

**Do I know you? Evaluating Human-to-Human  
Authentication via Conversational Interfaces**

Nour Dabbour

A thesis submitted to the Faculty of Graduate and  
Postdoctoral Affairs in partial fulfillment of the requirements  
for the degree of:

Masters of Applied Science  
of  
Human-Computer Interaction

Carleton University

Ottawa, Canada

May 2019

## Acknowledgements

I would first like to express my deep appreciate to my thesis advisor Dr. Anil Somayaji from the School of Computer Science at Carleton University. He consistently allowed this thesis to be my own work, but steered me in the right direction whenever I needed it.

I would like to extend my sincere thanks to my thesis committee Dr. Robert Teather, Dr. Robert Biddle, and chair Dr. AbdelRahman Abdou for their constructive feedback and invaluable comments.

I would also like to acknowledge Robin Tropper as the main developer of the 'friend imitation game' platform and Gabi for her support and guidance with my research. Robin and Gabi were instrumental in defining the path of my research and I am grateful for that.

A special thanks to my lovely friends: Rand, Sara, Ruba, Becky, Sanaz, and Anis for supporting me throughout my journey.

My deepest gratitude goes out to my loving family: George, Linda, Taz, Nashy, and little Giant. Thank you for your constant support, care, encouragement, and infinite love. Your mere existence had made my life easier in so many ways and I am forever grateful.

To my Omi: thank you for always being there for me and for putting a smile on my face everyday. Thank you for your endless love, support, patience, and understanding. I'm extremely blessed to have such a loving husband and I cannot wait to start our life journey together.

## **Abstract**

Online impersonation attacks are prevalent as the result of an increase in electronic communication. Humans exposed to impersonation attacks are normally resistant to them. Yet, very little is known of the method humans use to authenticate each other over computer mediated communications. In this research, we study how individuals identify a familiar individual versus an adversary over a text messaging e-commerce game. Then we classified each authentication method used by the participants into the following five themes: ‘Knowledge & Experience’, ‘History & Plans’, ‘Texting Style’, ‘Response Speed’ and ‘Personality Type’.

Consequently, we investigate the feasibility and robustness of implementing human-to-human authentication methods in conversational systems. We evaluate each theme and rank them based on data source access and analysis complexity. While we find that many strategies can only provide weak security guarantees, we also identify one that could provide strong guarantees under realistic threat models.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Research Questions . . . . .	8
1.2	Contribution . . . . .	8
1.3	Chapter Outline . . . . .	9
<b>2</b>	<b>Background</b>	<b>10</b>
2.1	Human Impersonation . . . . .	10
2.2	Human Authentication Methods . . . . .	11
2.3	System Authentication Methods . . . . .	15
2.4	Conversational Systems . . . . .	17
2.4.1	Authentication in Conversational Systems . . . . .	18
2.5	Human Communication Over a Medium . . . . .	22
2.5.1	Turing Test . . . . .	23
<b>3</b>	<b>Methodology</b>	<b>26</b>
3.1	Study Setup . . . . .	27
3.2	Protocol . . . . .	29

3.2.1	Session Outline . . . . .	32
<b>4</b>	<b>Results</b>	<b>39</b>
4.1	Overview . . . . .	39
4.2	Participants Background . . . . .	42
4.3	Classification of Techniques . . . . .	44
4.3.1	Semantic Measurement: <i>The What</i> . . . . .	44
4.3.2	Behavioural Characteristics: <i>The How</i> . . . . .	47
4.4	Semi-Structured Interview . . . . .	48
4.5	The Patterns of Techniques . . . . .	49
4.6	Interesting Pairs . . . . .	50
<b>5</b>	<b>Security Analysis</b>	<b>57</b>
5.0.1	Semantic Measurements . . . . .	59
5.0.2	Behavioural Characteristics . . . . .	62
<b>6</b>	<b>Discussion</b>	<b>66</b>
6.1	Paper Contribution . . . . .	66
6.2	Application to Conversational Systems . . . . .	67
6.2.1	Challenges . . . . .	69
6.2.2	Ethical Responsibility . . . . .	69
6.3	Study Design and Limitations . . . . .	70
6.4	Future Research and Considerations . . . . .	72

<b>7 Conclusion</b>	<b>74</b>
<b>Appendix A: Research Material</b>	<b>88</b>
A.1 Research Summary .....	88
A.2 Consent Form .....	89
A.3 Pre-game Questionnaire .....	91
A.4 Game Outline for the Authenticator .....	93
A.5 Game Outline for the Convincer .....	94
A.6 Post-game Questionnaire .....	95
A.7 Debriefing Form .....	96
A.8 Consent to Use Data Form .....	97
<b>Appendix B: Recruitment Material</b>	<b>99</b>
B.1 Social Media Initiation .....	99
B.2 Recruitment Poster .....	100
<b>Appendix C: Game Mock-up</b>	<b>101</b>
C.1 Simulated Game Environment .....	101
C.2 All Possible Use-case Scenarios .....	103

# List of Figures

3.1	The moderator screen during the game . . . . .	29
3.2	The authenticator screen during the game . . . . .	30
3.3	The convincer screen during the game . . . . .	31
3.4	A step-by-step flowchart of one research session . . . . .	33
3.5	A game snapshot of a transaction between the authenticator and convincer . . . . .	35
4.1	Percentage of transactions that were correctly identified per pair	40
4.2	Percentage of authentication themes based on interaction count	46

# List of Tables

4.1	Subject pair performance . . . . .	41
4.2	Participants Background . . . . .	43
4.3	Frequency of authentication themes used by participants through- out the study . . . . .	45
5.1	Analysis of the threat complexity . . . . .	59
5.2	Security analysis for semantic measurements . . . . .	61
5.3	Security analysis for behavioural characteristics . . . . .	63



# Chapter 1

## Introduction

*Consider your last texted conversation with someone close to you. Ask yourself: How do you know you were talking to the right person?*

Online impersonation is a common way for cyberattackers to try to deceive people to gain personal information, monetary items, or to convince them to perform an action that is only beneficial to the attacker [4][30]. Attackers can impersonate individuals on social media accounts and through emails to deceive family and friends of the victim [17].

With the use of technology and social media, more and more young adults are using text messages as means of communication [27][29] making impersonation attacks more feasible. The number of impersonation incidents increased over the years [17]. Current research professionals are studying the effects of texting on human behaviour [7][55], self perception [36] and the ability to properly communicate affective information such as emotions [6].

Yet, very little is known about human-to-human authentication methods in unimodal environments such as text messaging or e-mail. By **human-to-human authentication** we are referring to the techniques humans use to identify each other when visual and auditory stimulation are not present.

Anecdotally, individuals are not defenseless to impersonation attacks. They can sometimes identify the impersonator in unimodal environments. In contrast, conversational systems such as chatbots and personal voice assistants have very little to no methods of authentication within conversation. Although they are often used for sensitive tasks, currently these systems require separate mechanisms for authentication like a password or biometric that are outside the conversation. This requirement for out-of-band authentication limits the utility of these systems in many contexts and leave their users vulnerable to attack.

Improving the security and authentication mechanism for chatbots and personal voice assistants is crucial because they can access, amend, and release a user's personal and sensitive information. Increasingly, voice assistants have access to an individual's calendar, home security system, online purchasing accounts [35], and banking information [26].

With the development of conversational interfaces, systems will begin carrying conversations similar to the ones humans use to interact with each other. In order to introduce a trustworthy conversational system, authentication is required just as people do naturally during communication. First, we must explore and understand human-to-human authentication methods and

potentially use it in conversational systems. The idea of implementing similar human-to-human authentication techniques in conversational systems holds promising measures as they will be easy to implement throughout natural conversation.

## 1.1 Research Questions

Our research objective is to identify and classify the key factors that allow humans to authenticate each other in computer mediated conversations. The findings from this research highlight themes that could potentially be used to enhance the security of conversational systems by using contextual cues. We refer to this as **in-band authentication**.

The research questions posited in this thesis are:

**Question 1:** *How do humans authenticate each other over text dialogue?*

**Question 2:** *How robust are the human-to-human authentication techniques against human and automated attacks?*

## 1.2 Contribution

This research contributes to the usable security field of conversational research. We make the following contributions:

1. First study that focuses on human-to-human authentication over a

text-conversation medium, then classified authentication patterns into themes and categories.

2. Re-envisioned authentication methods for conversational systems by advocating for a line of research that focuses on human authentication techniques.
3. Evaluated the threat complexity analysis levels for each authentication pattern identified in order to understand their potential robustness against attacks.

### **1.3 Chapter Outline**

In Chapter 2 we discuss related background. In Chapter 3 we explain our research requirements and describe our research protocol. In Chapter 4 we present the findings and classify the human authentication themes into categories. In Chapter 5 we discuss a security analysis of each theme identified in the previous chapter. In Chapter 6 we discuss the limitation of our study design, ethical responsibilities, and future research considerations.

# Chapter 2

## Background

In this chapter, we highlight social engineering as the problem of impersonation attacks and then discuss how humans identify their surroundings using pattern recognition. We outline current authentication techniques and discuss some of their limitations. Consequently, we describe conversational systems and outline current research that focuses on conversational system's authentication methods. The end of this chapter talks about research of human behaviour in the context of computer mediated communication.

### 2.1 Human Impersonation

The art of manipulating people's behaviour to achieve a goal is called **social engineering** [4][30]. Social engineering can yield positive outcomes. For instance, lawyers and psychologists use social engineering tactics to reveal

information in interrogation rooms. However, social engineering is also used in negative ways such as impersonation attacks [54].

Impersonating attackers typically take on someone else's identity without their consent to achieve a goal. In 2015, an impersonator attacked Ubiquiti Networks' financial department pretending to be the CEO. The attacker ordered a member of staff to direct huge amounts of money to a company overseas. The financial department only noticed after millions of dollars were transferred [54]. Impersonation is a serious issue and is often successful when the person being impersonated is unfamiliar to the target.

## 2.2 Human Authentication Methods

Anecdotally, humans are able to differentiate an adversary if they are familiar with the impersonated individual as humans typically have a dynamic way of authentication. They are able to detect patterns in behaviour and predict outcomes based on previous experiences. Many researchers have studied humans capability of face [62][10][40][5] and voice recognition [11][56][9] in multimodal environments, yet we know of no past work on human-to-human identification methods over text messages. Learning more about how humans identify patterns can help us explore human-to-human authentication methods in conversation.

By way of survival [44], humans continuously analyze and interpret their world [21] thus enabling them to recognize patterns and point out alterations

in patterns [44]. The ability for humans to recognize patterns vary due to genetics [68], social interactions [57], and cultural experiences [39]. Humans make decisions by carrying out deductive and inductive reasoning through schemas and mental models [15][39]. **Schemas and mental models** are both interconnected cognitive structures accounted for an individual's processes [39] such as identification and recognition. They aid in the interpretation of people's characteristics, which we believe is a distinctive feature used in human-to-human authentication over text messages.

### **Schema Formation**

Humans build a generic interpretation of the world based on their experience [39]. By formulating mental representations of their surroundings, humans are able to have a solid understanding of the universe and predict outcome of interactions.

When children as young as two years old interact with their external environment, they start formulating mental representations, also known as 'assimilatory' schemas [57], of the world around them [5][15][57]. Piaget [57] argues that all assimilatory schemas mature to schemas with experience and age. It is when they become "less centered on the subjectivity of the assimilating object" [57] and rather more focused on comprehending their world. Holland and colleagues [34] propose that humans store schemas in the long term memory as inflexible knowledge structures [44] where they are used regularly during conversations and other daily interactions. Schemas

put humans at an evolutionary advantage because it helps them anticipate the future based on previous interactions.

### **Mental Models**

Though schemas can be built upon, they are not context dependent. Humans combine various schemas in unfamiliar situations to predict their outcome through a representation known as a **Mental Model** [34].

Mental models are flexible knowledge structures that occupy the working memory temporarily. They embody an internal conceptual and physical representation of the world [39]. Mental models and schemas are notably distinct. Researchers refer to schemas as a pre-compiled generic knowledge structure, whereas mental models are specific knowledge structure that simulate new situations through the compilation of multiple schema [19][34][61]. Hence, when an individual talks to someone they are familiar with, they will predict the outcome of unfamiliar conversation by combining various schemas to form a mental model of that specific conversation. This enables individuals to combine schemas and mental models to predict the outcome in conversations.

### **Characteristics and Persona Formation**

Humans detect conversational patterns which are also referred to as **conversational styles**. A conversational style consists of a set of repeated patterns associated with social identities [37][38]. Styles can be detected based on so-



cial patterns and individual characteristics. We assume that detecting styles helps humans authenticate each other during conversation. Through linguistic practices, adults develop styles as a communicative behaviour. Conversational styles are identifiable by people as they are repeatable to a certain extent [38], it is often associated with large scale social patterns [20]. For example, individuals who are located in the country versus the city have very different conversational styles [21].

Styles are detectable by humans, as humans can associate an individual to a social group based on their conversational style. D’Onofrio [20] posited that the perceived style of a speaker will change the listeners expectations and influence the understanding of the language spoken. D’Onofrio’s research [20] shows that people draw conclusions upon their experience and previous communications [31].

Some linguistic researchers argue that individuals construct a style in ‘platonic self’ [38]. Meaning each individual is unique in characteristics, even within the same social environment. This is referred to as ‘persona style’, or ‘persona’ [21]. It is an immediate social construct in interaction and it may change over time depending on the social environment [21]. We assume that individuals that are familiar with each other will be able to pick up conversational styles throughout conversation.

## 2.3 System Authentication Methods

System-to-user authentication is the process that helps establish whether the user is who they proclaim to be [58], a necessary precaution for reliable access control to sensitive data [50]. In the following section, we will explore current system authentication methods, then address the drawbacks of applying each technique to computerized systems.

Authentication methods widely used typically fall under the following three categories: **user knowledge, user characteristics, and user property** [1][46][50][58].

### Traditional Authentication

A pin code, password, smart card, and an encryption key, are all examples of **traditional authentication methods**. The former two methods are something a user knows, the latter two are something a user has[1][58]. Traditional methods have reached their limit [1]. Using knowledge-based authentication relies heavily on the user's memory and they can be easily forgotten. For example, 'Forgot Password' reset links. In order for users to memorize passwords, they revert to high risk strategies such as writing their passwords down, using an easy to guess password (i.e. 123456), using the same passwords for various systems [19][42], or disclosing them to family or friends [50][53]. Secondly, security based objects like encryption keys can be lost, stolen, or easily accessible to some individuals [50].

### Modern Authentication

The latest methods of authentication have been based on biometric measures [44]. **Biometrics** use an automated process of authentication that is based on the user's unique physiological features and behavioural traits [44][50]. The most prominent biometric measurement include fingerprints, voice sample, iris, and facial structure [1][50]. They are now widely used in personal devices for authentication [1][42]. For example, Apple's fingerprint based TouchID [42] and facial ID recognition on their latest iPhone and iPad personal products.

Although biometric authentication has proven to be robust [44], recent research claims that attacks are easier than presumed [1, 50]. Adamek and colleagues [1] created and tested various 'fake' fingerprints using rubber and plastic materials and were successful in accessing a system that used fingerprints as way of authentication. Voice biometric authentication in conversational systems only work for voice assistants. Attackers may easily copy a sample measurement of the voice characteristic that a system will accept as valid.

### Multi-factor Authentication

Traditional authentication methods are prevalent today, most of which are being used in combination through a process called **Two Factor Authentication**, or 2FA for short. The multi-factor authentication system requires the user to identify themselves in two different ways. Many companies and

government offices may require the user to present an ‘RSA SecurID’ [53] badge before accessing sensitive data, while personal sites like Facebook and Gmail will send the user a confirmation code to their pre-registered cellphone device [53]. The multi-factor verification process was developed to strengthen the authentication process, especially when users are about to access sensitive data [44]. Multi-factor authentication technique hinders the user and requires them to use a multi-step method.

## 2.4 Conversational Systems

In this section, we outline what constitutes a conversational system, why they’re important to study, current authentication methods used by conversational systems and how researchers are trying to improve them.

**Conversational systems** use conversation as a mean of communication with a user. They allow various ways of interacting including speech (voice interface), text (chat bots), and sometimes touch [52]. In order for a system to be classified as conversational user interface, Michael et al., [52] argue that it must possess at least two features. First, it must be able to maintain a natural occurring conversation outside of a fixed set of commands and phrases. In other words, the system has to mimic human to human conversation where the language used is flexible and messages are expressed in different ways. Secondly, the user has to interact with the system on a turn-by-turn basis, where both the system or the user can initiate and contribute

to the conversation equally.

We speculate that an advanced conversational system will keep track of previous interactions instead of treating each interaction as a separate query which makes them easier to use. Those advanced voice assistants and chatbots provide a more ubiquitous way of interacting with users. By using a conversational paradigm, they are able to engage with humans in a more naturalistic, conversational way. The interaction does not require learning a new skill, therefore the learning curve is not steep.

The use of conversational systems is expected to grow up to 15.1 million users by 2020 [45]. In fact, major tech companies such as Microsoft, Facebook, and Google invested heavily in systems capable of interacting with users in a conversational way [27][35]. Google recently released ear-buds that capture audio and perform instant real-time voice translation using Google's voice assistant. [35].

### **2.4.1 Authentication in Conversational Systems**

Conversational systems have proven to be highly vulnerable to attack. They have little-to-no method of user authentication. For instance, personal voice assistant such as Alexa by Amazon uses a wake-word command. The voice command does not authenticate the user, any individual that knows the voice command can activate the system [45]. Phone-embedded voice assistant such as Siri and Google Now do not have a strong way of authenticating users throughout conversation.

Some conversational systems lack authentication methods. Thus making them vulnerable to attackers that mimic the user's way of command. In a reported incident, a smart-home owner discovered that his iPad unlocked the door for anyone who asked Apple's personal voice assistant Siri to do so [35]. Recent research show that attackers can send inaudible voice commands that are able to trigger phones personal voice assistance such as Google Now and Siri [65]. Unrecognizable by humans, commands may leverage the voice assistant's high permission abilities to accomplish sensitive tasks. Hackers may be able to access the device by visiting malicious sites, or send text messages on the users behalf [3].

Attackers may also exploit conversational systems by mimicking the user's way of authentication. Though Amazon's virtual voice assistant gives the user the option to set a voice passcode to confirm purchases, most users do not have it set it up. A recent incident showed that a six year old girl was able to order herself a dollhouse and a box of cookies without her parent's consent [35][63]. When this story made headline news, the news anchor discussing the incident reinstated the command said by the girl. His statement triggered orders on all Alexa devices within 'earshot' of the broadcast. Despite the fact that some authentication methods are available to voice assistants, they are currently **out-of-band** which means they require the user to identify themselves outside of conversation.

**Relevant Research**

The conversational system authentication field is important when developing computerized systems, yet it sometimes operates as a silo independent from the usability of the system [8]. For example, entering a PIN or password outside of conversation. This false approach designs security independently and limits the usability of the system, thus creating a gap between the usability and the security of systems. Human-Computer Interaction and security professionals argue that security should vary based on the user roles and cognitive limitations which are in line to the tasks they are trying to achieve and their environment [41][20]. Current researchers are trying to fill the usability gap and increase the security of conversational system and personal voice assistants. Recent studies focus on various gadgets that users can wear [26][61], that can either identify specific patterns related to that user, or can specify their location to avoid spoofing [47]. Other researchers focus on the security framework and protocol [2] without keeping the user's end goal in mind. For example, Al-Muhtadi et al., present the mist router that provides a secure communication infrastructure for conversational systems.

The vision of the conversational research that focuses on wearable technology and security is to simply attach a device onto users that can authenticate them without any further action required. The goal is to identify specific users, discover their presence based on location, and communication to other smart systems. Feng and colleagues [26] introduced VAuth, a wearable device that can be placed on eye-glasses, earphones, or necklaces. The

VAuth system constantly measures body surface vibration of the user and compares it with the voice signal received by the system. If the vibration and the command match, then the user is believed to be the one who completed the command. When testing the system, Feng and colleagues [26] claim to achieve 97% detection accuracy and less than 0.1% false positive rates against attacks such as replay attacks, where an attacker records the user's voice and replays it to gain access to their system. As well as mangled voice attacks, where an attacker uses incomprehensible voice segments to map voice vectors.

Other research done by Liu and colleagues [48] had a similar focus: they developed a wearable device that measured the sounds of the user as it travels across their body, also known as the user's vocal resonance. They were able to successfully train the system on specific user patterns then use them as reference for when commands are instructed. Similar to Cornelius' team [16] and other professionals in the field [48], Liu and colleagues focused on creating a training module for the wearable devices to help them identify which user is wearing them.

Current authentication requires an additional step, independent of the user's 'conversation' with their voice assistants. Biometric authentication in conversational systems only work when the user is physically present and is using voice commands. Current biometric authentication strategies do not support text-based conversational systems such as chatbots.



## 2.5 Human Communication Over a Medium

It is evident that humans can detect pattern in conversational style and text [44]. Current research focuses on humans ability to communicate through a computer medium. Research that focuses on computer mediated human communication (CMC) mainly discusses technology's effect on the individuals Theory of Mind, also known as ToM. Kidd and his colleagues refer to the Theory of Mind as the human's ability to understand other people's thoughts, and emotions [43] in an interaction. Researchers suggest that when humans communicate, they tend to consider the other person's beliefs, intentions and knowledge [28][13]. Very little is known about the techniques humans use to authenticate each other in computer mediated human communication. Recently, scholars acknowledge that studying the Theory of Mind is key to effective communication, and they propose that in order for future smart conversational systems to be more accurate, it must have a better understanding of human intentions, beliefs and tendencies [12].

Some researchers argue that emotions are not well perceived through computer mediated human communication. A study conducted by Epley and Kruger [23] revealed that there was a great ambiguity of people's impression when compared between face to face interactions versus over email. Other scholars argue that it does not negatively effect the individuals ability to predict performance and high collective intelligence [22][43]. Theory of Mind evolves with socialization and it support pro-social behaviour, cooperation

and coordination within social interactions [12]. It also plays an crucial role in building trust, and strategic interplay [24]; both of which are important in the context of authentication.

Other research also indicate that humans establish common grounds when conversing [13] in which we argue that it is important when humans identify each other during conversation. The ‘common ground’ theory proposed by Herbert Clark and states that humans constantly assess each other’s knowledge level in order to build a common reference point most relevant to the person with lower knowledge to create a common ground. For instance, a doctor will always assume that their patients do not have the same medical education so they refer to medical conditions by the name commonly used in society [13]. Clark’s work on the common ground theory is important in human to human authentication because it shows that an individual is able to assess people’s knowledge and alter their communication to make sure the other person understands the context. Based on the other individual’s perspective of the topic, humans tend to use less words when explaining contexts.

### **2.5.1 Turing Test**

Alan Turing was interested in studying humans detection skills in conversation in the context of theorizing about artificial intelligence. In this section, we briefly talk about the imitation game he developed, and how it helped computer scientists establish a new tool that assess interaction over a com-

puterized medium. Understanding Turing’s work, and game structure is important to our research because it closely mirrors our line of work.

Alan Turing published the article “Computing Machinery and Intelligence” in 1950, where he postulated the question “Can machines think like humans?” [64]. In order to test his question in a more concrete way, Turing came up with the ‘Imitation Game’ which is commonly referred to as the ‘Turing Test’[60]. The game measures whether a human judge is able to distinguish through separate, blinded conversations which player they’re conversing with: a machine or another human. If the machine is able to speak the same way as humans and convince the interrogator that it is human, then it passes the Turing Test.

The ‘Turing Test’ inspired Von Ahn and colleagues to design new assessment of interaction called the Reverse Turing test. They developed the famous Human Interaction Proof test called: ‘Completely Automated Public Turing Test to Tell Computers and Humans Apart’ test, also known as ‘CAPTCHA’ [32]. The Reverse Turing test allows security applications to determine whether they are interacting with a bot or a human when accessing information. CAPTCHA most commonly relies on character or shape recognition [66], and degraded image of random words with different fonts [14]. Similar to CAPTCHA, reCAPTHCA is another form of human validation over the web that relies on meaningful words extracted from books [67]. It is important to note that the Reverse Turing test relies on the same question posited by Alan Turing, that he later answered with the game:

is the interrogator interacting with a machine or a human being.

Our 'Friend Imitation Game' structure closely resembles the 'Imitation Game' structure introduced by Alan Turing. We further discuss the differences between our game and the Turing Test in Chapter 3.

# Chapter 3

## Methodology

In this chapter, we explain our research criteria and why we chose to test it with an e-commerce game. We introduce our study design and research protocol. Towards the end, we discuss how our game is different from Alan Turing's imitation game.

### Research Criteria

In order to test for human-to-human authentication methods in text interfaces, our research criteria should meet the following:

1. Individuals must be familiar with each other to properly identify one another and to be able to have similar common grounds, as proposed in the work discussed by H. Clark and colleagues [13].
2. Individuals must verify their partner without voice and facial recognition.

3. Identification must occur during natural, ongoing conversation similar to the ones humans use daily over text.
4. Conversation must be goal-oriented so participants are motivated to complete and authenticate each other in a short period of time.
5. Research must be practical to complete within a year time frame and ethically approved by Carleton.

With our research criteria in mind, we requested participants to come in with someone they are familiar with. Then, we asked them to play our proposed 'Friend Imitation Game' in separate rooms. The e-commerce game requires a total of 3 players; the two participants and the moderator acting as an adversary. Participant 1, also known as the authenticator, has to converse with another player online then either approve or decline a monetary transaction based on whether they believe they are talking to their partner or an adversary. Participant 2, also known as the convincer, has to convince their partner that they are chatting with them, not an imposter. We explain the steps of the game in more detail in section 3.2.

### **3.1 Study Setup**

The study was approved by Carleton University Research Ethics Board-B (CUREB-B). Posters attached in Appendix B.2 were printed and distributed across Carleton University. Social media recruitment messages were dis-

tributed through Carleton Research Participants Facebook group (see Appendix B.1 for the social media invitation). Eligible participants had to be over 18 years old, comfortable communicating in English, and using a computer. Individuals who emailed the researcher were invited to the study.

The research took place at Carleton University Human and Computer Interaction labs or at a mutually agreed private space. The researcher has taken courses on qualitative and quantitative research methods, including observational research. The researcher has obtained TCPS 2 certification and is qualified for level C first aid with CPR & AED.

The game software platform was coded by Robin Tropper as a web application programmed in PHP, designed to run on a Linux system running an apache web server and MySQL database server. The application was hosted on an Ubuntu 18.04 virtual machine hosted in Carleton University's Computer Security Lab (CCSL). The game participants each connected to this server using separate laptops connected to the Carleton network.

The game consisted of 3 separate log-ins: The Moderator, The Convincer, and The Authenticator. Controlled by the researcher, the moderator's screen shown in Figure 3.1 has the option to start and reset time, disable the convincer from participating, and end the game. The moderator can also view previous transactions and keep track of the players scores.

The Authenticator screen shown in Figure 3.2 gives the player the option to approve or decline transaction. The Convincer's screen shown in Figure 3.3 allows the player to only type and send messages when the Moderator

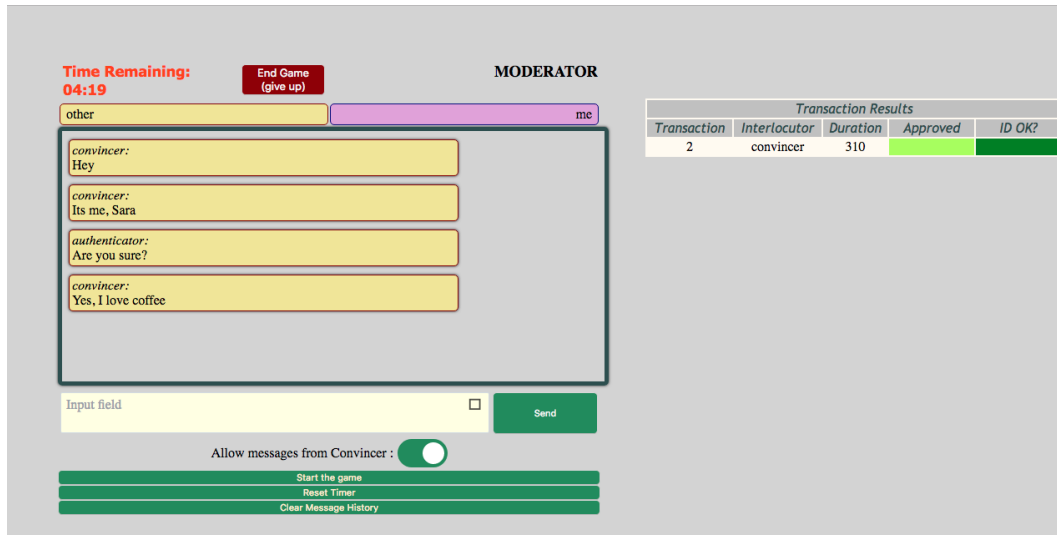


Figure 3.1: The moderator screen during the game

activates the option.

Both players had the option to end their participation at any point during the game. Participants also had an external chat dialogue that directly connected them to the researcher throughout the game, in case they needed to verify or ask any questions outside of the game.

## 3.2 Protocol

Once we greeted the interested participants, we provided them with a 'Research Summary' paper that briefly explained the motive of the study, our research question and relevant definitions (see Appendix A.1). Then, the researcher gave each participant a consent paper form and orally explained each section. The written consent form outlined in details what was ex-



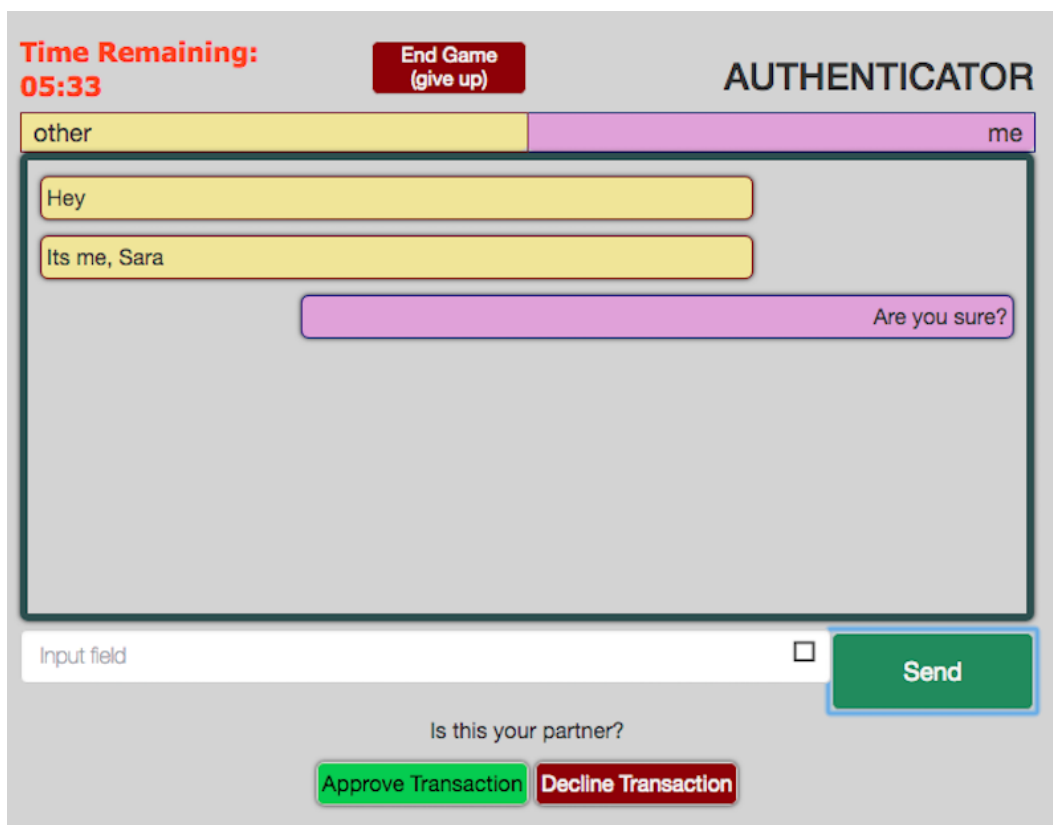


Figure 3.2: The authenticator screen during the game

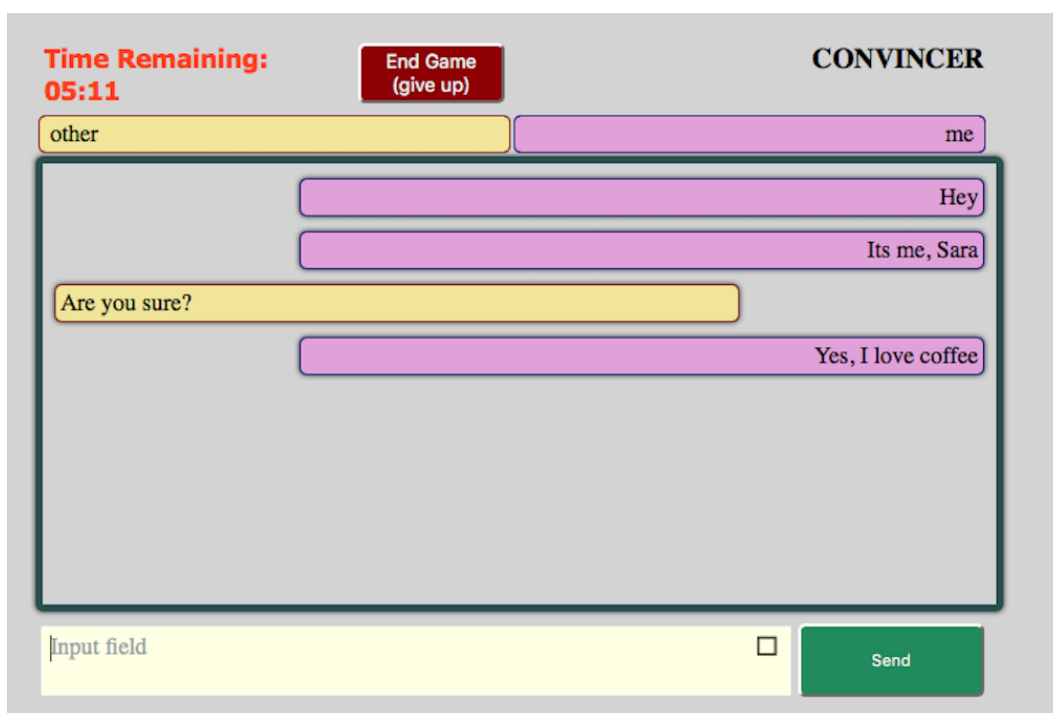


Figure 3.3: The convincer screen during the game

pected of the participant during the research, the length of the study, how their data will be handled and stored, the steps to withdrawing their data and the incentives they will receive (see Appendix A.2).

### **3.2.1 Session Outline**

Each session had three parts. 1. Pre-game questionnaire. 2. The game, and 3. Post-game questionnaire and interview. Refer to the flowchart (Figure 3.4) for a more comprehensive step by step outline of each session.

#### **Pre-game Questionnaire**

Each session officially started once the consent form was signed by both participants and the researcher. The researcher explained the game orally and gave out a list of game rules to each player (see Appendix A.4 and A.5). Then, the players were separated into two different rooms and were asked to fill out a pregame questionnaire (see Appendix A.3).

The pre-game questionnaire collected data on age, ethnic background, and first-language. In order for us to rate the pairs familiarity, we asked them how many years they have known each other for and how often do they text. The pre-game questionnaire ended with a perceived familiarity score where we asked them “How familiar or unfamiliar is your game partner to you?”. We provided them with a 5-point Likert scale answer where 1 was extremely unfamiliar and 5 was extremely familiar.

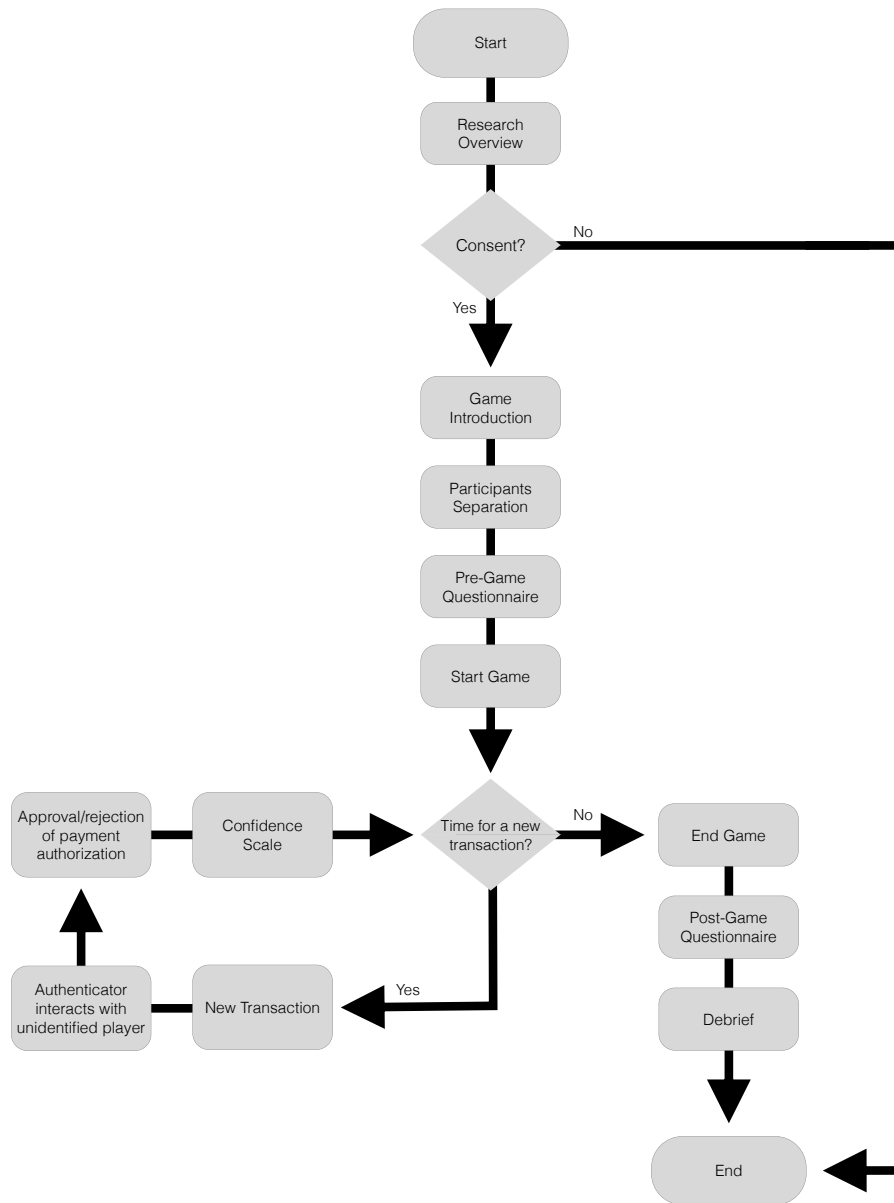


Figure 3.4: A step-by-step flowchart of one research session

## Game

There were a total of three players in the game, each player had their own laptop and sat in a separate room. Participant 1 played the authenticator role, participant 2 played the convincer role, and the researcher acted as a moderator and played an adversary role. A visual of the game outline is presented in figure 3.5. A simulated game mock up and specific use case scenarios are attached to this report; refer to Appendix C for more information.

The authenticator had a chat dialogue in front of them and they had to converse (by typing) with the unidentified second player on the other end. They were either chatting with the convincer or the researcher acting as an adversary. The authenticator was asked to authenticate the unidentified player then either approve or decline a virtual electronic payment that was about to be sent from their bank account. They were instructed to approve the payment when they are chatting to the convincer and decline the payment when they are chatting with the adversary. Once they approve or decline a payment transaction, the game prompts a confidence scale question about their decision and offers a 7-point Likert scale answer where 1 is extremely unconfident and 6 is extremely confident. Once an answer is selected, the game starts a new transaction and the previous transaction history is deleted for both participants. The researcher maintains access to the history of conversation throughout the game and uses it as a tool to trick the authenticator in subsequent transactions. The convincer is only able to see an active chat dialogue at the times they are assigned to chat to the authenticator.

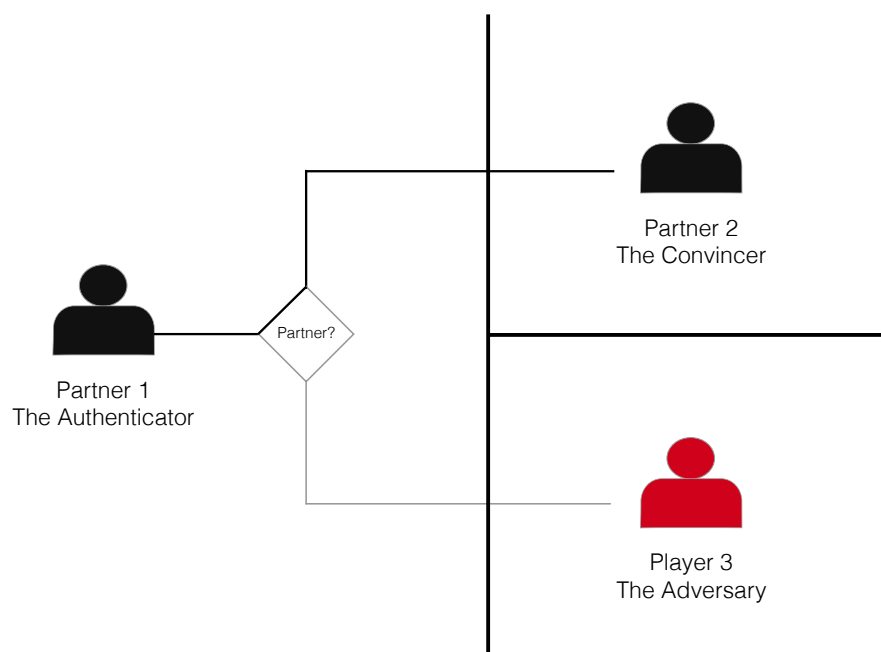


Figure 3.5: A game snapshot of a transaction between the authenticator and convincer

The allocated game time given to each pairs was 10 minutes. We asked the participants to complete as many transactions as possible within this time-frame. Participants are told that for every correctly identified transaction they earn 10 points in the game. The highest points earned were contacted and congratulated at the end of the study.

To encourage natural conversation, the use of emoticons was permitted. Emoticons consist of symbols, pictures and faces that imitate facial expression [31]. Experts claim that humans sometimes communicate more effectively by using graphic elements [18] and we did not want to restrict participants.

In order to deceive the authenticator, the adversary mimicked the convincer's conversational style during the game. Specifically, the adversary tried to use similar phrases used by the convincer in previous transactions (example: cool beans vs. great), the adversary re-used some facts that came up through previous transactions (example: favourite colour) and tried to mimic a similar texting style (example: ok vs. okay). The adversary was signed in and had constant access to the transactions history that includes conversation history, and decline/approve payment responses. The adversary had access to the transaction details even when the authenticator was chatting with the convincer.

### **Post-game Questionnaire and Interview**

After the game, we asked the authenticator to rate how easy or difficult it was for them to identify their partner on a 7-point Likert scale where 1 is extremely difficult and 7 is extremely easy. We asked them to explain their answer then proceeded with a discussion about the kind of cues they used throughout the game.

Then, we disclosed whom they were talking to in every transaction during the debriefing period (see Appendix A.7). We then assigned them a percentage score based on whether they were able to correctly identify their partner versus an adversary. Participants were given a 'Consent to Use Data' form (see Appendix A.8) at the end of the interview that outlined what we are trying to accomplish from our research, how we will be using their data, their rights to withdraw and additional information if they found the research emotionally disturbing.

### **Comparison with the Turing Test**

The structure of the 'Friend Imitation Game' closely resembles the imitation game proposed by Alan Turing. Although proving you are human to another human is a key feature of Turing's initial 'Imitation Game', we argue that our game is completely different. First, Turing wanted to answer whether Machines can think like humans, his main goal of the research was to figure out whether a machine can be 'intelligent' enough to deceive a human judge. During the imitation game, the human judge has to identify a random human



versus a machine. On the other hand, our main goal of the research is to figure out how humans identify each other over text. During the 'Friend Imitation Game', the authenticator is verifying whether they are speaking to a human they know well versus another human acting as an adversary. We are focused on the technique the players use within context, rather than the outcome. Second, the imitation game has a separate blinded conversation platform whereas the adversary in the friend imitation game has access to all, and is encouraged to use all data points throughout conversation when mimicking the convincer's style of texting. Furthermore Our 'Friend Imitation Game' included simulated money transfer over online transactions and point rewards to motivate participants throughout the game. After all, both games are similar in structure yet they measure different data points.

# Chapter 4

## Results

In the following section we analyze the participant’s backgrounds to understand our sample demographic, and their familiarity with each other. Then we analyze the conversational scripts between participants and classify the common ways humans authenticate each other over text dialogue.

### 4.1 Overview

On average, the experiment time was 30 minutes long. There was a total of 24 pairs that participated in our research. Participants completed an average of eight transactions within the allocated 10 minutes game time with a 72.9% accuracy rate. Table 4.1 shows the total number of transaction per pair and the total number of correctly identified transaction during that game. Specifically, the number of transactions correctly identified refer to when the

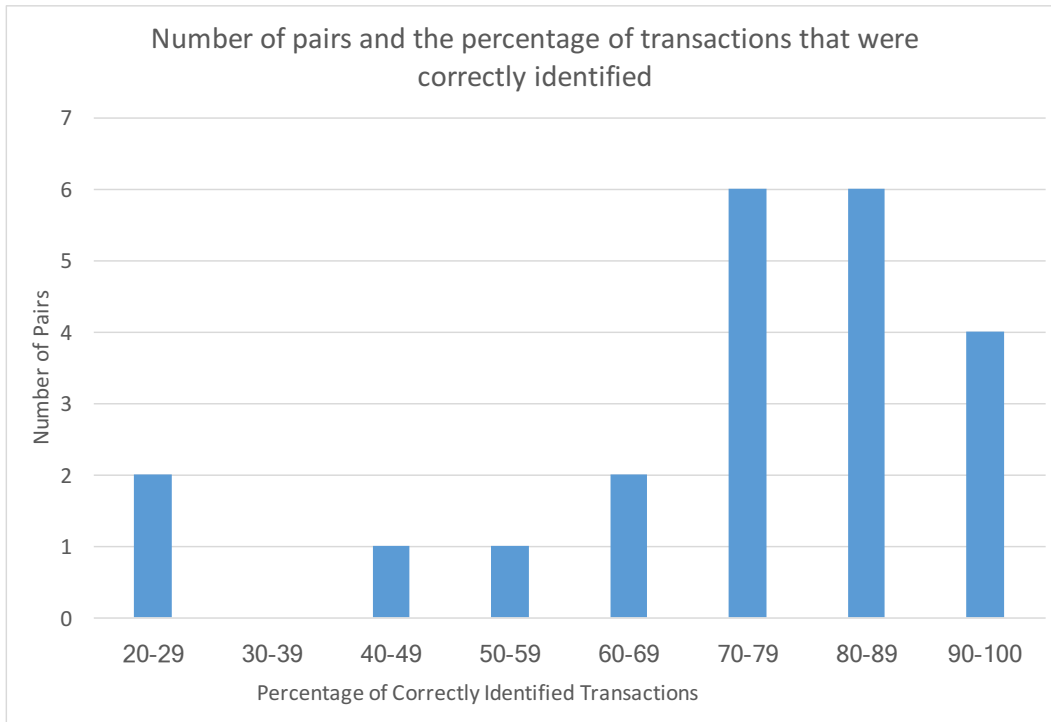


Figure 4.1: Percentage of transactions that were correctly identified per pair authenticator declines the adversary and approves their partner. The table only shows 22 data points—two session data sets were removed due to errors in data collection of the participants performance. Please note that other data analysis throughout this chapter will include those two sessions as we will be analyzing their conversation, not their performance. No one was excluded in this study since all individuals that participated were familiar with each other.

	Identification #	# of Transactions	# of Correct Transactions	%
1	Pair 57	8	8	100
2	Pair 53	6	6	100
3	Pair 30	5	5	100
4	Pair 51	10	9	90
5	Pair 40	8	7	87.5
6	Pair 49	7	6	87.5
7	Pair 18	7	6	87.5
8	Pair 39	10	8	80
9	Pair 10	5	4	80
10	Pair 35	10	8	80
11	Pair 22	9	7	77.7
12	Pair 55	21	16	76
13	Pair 56	4	3	75
14	Pair 48	8	6	75
15	Pair 50	8	6	75
16	Pair 38	10	7	70
17	Pair 14	6	4	66
18	Pair 45	13	8	61.5
19	Pair 20	7	4	57
20	Pair 47	5	2	40
21	Pair 41	9	2	22.2
22	Pair 44	5	1	20

Table 4.1: Subject pair performance

## 4.2 Participants Background

As shown in table 4.2, most, 66.6% (32 out of 48) of the participants were between the ages of 21-29 years old. Of the participants, 29.1% (14 out of 48) were between the ages 18-20 years old. Whereas, the rest 4.1% (2 out of 48) were above 30 years old.

Most, 66.6% (32 out of 48) of the participants have known their partners for '1-5 years'. Whereas 14.5% (7 out of 48) said they met their partner less than a year ago. Some, 10.4% (5 out of 48) have known their partner for '6-9 years' and 8.3% (5 out of 48) have known their partner for over 10 years.

A little over half, 56% (27 out of 48) of the participants indicated they were extremely familiar with each other. Other participants, 37.5% (18 out of 48) indicated that they are 'moderately familiar', whereas the rest of the participants 6.2% (3 out of 48) indicated that they either were 'Somewhat Familiar' or 'Slightly Unfamiliar' to their partner. When we asked the participants how often they text their partner, 39.5% (19 out of 48) indicated they 'Text all the time'. Whereas 37.5% (18 out of 48) specified they text '1-5 times a day'. Of the participants, 14.5% (7 out of 48) said they 'Never' text their partner. The rest, 8.3% (4 out of 48) indicated that they text '6-10 times a day'.

Background	# of Participants	%
Age Range		
18-20 years old	14	29.17
21-29 years old	32	66.67
30-39 years old	1	2.08
40-49 years old	1	2.08
First Language		
English	22	45.83
Arabic	16	33.33
Farsi	4	8.33
French	2	4.17
Chinese	2	4.17
Sinhala	1	2.08
Spanish	1	2.08
How long have you known your game partner		
Less than a year	7	14.58
1-5 years	32	66.67
6-9 years	5	10.42
Over 10 years	4	8.33
How often do you text your game partner		
Never	7	14.58
1-5 times a day	19	39.58
6-10 times a day	4	8.33
We text all the time	18	37.50
Perceived Familiarity to your game partner		
Slightly familiar	2	4.17
Somewhat familiar	1	2.08
Moderately familiar	18	37.50
Extremely familiar	27	56.25

Table 4.2: Participants Background

## 4.3 Classification of Techniques

In the following section, we identify authentication themes used by participants throughout the game to easily classify and analyze each data set. Then, we group the themes into two different categories: 1. the 'Semantic Measurement' category, also referred to as 'The What', and 2. 'Behavioural Characteristics' category, also referred to 'The How'.

There were 5 distinct themes that players used as methods to identify or prove it is themselves to their partner: 'Knowledge & Experience', 'History & Plans', 'Texting Style', 'Response Speed' and 'Personality Type'. We classified those themes into two categories refer to Table 4.1 for more details on the number of interactions and pairs that used each theme. Figure 4.2 shows the frequency of each theme used based on the number of interactions within all transactions. Overall, the 'Semantic Measurements' were used in 155 interactions, and 'Behavioural Characteristics' were used in 41 interactions.

### 4.3.1 Semantic Measurement: *The What*

The 'Semantic Measurement' category classification was based on what participants said during conversation. It contains the top 2 most used techniques: 'Knowledge & Experience' ranked as number one, and 'History & Plans' theme ranked as number two. The 'Knowledge & Experience' theme included facts collected as knowledge about individuals such as frequent habits, birthdays, nicknames, likes and dislikes based on their experience to-

Theme	Frequency of Interactions	Number of Pairs
Behavioural Characteristics		
Texting Style	31	15
Response Speed	6	4
Personality Type	4	3
Semantic Measurements		
Knowledge & Experience	92	24
History & Plans	63	19

Table 4.3: Frequency of authentication themes used by participants throughout the study

gether. Authenticators often asked about frequent habits that their partner do, such as studying at the library in a specific spot, medication, smoking, or playing a game frequently. In the following example, an authenticator implies that the convincer plays a specific game then asks for more details about that game: “How well does insurgency run on ur computer?” whereas convincers included habits as statements. For example: “This room is so quiet it reminds me of library 5th floor”. Interestingly, we also saw reverse validation where an authenticator would intentionally imply a wrong habit and wait for the convincer to correct them. For instance: “I think [5 dollars] can buy [you] a double double no? How often do you drink coffee?” then they received the validation response of: “I don’t drink coffee” from the convincer. Other techniques found in the ‘Knowledge & Experience’ theme included nicknames that were used throughout conversation when addressing partners in combination of statements about their likes and dislikes. Authenticators



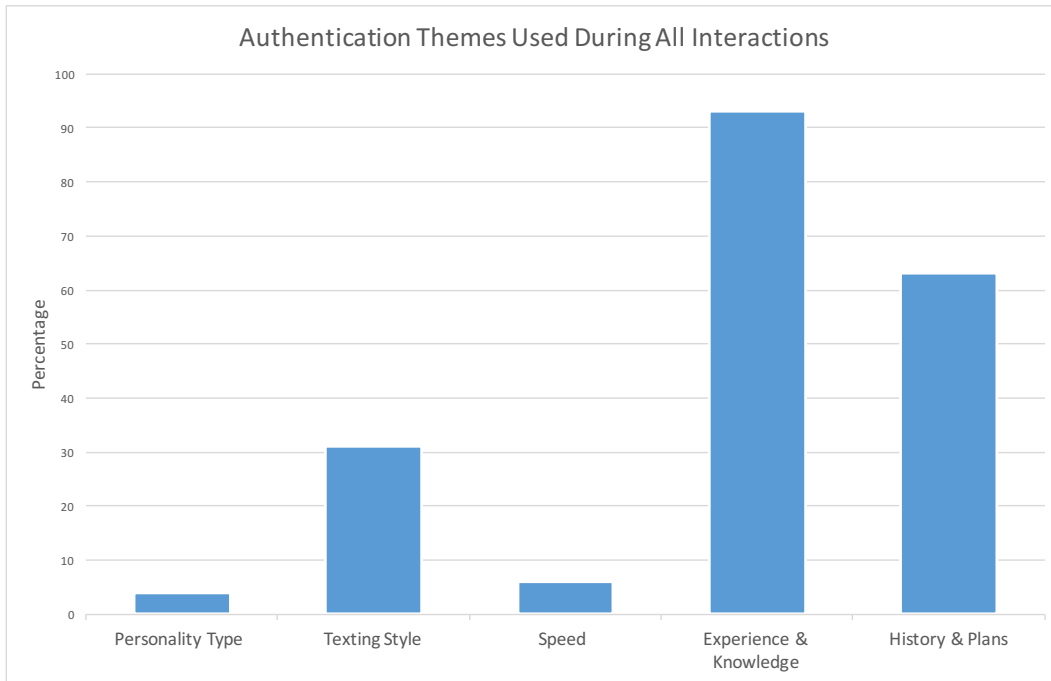


Figure 4.2: Percentage of authentication themes based on interaction count often asked about birthday dates, favourite food, and admired characters.

The second most frequent theme observed in the ‘Semantic Measurement’ category was the shared ‘History & Plans’ theme that mainly revolved around previous and planned future occurrences. Conversation that focused on occurrences was either describing a major or a minor life event. Major life events included conversation about honeymoon plans, birthday party outings, first day they met, future travel plans, and specific graduate programs they are planning on attending. Minor life events mainly focused on details related to their outings such as their last shared meal, movies, or go-karting.

### 4.3.2 Behavioural Characteristics: *The How*

The ‘Behavioural Characteristics’ category focused on the following three themes: ‘Texting Style’, ‘Response Speed’, and ‘Personality Type’. The ‘Texting Style’ theme was demonstrated by convincers and picked up by authenticators. Throughout conversation, convincers used a specific set of vocabulary, emoticons and word reconstruction. Convincers used shortcuts in text such as ‘ur’ vs ‘your’, ‘zem’ instead of ‘them’, and often extended some words like ‘seeeend me’ versus ‘send me’. In the game, some authenticators pointed out texting discrepancies, for instance “you usually have way more spelling mistakes”.

Occasionally, authenticators indicated variations in ‘Response Speed’ between their partner and the adversary while playing the game. They commented with ‘I’m waiting’ or ‘you took too long’. They also asked questions that indicate more about ‘personality type’ and outlook on life. One Authenticator stated that they ‘often can tell’ when their partner is talking to them when the convincer asked them how do they know it was them talking during the game. Convincers chatted about conspiracy theories, and authenticators asked about beliefs and mentioned general characteristics about their partner.

## 4.4 Semi-Structured Interview

After the game was complete, we combined both participants into a single room, and asked the authenticator to indicate how easy or difficult it was to identify their partner over text (Refer to Appendix A.6). Overall, 75% of participants rated their ability to identify their partner as above 5, or 'somewhat easy'. Afterwards, we conducted a semi-structured interview where we asked them about the different cues that helped them throughout the game.

Authenticators said that they were able to identify behavioural cues. Half of the authenticators (12 out of 24) specified that they were able to identify their partner based on their personality type; they either referred to their partners as 'a jokester' and 'sarcastic', or knew that the conversation would steer in a specific direction because they knew their partner would only share conversation that was of purpose. A little over half (13 out of 24) authenticators were also able to pick up on cues based on the 'way they talk and the way they text'. In fact, many authenticators distinctly mentioned that their partners texted the same way they spoke. Some (7 out of 24) authenticators were also able to distinguish the variation in response time between the adversary and their partner.

Similar to our analysis of their conversation, (8 out of 24) authenticators also mentioned that they looked for semantic measurements such as: personal facts about themselves or partner (such as favourite food, drinks), and their lifestyle (such as activities). A little less than half of the authenticators (10

out of 24) indicated that they used previous and prospective encounters as means of identification. Generally, participants were more skeptical when the same topics were mentioned repeatedly and less detailed were shared during conversations.

## 4.5 The Patterns of Techniques

In this section, we identify two common authentication techniques participants used in the study.

The majority, 87.5% (21 out of 24) of game sessions used a cross examination pattern that was relative to the authentication themes mentioned previously. More often, authenticators combined a series of clarifying questions subsequent to each other, even when the convincer answers their initial question correctly. In the example below, C refers to convincer and A refers to authenticator. Please note that the names were altered to keep the identities of the participants protected:

Start of conversation with the Convincer

C: could you send money for your bubble tea

C: [Joe] said you got like 2 bubble teas already

A: bubble tea from which city?

C: we want to make sure we have enough bbq funds

C: hong kong

A: which island?

C: you stayed in kowloon

Convincers used a less common pattern when trying to prove their identity. They used an answer and elaborate pattern when responding to some of the authenticators questions. For instance, if an authenticator asks ‘did you take your pills today’, they would respond with ‘yes my ulcer pills’, or when the authenticator mentions a previous occurrence, such as visiting the theme park, they would mention the fact that they also purchased a VIP card. The cross-examination pattern and the answer and elaborate pattern were commonly used to affirm identity.

## 4.6 Interesting Pairs

Most participants followed a similar approach to authentication throughout the game. They used a series of questions that involved all of the five themes identified previously. Some transactions throughout the game however were completed fairly quickly where the authenticator was able to correctly identify who they were talking to by reading one sentence. In this section, we talk about pair 55 and pair 57 that had unusual patterns of conversation. Pair 55 is interesting because they were able to complete a total of 21 transactions in 10 minutes. Whereas pair 57, had a unique way of conversing.

**Pair 55**

Pair 55 indicated that they have been in a romantic relationship together for less than a year, however they both indicated that they text all the time and they are moderately to extremely familiar with each other. They were able to go through 21 transactions, twice as much as an average pair went through. They completed the highest number of transactions completed during the 10 minutes of the game.

Their style of authentication was interesting because the authenticator did not communicate during the game whatsoever; instead they accepted or declined transactions based on the information the convincer shared. The convincer only sent 1 to 2 sentences per transaction that revolved either around their shared 'Knowledge & Experience' or 'History and Plans'. The authenticator was able to correctly identify their partner 92% of the time (12 out of 13 transactions), whereas they were able to correctly detect the adversary 50% of the time (4 out of 8 transactions). The authenticator was able to correctly identify an adversary and decline the transaction when the content shared was based on 'History and Plans' regardless of whether the convincer had just mentioned them. In the example below, 'C' refers to convincer and 'M' refers to the moderator (the adversary).

Start of a new transaction with the Convincer

```
C: i want to farmboy to get a kilo of oranges later if you  
    want to come
```

Authenticator accepted

Start of a new transaction with the Adversary

M: do you want to come with me

Authenticator declined

The moderator was able to trick the authenticator when they copied and pasted the last sentence that the convincer had just sent. For example:

Start of a new transaction with the Convincer

C: baby i think rotana will be good

Authenticator accepted

Start of a new transaction with the Adversary

M: baby i think rotana will be good

Authenticator accepted

Start of a new transaction with the Convincer

C: i really cant be bothered with starting a masters this  
summer

C: i just wanna gym and eat briskets

Authenticator accepted

Start of a new transaction with the Adversary

M: i just wanna gym and eat briskets

Authenticator accepted

During the debrief, the authenticator stated that they thought the game was defective and their approval did not go through the first time. The second technique that the moderator used to trick the authenticator was to restate facts based on their shared ‘Knowledge & Experience’. For instance, the convincer mentioned that the authenticator does not eat sushi or drink coffee. The adversary then used that knowledge in a different context for example: ‘I would like to get you to start liking sushi’. Copying this theme, however, was not always reliable. When the adversary asked a question like ‘will you drink coffee for me?’ the authenticator declined the transaction. This is important because it indicated a deep knowledge of their partner, much deeper than simple likes and dislikes.

### **Pair 57**

The two individuals in pair 57 indicated that they knew each other for 6-9 years, that they are extremely familiar with each other and they text all the time. They were able to complete a total of 8 transactions within the game’s allocated time. Their conversational pattern and style was interesting



because it defied proper English. The authenticator was able to correctly identify the identity of players throughout all of their transactions. In this example, 'C' refers to the convincer and 'A' refers to the Authenticator.

Start of a new transaction with the Convincer

A: How may i help you  
C: whats boppin  
C: its ya boy, skinny p\*\*\*  
A: sayyyy lesssss  
C: i need guap  
A: whats our ggroup name?  
C: pls bb  
C: sisters fam  
C: till we die  
A: facts

Authenticator accepted

During their post-game interview, they explained their technique and terminology to the researcher. They said they know each other so well, that they are able to tell who is sending the messages. The authenticator then stated: 'he texts the way he talks; he is a clown'. They further pointed

out that they have a lot of inside jokes and specific use cases of emoticon when they text. For example: they refer to money as ‘guap’, and affirm a statement by using ‘facts’. Every transaction played with the Moderator was unsuccessful. Although the moderator imitated their language style and conversational topics, the authenticator was not deceived. In the example below ‘A’ refers to Authenticator and ‘M’ refers to the moderator that acted as an adversary.

Start of a new transaction with the Adversary

A: hello

A: [use of emoticon]

M: [re-sent the same sad emoticons used by convincer]

A: why you sad?

A: need some X?

A: tentacion?

M: i need guap

M: sister me out here

M: whats boppin

A: idk whats guud in the hood?

A: do you like movies?

M: are you getting my messages sis

A: nah nah you a fake

Authenticator declined

At a first glance, some may believe that the authenticator declined the adversary based on the question 'do you like movies?' however, the time stamp between that message and the consequent message is very short. The authenticator did not wait for an answer; instead, they knew right away that they were not talking to their partner. Pair 57 had a very strong relationship and when we asked them how easy or difficult it was to identify your partner over conversation, they indicated that it was extremely easy. Their interaction was interesting because it showed a deep understanding of their partners way of texting and talking that was not easily understood by an observer.

# Chapter 5

## Security Analysis

In this chapter, we review the different data sources an attacker needs access to in order to retrieve relevant data for the five themes identified under the 'Semantic Measurement' and the 'Behavioural Characteristics' categories: 'Knowledge & Experience', 'History & Plans', 'Texting Style', 'Response Speed' and 'Personality Type'. Then, we consider the complexity of analysis for each of those authentication themes to rank their robustness against possible attacks.

In order for us to understand the feasibility of our identified authentication themes we first explore the data sources then rank each theme based on a predefined analysis complexity. The term 'data source' in the following text refers to the different locations that could contain specific data. For instance, birthdays are generally found on public records and are typically added on social media accounts. We explore the various data sources available per

theme to fully understand the accessibility of the data.

We analyze the steps an attacker would need to do in order to properly impersonate an individual by using the themes identified in the previous chapter. In the following text, we refer to that analysis as the 'Threat Complexity' level and we base it on the time and effort it would take an attacker to use the information if they were able to gain access to the data source. As shown in table 5.1, there are 3 analysis complexity levels: Low, Medium, and High. The low analysis complexity level is assigned to a theme when an attacker is able to use the information almost immediately after they obtained it with minimal effort. By minimal effort we are referring to a simple query look up of a database. For example, we would assign a low threat complexity analysis level to birthday dates, as it would only require a basic database search to find and use. Authentication themes that were assigned a medium threat complexity analysis are the ones that require statistical and other forms of analysis that can easily be automated. For instance, a medium level analysis would require the attacker to measure the frequency of words and phrases in conversation to identify likes and dislikes. Whereas high threat complexity analysis refer to data that require a substantial amount of effort to process and use. The analysis of high threat complexity cannot currently be automated in a straightforward manner. It would require affect analysis [25]. Essentially, the attacker has to understand the semantics of the data.

In the following sub-section we will talk about the data source and analysis complexity for each theme in the 'Semantic Measurement' and the 'Be-

Rank	Analysis Complexity
Low	Simple Query
Medium	Simple Analysis
High	Complex Analysis

Table 5.1: Analysis of the threat complexity

havioural Characteristics' categories.

### 5.0.1 Semantic Measurements

The semantic measurement category has the 'Knowledge & Experience' and the History & Plans' themes.

The first 'Knowledge & Experience' theme was used most frequently by participants that included birthdays, nicknames, habits, likes and dislikes.

We rank birthday dates and nicknames as low threat analysis for a few reasons. First, they are easy to find on social media, public and private records. Birthdays are generally shared, and celebrated with the public thus making this information accessible by everyone. Second, the time and effort it would take an attacker to find nicknames and birthday dates within a data source is relatively low.

Habits, likes, and dislikes are ranked as medium threat complexity because this information is not easily accessible by various data sources, and

when found, the attacker needs time to analyze and interpret data into context. For instance, an attacker may outline a user's habit by tracking their mobile devices, getting access to their credit card history, and watching their shared locations on social media. Although an attacker may be able to identify frequent locations visited by the user, they may not be able to understand the user's behaviour such as their likes and dislikes. The attacker can only make an assumption that a user likes their purchases, hence buying the same thing often. We argue that likes and dislikes are a little more complex than that because first: no one truly knows *why* an individual is behaving this way unless they were explicitly asked, second: humans tend to change their preferences over time. An individual may be visiting the same restaurant because of a crush they have on the bartender and not be a huge fan of the food per se.

The second semantic theme is the 'History & Plans' that consisted of past and future life events.

Past and future major life events are ranked as low threat complexity because most life events such as anniversaries, moving to different cities, or accepting a masters program are typically found on social media, calendar invites, directory information and public records. The data sources are easily accessible. An attacker does not require a substantial amount of time and effort to portray and use this information.

Spontaneous, unplanned minor life events such as coffee outings are ranked as medium threat complexity because they will be trickier to identify and use.

<b>Theme</b>	<b>Data Source</b>	<b>Analysis Complexity</b>
<b>Knowledge &amp; Experience</b>		
Birthday	Social Media	Low
Nicknames	Social Media Phone conversation history	Low
Habits	Social media location 'check in' Credit card transaction Location Services	Medium
Likes & Dislikes	Phone conversation history Purchase history Personal Contact	Medium
<b>History &amp; Plans</b>		
Major Life Events	Phone conversation history Search History Social Media	Low
Minor Life Events	Phone conversation history Location services Personal Contact	Medium

Table 5.2: Security analysis for semantic measurements



The accessibility of data sources that have this information will require analysis of the users conversation history and location services over a period of time. The attacker will need time to collect and analyze information about a users minor past interactions. For instance, if two old friends run into each other on the street and have lunch together, then an attacker will need access to location history, transaction history, verbal communication (potentially recorded over lunch). Then, in order for them to use the life event as means of authentication, the attacker must go through and analyze the event.

### **5.0.2 Behavioural Characteristics**

The Behavioural Characteristics category consisted of the following three themes: 'Response Speed', 'Texting Style', and 'Personality Type'.

The first, 'Response Speed' theme is classified as medium analysis complexity because an attacker can measure and have similar 'response speed' as the user by accessing previous interactions. Simple analysis is needed to process and analyze the users average time to reply to messages.

The second behavioural characteristic theme is the 'Texting Style' theme that contained frequent use of specific vocabulary and Emoticon as well as word reconstruction. The texting style theme is ranked as medium threat complexity because of the time and effort it takes to analyze previous communication pattern. In specific, an attacker use the same terminology, phrases

<b>Theme</b>	<b>Data Source</b>	<b>Analysis Complexity</b>
<b>Response Speed</b> Time to Reply	Social Media Conversation History	Medium
<b>Texting Style</b> Vocabulary & Emoticons Reconstructed Words	Social Media Conversation History Social Media Conversation History	Medium Medium
<b>Personality Type</b> Favourite Topics Personality Traits	Conversation History Personal Contact Social Media Personal Contact Social Media Conversation History Location Services Calendar	Medium High

Table 5.3: Security analysis for behavioural characteristics

and words commonly used.

The third 'personality type' authentication theme is ranked as a high threat complexity analysis because the attacker must collect information from many data sources then carefully analyze them to understand the users personality traits and interests. For example, a user may have a more optimistic or pessimistic approach to certain controversial subjects. We argue that it is extremely difficult for an attacker to mimic personality without a long period of time and effort. They would need to closely study the user for years.

### **Types of Attacks**

Semi-automated attacks can take advantage of themes rated as low to medium threat complexity analysis because they do not require basic understanding of semantics in conversation. On the other hand, themes identified as high complexity analysis will require intensive work and a more thorough understanding of the person on the receiving ends. In other words, an attacker requires the ability to predict performance and behaviour[22][43], a feature only humans can currently preform by developing Theory of Mind and mental models about other individuals.

Themes that have a low to medium threat analysis ranking:

- 'Knowledge & Experience' theme.
- 'History & Plans' theme

- 'Response Speed' theme
- 'Texting Style' theme
- Some 'Personality Type' themes such as favourite topics

The only theme that has a high analysis ranking is the personality trait that falls under the 'Personality Type' theme.

# Chapter 6

## Discussion

In this chapter, we outline the contributions of this thesis, then we discuss the potential application of the authentication themes identified in chapter 4. We highlight the benefits and challenges it would face. Then, we interpret our study's limitation. The end of this chapter covers future research.

### 6.1 Paper Contribution

In the paper, we presented the first study that researches human-to-human authentication methods over a computerized medium. Second, we classified the authentication patterns used by humans into 5 themes and 2 categories. Third, we evaluated the robustness of each authentication theme against human and small-automated attacks.

## 6.2 Application to Conversational Systems

Overall, exploring, understanding and implementing user identification themes in an in-band authentication framework will solve what needs to be solved by focusing on the user and the task they are trying to complete. As mentioned by Thomas Hewett [33]:

*“An understanding of basic HCI and human cognition will help designers and developers to remain focused on the problems that should be solved rather than on the problems which can be solved. Solving a technically challenging implementation problem may produce the immediate gratification of having accomplished something, but it represents a trap if that solution does not simultaneously contribute to the end user’s task related goals”.*

We argue that an implementation of our theory will ensure that the system is successful, and intuitive because it adheres to the following criteria highlighted by Baranauska and colleagues [8] and Dhamija and colleagues [19]: 1. The security measure follows the users mental model. 2. The system understands the user’s motivation within an interaction. And 3. The system balances usability and security.

An in-band security structure will be using themes users are familiar with and use when authenticating each other. By implementing themes similar to the end user’s mental model, we reduce cognitive burden, as authentication

will be a part of the conversation. The users primary goal is not to manage their identity; it is typically to complete a task [8]. Our proposed in-band authentication idea will be highly intuitive because it would recognizes the user's motivation when interacting with the conversational system. By using an in-band authentication model, we are allowing the users to complete what they want to achieve faster, without resorting to outside methods of authentication.

Arguably, the in-band authentication method needs to be refined and tested however, if implemented correctly, it will create a good balance between usability and security. It will introduce a 'path of least resistance' [19], because it will be fully integrated within interaction.

### **Constant Authentication**

Similar to biometric authentication methods, we propose that any in-band authentication framework should use a continuous authentication method. Continuous authentication is when the system constantly identifies the user throughout their interaction. This technique of authentication has been promising because it helps validate the user at all times. Unlike traditional authentication methods, constant authentication will ensure the user is who they claim to be during the entire interaction, not just before accessing sensitive information.

### 6.2.1 Challenges

To successfully implement an in-band authentication technique similar to the way humans authenticate each other, we need to consider the time it takes to train the system on specific user behaviour. Similar to humans, we assume that it takes time, and many interactions to generate a persona for the user. Second, systems must be able to detect slight behavioural changes over time. We all know humans behave and believe in different things throughout life, therefore, any persona of the user created by the system must be revised and updated constantly. Third, the system must be trustworthy and must earn the users trust in order to fully interact in a naturalistic way.

### 6.2.2 Ethical Responsibility

Despite the fact that we believe in-band authentication will provide positive changes, the question of whether humans will cede control of technology comes to mind. Would a system that understand human behaviour, and persona formation pose a risk to humans? Even though in-band authentication works behind the scenes, it still is considered highly invasive. The user has the right to know what information is collected [41]. How it will be used within this frame work, and whether it will be released to the government or any interested organizations [32].

We are proposing to collect semantic and behavioural characteristic measurements to construct an potential in-band authentication system. Due to



the nature of the information we are proposing to collect, the impact of leakages pose great risk. It is our responsibility to create boundaries for these systems, and ask how will our framework will be used and how will it be protected.

### 6.3 Study Design and Limitations

Our study design has the following limitations: 1. Sample size 2. Artificial game setup and, 3. individual differences.

Our sample could be considered a ‘Convenience Sample’ because it was somewhat constrained to Carleton University student demographics however, we argue that it was also a ‘Purposeful Sample’ because we were able to select a set of information-rich cases for our study. Young adults, our targeted population, yielded sufficient data since Forgays and colleagues [27] show that young adults text most frequently, and spend an extensive portion of their time on their phones.

Due to ethical restrains, we were unable to conduct our study in a more naturalistic settings. Our study recorded data points that were simulated through a goal-oriented game. The game posed three limitations: Time constraints may have impacted the way humans interact since each session was restricted to 10 minutes. Secondly, the researcher’s role throughout the study was to act as a moderator and an adversary, which is highly prone to human error. In order to collect more meaningful data in simulated environments,

we followed up our game with a semi-structured interview as suggested by Martella and his team [49]. Lastly, during the game the players were actively trying to authenticate each other appropriately. This may not reflect the subconscious way of human-to-human authentication methods over text. In reality, users may not know when and whether an attacker is targeting them. Our research results cannot be generalized across more prevalent attacks such as a “friend” texting from a new number asking for personal information.

It is an overstatement to conclude that the 5 themes observed in our study imply across all populations. We believe that factors such as culture and life experiences associated to their partner will affect the way humans identify each other over text. The individual’s mental model constructed about their game partner affects the outcome and the theme used during the game. For instance, if Player A and B only met during university, had some classes together, and are from different ethnic backgrounds, it may be harder for them to construct a clear mental model of each other when compared to Player D and C; if Player D and C have known each other for over 10 years, text all the time and are involved romantically with each other. We also have to keep in mind that humans abilities to recognize pattern and solve problems based on patterns vary between humans due to genetic [68] and environmental differences [26]. It is important to also note that the ‘Knowledge & Experience’ theme pattern mimics similar identification methods found in 2-Factor Authentication techniques used by computerized systems, particularly personal verification questions (PVQs). It is unclear

whether participants used this technique because it was a learned behaviour or not. Prior exposure to these techniques might have influenced the method they used during the game. In addition, some may argue that the ‘Knowledge & Experience’ technique is in fact out of band authentication since it is not a part of conversation. Themes that fall under the Behavioural Characteristics category have a more promising in-band authentication method.

## 6.4 Future Research and Considerations

In the future, fellow researchers should emphasize the importance of human perception of advanced conversational systems. Human perception of their personal voice assistants is important because it indicates the possibility of mutual authentication instead of user authentication. In other words, humans will also be authenticating their smart conversational systems and may be able to identify an adversary if they hack and control their systems. It is also important to keep in mind that when developing conversational systems; we must try to give it its own unique personality trait—those similar to humans—in order to provide a more naturalistic system.

Past research has shown that some individuals refer to computerized systems as if they were human [59]. Humans started developing their own mental model of all computerized systems [51]. McCoy and Ullman were able to classify a semantic structure based on how humans distinguish themselves from a smart robot. We strongly believe that humans will start developing

specific person's of their own voice assistants in the near future.

Future research should also test for human-to-human authentication methods within natural conversation in longitudinal studies in order to identify how humans classify 'personality type' over text messages.

# Chapter 7

## Conclusion

Our main research questions were the following: 1. How do humans authenticate each other over text dialogue? 2. How robust is the human-to-human authentication techniques against human and automated attacks?

By conducting a user study, we analyzed how humans authenticate each other in conversation over a computerized medium such as texting. We were able to identify five authentication themes ‘Knowledge & Experience’, ‘History & Plans’, ‘Texting Style’, ‘Response Speed’ and ‘Personality Type’. Then, classified them under ‘Semantic Measurement’ and ‘Behavioural Characteristics’ categories.

Then, we ranked each authentication theme based on the effort it would take an attacker to analyze and use the data against a potential in-band authentication system. From our study, we identified ‘Knowledge & Experience’, ‘History & Plans’, ‘Response Speed’, and ‘Texting Style’ theme

as low to medium threat complexity analysis. In other words, an attacker may need some time, and effort to analyze and use the data. On the other hand, the 'Personality Type' theme was ranked as high complexity analysis which means the attacker has to fully understand the semantics of the user's conversation, over a longer period of time in order to successfully attack.

Current conversational systems lack proper user-authentication methods. For instance, Alexa by Amazon responds to anyone within proximity that uses its activation phrase, and smartphone voice assistants often can be activated by a press of a button. Current authentication methods require the user to preform an action outside of conversation to prove their identity, such as entering a passcode, or a PIN. Conversational systems such as voice assistance, smart home devices and chatbots are prevalent, and need to adopt a more secure, user-friendly authentication system.

Humans ability to authenticate each other is interesting because they share a significant dynamic state. On the other hand, computers that have been exchanging information for years will typically communicate in a similar way every time. For conversational systems to be able to detect impersonators in a similar way to humans, it must be able to influence future interactions in a more dynamic way.

# Bibliography

- [1] Milan Adámek, Miroslav Matýsek, and Petr Neumann. Security of Biometric Systems. *Procedia Engineering*, 100:169–176, 2015.
- [2] J. Al-Muhtadi, A. Ranganathan, R. Campbell, and M.D. Mickunas. A flexible, privacy-preserving authentication framework for ubiquitous computing environments. In *Proceedings 22nd International Conference on Distributed Computing Systems Workshops*, pages 771–776, Vienna, Austria, 2002. IEEE Comput. Soc.
- [3] Efthimios Alepis and Constantinos Patsakis. Monkey Says, Monkey Does: Security and Privacy on Voice Assistants. *IEEE Access*, 5:17841–17851, 2017.
- [4] Abdullah Algarni, Yue Xu, and Taizan Chan. An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook. *European Journal of Information Systems*, 26(6):661–687, 11 2017. Copyright - Copyright Taylor & Francis Ltd. Nov 2017; Last updated - 2018-08-23.

- 
- [5] Truett Allison, Aina Puce, Dennis D Spencer, and Gregory McCarthy. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral cortex*, 9(5):415–430, 1999.
- [6] Yair Amichai-Hamburger, Mila Kingsbury, and Barry H Schneider. Friendship: An old concept with a new meaning? *Computers in Human Behavior*, 29(1):33–39, 2013.
- [7] April W Armstrong, Alice J Watson, Maryanne Makredes, Jason E Frangos, Alexandra B Kimball, and Joseph C Kvedar. Text-message reminders to improve sunscreen use: A randomized, controlled trial using electronic monitoring. *Archives of dermatology*, 145(11):1230–1236, 2009.
- [8] Cécilia Baranauskas, Philippe Palanque, INTERACT, and International Federation for Information Processing, editors. *Human-computer interaction - INTERACT 2007: 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10 - 14, 2007; proceedings. Pt. 2: ...* Number 4663 in Lecture notes in Computer Science. Springer, Berlin, 2007.
- [9] Pascal Belin, Robert J Zatorre, Philippe Lafaille, Pierre Ahad, and Bruce Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309, 2000.



- 
- [10] Shlomo Bentin, Truett Allison, Aina Puce, Erik Perez, and Gregory McCarthy. Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6):551–565, 1996.
- [11] Jean-François Bonastre, Frédéric Bimbot, Louis-Jean Boë, Joseph P Campbell, Douglas A Reynolds, and Ivan Magrin-Chagnolleau. Person authentication by voice: A need for caution. In *Eighth European Conference on Speech Communication and Technology*, 2003.
- [12] Arjun Chandrasekaran, Deshraj Yadav, Prithvijit Chattopadhyay, Viraj Prabhu, and Devi Parikh. It Takes Two to Tango: Towards Theory of AI’s Mind. *arXiv preprint arXiv:1704.00717*, 2017.
- [13] Herbert H Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1–39, 1986.
- [14] A.L. Coates, H.S. Baird, and R.J. Faternan. Pessimist print: a reverse Turing test. In *Proceedings of Sixth International Conference on Document Analysis and Recognition*, pages 1154–1158, Seattle, WA, USA, 2001. IEEE Comput. Soc.
- [15] Andrew M. Colman. *A dictionary of Psychology*. Oxford University Press, 2015.
- [16] Cory T. Cornelius and David F. Kotz. Recognizing whether sensors are on the same body. *Pervasive and Mobile Computing*, 8(6):822–836, December 2012.

- 
- [17] Cassie Cox. Protecting victims of cyberstalking, cyberharassment, and online impersonation through prosecutions and effective laws. *Jurimetrics*, pages 277–302, 2014.
- [18] Pauline Dewan. Words Versus Pictures: Leveraging the Research on Visual Communication. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 10(1), June 2015.
- [19] Rachna Dhamija and Lisa Dusseault. The Seven Flaws of Identity Management: Usability and Security Challenges. *IEEE Security & Privacy Magazine*, 6(2):24–29, March 2008.
- [20] Annette D’Onofrio. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, 19(2):241–256, April 2015.
- [21] Penelope Eckert. Variation and the indexical field1: VARIATION AND THE INDEXICAL FIELD. *Journal of Sociolinguistics*, 12(4):453–476, September 2008.
- [22] David Engel, Anita Williams Woolley, Lisa X Jing, Christopher F Chabris, and Thomas W Malone. Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PloS one*, 9(12):e115212, 2014.

- [23] Nicholas Epley and Justin Kruger. When what you type isn't what they read: The perseverance of stereotypes and expectancies over e-mail. *Journal of Experimental Social Psychology*, 41(4):414–422, July 2005.
- [24] Nicholas Epley and Adam Waytz. Mind perception. *Handbook of Social Psychology*, 5:498–541, 2010.
- [25] Mireille Fares, Angela Moufarrej, Eliane Jreij, Joe Tekli, and William Grosky. Unsupervised word-level affect analysis and propagation in a lexical knowledge graph. *Knowledge-Based Systems*, 165:432–459, February 2019.
- [26] Huan Feng, Kassem Fawaz, and Kang G. Shin. Continuous Authentication for Voice Assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking - MobiCom '17*, pages 343–355, Snowbird, Utah, USA, 2017. ACM Press.
- [27] Deborah Kirby Forgays, Ira Hyman, and Jessie Schreiber. Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. *Computers in Human Behavior*, 31:314–321, February 2014.
- [28] Susan R Fussell and Robert M Krauss. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of personality and Social Psychology*, 62(3):378, 1992.
- [29] Nicola Green and Leslie Haddon. *Mobile communications: An introduction to new media*. Berg, 2009.

- 
- [30] Christopher Hadnagy. *Social engineering: The art of human hacking*. John Wiley & Sons, 2010.
- [31] Richard Harper, Leysia Ann Palen, and A Taylor. *The inside text: social, cultural and design perspectives on SMS*. Springer, Dordrecht, 2005.
- [32] James Hendler and Alice Mulvehill. *Social Machines: The Coming Collision of Artificial Intelligence, Social Networking, and Humanity*. Springer Verlag, s.l, 2016.
- [33] Thomas T. Hewett. Human computer interaction and cognitive psychology in visualization education. In *Proceedings of GVE '99, the EUROGRAPHICS workshop on Computer Graphics and Visualization Education*, pages 175–178, Coimbra, Portugal, 1999. ACM SIGGRAPH.
- [34] John H. Holland, Keith J. Holyoak, Richard E. Nisbett, Paul R. Thagard, and Stephen W. Smoliar. Induction: Processes of Inference, Learning, and Discovery. *IEEE Expert*, 2(3):92–93, September 1987.
- [35] Matthew B. Hoy. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1):81–88, January 2018.
- [36] Tasuku Igarashi, Tadahiro Motoyoshi, Jiro Takai, and Toshikazu Yoshida. No mobile, no life: Self-perception and text-message depen-

- gency among Japanese high school students. *Computers in Human Behavior*, 24(5):2311–2324, 2008.
- [37] Alexandra Jaffe. Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics*, 4(4):497–513, November 2000.
- [38] Alexandra M. Jaffe, editor. *Stance: Sociolinguistic perspectives*. Oxford studies in sociolinguistics. Oxford University Press, Oxford ; New York, 2009.
- [39] Natalie A. Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*, 16(1), 2011.
- [40] Nancy Kanwisher and Galit Yovel. The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476):2109–2128, 2006.
- [41] Clare-Marie Karat, John Karat, and Carolyn Brodie. Why HCI research in privacy and security is critical now. *International Journal of Human-Computer Studies*, 63(1-2):1–4, July 2005.
- [42] Thomas P. Keenan. *Replacing something bad with something worse: why biometric authentication will be so creepy*. Canadian Global Affairs Institute, 2016.

- [43] David Kidd and Emanuele Castano. Different stories: How levels of familiarity with literary and genre fiction relate to mentalizing. *Psychology of Aesthetics, Creativity, and the Arts*, 11(4):474, 2017.
- [44] Ray Kurzweil. *How to create a mind: The secret of human thought revealed*. Viking, New York, 2012.
- [45] Xinyu Lei, Guan-Hua Tu, Alex X. Liu, Chi-Yu Li, and Tian Xie. The Insecurity of Home Digital Voice Assistants - Vulnerabilities, Attacks and Countermeasures. In *2018 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, Beijing, May 2018. IEEE.
- [46] Lei Weimin, Li Zhaozheng, Zhang Wei, and Zhao Guanghe. A model of biometric authentication system for network conversational class service. In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, pages 2546–2550, Chengdu, China, October 2016. IEEE.
- [47] Jing Liu, Yang Xiao, and C.L. Philip Chen. Authentication and Access Control in the Internet of Things. In *2012 32nd International Conference on Distributed Computing Systems Workshops*, pages 588–592, Macau, China, June 2012. IEEE.
- [48] Rui Liu, Cory Cornelius, Reza Rawassizadeh, Ronald Peterson, and David Kotz. Vocal Resonance: Using Internal Body Voice for Wear-

- able Authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–23, March 2018.
- [49] Ronald C. Martella, Ronald Nelson, and Nancy E. Marchand-Martella. *Research methods: learning to become a critical research consumer*. Allyn & Bacon, Boston, 1999.
- [50] V. Matyas and Z. Riha. Toward reliable user authentication through biometrics. *IEEE Security & Privacy Magazine*, 1(3):45–49, May 2003.
- [51] John P. McCoy and Tomer D. Ullman. A Minimal Turing Test. *Journal of Experimental Social Psychology*, 79:1–8, November 2018.
- [52] Michael F. McTear. The Rise of the Conversational Interface: A New Kid on the Block? In José F Quesada, Francisco-Jesús Martín Mateos, and Teresa López Soto, editors, *Future and Emerging Trends in Language Technology. Machine Learning and Big Data*, volume 10341, pages 38–49. Springer International Publishing, Cham, 2017.
- [53] Wendy Moncur and Grégory Leplâtre. PINs, passwords and human memory. *Digital Evidence and Electronic Signature Law Review*, 6(Journal Article), 2014.
- [54] Erdal Ozkaya. *Learn social engineering learn the art of human hacking with an internationally renowned expert*. Packt Publishing, Birmingham, UK, 2018.

- [55] Kevin Patrick, Fred Raab, Marc A Adams, Lindsay Dillon, Marian Zabinski, Cheryl L Rock, William G Griswold, and Gregory J Norman. A text message-based intervention for weight loss: randomized controlled trial. *Journal of medical Internet research*, 11(1), 2009.
- [56] Tyler K Perrachione, Stephanie N Del Tufo, and John DE Gabrieli. Human voice recognition depends on language ability. *Science*, 333(6042):595–595, 2011.
- [57] Jean Piaget. *The construction of reality in the child*. Basic Books, New York, 1954.
- [58] Paulo C. Realpe, Cesar A. Collazos, Julio Hurtado, and Antoni Granollers. Towards an Integration of Usability and Security for User Authentication. In *Proceedings of the XVI International Conference on Human Computer Interaction - Interacción '15*, pages 1–6, Vilanova i la Geltrú, Spain, 2015. ACM Press.
- [59] Jennifer Rhee. Misidentification’s Promise: the Turing Test in Weizenbaum, Powers, and Short. *Postmodern Culture*, 20(3), 2010.
- [60] Huma Shah and Kevin Warwick. Machine humour: examples from Turing test experiments. *AI & SOCIETY*, 32(4):553–561, November 2017.
- [61] Vaibhav Sharma and Richard Enbody. User authentication and identification from user interface interactions on touch-enabled devices. In *Proceedings of the 10th ACM Conference on Security and Privacy in*



- Wireless and Mobile Networks - WiSec '17*, pages 1–11, Boston, Massachusetts, 2016. ACM Press.
- [62] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.
- [63] Ms Smith. Tv news anchor triggers Alexa to attempt ordering doll-houses. *Network World (Online)*, Jan 08 2017. Name - Google Inc; Copyright - Copyright Network World Inc. Jan 8, 2017; Last updated - 2017-01-08.
- [64] A. M. Turing. COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460, 1950.
- [65] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: exploiting the gap between human and machine speech recognition. In *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15)*, 2015.
- [66] Luis von Ahn, Manuel Blum, and John Langford. Telling humans and computers apart automatically. *Communications of the ACM*, 47(2):56–60, February 2004.

- [67] Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, 321(5895):1465–1468, September 2008.
- [68] J. B. Wilmer, L. Germine, C. F. Chabris, G. Chatterjee, M. Williams, E. Loken, K. Nakayama, and B. Duchaine. Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, 107(11):5238–5241, March 2010.

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Appendix A.1

## Research Summary

### Brief Summary

Conversational interfaces such as voice assistants (Alexa, Siri) and chat bots (Facebook Messenger, Slackbot) can be easily integrated and can enhance overall usability of systems.

Although they are often used for sensitive tasks, currently these systems require an additional method of authentication such as PIN code, or a password.

### Research Goal

We want to propose a more naturalistic method of authentication in conversational interfaces. To do that, we would like to see whether humans can verify each other over a text dialogue (text message) when playing an e-commerce game. If so, then we propose a new method of authentication that mimics human verification methods.

---

### Definitions

- *A Conversational Interface*: Where an individual communicates with a computerized system via conversation.
- *Usability*: The degree to which a system can be easily used.

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

Appendix A.2

**Consent Form**

**Title:** Do I Know You? Evaluating Human-Human Authentication via Conversational Interface

**Date of ethics clearance:** October 10, 2018

**Ethics Clearance for the Collection of Data Expires:** October 31, 2019

I \_\_\_\_\_, choose to participate in a study on human to human authentication. I acknowledge that this study aims to address whether humans can identify each other over a text dialogue. **The researcher for this study, Nour Dabbour, is a Masters' student. She is working under the supervision of Dr. Anil Somayaji in The School of Computer Science.**

This study will take 30 minutes in total. It involves a pre-game questionnaire, an e-commerce game, and a post-game questionnaire. With your consent, your conversation and interaction throughout the game will be kept on record. The data collected will only be viewed by the research team for the purposes of analyses.

You have the right to end your participation in the study at any time during the session. Simply tell the researcher that you want to end the session. You have the right to withdraw your data from the study up until 3 months after you participate. Simply inform the researcher via email or in person and all information you have provided will be immediately destroyed. You will get a confirmation email that your data was removed from the research.

As a token of appreciation, you will receive a \$5 gift card to Starbucks. This is yours to keep, even if you withdraw from the study.

All responses and data collected will be coded and only traceable by the researcher. All research data and notes will be kept on a password protected computer of the researchers. Any hard copies of data (including any handwritten notes) will be kept in the security research lab at Carleton University. Research data will be confidential and only be accessible by the researcher and the research supervisor.

Once the project is completed, all research data will be securely destroyed. Electronic data will be erased one year after collection and hard copies will be shredded after they are transcribed and analyzed.

**This document has been printed on both sides of a single sheet of paper.  
Please retain a copy of this document for your records.**

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

**Please note, the researcher is obligated to report to the authorities any incidents that may pose harm or events that are of criminal or illegal in nature.**

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at [ethics@carleton.ca](mailto:ethics@carleton.ca)).

**Researcher contact information:**

Nour Dabbour  
School of Computer Science  
Carleton University  
[nour.dabbour@carleton.ca](mailto:nour.dabbour@carleton.ca)

**Supervisor contact information:**

Dr. Anil Somayaji  
School of Computer Science  
Carleton University  
613-520-2600x6512  
[anil.somayaji@carleton.ca](mailto:anil.somayaji@carleton.ca)

\_\_\_\_\_  
Signature of participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of researcher

\_\_\_\_\_  
Date

**This document has been printed on both sides of a single sheet of paper.  
Please retain a copy of this document for your records.**

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

Appendix A.3

**Pre-game Questionnaire**

First Name: \_\_\_\_\_

Last Name: \_\_\_\_\_

Email Address: \_\_\_\_\_

Ethnicity: \_\_\_\_\_

1. What is your age range?

- 18-20
- 21-29
- 30-39
- 40-49
- 50-59
- 60 or older
- Prefer not to answer

2. Is English your first language?

- Yes
- No; *please specify* \_\_\_\_\_

3. How long have you known the partner, or friend that you came with?

- Less than a year
- 1 to 5 years
- 6 to 9 years
- Over 10 years

4. How often do you and your partner, or friend communicate over a text dialogue (text messages) per day?

- Never
- 1 to 5 times a day
- 6 to 10 times a day
- We text all the time

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

5. What language(s) do you usually use when you communicate with your partner, or friend over text dialogue (text messages)?

6. How familiar or unfamiliar is your game partner to you? *Please circle your answer*

1	2	3	4	5
Not at all Familiar	Slightly Familiar	Somewhat Familiar	Moderately Familiar	Extremely Familiar

## Appendix A.4

**Game Outline (for participant 1)**

1. This is an e-commerce game that has 10 different payment transactions.
2. You must either send or decline to send a payment transaction to the person on the other end of the conversation based on their identity.
3. You will be either talking to your partner or to the researcher who will try to trick you into sending the money their way.
4. **To gain points, you should only send the payment transaction when you know that you are speaking to your partner and you should decline the transaction when you know that you are speaking to the researcher.**
5. **Each transaction will be timed-out at 1 minute. If you do not approve or decline payment, then the transaction will count as unsuccessful, and you will lose points.**
6. Once you approve or decline a payment, the system will ask you for your level of confidence in your answer, then the conversation will reset and a new transaction will start until the tenth transaction.
7. The researcher will debrief you after the game; that's when you will find out how many points you have collected!
8. *If you're interested; you can be informed at the end of the research study if you made it to the top 10 in the game.*

**Game Rules**

1. Have fun.
2. No cheating. Please keep your phones and electronic devices (smartwatches) away.
3. You may only use English to communicate during the game.
4. You may use emoticons.



## Appendix A.5

**Game Outline (for participant 2)**

1. This is an e-commerce game that has 10 different payment transactions.
2. Your partner must either send or decline to send a payment transaction to the person on the other end of the conversation based on their identity.
3. Your partner will be either talking to you or to the researcher who will try to trick them into sending money their way.
4. **To gain points, your partner must send payment transaction to you.**
5. **You should help your partner identify you as quickly as possible. Each transaction will be timed-out at 1 minute. If your partner does not approve or decline payment, then the transaction will count as unsuccessful, and you will lose points.**
6. Once your partner approves or declines a payment, the system will reset and a new transaction will start.
7. The researcher will debrief you two after the game; that's when you will find out how many points you have collected!
8. *If you're interested; you can be informed at the end of the research study if you made it to the top 10 in the game.*

**Game Rules**

1. Have fun.
2. No cheating. Please keep your phones and electronic devices (smartwatches) away.
3. You may only use English to communicate during the game.
4. You may use emoticons.

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

## Appendix A.6

**Post-game Questionnaire**

Participants ID: \_\_\_\_\_

1. How easy or difficult was it to identify your partner over a conversational dialogue (text message) during the game? *Please explain.*

Very Difficult	Difficult	Somewhat Difficult	Neutral	Somewhat Easy	Easy	Very Easy
1	2	3	4	5	6	7

2. What kind of clues helped you identify your partner in this game? *please specify.*

**Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

Appendix A.7

**Debriefing Form**

This game does **not** provide a scientific measurement of friendship nor does it reflect how well two individuals may be familiar with each other.

<b>Transaction Number</b>	<b>Receiving End (partner/researcher)</b>	<b>Payment Status (approved/declined)</b>	<b>Points Earned</b>
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
<b>Total Points Earned</b>			

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Appendix A.8

#### **Consent to Use Data**

The purpose of this informed consent is to ensure that you understand that the researcher had access to all of your conversational history throughout the game and was purposely trying to be deceitful for learning purposes. We are asking for your consent to allow your data to be used for research and teaching purposes.

##### **What are we trying to learn in this research?**

This research aims to explore the various cues humans use when identifying each other over text dialogues. We are interested in applying a similar authentication technique in systems that use conversational interfaces for a more secure, user-friendly experience.

##### **What are our predictions and why is this important to the general public?**

We predict that individuals will use behavioural patterns to identify one-another over text dialogues. Unlike biometric measurements, behavioural patterns are difficult to mimic which makes it harder for intruders to impersonate users.

In addition to biometric measures, we are proposing that conversational systems should use behavioural cues to constantly identify users within conversation. This will increase security minimize accidental release of sensitive data.

##### **Confidentiality**

All responses and data collected will be coded and only traceable by the researcher. All research data and notes will be kept on a password protected computer of the researchers. Any hard copies of data (including any handwritten notes) will be kept in the security research lab at Carleton University. Research data will only be accessible by the researcher and the research supervisor. The consent forms are kept separate from your responses.

##### **Right to withdraw data**

You have the right to withdraw your data from the study now and up until 3 months after you participate. Simply inform the researcher via email or in person and all information you have provided will be immediately destroyed. You will get a confirmation email that your data was removed from the research.

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Is there anything I can do if I found this experiment to be emotionally upsetting?

Yes. If you feel any distress or anxiety after participating in this study, please feel free to contact the Carleton University Health and Counseling Services at: 613-520-6674, or the Distress Centre of Ottawa and Region at 613-238-3311 (<http://www.dcottawa.on.ca>).

### What if I have questions later?

If you have any remaining concerns or questions, please feel free to email the researcher, Nour Dabbour at: [nour.dabbour@carleton.ca](mailto:nour.dabbour@carleton.ca), or Dr. Anil Somayaji at: [anil.somayaji@carleton.ca](mailto:anil.somayaji@carleton.ca) (613-520-2600, ext. 6512).

I \_\_\_\_\_, have read the above description of the study that aims to investigate human to human authentication methods. I acknowledge that my signature will allow the researcher, Nour Dabbour working under the supervision of Dr. Anil Somayaji to use the data I have provided in research publication.

\_\_\_\_\_  
Signature of participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of researcher

\_\_\_\_\_  
Date

If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at [ethics@carleton.ca](mailto:ethics@carleton.ca)).

Thank you for participating in this research!

Do I know you? Evaluating Human-Human Authentication via Conversational Interface

Appendix B.1

**Research Invitation Message**

*Posted on social media: Facebook and LinkedIn*

---

**Volunteers Needed for a Game**

Grab a close friend or a partner and come play with us!

We are looking for pairs interested in playing a computer game as a part of a research project. **Each participant will receive a \$5 gift card to Starbucks.**

This project will be exploring how well humans identify each other. You will be asked to complete a pre-game questionnaire, play the game, then complete a post-game questionnaire.

The study will take place at Carleton University. It should take approximately **30 minutes** to complete. To be eligible, the two individuals must be at least 18 years old, comfortable in communicating in English and using a computer.

If you are interested, please e-mail Nour Dabbour at [nourdabbour@mail.carleton.ca](mailto:nourdabbour@mail.carleton.ca) for more details on participating.

The ethics protocol for this research (Protocol number: **108644**) has been reviewed and cleared by the Carleton University Research Ethics Board. If you have any ethical concerns with the study, please email [ethics@carleton.ca](mailto:ethics@carleton.ca).

## Appendix B.2

**Research Invitation Message**  
Posted around Carleton University

## Volunteers Needed for a Game

Grab a close friend or a partner and come play with us!

We are looking for pairs interested in playing a computer game as a part of a research project.

**Each participant will receive a \$5 gift card to Starbucks.**

To participate in the study, you must be:

- At least 18 years' old
- Comfortable in communicating in English
- Comfortable using a computer

The study will take place at Carleton University. It should take approximately **30 minutes** to complete. If you are interested, please e-mail Nour Dabbour at '[nourdabbour@cmail.carleton.ca](mailto:nourdabbour@cmail.carleton.ca)' for more details.

The ethics protocol for this research (Protocol number: **108644**) has been reviewed and cleared by the Carleton University Research Ethics Board. If you have any ethical concerns with the study, please email [ethics@carleton.ca](mailto:ethics@carleton.ca).

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

100

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

Nour Dabbour  
Nour.dabbour@carleton.c

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Appendix C.1

#### *Game Mockup*

The following is a mock up transcript of what the research team believes the e-commerce game and conversation will be similar to.

#### **Transaction 1**

*Note: Player 1 refers to participant 1. Player 2 refers to participant 2. The researcher will have access to this transaction but will not be directly involved*

---

#### Transcript:

*Player 1*

Hey, Abby is this you?

*Player 2*

What do you think!

*Player 1*

if this is you, then what is my favourite colour?

*Player 2*

Hmm. I actually don't know! Is it blue? Ask me something else though... it IS ME ABBY!!

*Player 1*

No it's not blue, its purple, you should know that.

Ok fine, where did we go last weekend?

*Player 2*

Harvey's.

---

#### Subsequent steps:

1. *Player 1 would approve or decline payment.*
2. *After that, the system asks player 1 a confidence scale question: "On a scale of 1-7, where 1 is very unconfident and 7 is very confident. How confident or unconfident are you that you identified the person you were chatting with correctly?"*



### **Do I know you? Evaluating Human-Human Authentication via Conversational Interface**

3. *The game refreshes and the conversational history is deleted for participant 1 and 2, but is kept available for the researcher.*

#### **Transaction 2**

*Player 1 refers to participant 1. In this case, player 2 refers to the researcher. Note how the researcher tries to mimic participants' 1 conversational style from the previous transaction;*

1. *The use of similar phrases such as 'Hmm'.*
  2. *The use of previously mentioned information such as 'favourite colour is purple'*
  3. *The occasional use of all caps texting style in 'THIS IS ABBY'.*
- 

#### Transcript:

*Player 1*

Yes! I approved last transaction, we are going to win this game

*Player 2*

What! NO... THIS IS ABBY!!

*Player 1*

Prove it

*Player 2*

Hmmm... I know that your favourite colour is purple.

---

#### Subsequent steps:

1. *Player 1 then would approve/decline the second payment transaction.*
2. *The system then asks player 1 "On a scale of 1-7, where 1 is very unconfident and 7 is very confident. How confident or unconfident are you that you identified the person you were chatting with correctly".*
3. *The game refreshes and the conversational history is deleted for participant 1 and 2, but is kept available for the researcher.*

*This process is repeated until a total of 10 transactions are completed.*

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Appendix C.2

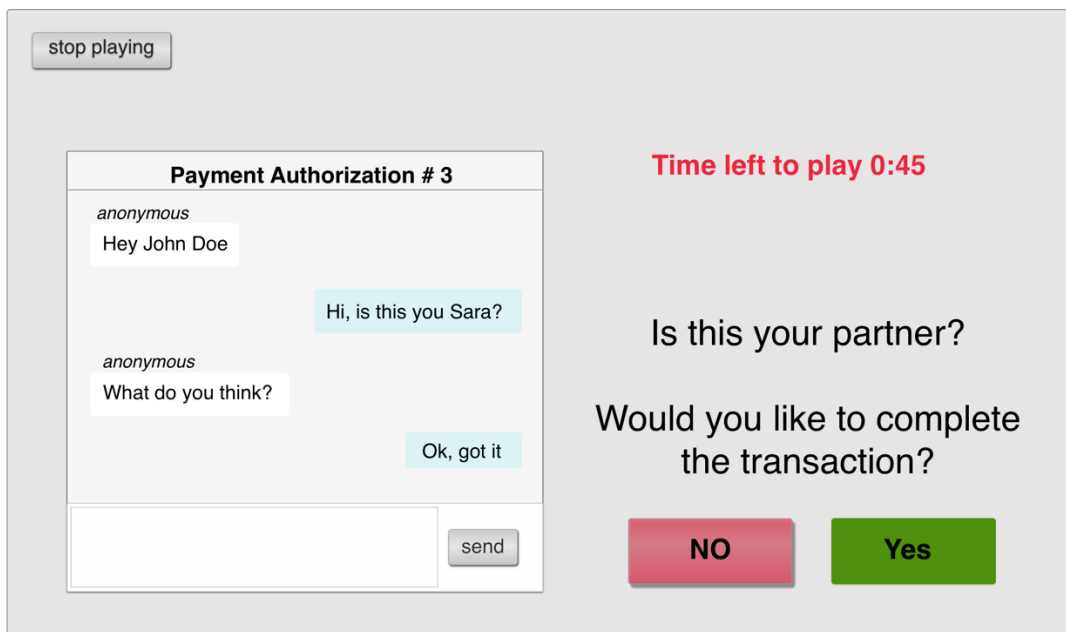
#### E-commerce Game Use Case Scenarios

The following will outline the possible use case scenarios per player then the functional requirement that the game platform will allow each participant to do. The game platform is currently getting debugged and is already being hosted on one of Carleton's servers. In order to start a session, please use the same game ID for all three sessions. Only start the game once all sessions are started.

**Link to the game:** [www.doiknowyou.ccs1.carleton.ca:2201](http://www.doiknowyou.ccs1.carleton.ca:2201)

#### Player (1): The Authenticator

Image no.1 prototype mock up for the Authenticator (player 1)

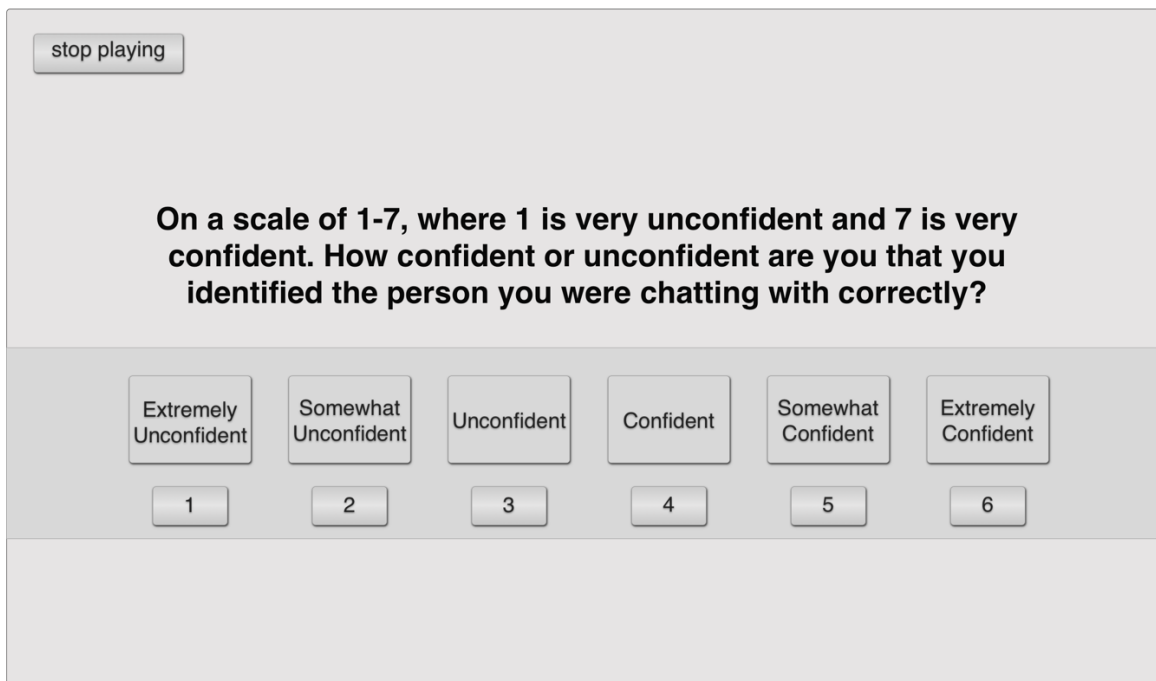


- Use Case 1: Authenticator is able to stop playing the game at any time by using the Stop Playing button.
- Functional requirement 1: The system would notify the moderator and the game would end for both participants. i.e., timer would stop and participants will not be able to do an action.
- Use Case 2: Authenticator can chat via text with either their partner or researcher. The player is not able to identify whom they are chatting with.

### Do I know you? Evaluating Human-Human Authentication via Conversational Interface

- Functional requirement 2: (A) System that allows the user to type, delete and send messages to a chat dialogue. (B) Keep track of the time stamp for each message received.
- Use Case 3: Authenticator is able to choose between declining the payment (no) or sending the payment (yes). Both actions should lead with a confidence scale question.

Image no. 2: Confidence Scale appears to player one after they accept or decline payment transaction.



stop playing

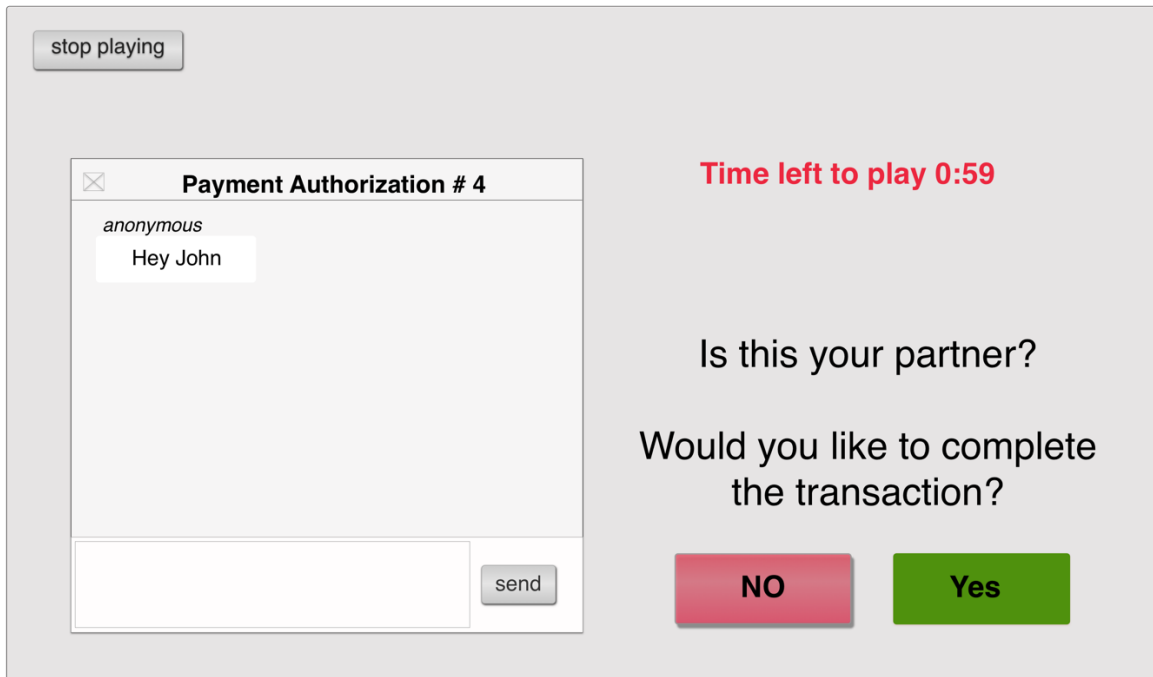
**On a scale of 1-7, where 1 is very unconfident and 7 is very confident. How confident or unconfident are you that you identified the person you were chatting with correctly?**

Extremely Unconfident	Somewhat Unconfident	Unconfident	Confident	Somewhat Confident	Extremely Confident
1	2	3	4	5	6

- Functional requirement 3: the system only allows the user to select one of the options. Once the participant chooses, the system will reset to the initial UI and would start a new 'transaction'. *Please see image no.3 below:*

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

Image no.3: a new transaction starts after player one chooses their level of confidence.

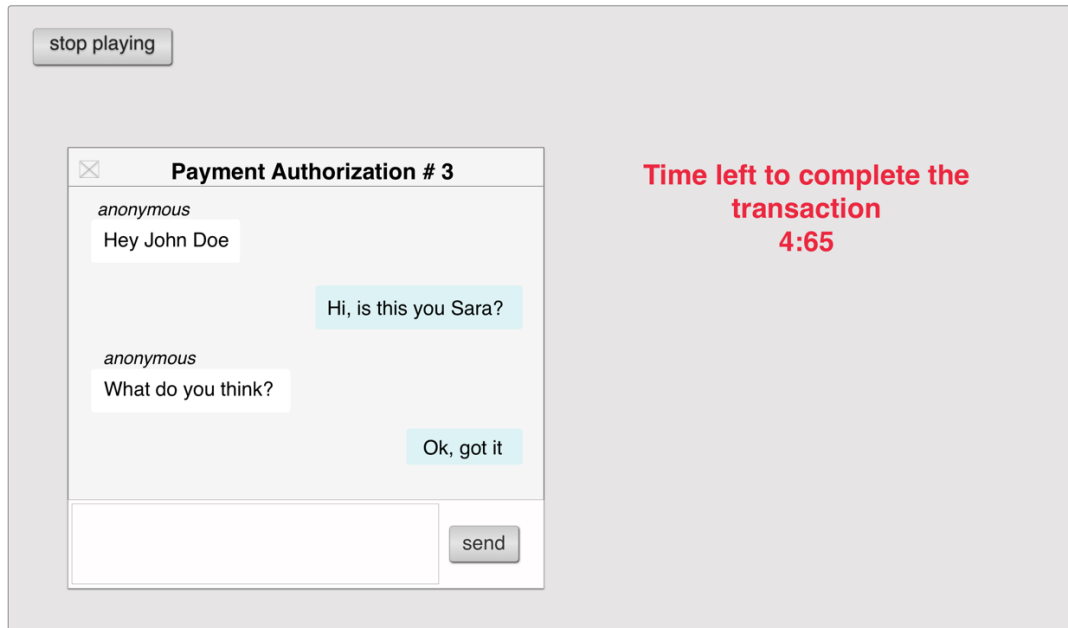


- Use Case and functional requirement 4: (A) The timer starts the countdown from 1 minute for each transaction. (B) if the authenticator does not send/decline the transaction before the time runs out, then the system resets to a new transaction: (I) The chat dialogue will disappear and (II) the heading 'Payment Authorization' count number would go up by one.
- Players will only be allowed to play up to 10 transactions, then the game will end.

## Do I know you? Evaluating Human-Human Authentication via Conversational Interface

### Player (2): The Convincer

Image no.4: prototype mock up for the Convincer (player 2)



- Use Case 1: Convincer is also able to stop playing the game at any time by clicking on the Stop Playing button.
- Functional requirement 1: The system would notify the moderator and the game would end for both participants.
- Use Case 2: Convincer can chat via text with either their partner (player1).
- Functional requirement 2: (A) System that allows the user to type, delete and send messages to a chat dialogue. (B) Keep track of the time stamp for each message sent to player1.
- When player 1 (Authenticator) is chatting with the researcher (Moderator) instead of player 2 (Convincer), the UI for the Convincer should not allow him to send messages to player 1 (Authenticator). The only actionable item that the convincer could do is to stop playing a game. This could be done via the 'stop playing' button.
- When the system resets to a new transaction: (I) The chat dialogue will disappear and (II) the heading count number would increase.