Towards In-Band Non-Cryptographic Authentication

Nour Dabbour
Carleton University
Ottawa, Canada
nour.dabbour@outlook.com

Anil Somayaji
Carleton University
Ottawa, Canada
soma@scs.carleton.ca

ABSTRACT

Robust, secure authentication is essential in online interactions. Current best practice is to combine factors communicated using different channels; however, in some contexts multi-factor authentication may not be feasible or appropriate. Thus there is a need for authentication strategies that do not rely on classic multiple factors. While people normally rely upon multiple factors to authenticate each other, there is anecdotal evidence that such factors are not needed to authenticate close relationships, and that in fact they can recognize each other over an extremely low-bandwidth channel: texting.

In this work we examine whether people who know each other well can, in fact, authenticate each other while texting in an adversarial context. We present results from a "friend imitation" game that has many similarities to Turing's Imitation Game. Results from this user study indicate that people use a variety of syntactic and semantic techniques to authenticate each other when texting. While some of the observed techniques are not secure against adversaries with access to social media and other data sources, others leverage sophisticated mental models of the other person's expected behaviour that can quickly be used to detect impersonation attempts. We also explore to what extent these insights could inform mechanisms for in-band non-cryptographic authentication in computer-to-human, human-to-computer, and computer-to-computer communication contexts.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**.

KEYWORDS

conversational interfaces, non-cryptographic authentication, computer security

ACM Reference Format:
Nour Dabbour and Anil Somayaji. 2020. Towards In-Band Non-Cryptographic Authentication. In *New Security Paradigms Workshop 2020 (NSPW '20), October 26–29, 2020, Online, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442167.3442180

1 INTRODUCTION

Multi-factor authentication is the current best practice for secure authentication. Individual factors, on their own, are subject to compromise: passwords can be phished, tokens can be stolen, and biometrics can be faked. Multiple factors together improve security because now an attacker must compromise all of them to gain access. However, each additional factor increases deployment cost and decreases usability. As attacks improve, factors individually and collectively become more complex, making the situation worse.

Secure authentication methods normally require a combination of properties, each of which will generally not be fully achieved:

- Authentication input should not be "guessable," it should have sufficient entropy.
- Authentication input should be bound to a specific entity, i.e., no password sharing.
- Replay attacks should be infeasible. For example, ideally a copy of a fingerprint should not be a substitute for a real finger with that fingerprint.

Specific attack strategies target one or more of these properties. Phishing attacks destroy the binding between a user and her authentication secrets, rainbow tables [42] make guessing long passwords feasible, and biomimicry attacks such as those using fake fingers [40, 56] are effectively a combination of a binding break and a replay attack.

While the above requirements are essential for current human-to-computer authentication, human-to-human authentication may operate using different principles. When face to face, people use a variety of perceptual input modalities (senses) and analysis techniques to identify each other: we recognize faces, voices, smells, and even gait patterns. As shown in research on biometrics, these all can serve as relatively strong authenticators, especially when performed by humans who can trivially make sure the "liveness" requirement is satisfied. When communicating by text, however, the situation is rather different.

Consider the problem of texting impersonation. When a device or an account is compromised, an attacker can send messages that originate from a legitimate, often extremely trusted source. Even though all technical authentication methods have been bypassed, targets can sometimes detect the impersonation with minimal effort, as we have found from personal experience. We and people we know have experienced situations where we get messages from a social media contact that turn out to be fraudulent because our contact's account had been compromised. These "Facebook money scam"-like attacks [47] were a mere nuisance because it was "obvious" that something was wrong. The attackers had full access to our contacts' account, including past postings and chats—and this information was used to craft some of the messages they sent. All

NSPW '20, October 26–29, 2020, Online, USA

ACM ISBN 978-1-4503-8995-2/20/10...$15.00
https://doi.org/10.1145/3442167.3442180

three properties listed above arguably didn't hold. Nevertheless, the deceptions were transparent.

In the context of traditional authentication methods, unmasking an attacker so quickly is a remarkable feat. Note however that identity here is verified not through a password, cryptographic key, or biometric; instead, it is verified through the syntax and semantics of the communication itself. The authentication is in-band—it is integrated into the communication stream rather than being communicated through a separate channel. Unlike in-band signalling (such as the spoofable tones in older versions of the telephone network [37]), there is no straightforward way to separate the authentication messages from the rest—all message syntax and semantics are potentially part of the ongoing authentication process. If we could develop in-band authentication strategies for human-to-computer or even computer-to-computer authentication that are similar in strength to the ones used by people to recognize each other, they could complement existing authentication strategies, potentially increasing security and usability at the same time.

Before we can even begin to build such systems, we first need to better understand human-to-human authentication. As a first step, we conducted a user study designed to examine how people authenticate each other when texting in an adversarial context. In this study pairs of subjects who have prior close social relationships play a "friend imitation game" in which an attacker (the experimenter) pretends to be one of the subjects and the other must decide whether they are texting their friend (or relative or lover) or a stranger. Our study design has many similarities to the classic Turing Test [54]; however, it is a three-party game where one party is trying to impersonate the other, rather than having a computer pretend it is a human.

Our results suggest that people make use of a variety of syntactic and semantic cues to authenticate each other. Some of these are quite superficial, such as word choice and the timing of responses. Others, however, seem to be much deeper—they can assert that "my friend would never say that" when to an outside observer the exchange seems perfectly reasonable. Authentication performance on this task thus appears to be more related to the richness of shared models of behaviour between the parties than the complexity of the actual exchanged messages. Thus, a party who does not share the same behaviour model (which in fact may be specific to a relationship rather than an individual) is unlikely to be able to choose responses that would fit within that model. We believe this insight into person-to-person authentication suggests potential design strategies for human-to-computer in-band authentication and potentially even computer-to-human and computer-to-computer authentication systems.

The rest of this paper proceeds as follows. In Section 2 we examine what is known about how humans authenticate each other. Section 3 presents the rationale and design of the user study we performed on human-human authentication in text-based conversations. Section 4 analyzes the study results and identifies the authentication techniques we observed, and Section 5 analyzes the security of these techniques against automated and semi-automated adversaries. In Section 6 we speculate about the applicability of human-to-human authentication techniques to computer-to-computer authentication systems. Section 7 outlines related work such as conversational interfaces and the seminal Turing test. We discuss the implications of our results in Section 8. Section 9 concludes.

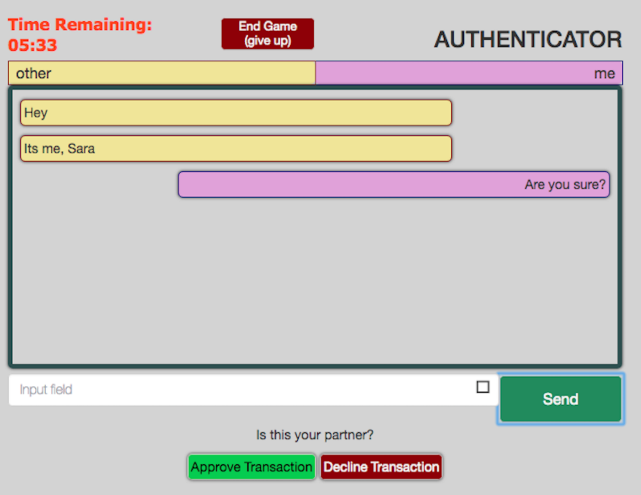
2 AUTHENTICATION BETWEEN HUMANS

While there exists an extensive literature in human-to-computer authentication, encompassing text password authentication [30], biometrics [29], and alternative schemes such as graphical passwords [9], the problem of human-to-human authentication has not been extensively studied directly. Instead, researchers have addressed how people identify each other as part of work in fields such as neuroscience, psychology, and sociolinguistics, where the issues of active attacks and deception are not central. Also, work on social engineering documents how to subvert social norms to compromise the security of individuals and organizations. We review these works below.

Neuroscientists have demonstrated that our brains have specialized areas devoted to facial [3, 8] and vocal [7] processing. These areas appear to help people distinguish between individuals by extracting a variety of specific features, as can be seen through psychological studies of facial recognition [51]. Actual recognition of people does not rely purely on these sort of specialized mechanisms, however, as people with specific deficits can still identify other people. In fact, people with prosopagnosia (inability to recognize familiar faces) can learn techniques to recognize faces [15] and people with phonagnosia (inability to recognize familiar voices) [21] seem to often not recognize they have a problem.

Research in psychology has shown that humans continuously analyze and interpret their world [17] enabling them to recognize patterns and point out alterations in patterns [36]. Humans build a generic interpretation of the world based on their experiences that allow them to make predictions regarding future interactions [31]. In the psychology literature, human pattern recognition is abstracted as working through schemas and mental models [13, 31]. Humans combine various schemas in unfamiliar situations to predict their outcome through mental models [24]. Mental models are temporary, flexible knowledge structures that occupy working memory embodying an internal conceptual and physical representation of the world [31]. 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. When children as young as two years old interact with their external environment, they start formulating mental representations ('assimilatory' schemas) that mature into schemas with experience and age as they move from understanding their subjective experience to understanding the world [45].

Conversation style also seems to be a very important part of establishing identity in social contexts. In sociolinguistics, a conversational style consists of a set of repeated patterns associated with social identities [27, 28]. The perceived style of a speaker will change the listener's expectations and influence their understanding of the language spoken [16] and people adjust their expectations based on their experiences and previous communications [23].



Hey

Its me, Sara

Are you sure?

Figure 1: The authenticator's screen during the game.

Code switching—switching between different languages or dialects depending upon the social context—helps people establish their social identity by signalling and reinforcing group membership [6]. Some linguistic researchers argue that individuals construct a style in 'platonic self' [28], meaning that each individual is unique in characteristics even within the same social environment. This is referred to as 'persona style', or 'persona' [17]. Personas may change over time depending on the social environment [17].

While human recognition works reliably for most people under normal conditions, people can be deceived using a variety of techniques. In particular, social engineering attacks [22] allow untrusted individuals to obtain privileged access, typically to gain physical access to a restricted area or to obtain confidential information through in-person interactions; however, social engineering can also be used in purely online attacks through email or social media [2]. Such attacks exploit people's trust in others whom they do not know personally but who appear to fit into a trusted social role (e.g., a CEO or a service technician). Social engineering is often used by red teams to demonstrate how technical security mechanisms can be bypassed by targeting the people of an organization [53]. Social engineering attacks can be extremely damaging, allowing attackers to steal millions of dollars [43]. Note that spear phishing, a kind of social engineering, often only requires a target to click on on a link [12].

To summarize, past work in neurology, psychology, and sociolinguistics give us evidence that people have sophisticated means of identifying each other, while work in social engineering shows that these techniques are far from perfect, particularly when interacting with people we don't know well. Our focus here, however, is on how two people who know each other well can detect when someone tries to impersonate one to the other. To address this, we conducted a user study, as described in the following section.

3 THE FRIEND IMITATION GAME

In order to study how humans authenticate each other while texting, we designed a study around a "friend imitation" game in which two participants who have an existing close relationship attempt to carry out a set of cooperative texting-based tasks while the experimenter attempts to subvert those tasks by pretending to be one of them.[1] Participants are thus given an incentive to identify each other reliably, allowing us to observe their approaches to the problem. Note that our goal was not to determine the relative frequency or utility of identification techniques, as such an effort would require a much larger study size with a more representative sample. Further, such quantification would not aid us in the goal of understanding whether these techniques could be used in human-to-computer and computer-to-computer contexts.

In our study we requested participants to come in with someone they were familiar with. Then, we asked them to play our 'friend-imitation game' in separate rooms. The game, implemented using a simple web application, requires a total of three players: the two participants and the adversary (the experimenter). One participant plays the role of the authenticator while the other plays the role of the convincer, with each participant pair choosing amongst themselves which role each would take. The authenticator has to converse with another unidentified player through text messages. If they believe they are communicating with the convincer, they should approve a fictional monetary transaction; if they believe they are instead communicating with the adversary, they should decline the transaction. Figure 1 shows what the authenticator's screen looks like while playing the game. The allocated game time given to each pair was 10 minutes. We asked the participants to complete as many transactions as possible within this timeframe.

Each research session had three parts: a pre-game questionnaire, the game, and the post-game questionnaire and interview. Participants were separated during the game so the only communication possible was via texting. We recorded participants' answers to the questionnaires, text messages exchanged during the game, and transcripts of the interviews for later analysis. Figure 2 shows a flowchart of a session.

In order to deceive the authenticator during the game, the adversary mimicked the convincer's conversational style and behaviour observed during the game. Specifically, the adversary tried to use similar phrases used by the convincer in previous transactions (example: cool beans vs. great), 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). Through having special privileges in the web application, the adversary had constant access to the transactions history that included conversation history and decline/approve payment responses. The adversary had access to the transaction details even when the authenticator was chatting with the convincer.

After the game, we asked the authenticator to rate how easy or difficult it was for them to identify their partner. We then asked them to explain their answer then proceeded with a discussion about the kind of cues they used throughout the game. All participants were given a small gift card at the end of the session; the same gift card was given no matter their performance.

Our study was reviewed and approved by the Carleton University Research Ethics Board (CUREB-B Clearance # 108644). There was

[1] Full details on the study are available in [14]; what follows here is a summary of the study methodology and results.

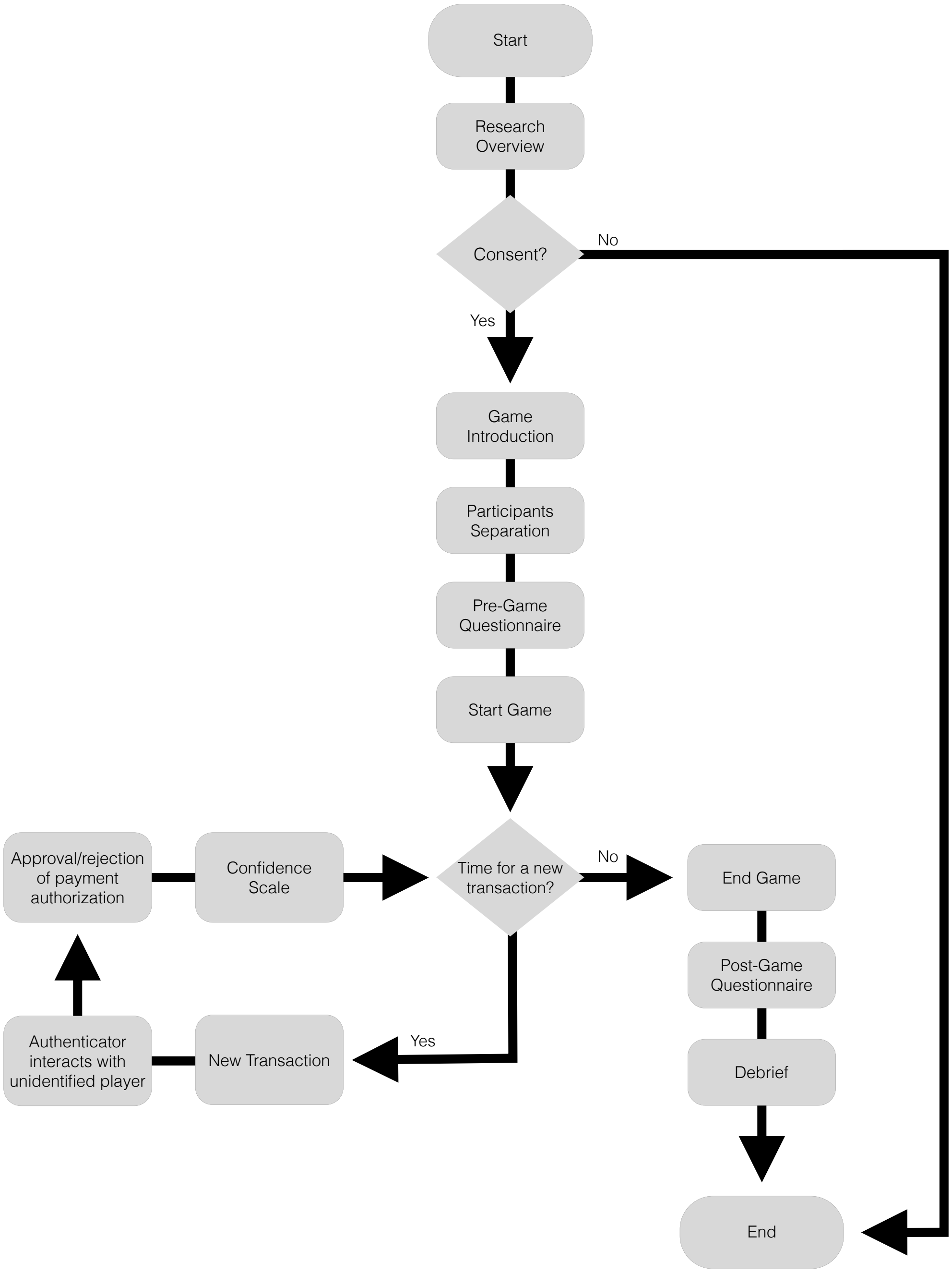


Figure 2: A step-by-step flowchart of one research session.

Table 1: Participants' background.

Background	# of participants	%
Age Range		
18-20 years old	12	27.27
21-29 years old	30	68.18
30-39 years old	1	2.27
40-49 years old	1	2.27
Time they knew each other		
Less than a year	7	15.91
1-5 years	28	63.64
6-9 years	5	11.36
over 10 years	4	9.09
Texting Frequency		
Never	7	15.91
1-5 times a day	19	43.18
6-10 times a day	4	9.09
We text all the time	14	31.82
Perceived Familiarity		
Slightly familiar	2	4.55
Somewhat familiar	1	2.27
Moderately familiar	18	40.91
Extremely familiar	23	52.27

Table 2: Frequency of authentication techniques.

Technique	# of interactions	# of pairs
Behavioural Characteristics		
Texting Style	31	15
Response Speed	6	4
Personality Type	4	3
Semantic Measurements		
Experience & Knowledge	92	24
History & Plans	63	19

a total of 22 pairs that participated in our research recruited from the university community.

4 RESULTS

Backgrounds of the participants is summarized in Table 1. On average, the experiment time was 30 minutes long. Participants completed an average of 8 transactions within the allocated 10 minutes game time with a 72.9% accuracy rate. Note that we specifically are not reporting more on how successful the participants were on the task, as performance was very context specific and is unlikely to generalize. Instead, we focus on reporting the authentication techniques that were observed, as that tells us how people identify each other through text-based communication in the presence of an adversary.

4.1 Classification of techniques

After analyzing the conversational scripts between participants, we identified five distinct techniques that players used to identify or prove it is themselves to their partner: 'Experience & Knowledge', 'History & Plans', 'Texting Style', 'Response Speed' and 'Personality Type'. We classified these techniques into two categories, Behavioural Characteristics and Semantic Measurements. Table 2 shows the frequency of use of each authentication technique, in terms of how many interactions or pairs used each. Overall, the 'Semantic Measurements' were used in 155 interactions and 'Behavioural Characteristics' were used in 41 interactions.

The 'Semantic Measurement' category classification was based on what participants said during conversation. It contains the top two most used techniques: Experience & Knowledge ranked first and History & Plans theme ranked second. The 'Experience & Knowledge' theme included facts collected as knowledge about individuals such as frequent habits, birthdays, nicknames, likes, and dislikes based on their experience together. 'History & Plans' technique revolved around previous and planned future occurrences. Conversations that focused on occurrences were either describing a major or a minor life event.

The 'Behavioural Characteristics' category includes three techniques: 'Texting Style', 'Response Speed', and 'Personality Type'. The 'Texting Style' technique was demonstrated by convincers and picked up by authenticators. Throughout the conversations, convincers used a specific sets of words, letter patterns, and emoticons. Convincers used shortcuts in text such as 'ur' vs 'your', 'zem' instead of 'them', and often extended some words like 'seeeeend 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.

4.2 Patterns of techniques

The majority of participants used a cross examination pattern that was relative to the techniques mentioned previously. More often, authenticators combined a series of clarifying questions subsequent to each other, even when the convincer answered 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 protect the identity of the participants.

```
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 also used a reply and elaborate pattern when answering 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. This kind of pattern was shown throughout conversation as a way to confirm one's identity.

4.3 Pairs with notable interactions

Each pair of participants had their own style of conversation they used throughout the game. Some, however, triggered a slight interest in us so we would like to discuss them in more detail.

4.3.1 Pair 55. Pair number 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 many as an average pair went through. They completed the highest number of transactions 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 one or two sentences per transaction that revolved either around their shared 'Experience & Knowledge' 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 X refers to the adversary:

```
--start of 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 new transaction with the Adversary--
X: do you want to come with me
Authenticator declined
```

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

```
--start of new transaction with the Convincer--
C: baby i think rotana will be good
Authenticator accepted

--start of new transaction with the Adversary--
X: baby i think rotana will be good
Authenticator accepted

--start of 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 new transaction with the Adversary--
X: 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 adversary used to trick the authenticator was to repeat facts based on their shared 'Experience & Knowledge'. 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.

4.3.2 Pair 57. The two individuals in pair 57 indicated that they knew each other for 6–9 years, were extremely familiar with each other, and 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 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 emoticons when they text. For example: they refer to money as 'guap' and affirm a statement by using 'facts'. Every transaction played with the adversary was unsuccessful. Although the adversary imitated their language style and conversational topics, the authenticator was not deceived.

```
--start of new transaction with the Adversary--
A: hello
A: [use of emoticon]
X: [re-sent the same sad emoticons used by convincer]
A: why you sad?
A: need some X?
A: tentacion?
X: i need guap
X: sister me out here
X: whats boppin
A: idk whats guud in the hood?
A: do you like movies?
X: are you getting my messages sis
A: nah nah you a fake
Authenticator declined
```

At a first glance, it may appear that the authenticator declined the adversary based on the question 'do you like movies?' however, the timestamp 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 outsiders (such as the experimenter). This mutual understanding was in part based in a shared sense of humour. We discuss humour and computer security more in Section 8.

5 SECURITY ANALYSIS

A truism in computer security is that attacks get better over time. We now evaluate how robust the observed authentication strategies would be given attackers with access to significantly more data and computational resources than we had in the user study. Specifically, we review different data sources an attacker would need access to in order to retrieve relevant data for the five themes identified under the 'Semantic Measurement' and the 'Behavioural Characteristics' categories. Next, we consider the complexity of analysis for each of those authentication themes to rank their robustness against possible attacks. We then discuss what sort of parties in 'Attacker Characteristics'.

The term 'data source' in the following text refers to the different locations that could contain specific data. For example, birthdays are available through public records and social media accounts. We explore the various data sources available per theme to fully understand the accessibility of the data. We analyze 'Complexity of Analysis' levels based 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. We ignore the difficulty of gaining access to these data sources; if they exist, we assume an attacker can access them. We divide analysis complexity into three levels: low, medium, and high:

- Low analysis complexity is assigned to a theme when an attacker can obtain the information using a database lookup and can use it with minimal transformation. For example, we would assign a low threat complexity analysis level to birthdays.
- Medium threat complexity is assigned when using the data would require statistical and other forms of analysis that can easily be automated. For example, favourite restaurants could be identified through a statistical analysis of location information correlated with restaurant locations.
- High threat complexity refers to data that require a substantial amount of manual effort to process and use. The analysis cannot currently be (completely) automated in a straightforward manner. We use this category for patterns that people can recognize intuitively but that we cannot currently describe algorithmically. We are not asserting that such patterns cannot be recognized computationally. In fact, much as with work on breaking CAPTCHAs, we anticipate study would lead to significant progress; however, in the limit they may be AI-complete (i.e., solving them is equivalent to creating a fully sentient artificial intelligence).

5.1 Semantic Measurement

Our classification of semantic measurements is summarized in Table 3.

The 'Experience & Knowledge' theme was used most frequently by participants and included birthdays, nicknames, habits, likes, and dislikes. We rank birthday dates and nicknames as low analysis complexity for a few reasons. First, they are easy to find on social media, public, and private records. Second, the time and effort it would take an attacker to find nicknames and birthday dates within a data source is relatively low.

In contrast, habits, likes, and dislikes are ranked as medium threat complexity in part because such information is not so easily accessible. While some can be found in social media and advertising profiles, their accuracy and scope is limited. Further, even when a person might be able to infer the appropriate patterns without much effort, automated analysis will likely require sophisticated machine learning methods.

The second semantic theme is the 'History & Plans' that consisted of past and future life events. These are ranked as low threat complexity because most life events such as anniversaries, moving to different cities, or accepting a masters program are typically recorded in social media, calendar invites, directory information, and public records, all data sources that are easily accessible to many attackers. An attacker does not require a substantial amount of time or effort to extract and use such 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, as they will not necessarily be entered into calendars or status updates using the regular patterns of major life events. Location information across users can help identify meetings, and related metadata can be used to infer the (high level) nature of the meetings. The data will be noisy so the analysis will not be trivial; however, we expect such analysis to be automatable using standard methods.

5.2 Behavioural Characteristics

Our classification of behavioural characteristics is summarized in Table 4.

Overall, behavioural characteristics are more resistant to attack than ones based on shared knowledge, although there is significant variability. 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 user's average time to reply to messages. The second behavioural characteristic theme is the 'Texting Style' theme that contained frequent use of specific vocabulary and emoticons as well as word patterns. This theme is ranked as medium threat complexity because of the time and effort it takes to analyze previous communication patterns, although such analysis could definitely be automated. The third 'personality' authentication theme is ranked as a high threat complexity analysis overall because while it may be feasible to determine a person's favourite topics through a superficial analysis of online activity, generating a model that allows that allows for someone's personality to be mimicked is much, much more difficult. This classification is more of a conjecture; however, it is based on

Table 3: Security analysis for semantic measurements.

Technique	Data Source	Analysis Complexity
Knowledge & Experience		
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
History & Plans		
Major Life Events	Phone conversation history Search History Social Media	Low
Minor Life Events	Phone conversation history Location services Personal Contact	Medium

the observation that computers today are very bad at imitating generic human behaviour in an unconstrained context. Imitating a specific human should be even more difficult, but how much harder is an open question.

Semi-automated attacks can take advantage of themes rated as low to medium, such as knowing a target's birthday, favourite restaurant, or social acquaintances, because they do not require an understanding of semantics in conversation. Personality traits, however, will require intensive work and a more thorough understanding of the person on the receiving end. In other words, an attacker requires the ability to predict performance and behaviour [19, 33], a feature only humans can currently perform by developing mental models about other individuals.

5.3 Attacker Analysis

Now that we have some idea of what kind of knowledge is required to attack conversational authentication, we now consider what sorts of attackers would be in a position to mount imitation attacks where one person pretends to be another via text communication. We consider the following: long-term partners & family members, community members, strangers, major data-gathering companies, and government intelligence organizations.

Long-term partners and family members are in some ways the best equipped to mount impersonation attacks. They know the user's history and patterns, having a model of behaviour built up over years. Their ability to carry out an attack will depend, however, on how well they know the other party. For example, one child pretending to be another when interacting with a parent would have a huge number of advantages, because they share history and context. This advantage is mitigated some, however, by reciprocal knowledge. Not only must an imitator act like the target, they must also *not* act like themselves.

As connections get further away or as people share less context, imitation becomes harder. A sibling trying to imitate another when interacting with a college friend will face a challenge because they don't share the same experiences—they weren't witness to their sibling's relationship with the friend. This challenge increases for community members such as work colleagues, friends of friends, and acquaintances. In the age of social media, community members can learn a lot about us; however, that view is heavily filtered and biased. Direct imitation within a community (where A, B, and C

Table 4: Security analysis for behavioural characteristics.

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

are all part of the same community, and A pretends to be B when interacting with C) may be possible if members of that community don't know each other well; imitation in the context of closer relationships, though, will be much harder.

Major data-gathering companies and government intelligence organizations have differing levels of access depending upon context; however, in practice both Google and the NSA can potentially have extremely deep knowledge of any individual. They can access not just public data but all manner of private data, often including intimate text, voice, and video communication. Using such deep repositories, in principle rather complete models of personal behaviour could be constructed, and those models could be used to imitate people.

Of course, fully exploiting such data is AI-complete, so today any sort of reasonable imitation attack would be at best semi-automated. If we can create visual and auditory deepfakes, perhaps we could do the same thing with conversations. However, also note that that we cannot make chatbots that can carry on a full conversation in any style. Hence, artificially imitating a person believably is not a near-term threat.

6 CONVERSATIONAL AUTHENTICATION FOR COMPUTERS

Conversational authentication, as performed between people, makes use of fundamentally different mechanisms than traditional computer-to-computer authentication mechanisms. We see them as differing in four key dimensions, as summarized in Table 5. We now discuss each of these dimensions to examine the feasibility of making computers follow a more conversational approach to authentication. We then hypothesize how they could be brought together to make conversational authentication systems for computers.

6.1 Technical Mechanism

The first dimension of difference is the technical mechanism: people engage in a conversation using natural language, while computers exchange messages conforming to precise cryptographic protocols. If we take this comparison at face value, the comparison is almost nonsensical: natural language is much too imprecise and ill-defined to be used for computer-to-computer communication in most contexts.

Table 5: Comparison of human and conventional computer authentication along multiple dimensions.

Authentication Dimension	Humans	Computers
technical mechanism	conversation	cryptographic protocols
identity basis	behaviour & knowledge	random secrets
trust temporality	continuously variable	constant/binary
task semantics	integrated	separated

The real difference between cryptographic protocols and natural language, however, is not a matter of precision but of expressiveness. By design, cryptographic protocols can express very limited semantics related to the purpose of the protocol. Natural language can express almost anything. The flexibility of natural language is a key part of conversational authentication, because it allows for a very large space of possible communications. Cryptographic protocols are utterly predictable except for keys, nonces, and data, and the data transmitted has no connection to the trust assessment. While conversations can be repetitive, their normal variability is key to making conversational authentication work.

6.2 Identity Basis

In human communication, when we take away physical characteristics such as appearance and the sound of voices, we identify each other by what we say and how we say it. In contrast, cryptographic authentication protocols are based on secrets. Proving that one party controls the private key corresponding to a public key is a very different basis for identity than discussing what happened during that picnic in seventh grade. Even having a video record of that picnic wouldn't necessarily be enough to fake such an interaction, because the memory of it is intertwined with emotional affect and experiences that occurred before and after the event. There is a "secret" in the human case, but it comes down to the total amount of information that determines who a person is and how they relate to others. This information is much harder to steal than a 4096-bit key.

Of course, computers have much, much more state than cryptographic keys. To have a more conversational-like identity basis, computers would need to use more of the full complexity of their state and mechanism to identify themselves. Like people, system identity would be a function of its past actions and interactions.

6.3 Trust Temporality

While continuous authentication is used in some contexts for human-to-computer authentication [20], with computer-to-computer authentication, trust is a fixed quantity: connections are authenticated or not with a given set of credentials, and their trust level is determined by those credentials. Cryptographic protocols take elaborate steps to make sure that third parties cannot tamper or eavesdrop on communications, and remote attestation can demonstrate that code and data are authentic to a remote party. These mechanisms, however, do not allow for truly dynamic trust assessments. We can lose our trust in a good friend in a single second even if we can identify them with 100% certainty—they merely need to do something dangerous like point a gun at us. With computers, trust is maintained so long as the protocol is followed, even if one host is actively and obviously trying to exploit a vulnerability in another.

To have computers determine trust in a way closer to people, we would need to integrate more holistic measures of trust that take into account identity and observed behaviour, and these measures would need to be assessed on an ongoing basis.

6.4 Task Semantics

Human-to-human conversations involve complex syntax and semantics, as we have discussed previously. References to shared history and cultural references are mixed with emotive constructs, biased word choices, and idiosyncratic syntax. Identity verification is a small aspect of a larger communication flow. Cryptographic authentication protocols, however, have almost no content other than the authentication process itself. Application semantics are, by design, completely separate. To bridge the task semantics gap, we would need to integrate authentication and application-level communication so that each influenced the other on an ongoing basis.

6.5 Imagining Conversational Authentication for Computers

To summarize the points made previously, to make computer-to-computer authentication fully conversational, we would need a protocol that allowed for a wide range of semantics in communication, where identity was based on shared history and behaviour characteristics, where trust levels were continuously adjusted based on what was being communicated, and where the authentication process was integrated with application behaviour. The question we now ask is how difficult would it be to change existing systems to have these properties?

To incorporate all of these characteristics, we could get into AI-complete territory, which would clearly make it infeasible. However, if we are willing to be a bit more modest in our goals, we can get implementable systems that address real threats. Let us start with imagining how adding one small bit of semantics analogous to minor life events, application history, could allow for a simple form of conversational authentication between an email client and server.

We could still use TLS to provide a base level of security. The conversational authentication layer would happen at the application layer using state shared between the client and server. One straightforward bit of state to track would be the disposition of recent messages, say over the past two weeks. For each message the client and server (communicating over a standard email protocol such as IMAP) would record what email messages were read and what was done with them: were they deleted, marked as unread, or moved to a folder/archived, in a manner similar to our previous work on an email intrusion detection system [39]. After each conventional authentication event (using passwords, certificates, or some form of two-factor authentication), the server and client could compare their message read histories, perhaps using some type of zero-knowledge protocol, over a protocol extension or a separate network connection. If the client doesn't have the right history, it will be marked as unauthorized and the client will have to engage in some sort of secondary authentication process (such as using a special service provided by the organization's IT department). Of course people may use multiple clients. But if so, the server just needs to keep a read message disposition history per client. Most users make use of a limited number of email clients and devices, especially for sensitive email accounts, so this duplication of history should not be a significant burden.

This same basic strategy could be used for many web applications by recording any application-level history on both the client and server. (Note that lots of state can be stored client side persistently nowadays.) The key attack this prevents is credential reuse: a stolen password or authentication token is no longer sufficient to access the account, as they also need to steal a record of recent history (or generate it by other means). This conversational authentication layer, however, is also robust to server impersonation and client state copying. An impersonating server will be detected unless it also has a copy of much of the server's state. Copied client state can be detected as it diverges from the original. An attacker copying the client data on Monday could use it on Monday evening, say after the real user has stopped using the account for the day. But the next day the target will notice the attack when their client is marked as unauthorized by the server (because it does not have a record of what the attacker did). To mount a successful attack would now require the client system to be completely compromised, a significantly higher bar than a credential reuse attack.

Ultimately conversational authentication between people involves them verifying that they share some common context that is only shared by those two people. Applying conversational authentication as a strategy to computer applications would involve the same thing, except the context being compared would be patterns observable by programs rather than those observable by people. The more complex and more context-sensitive the patterns, the harder they would be for an attacker to imitate. Wholesale copying of one party's state is a risk, but it is mitigated by the need to keep that state synchronized. To work over an extended duration, the attacker has to, for all intents and purposes, *become* the compromised party.

Note that simple versions of these mechanisms are already used on the web. Cookies are used to maintain session state, but they are not considered to be strong enough authenticators to do privileged operations, such as purchasing goods or services, unless they have been verified recently (say, by requesting the user's password to be re-entered). In practice these sorts of mechanisms can be very annoying to users as they interrupt the user's task at critical junctures. If we took a more conversational approach to the problem, allowing past browsing activity and other client state to be explicitly exposed in a way that could be checked by the server in a privacy-preserving way, user-level authentication could be grounded in a more robust computer-to-computer authentication and trust relationship context. As the example discussed above shows, the technical barriers to implementing such mechanisms, at least in their simpler forms, are quite modest. Implementing better mechanisms, however, requires us to rethink the classic cryptography-focused authentication paradigm to see the utility of incorporating conversational authentication patterns.

7 RELATED WORK

In Section 2 we reviewed past work potentially related to human-to-human authentication. Here we discuss systems and techniques that share some characteristics with our study design and other related work.

In 1950, Alan Turing posed the question "Can machines think like humans?" [54]. To test this question, he proposed the 'Imitation Game' which is commonly referred to as the 'Turing Test' [50]. The game measures whether a human judge is able to distinguish (through text-based conversation) whether they are conversing with a human or a computer. If the machine is able to convince the judge that it is human, then it passes the Turing Test. While the Turing Test is not a practical path towards general artificial intelligence, it has inspired something very important for the modern Internet, Reverse Turing Tests. Better known as CAPTCHAs [58], these automated tests that help online applications verify whether they are dealing with a human rather than a bot. Our 'Friend Imitation Game' structure closely resembles the 'Imitation Game' structure introduced by Alan Turing; however, instead of a computer trying to convince a judge that it is human, an adversary attempts to impersonate a specific person when interacting with another.

Stylometry is the study of authorial style in the context of creative works, particularly using statistical analyses. It has a long history of use to identify the authorship of written works of questionable provenance [25], and the increasing availability of computation has led to extensive work in the field, including on de-anonymizing programmers [11]. A key strategy for the defenders to succeed in our friend imitation game is for the authenticator to be able to reliably identify the convincer's writing style, which means that the authenticator has to engage in a type of real-time stylometry, except that the authenticator can also issue targeted prompts with which to get more information.

Many have studied how to apply natural language processing techniques and theory to computer security. Attalah [5] discusses several approaches, including using NLP to help users memorize passwords using humorous phrases, watermarking documents, and sanitizing information. Conversations, however, are not a focus in such work.

Many researchers have recognized that key parts of security protocols are not actually part of the protocol, but instead involve the people using the systems engaged in the protocol. For example, TLS is predicated people verifying that they are connected to the correct service by checking the metadata associated with the remote certificate. We can study the "ceremony" of TLS, rather than just the protocol, by considering the role of people in it [18]. Past work on authentication ceremonies have generally found how difficult it is for people to complete them successfully, even when usability is taken into account [57].

Phishing can be seen as a fundamental failure to identify a remote service by a user. While anti-phishing tools are an active area of research [46], current approaches largely focus on either training users to recognize the key signs [35] or automatically identifying and restricting access to malicious sites [48]. Note that in these examples, either users have to decide based on information that is provided up-front, or they are not involved in the assessment at all. In both cases, the process is not at all like a conversation.

There has been extensive past work in human-to-computer authentication in the area of behavioural biometrics. In this field, people demonstrate their identity to computers using behavioural characteristics including typing patterns [4, 44] and touch patterns [20]. These methods are similar to how people can identify each other using the timing, vocabulary, and syntax of text messages, as we discussed previously, especially when done implicitly as part of another task [10]. Our past work on narrative authentication [52], where a person identifies themselves to a computer through a series of interactions with a text adventure-like system, is conversation-like in appearance but can only be implemented currently under very severe constraints.

Conversational authentication can only work in the context of a relationship, as that relationship is what creates the shared context that is leveraged for authentication. Others have recognized the potential of social relationships to improve security. In particular, social authentication [49] leverages existing social relationships to authenticate users, normally as a recovery method for when other methods have failed. One form of social authentication requires users to contact previously designated trusted contacts and interact with them to get a code, generally through a voice call. Alternately, social media photos or posts are shown to the user for them to classify as belonging to their friends or not [34]. When calling a trusted contact, presumably individuals would authenticate each other using the sound of their voices and other semantic cues embedded in the conversation; in such a context a human to human conversation is used to authenticate a human to a computer.

This work was originally inspired by the weak authentication present in current conversational systems, whether the interaction is via text or speech. Existing conversational interfaces only allow access to sensitive information after out-of-band authentication (e.g., after unlocking a device with a PIN, fingerprint, or facial recognition). Authentication using a verbal PIN or similar method is awkward and subject to easy replay attacks. Note that voice interfaces can be particularly vulnerable to attack, as attackers can sometimes send inaudible voice commands that are still able to be recognized by the system [55] allowing attackers to even send text messages on the user's behalf [1].

While this work may suggest potential mechanisms for improving conversation interface security, we should mention that current voice assistants such as Amazon's Alexa, Apple's Siri, and Google Assistant barely qualify as conversational systems, as they require very rigid forms of interaction to work and full conversational systems should be much more flexible [41], and that flexibility would be required to implement the techniques people were observed to use in our study. Systems in development such as Google Duplex [38] are able to carry on much more natural conversations, albeit within very limited domains such as managing restaurant reservations. Human-like in-band authentication strategies may be more feasible with such systems. Fundamental problems of what to communicate and how to do so in a secure and usable way, though, remain to be solved, as we discuss in the next section.

8 DISCUSSION

Our study results suggest that people can identify each other when interacting via text using a mix of syntactic and semantic cues that range from trivial to imitate to ones almost inexplicable to a third party. These cues can allow many pairs of people to mutually authenticate each other reliably if they know each other well. While these results suggest that it may be possible to develop analogous techniques for in-band authentication between humans and computers, we must note that limitations in our original study may have misled us.

First, our study was not large or unbiased enough to give any indications about how reliable person-to-person recognition is; to get better accuracy we would need many more participants and more than one adversary. Even if we obtained more accurate results, our study was quite artificial in its construction. We suspect that a more ecologically valid design, one in which deception was included as part of everyday interactions, would likely lead to different communication patterns and would have a much lower defence success rate simply because in normal life most people aren't expecting impersonation attacks. Phishing attacks of all kinds rely on people not thinking too much about whether a message is legitimate. Similar inattention would greatly facilitate impersonation attacks.

Even if the results from our study are accurate they may not be so helpful in computational contexts, whether it is for human-to-computer, computer-to-human, or computer-to-computer authentication. As discussed in Section 6, people identify each other in part by verifying shared context, and we can envision computational systems that would do the same. What is not clear, however, is how much of the security of human-to-human conversational authentication is based on simply verifying shared context. For example, humour appeared to be a significant factor in the exchanges between at least some of our notable pairs. While there has been significant work on computational models of humour generally [32], humour remains extremely subjective and hard to quantify. Indeed, the study of humour has been identified as a possible path to understanding human intelligence [26]. We should expect, then, that we won't be able to get computers to be able to recognize people or each other in the way people do for a very long time.

Another challenge lies in the semantics of identification and trust. In computational authentication schemes trust normally starts out at zero and only increases as specific tests are passed, i.e., a correct password is entered or a biometric is matched. In contrast, social interactions outside of armed conflict start with a significant level of trust: we expect most people we encounter to behave in a civil fashion, to abstain from violence, and generally be reasonable to interact with. Positive recognition is one means by which this trust level can be increased or decreased, but it far from the only way. People also look for indicators in appearance and behaviour that can lead to distrust. These mechanisms can be superficial and often inaccurate—indeed, systemic racism can be seen as a pattern of behaviour in which skin colour is used as a signal for trust or distrust. Others, however, are much more reliable, such as recognizing hostile facial expressions and the brandishing of weapons. Conversation is a key means by which people negotiate trust in social contexts. Computational approaches to conversational authentication potentially lend themselves towards measuring more fluid trust relations; without more flexible and dynamic access control methods, however, it will be hard to take advantage of the potential benefits of conversational authentication.

If we step back, though, adding even simple mechanisms for remembering and sharing significant events—the easiest to spoof technique that humans use to identify themselves—to computer systems puts us in strange territory. Identifying important occurrences in our past requires us to reflect on our history and build a narrative about how we got to where we are. We tell stories about falling in love and losing people because these events change our lives *and we know that they changed us*. Could we make computers take note of significant software updates, changes in data repositories, and major configuration changes? No question. But if we make systems identify themselves to each other based on these sorts of "major events", our systems become a tiny bit self aware. In small amounts, the differences from conventional systems is hardly notable. However, when scaled up, perhaps we start taking steps towards machine consciousness, as what else is consciousness except the ability to reflect on ourselves, each other, and the wider world? If so, perhaps human consciousness is simply a sophisticated solution to the problem of establishing trust in a potentially hostile world.

9 CONCLUSION

In this paper we study how humans can identify and authenticate each other while texting in hostile contexts using a friend imitation game. Analysis of results from our user study showed that people employ a variety of strategies including verifying knowledge of shared life events, recognizing styles of communication, and confirming the other party's overall personality. Some of these strategies could in principle be subverted using online and social media data; others, however, would potentially require deep knowledge of a person that even close family members might not have.

While these results indicate that it might be possible to improve human-to-computer authentication using conversational authentication techniques, many usability and security issues would need to be addressed. A more promising direction may be to develop computer-to-computer authentication mechanisms based on past application-level activity. While we see no technical barriers to implementing simple mechanisms, more complex ones may raise some interesting philosophical questions.

Having said this, our purpose here is really to show that existing authentication approaches, particularly those based on cryptography, only explore a small part of the authentication design space. Human-to-human authentication shows that there are many strategies, and these strategies could in principle be applied in the context of virtually any communications protocol. We don't necessarily need to even change the communication protocols themselves—we just need to better model the data that we already have and develop strategies for comparing models efficient in bandwidth and secure in the presence of passive and active adversaries.

ACKNOWLEDGMENTS

The authors would like to acknowledge the valuable feedback of the anonymous reviewers, workshop participants, and our shepherds Mary Ellen Zurko and Elizabeth Stobert. This paper has evolved significantly through the NSPW process and, we believe, is much better as a result.

REFERENCES

[1] Efthimios Alepis and Constantinos Patsakis. 2017. Monkey Says, Monkey Does: Security and Privacy on Voice Assistants. *IEEE Access* 5 (2017), 17841–17851. https://doi.org/10.1109/ACCESS.2017.2747626

[2] Abdullah Algarni, Yue Xu, and Taizan Chan. 2017. An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook. *European Journal of Information Systems* 26, 6 (11 2017), 661–687. https://doi.org/10.1057/s41303-017-0057-y

[3] Truett Allison, Aina Puce, Dennis D Spencer, and Gregory McCarthy. 1999. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral cortex* 9, 5 (1999), 415–430. https://doi.org/10.1093/cercor/9.5.415

[4] Lívia CF Araújo, Luiz HR Sucupira, Miguel Gustavo Lizarraga, Lee Luan Ling, and Joao Baptista T Yabu-Uti. 2005. User authentication through typing biometrics features. *IEEE transactions on signal processing* 53, 2 (2005), 851–855. https://doi.org/10.1109/TSP.2004.839903

[5] Mikhail J Atallah, Craig J McDonough, Victor Raskin, and Sergei Nirenburg. 2001. Natural language processing for information assurance and security: an overview and implementations. In *NSPW '00: Proceedings of the 2000 workshop on New security paradigms*. 51–65. https://doi.org/10.1145/366173.366190

[6] Peter Auer. 2013. *Code-switching in conversation: Language, interaction and identity.* Routledge. ISBN 978-1134606733

[7] Pascal Belin, Robert J Zatorre, Philippe Lafaille, Pierre Ahad, and Bruce Pike. 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 6767 (2000), 309. https://doi.org/10.1038/35002078

[8] Shlomo Bentin, Truett Allison, Aina Puce, Erik Perez, and Gregory McCarthy. 1996. Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience* 8, 6 (1996), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

[9] Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot. 2012. Graphical Passwords: Learning from the First Twelve Years. *ACM Comput. Surv.* 44, 4, Article 19 (Sept. 2012), 41 pages. https://doi.org/10.1145/2333112.2333114

[10] Michael John Kendal Bingham. 2016. *Towards Effective Behavioural Biometrics for Mobile Devices.* Master's thesis. Carleton University. https://curve.carleton.ca/503e0b00-9df9-4420-ab69-2e4c3d89e786

[11] Aylin Caliskan-Islam, Richard Harang, Andrew Liu, Arvind Narayanan, Clare Voss, Fabian Yamaguchi, and Rachel Greenstadt. 2015. De-anonymizing Programmers via Code Stylometry. In *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, Washington, D.C., 255–270. ISBN 978-1939133113 https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/caliskan-islam

[12] D. D. Caputo, S. L. Pfleeger, J. D. Freeman, and M. E. Johnson. 2014. Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security Privacy* 12, 1 (2014), 28–38. https://doi.org/10.1109/MSP.2013.106

[13] Andrew M Colman. 2015. *A dictionary of psychology.* Oxford University Press. ISBN 978-0199657681

[14] Nour Dabbour. 2019. *Do I know you? Evaluating Human-to-Human Authentication via Conversational Interfaces.* Master's thesis. Carleton University. https://curve.carleton.ca/6309e2ee-55d2-4d3a-a97c-db44b726abf4

[15] Joseph M. DeGutis, Christopher Chiu, Mallory E. Grosso, and Sarah Cohan. 2014. Face processing improvements in prosopagnosia: successes and failures over the last 50 years. *Frontiers in Human Neuroscience* 8 (2014), 561. https://doi.org/10.3389/fnhum.2014.00561

[16] Annette D'Onofrio. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19, 2 (April 2015), 241–256. https://doi.org/10.1111/josl.12115

[17] Penelope Eckert. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12, 4 (Sept. 2008), 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

[18] Carl M Ellison. 2007. Ceremony Design and Analysis. *IACR Cryptol. ePrint Arch.* 2007 (2007), 399. https://eprint.iacr.org/2007/399.pdf

[19] David Engel, Anita Williams Woolley, Lisa X Jing, Christopher F Chabris, and Thomas W Malone. 2014. 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 (2014), e115212. https://doi.org/10.1371/journal.pone.0115212

[20] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2012. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE transactions on information forensics and security* 8, 1 (2012), 136–148. https://doi.org/10.1109/TIFS.2012.2225048

[21] Sascha Frühholz, Pascal Belin, Claudia Roswandowitz, Corrina Maguinness, and Katharina von Kriegstein. 2019. *Deficits in Voice-Identity ProcessingAcquired and Developmental Phonagnosia.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.39

[22] Christopher Hadnagy. 2010. *Social engineering: The art of human hacking*. John Wiley & Sons. ISBN 978-0470639535

[23] Richard Harper, Leysia Ann Palen, and A Taylor. 2005. *The inside text: social, cultural and design perspectives on SMS*. Springer, Dordrecht. ISBN 978-1402030604

[24] John H. Holland, Keith J. Holyoak, Richard E. Nisbett, Paul R. Thagard, and Stephen W. Smoliar. 1987. Induction: Processes of Inference, Learning, and Discovery. *IEEE Expert* 2, 3 (Sept. 1987), 92–93. https://doi.org/10.1109/MEX.1987.4307100

[25] David I. Holmes. 1998. The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing* 13, 3 (09 1998), 111–117. https://doi.org/10.1093/llc/13.3.111

[26] Matthew M Hurley, Daniel Clement Dennett, Reginald B Adams Jr, and Reginald B Adams. 2011. *Inside jokes: Using humor to reverse-engineer the mind*. MIT press. ISBN 978-0262518697

[27] Alexandra Jaffe. 2000. Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics* 4, 4 (Nov. 2000), 497–513. https://doi.org/10.1111/1467-9481.00127

[28] Alexandra M. Jaffe (Ed.). 2009. *Stance: sociolinguistic perspectives*. Oxford University Press, Oxford ; New York. ISBN 978-0195331646

[29] Anil K Jain, Patrick Flynn, and Arun A Ross. 2007. *Handbook of biometrics*. Springer Science & Business Media. ISBN 978-0387710419

[30] S. Jeyaraman and U. Topkara. 2005. Have the cake and eat it too—infusing usability into text-password based authentication systems. In *21st Annual Computer Security Applications Conference (ACSAC'05)*. https://doi.org/10.1109/CSAC.2005.28

[31] Natalie A. Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. 2011. Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society* 16, 1 (2011). https://doi.org/10.5751/ES-03802-160146

[32] Justine T Kao, Roger Levy, and Noah D Goodman. 2016. A computational model of linguistic humor in puns. *Cognitive science* 40, 5 (2016), 1270–1285. https://doi.org/10.1111/cogs.12269

[33] David Kidd and Emanuele Castano. 2017. Different stories: How levels of familiarity with literary and genre fiction relate to mentalizing. *Psychology of Aesthetics, Creativity, and the Arts* 11, 4 (2017), 474. https://doi.org/10.1037/aca0000069

[34] Hyoungshick Kim, John Tang, and Ross Anderson. 2012. Social authentication: harder than it looks. In *International Conference on Financial Cryptography and Data Security*. Springer, 1–15. https://doi.org/10.1007/978-3-642-32946-3_1

[35] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of Phish: A Real-World Evaluation of Anti-Phishing Training. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. Association for Computing Machinery, New York, NY, USA, Article 3, 12 pages. https://doi.org/10.1145/1572532.1572536

[36] Ray Kurzweil. 2012. *How to create a mind: the secret of human thought revealed*. Viking, New York. ISBN 978-0670025299

[37] P. Lapsley. 2013. Phreaking out ma bell. *IEEE Spectrum* 50, 2 (2013), 30–35. https://doi.org/10.1109/MSPEC.2013.6420138

[38] Yaniv Leviathan and Yossi Matias. 2018. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

[39] Yiru Li and Anil Somayaji. 2005. Securing email archives through user modeling. In *21st Annual Computer Security Applications Conference (ACSAC'05)*. https://doi.org/10.1109/CSAC.2005.50

[40] Tsutomu Matsumoto, Hiroyuki Matsumoto, Koji Yamada, and Satoshi Hoshino. 2002. Impact of artificial "gummy" fingers on fingerprint systems. In *Optical Security and Counterfeit Deterrence Techniques IV*, Rudolf L. van Renesse (Ed.), Vol. 4677. International Society for Optics and Photonics, SPIE, 275 – 289. https://doi.org/10.1117/12.462719

[41] Michael F. McTear. 2017. The Rise of the Conversational Interface: A New Kid on the Block? In *Future and Emerging Trends in Language Technology. Machine Learning and Big Data*, José F Quesada, Francisco-Jesús Martín Mateos, and Teresa López Soto (Eds.). Vol. 10341. Springer International Publishing, Cham, 38–49. https://doi.org/10.1007/978-3-319-69365-1_3

[42] Philippe Oechslin. 2003. Making a Faster Cryptanalytic Time-Memory Trade-Off. In *Advances in Cryptology—CRYPTO 2003*, Dan Boneh (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 617–630. https://doi.org/10.1007/978-3-540-45146-4_36

[43] Erdal Ozkaya. 2018. *Learn social engineering learn the art of human hacking with an internationally renowned expert*. Packt Publishing, Birmingham, UK. ISBN 978-1788838009

[44] Alen Peacock, Xian Ke, and Matthew Wilkerson. 2004. Typing patterns: A key to user identification. *IEEE Security & Privacy* 2, 5 (2004), 40–47. https://doi.org/10.1109/MSP.2004.89

[45] Jean Piaget. 1954. *The construction of reality in the child*. Basic Books, New York. https://doi.org/10.1037/11168-000

[46] Issa Qabajeh, Fadi Thabtah, and Francisco Chiclana. 2018. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review* 29 (2018), 44 – 55. https://doi.org/10.1016/j.cosrev.2018.05.003

[47] Linda Roeder. 2019. The Facebook Money Scam: What It Is And How To Protect Yourself From It. https://www.lifewire.com/i-need-money-facebook-scam-2654773

[48] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. 2019. Machine learning based phishing detection from URLs. *Expert Systems with Applications* 117 (2019), 345 – 357. https://doi.org/10.1016/j.eswa.2018.09.029

[49] Stuart Schechter, Serge Egelman, and Robert W Reeder. 2009. It's not what you know, but who you know: a social approach to last-resort authentication. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1983–1992. https://doi.org/10.1145/1518701.1519003

[50] Huma Shah and Kevin Warwick. 2017. Machine humour: examples from Turing test experiments. *AI & SOCIETY* 32, 4 (Nov. 2017), 553–561. https://doi.org/10.1007/s00146-016-0669-0

[51] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell. 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* 94, 11 (2006), 1948–1962. https://doi.org/10.1109/JPROC.2006.884093

[52] Anil Somayaji, David Mould, and Carson Brown. 2013. Towards narrative authentication: or, against boring authentication. In *NSPW '13: Proceedings of the 2013 New Security Paradigms Workshop*. 57–64. https://doi.org/10.1145/2535813.2535820

[53] Jeremiah Talamantes. 2014. *The Social Engineer's Playbook: A Practical Guide to Pretexting* (1st ed.). Hexcode Publishing. ISBN 978-0692306611

[54] A. M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236 (October 1950), 433–460. https://doi.org/10.1093/mind/LIX.236.433

[55] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. 2015. Cocaine Noodles: Exploiting the Gap between Human and Machine Speech Recognition. In *9th USENIX Workshop on Offensive Technologies (WOOT 15)*. USENIX Association, Washington, D.C. https://www.usenix.org/conference/woot15/workshop-program/presentation/vaidya

[56] Ton van der Putte and Jeroen Keuning. 2000. Biometrical Fingerprint Recognition: Don't get your Fingers Burned. In *Smart Card Research and Advanced Applications: IFIP TC8 / WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications September 20–22, 2000, Bristol, United Kingdom*. Springer US, Boston, MA, 289–303. https://doi.org/10.1007/978-0-387-35528-3_17

[57] Elham Vaziripour, Justin Wu, Mark O'Neill, Daniel Metro, Josh Cockrell, Timothy Moffett, Jordan Whitehead, Nick Bonner, Kent Seamons, and Daniel Zappala. 2018. Action needed! helping users find and complete the authentication ceremony in signal. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, 47–62. https://www.usenix.org/conference/soups2018/presentation/vaziripour

[58] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. 2003. CAPTCHA: Using hard AI problems for security. In *International conference on the theory and applications of cryptographic techniques*. Springer, 294–311. https://doi.org/10.1007/3-540-39200-9_18