 shows the same comparison, but expands it to include all participants. Although the graph is visibly cluttered, the lack of agreement is apparent due to the spread of the data to both extremes of the rating scale for each participant and for each player.
Even though nine of ten players were bots, examining the 68 participant ratings for bot players shows that only 13/68 (19.1%) ratings were “very botlike”, while 20/68 (29.4%) were “botlike”. This indicates that only 33/68 (48.5%) could be considered “correct” detection of players as bots. As shown in Figure 4.6, this indicates that 26/68 (38.2%) ratings were considered “human” or “very human”, showing that a large number of ratings identified bots as humans. The remaining 9/68 (13.2%) ratings were neutral, meaning that participants were unable to qualify the player as a bot or human. If only non-neutral ratings are considered, participants were slightly more successful at identifying the bot players than random chance, with 34/59 (57.6%) correct ratings. This chance of success is likely influenced by the fact that each bot represented a different configuration, and that participants were told that at least one human was present in the game.

![Figure 4.6 Frequency of Each Botness Rating](image)

**Figure 4.6 Frequency of Each Botness Rating**

Due to this level of accuracy, and the fact that no participant neared the 85% accuracy seen by a participant in McGlinchey and Livingstone’s (2004) Pong study or the 100% accuracy from a single participant in Laird and Duchi’s (2001) Quake study, it is concluded that the data
supports the primary hypothesis: expert participants do not reach the same conclusions concerning the identification of players as bots or humans when observing a game of online poker.

4.2 Expert Evaluation of Skill Levels

The second hypothesis stated that expert players would come to similar conclusions when asked to identify the best and worst players in a game of online poker. Raw data for this hypothesis can be found in Table D-2.

4.2.1 Experimental Results

This hypothesis was evaluated using an Independent Samples Kruskal-Wallis Test to determine whether bot characteristics (as defined by seat number in Table 3.3) had a significant impact on skill rankings. The null hypothesis, $H_0$, was that there would be no difference between the skill ratings across all players. The alternate hypothesis, $H_a$, was that there would be a significant difference between the skill ratings across all players. An alpha value of 0.05 was used for this test.

The results of the test were significant [$H(9) = 18.370, p = 0.031$], indicating that experts viewed the players as having differing skill levels. Therefore, this rejects the null hypothesis and supports the second hypothesis. Participant skill ratings by player can be seen in Figure 4.7.
In examining the data, it appeared that Seat 9’s skill was ranked highly by participants, relative to other player rankings. When Seat 9’s skill ranking was separated from other seats, a very high level of agreement from most participants can be observed (Figure 4.8).
4.2.2 Traits of Skilled and Unskilled Players

The next set of data, shown in Table 4.6, contains the responses to Q1 from Appendix B.3, which concerned the best and worst player traits observed by participants. The worst traits can be summed up as a combination of poor betting habits and an unwillingness to cut further losses by folding a hand. The best traits consist of the opposite – knowing when to fold is identified as the most valuable trait, followed closely by playing position. Playing position is likely identified as a best player trait due to the experience required: it is no small feat to consider the state of the table, previous player behavior, and player turn order into account prior to taking an action.

<table>
<thead>
<tr>
<th>Traits of Skilled Players (Q1)</th>
<th>Traits of Unskilled Players (Q1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Folding unplayable hands (4)</td>
<td>• Bad or missed bets (5)</td>
</tr>
<tr>
<td>• Playing position (3)</td>
<td>• Staying in with bad/arbitrary cards (4)</td>
</tr>
<tr>
<td>• Trapping (1)</td>
<td>• “Calling station” / Calling everything (2)</td>
</tr>
</tbody>
</table>

Table 4.6 Player Traits, (#) = count of participants who made this comment, n=8

Since player skill ranking differed significantly by seat, Seat 9 was considered to be highly skilled by nearly all participants, and a limited set of positive and negative traits were identified by participants, it is concluded that the data supports the second hypothesis: expert participants came to similar conclusions when asked to identify the best and worst players in a game of online poker.

4.3 Relationship Between Skill and Botness

The third hypothesis stated that skill would not be a reliable indicator as to whether a given player was a bot or a human, and therefore there would be no relationship between player skill level and whether that player is identified as a bot.
4.3.1 Experimental Results

This hypothesis was evaluated using a Spearman correlation, which determined if there was a relationship between botness ratings and skill rankings assigned to each player by the participants. The null hypothesis, $H_0$, was that there would be no correlation between the skill and botness ratings across all players. The alternate hypothesis, $H_\alpha$, was that there would be a significant correlation between the skill and botness ratings across all players. An alpha value of 0.05 was used for this test.

The results were not significant [$r_s = 0.138, p = 0.236, ns$], indicating that player skill was not correlated to its botness. Therefore, this rejects the alternate hypothesis and supports the third hypothesis, as skill is shown to not be a reliable indicator of whether a player is a bot in the experimental environment. The average botness ratings and skill rankings for each player can be seen in Figure 4.9.

![Figure 4.9 Average Botness Rating and Skill Ranking (inverted) by Seat](image)

Figure 4.9 Average Botness Rating and Skill Ranking (inverted) by Seat
4.3.2 Relationship between Skilled Player Traits and Player Identifiers

Although there is insufficient quantitative data to explore this hypothesis further, the qualitative findings are still of value. In this particular case, Table 4.7 has been populated with three columns from tables outlined in Section 4.1.3 and Section 4.2.2. This table aligns traits associated with skilled play (Table 4.6) with identifiers used in differentiating bots from humans with identifiers of human and bot gameplay (Table 4.2). This is done to show how bot and human gameplay differ when compared against the best traits.

<table>
<thead>
<tr>
<th>Traits of Skilled Players (Q1)</th>
<th>Identifying a Player as a Bot (Q2)</th>
<th>Identifying a Player as a Human (Q2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Folding unplayable hands (4)</td>
<td>• Plays consistently, mathematical (2)</td>
<td>• Knowing when to fold (2)</td>
</tr>
<tr>
<td>• Playing position (3)</td>
<td>• Play position better (1)</td>
<td>• Play position better (2)</td>
</tr>
<tr>
<td>• Trapping (1)</td>
<td>• Consistent playstyle (1)</td>
<td>• Hesitant, bet/raise less (2)</td>
</tr>
</tbody>
</table>

Table 4.7 Participants’ Opinions on Best Traits vs Identifiers and Descriptors, (#) = count of participants, n=8

The overlap illustrated across each row identifies points concerning playing position and folding poor hands that are common to all three categories. These were both indicated in some way as either an identifier or gameplay descriptor for both bots and humans, providing further support for the first hypothesis (see Section 4.1). This overlap also provides additional support for the third hypothesis, as it demonstrates that the best player traits were not a reliable influence on whether a player was considered to be a bot or a human. This will be discussed further in the next Chapter.
5 Summary and Conclusions

The three hypotheses explored in this study examined the interaction of two specific factors in a poker environment: player botness and player skill. The results of each exploration support the overall thesis statement: expert poker players made similar observations about player behavior in a poker environment, yet used these observations to reach differing conclusions about whether a given player was a bot or a human. In the process of exploring each hypothesis, additional trends and questions were identified, as well as potential shortcomings in this study.

5.1 Identifying a Difference between Bots and Humans

Above all, it was found that expert players were unable to reliably differentiate bots from human players within the experimental environment; different observers were found to use the same rationale to support opposing player classifications. The ambiguity present in prior experimental environments (McGlinchey & Livingstone, 2004; Ryan, 2007; Hagelbäck & Johansson, 2010) has thus been shown to persist in this experimental environment as well. The lack of significance when analyzing participant botness ratings suggests that they were unable to reliably differentiate bots from humans through observation of the limited hands. In examining their rationale for player classifications, it became clear that the ambiguity of behavior was responsible for their lack of participant success and lack of agreement with one another – there were overlaps and contradictions throughout each participant’s rationale for player classification. Consequently, the results of the first hypothesis directly supported the thesis statement.

Startlingly, even though participant confidence levels varied concerning their identification of players as bots, the majority of participants were confident with the idea of automated detection and banning systems (Table 4.5). This suggests that participants assume automatic banning system to be more robust and reliable than manual detection methods. Some
are comfortable banning players after viewing a limited number of hands, or in some cases, after further investigation or observation.

5.2 **Experts Agree Concerning Player Skill**

Prior to examining the interaction of skill and botness, it was important to establish that skill was a reliable metric to use for expert observation. To that end, participants were asked to rank the skill of all players they had observed, as well as outline the traits of the best and worst players. It was found that 7 of 8 participants agreed that Seat 9 was one of the two best players, but agreement levels receded significantly for the players who were considered less skilled. This may be because it becomes difficult to assign a skill rating relative to other players when similar mistakes are made, as was the case with the bots in this study. For instance, if Seat 4, Seat 7 and Seat 8 all made similar ill-advised bets at different times, these events may be equally salient when the participant is trying to rank the players after the game, and yet they are forced to rank one above the others. Some participants also claimed that the lack of play from players hesitant to bet with non-optimal hands made it very difficult to evaluate their skill – they were simply described as risk adverse, and therefore likely bad players. The fact that expert participants could identify and agree upon skill for the best player, regardless of botness, suggests support for skill as an indicator of botness.

5.3 **Defining Player Skill in Poker**

The findings concerning the best and worst player traits were consistent with expectations (Section 4.2.2), in that they are centered on optimal betting and folding. They also demonstrate just how straightforward it can be to define skilled poker play. From participant responses, it became clear that the worst traits could be classified as actions typically taken by new players.
These actions included staying in with poor quality cards, placing bets at the wrong time, neglecting to place bets at opportune times, and calling every bet made (Table 4.6). The first point is tightly tied with the fourth: a player who stays in with poor quality cards is likely calling every bet made. The fourth point is additionally tied to the third; a player who is calling every bet made is likely missing out on additional winnings that might have been earned with betting more aggressively.

As a whole, these betting traits can be rolled up into a single concept of playing position as a desirable player trait. As previously described, playing position consists of modifying betting behavior based on order of play, as well as basing decisions and behavior interpretations on how much information is available at the time a player makes their move. The other positive trait was folding hands when they are truly unplayable – a direct opposite to staying in with the negative trait. This can be summed up to the following point: experts feel that a good poker player effectively plays position, both in terms of knowing which hands to play, and the appropriate time to bet. Those who do not play position correctly are classified as bad players.

Lastly, it appears that random chance, determined by the cards dealt to players, and overall success, determined by chip count, had an observable impact on participant opinion of player skill. This is best demonstrated by the difference in skill rankings for players 9 and 10, who were configured identically. It is proposed that this difference in skill ranking was consequently a direct result of the cards drawn and the position the player was in at that time, since other factors were held constant.

5.4 Lack of a Correlation between Skill and Botness

From initial examination of the data, it appeared as if skill would have a significant interaction with botness ratings, as participants seemed to associate skill with humanness. As an example,
many participants felt that player 9 must be human because of its usage of a “check-raise”
technique when holding a strong hand. This resembled the “rocket jump” example from Gorman
et al.’s 2006 study, wherein players were deemed more human if they engaged in “advanced”
actions.

By the conclusion of the experiment, however, other participants had reported that player
9 must be a bot because of its usage of the advanced technique, when most players at the table
were considered, at best, mediocre. Consequently, this result is similar to the “item gathering”
scenario from Ryan’s 2007 study, wherein players were deemed either more human or more
botlike, depending on the observer, if they sought out the most beneficial items first.

Upon statistical analysis of the quantitative data, no significant interaction between skill
and botness was found. Participants were also confident in their ratings when backed by such
rationale, which serves to further demonstrate how the same behavior could be attributed to both
a bot and a human. Rather than being a clear indicator of botness, skill became yet another
factor related to ambiguous behavior, and directly supports the thesis statement as a result.

5.5 Preconception of Player Types
It is also proposed that preconceptions concerning player types may have an influence upon how
expert poker players classify players based off of observed behavior, as most participants were
watching for “mistakes” or “successes”. They then classified a player as a bot or a human
depending on the player type they viewed as an inherently weaker or stronger player.
Participants also had strong opinions concerning how a bot could be more believable – the most
common recommendation was for bots to play like an inexperienced human, which suggests that
at least some participants consider bots to be more skilled than humans as a whole.
If an individual has a low opinion of bots, they may be more inclined to classify bad players as bots, and good players as human. The converse is a reasonable opinion as well – this helps to explain the lack of correlation between the two factors, and the general lack of agreement among participants concerning botness. This is evident when both ratings are compared in a single graph (Figure 4.9). In this case, botness ratings have been doubled in order to be directly comparable with the skill ranking scale. These scales differed as participants were asked to provide each player with a unique rank indicating their skill relative to the others, while botness ratings could overlap in order to reduce the set of possible values and encourage participants to assess players individually and make firmer decisions. This permitted a more granular evaluation of player skill while avoiding a situation in which participants were provided with too many options for botness, when the binary decision was what was desired.

5.6 Participants Relate Observations to Their Own Play Style

Another finding that arose from the session comments was that some participants actively sought to identify their own play style in each of the video players. Those who matched the play style were considered humans or “very good bots” – this suggests that the participants were framing their classifications around their own play style and experiences. Players who did not match an expert’s play style were typically classified as human, as their behavior was considered too typical of inexperienced human gameplay. It is therefore proposed that, when asked to evaluate other players, an expert player may have a tendency to compare observed tactics against their own – those who behave similarly are treated as highly skilled humans or bots. It is also possible that the similarity to their own play style leads the expert to classify a given player as human, as they are better able to relate to the decisions made by the player.
This effect was seen when examining the comments for seats 9 and 10, who were configured identically yet ranked differently in terms of skill. Based on comments made throughout the study, a number of participants felt that Seat 9 was more human than Seat 10, with some participants stating this was because its play style, given certain cards in certain positions, was very similar to their own.

5.7 Study Limitations

Five factors have been identified that may have reduced the quality of the study’s results: the use of limit poker, the limited number of hands observed by participants, the removal of decision time, the inability to have only bot players, and the inability to fully customize the bot configurations.

Limit poker, in which the maximum bet cannot exceed a certain amount per round of betting, was used in order to ensure that no player would be forced from the game for the duration of the recorded videos – however, this significantly changes the strategy of the game and limits the risk involved when compared to no-limit poker, where players are free to bet whatever amount they want at any time. Prior to viewing the videos, one participant jested that, in limit poker, “everyone plays like a robot” due to the limited gameplay strategies available.

Second, although it was necessary to limit the number of hands to keep a reasonable session length, this configuration affected participants’ classifications and reduced their confidence. The methodology was also highly dependent upon the bot attribute configurations in Poker Academy Pro – some of the configurations resulted in a play style that was extremely difficult for participants to evaluate as the bots were very hesitant to play high-risk hands.

Third, the necessary removal of decision time had both positive and negative impacts: on the one hand, it allowed for better examination of other factors that may communicate botlike
behavior, while on the other, it prevented conclusions from being made concerning bot
identification in typical online environments, where decision time varies from player to player.

Fourth, the software used for the experiment permitted a maximum of 10 players per
table, one of whom had to be human. The inability to completely remove humans from the game
resulted in a tenth of the data being irrelevant to the analysis. The loss of data would have been
mitigated if more bots could have been added to the table, but it would have been optimal to
have all players be bots as well.

Lastly, due to available resources, the experiment was bound to the behavior
configurations defined by the software environment. With more time and resources, this could
be explored with a multitude of bot types with more granular configurations that can be fully
quantified. As it was, the closed-source software and bot design prevented further dissection of
the behaviours for analysis.

5.8 Future Research

A number of opportunities for future research were identified during the literature review and
data analysis portions of this study. One key point, the lack of significance in the results of the
first hypothesis, raises questions concerning the validity of administrative review and community
“report”/self-policing systems. If unbiased observers are truly unable to reliably identify non-
human players based on behaviour, there may be little value in having such mechanisms in place
at all. As outlined in the 2009 study by Chen et al., these mechanisms can also lend themselves
to exploitation and collusion if they are poorly implemented. Additionally, Cornelissen and
Grootjen (2008) point out that any sort of judgment based off of policing mechanisms can have
disastrous effects if the consequences are severe and unwarranted. Thus, this study should be
extended with a larger number of participants to explore whether or not self-policing mechanics
can be effective. If experts cannot reliably identify non-human players over longer observation periods, these systems will continue to create additional work for administrators, who may have difficulty detecting bots as well, and will continue to be ripe for exploitation. If, on the other hand, experts are shown to reliably identify human players, there may be simple modifications that can be made to these systems to render them as effective as possible while preventing exploitation. This question could be optimally explored if online poker administrators were recruited, as they are trusted with the task of manually reviewing accounts suspected to be bots.

Another point that deserves future study would be the exploration of participants who are extremely confident in their detection of bots. It would be valuable to understand the source of such confidence, and determine whether they remain consistent after failing to accurately detect bots. Conducting multiple studies with the same participants could reveal whether a confident expert re-evaluates their decisions and rationale after being confronted with their errors. Unfortunately, there were not a sizable number of extremely confident participants in the study, and therefore this experiment was unable to explore this idea further.

A final set of proposals for future research are sourced from this study’s limitations. Although decision time was eliminated as a factor in order to examine other influences on perceptions of botness, it has been shown to be a significant influence in previous studies. However, these studies did not examine decision time’s impact in a strategy-rich environment like poker. Due to the general lack of urgency within the environment, and the use of decision time as a strategic tool, it was initially hypothesized that it would have less impact in the experimental environment. However, due to the desire to focus on other factors, time, and participant constraints, exploration of this question was not possible in this study.
A second proposal sourced from the limitations would be to modify the traits of the experimental environments. Rather than focusing on limit poker, in which player strategy is limited due to fixed bet amounts, this study could be extended to no-limit poker, in which players are free to bet any amount. This could increase the impact of both honesty and aggression traits, as limitless betting would favour those who can bluff strategically.

A third proposal would be to modify the criteria for the participants recruited. It would be interesting to see how well non-experts can detect bots, and what rationale they use for doing so. Since they lack the significant experience that an expert has, non-experts may use different rationale – perhaps not leveraging gameplay observation at all. This may or may not result in better bot detection, but it would certainly be a way to examine alternative approaches to the problem.

Non-experts may also be more influenced by certain behaviour traits over others, due to their limited experience. Removing the requirement for expert-level skill would also facilitate the recruitment of a larger number of participants. By recruiting a combination of expert, novice and intermediate players, additional conclusions could be reached concerning the impact of experience upon bot detection. This could be extended even further by recruiting from both online-only and offline-only groups, in an effort to determine which environment best develops the ability to detect bots.

Fourth and lastly, future research could explore additional environments. While this thesis has focused on a strategy-based environment, there are many other gaming environments that can be explored: asynchronous games, strategy-based games with observable avatars, text-based games, etc. All of these environments are likely to experience similar issues with bot players and may rely on manual detection of these players by administrators and the community.
As indicators of bot-like behavior have been shown to vary widely across different online
gaming environments, it may be important to ask these same questions regarding bot behavior in
as many different types of gaming environments as possible. The results of such studies could
prove to lay the groundwork for new self-policing mechanisms for bot detection, and the
identification of environments in which they would prove to be consistently effective and
beneficial to the community.
References


Appendices

APPENDIX A – Consent Forms

Appendix A.1  Consent Form #1 – Pre-session

CONSENT TO PARTICIPATE IN RESEARCH

Bot Identification in Poker Games: How accurately can humans detect bots in a poker environment, and which characteristics are considered “botlike”?

You are asked to participate in a research study conducted by Dr. Blair Nonnecke and Ben Altman, from the department of Computing & Information Science at the University of Guelph. The results of this project will be contributed towards Ben Altman’s Master’s thesis.

If you have any questions or concerns about the research, please feel free to contact Dr. Blair Nonnecke at (519)824-4120 x56407.

PURPOSE OF THE STUDY

This study seeks to determine how accurately humans can identify computer players, known as “bots”, within a poker game environment, and establish which player traits are considered to be the most and least “botlike”.

PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Fill out a brief questionnaire (<5 minutes)
- Watch one series of video recordings of online poker games and take notes on each player (30 minutes)
- Answer some additional interview questions (20 minutes)

Research findings will likely be available online for a short period of time after the completion of the thesis project.

POTENTIAL RISKS AND DISCOMFORTS

We do not anticipate any significant potential risks or discomforts, except perhaps mild negative reactions to incorrectly identifying players as bots.
POTENTIAL BENEFITS TO PARTICIPANTS AND/OR TO SOCIETY

Participants in this study will primarily benefit from the opportunity to test their observational skills against some of the highest quality poker bots available.

It is expected that the results of this study will be primarily of interest to bot developers and game administrative and security staff.

PAYMENT FOR PARTICIPATION

Participants in this study will not receive any payment for their participation.

CONFIDENTIALITY

Every effort will be made to ensure confidentiality of any identifying information that is obtained in connection with this study.

Participant names will not be recorded on any of the session logs or schedules. In the data, participants will simply be referred to by a given number (P1, for instance). All softcopies of data will be stored on password-protected computers while hardcopies will be stored under lock-and-key in the Reynolds Building. The raw data of this study will not be publicly released.

PARTICIPATION AND WITHDRAWAL

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may exercise the option of removing your data from the study. You may also refuse to answer any questions you don’t want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise that warrant doing so.

RIGHTS OF RESEARCH PARTICIPANTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. This study has been reviewed and received ethics clearance through the University of Guelph Research Ethics Board. If you have questions regarding your rights as a research participant, contact:

Research Ethics Coordinator          Telephone: (519) 824-4120, ext.
University of Guelph                 56606
437 University Centre               E-mail: sauld@uoguelph.ca
                                      Fax: (519) 821-5236
Guelph, ON  N1G 2W1

SIGNATURE OF RESEARCH PARTICIPANT/LEGAL REPRESENTATIVE

I have read the information provided for the study “Bot Identification in Poker Games” as described herein. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

_______________________________________
Name of Participant (please print)

_______________________________________  ________________
Signature of Participant             Date

SIGNATURE OF WITNESS

____________________________________________________
Name of Witness (please print)

_______________________________________  ________________
Signature of Witness             Date
CONSENT TO PARTICIPATE IN RESEARCH

Bot Identification in Poker Games: How accurately can humans detect bots in a poker environment, and which characteristics are considered “botlike”? You have participated in a research study conducted by Dr. Blair Nonnecke and Ben Altman, from the department of Computing & Information Science at the University of Guelph. The results of this project will be contributed towards Ben Altman’s Master’s thesis.

If you have any questions or concerns about the research, please feel free to contact Dr. Blair Nonnecke at (519)824-4120 x56407.

TRUE PURPOSE OF THE STUDY

This study seeks to determine how accurately humans can identify computer players, known as “bots”, within a poker game environment, and establish which player traits are considered to be the most and least “botlike”. You were initially led to believe that some players were human, and some players were bots. This was a deception. All of the players were bots – no human players are recorded in the videos. Consequently, determining “botlike” and “human” traits is the primary focus of this study, while the accuracy of “bot” classification is secondary.

Participants may experience minor embarrassment or anger due to the deception involved in this study. Deceiving participants into believing some players are human is necessary in order to guarantee we are comparing “bot” behaviour to human behaviour rather than “bot” behaviour to “human bot” behaviour.

This second consent form serves to provide you with a way to explicitly refuse the use of your data in our study after fully understanding the purpose of the study.

SIGNATURE OF RESEARCH PARTICIPANT

I have read the information provided for the study “Bot Identification in Poker Games”. I realize that I was deceived during the course of this project, but my questions have now been answered to my satisfaction, and I agree to allow my data to be used in this study. I have been given a copy of this form.

Name of Participant (please print)

Signature of Participant _____________________________ Date _____________________________
SIGNATURE OF WITNESS

Name of Witness (please print)

______________________________________
Signature of Witness

______________________________________
Date
APPENDIX B – Questionnaires

Appendix B.1 Questionnaire #1 – Pre-session

Section 1: General Information
Age: _____
Sex: [ ] Male [ ] Female [ ] Transgender

Section 2: Poker Information

How often do you watch Texas Hold’Em Poker games (TV/Online/etc.)?
* X box: Current rate  O box: Peak rate *
[ ] Never [ ] Every day [ ] Every few days
[ ] Every week [ ] Every 2 weeks [ ] Every month

Offline Information
Do you play Texas Hold’Em Poker offline/in-person?
[ ] Yes [ ] No (skip to Online Information)

How long have you been playing Texas Hold’Em Poker offline?
[ ] 1 year or less [ ] 2 years [ ] 3 years
[ ] 4 years [ ] 5+ years

How often do you play Texas Hold’Em Poker offline?
* X box: Current rate  O box: Peak rate *
[ ] Every day [ ] Every few days [ ] Every week
[ ] Every 2 weeks [ ] Every month [ ] Every few months

Which type of Texas Hold’Em Poker do you typically play offline?
[ ] Ring/”Cash” [ ] Tournament

Online Information
Do you play Texas Hold’Em Poker online?
[ ] Yes [ ] No (skip to Game Information)

How long have you been playing Texas Hold’Em Poker online?
[ ] 1 year or less [ ] 2 years [ ] 3 years
[ ] 4 years [ ] 5+ years

How often do you play Texas Hold’Em Poker online?
* X box: Current rate  Circle box: Peak rate *
[ ] Every day [ ] Every few days [ ] Every week
[ ] Every 2 weeks [ ] Every month [ ] Every few months
Which type of Texas Hold’Em Poker do you typically play online?
[ ] Ring/”Cash”       [ ] Tournament

Approximately what percentage of your Texas Hold’Em Poker playing is done online?
___ % Online

Section 3: Game Information

Have you ever used the program called “Poker Academy”?
[ ] Yes       [ ] No

If yes, how would you describe the computer opponents?
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________

What games have you played against computer opponents? Check all that apply.
[ ] Strategy Games       [ ] Shooters       [ ] Poker
[ ] Other Card Games     [ ] Role-Playing Games

Which type of game has the best computer opponents? Why?
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
Appendix B.2  Questionnaire #2 – Post-session Rankings

Please rank the players from best to worst. If possible, briefly describe your reasoning for each player’s ranking.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Seat #</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B.3  Questionnaire #3 – Post-session Interview

1. Please describe how you decided on the best and worst players in the game:

2. Please describe how you decided whether players were bots or humans:

3. How confident are you in your classification of players as bots?

4. As an administrator, would you feel comfortable banning the players that you identified as bots? Why or why not?

5. If bots could be identified by the system with a high degree of accuracy, would you feel comfortable having them banned automatically? Why or why not?

6. How would you describe the overall gameplay used by the bots?

7. How would you describe the overall gameplay used by the humans?

8. How do you feel the bots could be improved?

9. How do you feel the bots could be made to be more believable (more human)?

10. If you were training new administrators to spot bots in poker games, what would you have them look for?

11. Please rank the following factors according to how well they indicate that a player is a bot. If possible, specify why you believe this to be the case.

   a. Decision Speed
   b. Likelihood to Fold
   c. Likelihood to Bet
   d. Likelihood to Re-raise
   e. Likelihood to Bluff
   f. Adapting to opponents

This concludes the study, thank you for your time.

Next, we will describe what the purpose of this study was, and how your data will help us in our research.
# Terminology

Sourced from holdemtight.com/pgs/dc/Dic/holdempokerdictionary.htm

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check</td>
<td>[Action] Not betting, aka <em>pass</em>. Possible only if no one has bet before you. If someone has bet out, you must either call, raise, or fold.</td>
</tr>
<tr>
<td>Call</td>
<td>[Action] Matching, but not raising, previous bets. This allows a player to stay in the hand.</td>
</tr>
<tr>
<td>Fold</td>
<td>[Action] Giving up a hand rather than putting money in the pot.</td>
</tr>
<tr>
<td>Raise</td>
<td>[Action] Bet where the player puts more money in the pot than the previous bettor. Opponents must either fold, match (call), or reraise.</td>
</tr>
<tr>
<td>Re-raise</td>
<td>[Action] To response to a raise action with another raise, forcing opponents to either fold, call or reraise.</td>
</tr>
<tr>
<td>Check-raise</td>
<td>[Action] First checking, then raising anyone who bets after you.</td>
</tr>
<tr>
<td>River</td>
<td>The fifth and final community card in Hold’Em, dealt face up in the middle of the table, which anyone can use as a part of a hand.</td>
</tr>
<tr>
<td>Turn</td>
<td>The fourth community card in Hold’Em, dealt face up in the middle of the table, which anyone can use as part of a hand.</td>
</tr>
<tr>
<td>Flop</td>
<td>The first three community Hold’Em cards dealt face up in the middle of the table for everyone to see and use.</td>
</tr>
<tr>
<td>Pre-Flop</td>
<td>The round of betting that occurs after the deal, but before the Flop occurs.</td>
</tr>
<tr>
<td>Deal</td>
<td>The portion of the hand in which two cards are dealt to each player at the table.</td>
</tr>
<tr>
<td>Calling Station</td>
<td>Passive player who calls constantly with weak hands, and often calls when they should raise.</td>
</tr>
</tbody>
</table>
## APPENDIX D – Raw Data

### Table D.1 Botness Rating per Seat per Participant

<table>
<thead>
<tr>
<th>Participant</th>
<th>P1</th>
<th>P2</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seat</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
<td>S4</td>
<td>S5</td>
<td>S6</td>
<td>S7</td>
<td>S8</td>
</tr>
<tr>
<td>S1</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>S3</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>S4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>S5</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>S6</td>
<td>6</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>S7</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>S8</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>S9</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>S10</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

### Table D.2 Skill Ranking per Seat per Participant

<table>
<thead>
<tr>
<th>Participant</th>
<th>P1</th>
<th>P2</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seat</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
<td>S4</td>
<td>S5</td>
<td>S6</td>
<td>S7</td>
<td>S8</td>
</tr>
<tr>
<td>S1</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>S2</td>
<td>6</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>S3</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>S4</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>S5</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>S6</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>S7</td>
<td>10</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>S8</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>S9</td>
<td>8</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>S10</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>