Let's Get Cracking: Leveraging Gameplay from an Adversarial Perspective to Teach Password Security Concepts

Dylan Lins Brasiliense Drucker

MInf Project (Part 2) Report
Computer Science
School of Informatics
University of Edinburgh

2025

Abstract

Text-based passwords remain the most commonly used method of authentication on the internet. Despite their widespread use, a significant gap persists between the general public's understanding of password security and the realities posed by adversarial threats. This dissertation presents "Let's Get Cracking" (an updated version from Minf 1), a serious game designed to enhance users' understanding of password security by placing them in the role of an adversary attempting to crack passwords.

The new version of the game features an improved user interface, an enhanced tutorial to better guide players, new gameplay mechanics including upgrades and rainbow tables, and various quality-of-life updates to increase accessibility and player engagement. The game continues to simulate real-world password-cracking techniques, including brute-force attacks, dictionary attacks, and the exploitation of common password patterns. Rather than explicitly telling players what makes passwords weak, the game allows them to implicitly learn through gameplay, providing valuable insights into password security.

The study demonstrates the effectiveness of these improvements to the game through a comprehensive literature review, game design documentation of both versions of the game, and an empirical evaluation involving user questionnaires. The findings indicate that the updated version of the game not only maintains its engaging and enjoyable nature but also further improves players' ability to identify weaknesses in their own password practices and deepens their understanding of the factors involved in password security.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 987681
Date when approval was obtained: 2024-12-20
The participants' information sheet and consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Dylan Lins Brasiliense Drucker*)

Acknowledgements

I would like to thank all of the participants of the study for their help and enthusiasm. Thank you for volunteering your time. I would also like to thank all of the game's initial play-testers, who helped identify bugs and typos, and help me tweak the game's balance. Lastly, I would like to thank my supervisor, Borislav Ikomonov. I greatly appreciate the support and guidance you have provided over these last two years. I have truly enjoyed working on this dissertation and it is likely the project I am most proud of in my time at University. Thank you for everything, I couldn't have done it without you!

Table of Contents

Chapter 1

Introduction

Passwords play a fundamental role in digital security, acting as the primary line of defense in user authentication. Despite emerging alternatives, text-based passwords are still the dominant authentication method in computer and web systems (Kelley et al., 2012). Systems for cracking passwords have advanced, leading to the implementation of stricter password requirements. Despite the increasing use of password managers that generate complex passwords, a significant portion of users still opt to create their own (Pearman et al., 2019). Unfortunately, these user-generated passwords tend to be less secure, as they often prioritise memorability over strength.

Password data breaches are frequent and have potentially severe consequences, yet, multiple studies have identified a fundamental gap between the average person's understanding of password security and the practices that are exploited by the attacker model, leading to unsafe password practices (Ur et al., 2016) (Pearman et al., 2019). Password-strength meters, commonly using LUDS (lowercase, uppercase, digit, symbols) based on length and character diversity, are the prevalent method of measuring password strength. This system is considered somewhat ineffective as it can mislead users into creating passwords that, while considered strong by these criteria, are highly susceptible to being cracked (e.g. P@$$w0rd123) (Wheeler, 2016). It also fails to capture heuristics such as the use of common keyboard patterns, character substitutions, dictionary words, and predictable positioning of characters. The problem introduced by ineffective password strength estimators is exacerbated by the gap in knowledge regarding password storage practices and the effectiveness of brute-force attacks.

Recognising the limitations of traditional password strength metrics and the existing knowledge gap, this leads us to examine the use of serious games as a more effective method of teaching these concepts.

1.1 Limitations of Existing Serious Games

The use of serious games - games that do not have entertainment as their primary purpose - has been proven to be an effective teaching method. By immersing players in interactive and engaging environments, they facilitate the retention of information,

enabling players to acquire and apply new knowledge and skills in practical situations (Backlund and Hendrix, 2013).

While serious games aimed at teaching important password security practices exist, they often use an instructional approach, explicitly stating password characteristics to avoid, without explaining the underlying principles such as brute-force, hashing, and dictionary attacks. By instead focusing on improving the player's understanding of the attacker model, these games could empower players to come to these conclusions independently.

1.2 Proposal and Rationale for "Let's Get Cracking"

These insights led to the creation of the serious game detailed in this project, "Let's Get Cracking". Designed to place players in the role of an adversary, the game challenges them to crack as many passwords from a leaked database as possible within a set time limit. By adopting this unique adversarial perspective, players are encouraged to think creatively about exploiting vulnerabilities in password security to maximise their score. As a result, they are able to recognise the characteristics of weak passwords intuitively. This approach aims to teach the underlying mechanics behind password storage, expose players to the tools used by attackers, and ultimately help players intuitively understand the characteristics of weak passwords.

1.3 Aims and Organisation

This dissertation builds upon the successful previous version of "Let's Get Cracking," which was praised for its engaging and educational approach to password security. However, feedback indicated areas for improvement in gameplay and educational content. This study focuses on enhancing the game by incorporating user feedback and adding new features, including improved gameplay mechanics, an upgraded user interface, a more comprehensive tutorial, and a hint system (discussed in Chapter 3).

The aim of this dissertation is to evaluate how the improvements made to the game affect players' experience with the game, and its continued effectiveness in teaching password security from an adversarial perspective. This was evaluated with 22 participants using pregame and postgame questionnaires, mirroring the methodology of the previous study.

The dissertation is organised as follows: Chapter 2 reviews the existing literature on serious games and password security. Chapter 3 summarises the previous work done on this project. Chapter 4 details the design and implementation of the improved version of "Let's Get Cracking". Chapter 5 presents the methodology used in designing the questionnaire to evaluate the game, and reviews the results obtained from it. Finally, Chapter 6 concludes the dissertation by synthesising the results and proposing potential future extensions to further enhance the game based on the findings from the study.

Chapter 2

Background

This chapter provides background on the role of passwords in digital security and an overview of serious games. It introduces key terminology related to password storage, hashing, and the strategies used by attackers. It also explores the knowledge gap between public perception and the actual principles of secure password practices, as well as how to educate the public on creating strong, memorable passwords. The concept of serious games is introduced, and examples of existing serious games about password security are analysed and critiqued. These games usually bring attention to a very limited number of weak password characteristics and do not explain the underlying reasons for their vulnerabilities, particularly in relation to the attack model. These insights were used in the development of ”Let’s Get Cracking” with the aim of reducing the disparity in public perception and guiding their future behaviour in ways that previous work has not fully achieved.

2.1 Passwords in Digital Security

2.1.1 The Weakness of Passwords

As the internet has become a more integral part of people’s lives, passwords now protect increasingly sensitive data. With advancements in computational power and more efficient password-cracking techniques, service providers have grown more concerned about the security of online accounts. In an attempt to guide users to using stronger passwords, service providers enforce password composition policies, usually requiring a diversity of character types and a minimum length (Shay et al. (2016)).

Although this does increase password complexity, it has been shown that these requirements result in users creating passwords with unsafe practices, in an attempt to make them more memorable. These include, but are not limited to:

- Placing numbers, special characters, and uppercase letters in predictable locations.
- Using only a small fraction of possible symbols.
- Replacing letters with similar-looking numbers or symbols (e.g., `P@ssw0rd`)

- Using common dictionary words
- Using numeric or keyboard patterns

Attackers can use these patterns in making decisions on how to target passwords. The problem is exacerbated by the frequent reuse of passwords across different accounts.

Not all secure passwords have to have low memorability. Research by Shay et al. (2016) showcases techniques for creating memorable passwords resilient to attacks. These techniques leverage the increasing computational complexity of cracking longer passwords, encouraging passwords of lengths 12-16, created by combining words with non-alpha characters.

2.1.2 Password Storage

As passwords remain the most commonly used method of user authentication, organisations must carefully consider how they store them within their databases. Storing passwords in plaintext, as exemplified by notorious leaks such as the RockYou data breach discussed in section 2.1.4, is a critical vulnerability that should be avoided. Instead, the industry-standard employs hashing.

2.1.2.1 Hashing

Hashing involves using a hash function to convert plaintext passwords into a unique fixed-length string of characters. Hashing is a one-way function designed to be easy to compute but difficult to reverse. Hashing functions are deterministic in that the resulting hash value will always be the same for a given input. This predictability is necessary for authentication processes, as it allows systems to verify passwords by comparing the hash of the inputted password with the stored hash in the database.

While websites usually use hashing in combination with other security measures for their password storage mechanisms, these methods are not relevant to the dissertation as ”Let’s Get Cracking” focuses on emphasising the vulnerabilities that attackers exploit in common password practices, rather than the details of secure storage implementations.

2.1.3 Types of Attacks

When a password storage database is compromised, attackers typically gain access to a list of hashed passwords. They then employ computational resources to generate a roster of potential passwords and subsequently hash them using the same algorithm implemented by the password storage system. The attacker then matches the newly generated hashes with those stored in the compromised database. The detection of a match indicates a successful guess of the original password.

While attackers could potentially resort to brute-force attacks, systematically generating hashes for every conceivable password, this method is inefficient. Instead, they often exploit patterns used to create memorable passwords. These patterns include the incorporation of easily recognisable keyboard sequences (e.g., ”qwerty” and ”1234”), the substitution of characters with similar symbols (e.g., using ”@” instead of ”a”), and

exploiting predictable character placements (e.g. numbers and special characters tend to appear at the end of passwords). Attackers use what are called dictionary attacks, which involve making guesses using common words, phrases, keyboard patterns, and cracked passwords (Summers and Bosworth, 2004). Each of these techniques is aimed at exploiting the predictability of human password creation.

In scenarios where hashing is not used in combination with other security measures, attackers can potentially exploit precomputed tables of hashes known as rainbow tables. These tables contain pairs of plaintext passwords and their corresponding hash values. These allow password crackers to find matches without the extensive computational effort of generating these hashes themselves.

Once a password is obtained, attackers may unlock the compromised account and potentially other accounts across the internet that employ the same username and password combination. This is primarily due to the common practice of password reuse among individuals. A study by Poornachandran et al. (2016) found that 59% of regular internet users reuse passwords for multiple accounts and that 57% use slightly modified versions of existing passwords.

2.1.4 The RockYou Data Breach

RockYou, an advertising network best known for distributing mobile games to third-party platforms, such as Facebook, suffered a significant security breach in December 2009 due to a SQL vulnerability. 32 million user accounts were exposed, including many from third-party websites. It was revealed that RockYou was storing passwords in plaintext rather than securely hashing them. To this day, it remains the largest single data breach of plaintext passwords[1].

Due to its size and ease of access, the leaked database has been used extensively in password security research, providing data regarding password selection practices and patterns (Ur et al., 2016) (Kelley et al., 2012) (Tatlı, 2015).

While there may be ethical considerations with the usage of this password data, the passwords are not associated with any usernames, login details, or any other identifiable personal information. As the database is easily accessible, more than a decade old, and frequently used in research, the risks for the original users are negligible.

RockYou did not impose any password length or character requirements on their users, and given it was a low-stakes website, users typically used very weak passwords of low complexity. About 30% of the unique passwords from the dataset are equal to or below 6 characters in length, and 60% of the passwords do not contain a special character (Tatlı, 2015).

2.1.5 Public Perception of Password Security

Public perception of password security often deviates from established best practices. A study by Ur et al. (2016) revealed that many individuals underestimate the vulnera-

[1] Understanding RockYou.txt, https://www.keepersecurity.com/blog/2023/08/04/understanding-rockyou-txt-a-tool-for-security-and-a-weapon-for-hackers/

bilities associated with constructing passwords around common keyboard patterns and familiar phrases. Users frequently employ predictable characteristics in their passwords, such as incorporating words, easily recognisable keyboard sequences, and substituting characters with similar-looking symbols (also known as ”common l33t”).

It has also been observed that users tend to base passwords on easily guessable words, phrases, or concepts, including names, dates, and song lyrics. This stems from a lack of knowledge regarding the details of large-scale database attacks. Although 73% of participants in the Ur et al. (2016) study identified large-scale guessing attacks as a threat, there was a wide variance in the estimated number of adversarial guesses. 34% of participants believed a password to be secure if it could resist up to 50 guesses, while 67% believed that withstanding 50,000 guesses was a sign of security. In reality, a good password should be able to withstand between 10^{14} - 10^{20} guesses to be considered secure, depending on the type of hash function used (Ur et al., 2016).

2.2 Serious Games

Serious games are commonly defined as games that do not have entertainment, enjoyment, or fun as their primary purpose. The field of serious games has seen rapid growth in the past decade, with a market value in the billions (Laamarti et al., 2014). The field has also seen rapid growth in research due to its potential in various industries. Serious games leverage the interactive, immersive, and addictive nature of games to engage and motivate users to learn and improve skills (Laamarti et al., 2014). The research conducted by Backlund and Hendrix (2013) yielded positive outcomes for serious games and their impact on learning. Out of the 40 selected studies, 29 demonstrated a positive effect, while only 7 showed neutral results. This variance was attributed to the range of game-lengths with some games being too short to appropriately cover the educational content and others too long to maintain engagement.

2.2.1 Key Elements of Effective Serious Games

2.2.1.1 Enjoyment

Although serious games do not prioritise entertainment, player engagement is highly correlated with enjoyment and fun. To do this, the game should contain varied and interesting gameplay that embeds the educational goal. A balance between the educational and game parts must be carefully considered by ensuring all learning tasks are connected to in-game elements (Caserman et al., 2020).

2.2.1.2 Appropriate Rewards, Feedback, and a Clear Objective

The game’s goal should be aligned with its educational purpose and encouraged through positive reinforcement. For example, a very common form of in-game reward is through a point or score system. This can be further improved with a high-score table, allowing players to compare their performance against other players, further encouraging them to perform better (Caserman et al., 2020). The player’s score and/or status must always be visible, and they should receive immediate feedback as a result of their actions. This

is necessary not to break player concentration and to maintain a flow state (Sweetser and Wyeth, 2005).

2.2.1.3 Player Control

Players should possess a sense of autonomy over their decisions in-game rather than simply employing strategies intended by the developer. Through this, players are able to experiment and feel an increased sense of control. A game with too many constraints on player freedom can give the perception the player must follow a predetermined path, frustrating players and negatively impacting their experience. Player control must also be maintained in the case of player error. The game should continue to function if the player makes an error and help them recognise and fix the error through, for example, a warning message (Sweetser and Wyeth, 2005).

2.2.1.4 Adaptive Level of Difficulty

The level of challenge should match the skill level of the player. As the player naturally improves at the game, the difficulty should also increase in tandem. As a large discrepancy between challenge and skill level can negatively affect players' enjoyment of the game, care should be put into ensuring a minimal gap. The player should also be taught to play the game through an interactive tutorial. The tutorial should not feel very different from what actually playing the game is like (Sweetser and Wyeth, 2005).

2.2.1.5 Appropriate Graphics and Sound

Player engagement can be improved by stimulating different human senses through the use of appropriate audiovisual elements (Caserman et al., 2020). These include appropriately thematic visuals, sound effects, and background music. Care should be taken to ensure these game elements do not distract players from the goal and interfere with learning.

2.2.2 Examples of Serious Games and Related Work

2.2.2.1 Re-mission

The Re-Mission series of games are serious games designed for young cancer patients. In the game, you play as a nanobot designed to fight cancer and related infections in the human body (Figure 2.1). The goal of the game is to teach players about cancer treatments and how they work to promote adherence to self-care during treatment. A study conducted by Beale et al. (2007) found a significantly larger increase in the retained knowledge of re-mission players over a 3-month period compared to the control group.

2.2.3 Plague Inc.

"Plague Inc." is a popular strategy simulation game where players take the role of a pathogen with the goal of infecting and eradicating the global population. By placing

Figure 2.1: Gameplay of Re-Mission (Beale et al., 2007)

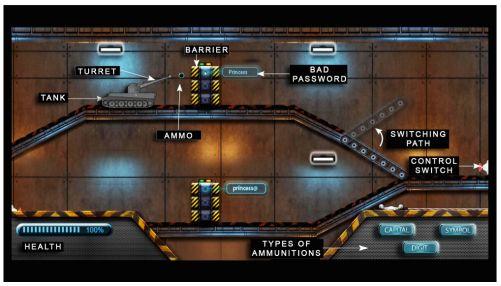


Figure 2.2: Gameplay of GAP (Tupsamudre et al., 2018)

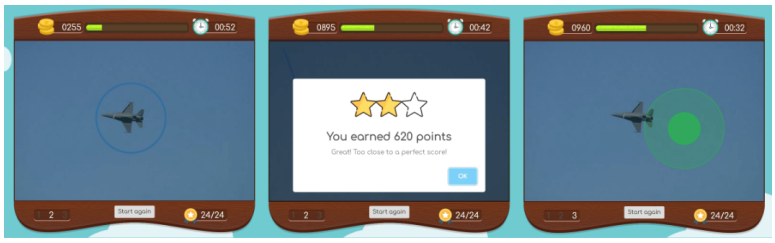


Figure 2.3: Gameplay of GamePass (Raptis et al., 2021)

players in this adversarial role, they gain a practical understanding of how diseases spread and how humanity attempts to combat them (Filho et al., 2023). This approach has been used in an undergraduate Microbiology module at the University of Derby, with students reflecting on pathogen evolution, transmission strategies, and deeper epidemiological insights (Robinson et al., 2018).

2.2.4 GamePass

GamePass is a mechanism to gamify the creation of graphical user authentication (GUA) passwords. GUA passwords are an alternative to text-based passwords, where users draw passwords on background images. GamePass encourages users to create more unpredictable passwords by guiding user attention to non-salient areas of authentication images. This behaviour is further encouraged with a reward system which gives higher scores based on how secure the password is (Figure 2.3). The research showed an improvement in the GUA passwords created by GamePass players. Although GUA passwords are much rarer than traditional text-based passwords, GamePass demonstrates how an in-game reward system can encourage users to generate better passwords (Raptis et al., 2021).

2.2.5 PASDJO

PASDJO is an online game where players rate the strength of text-based passwords under a time limit (Figure 2.4). Players are then scored based on how closely their perceived score is to the expected score. A results screen provides very brief justifications for the expected strength ratings of passwords. This helps users learn about the characteristics of weak passwords they may not have been familiar with. The study observed how

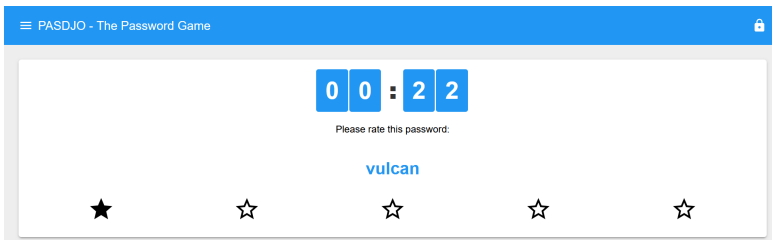


Figure 2.4: Gameplay of PASDJO (Seitz and Hussmann, 2017)

users underestimated passphrases by an average of 1.4 points on a 5-point strength scale. Although multiple play-throughs of the game were shown to improve password strength perception, only 27% of the participants chose to play more than one round. The game only used 4 types of passwords to test a few selected characteristics: random passwords, common passphrases, common alterations, and ”common l33t” (the substitution of letters for similar-looking characters). PASDJO falls short in explaining the reasons why these characteristics are considered weak, omitting any explanation of the attack model and the tools used to exploit these vulnerabilities (Seitz and Hussmann, 2017).

2.2.6 GAP

GAP is a serious game created with the goal of educating users about common and unsafe password habits. Players play as a tank that must navigate a maze filled with barriers representing different passwords (Figure 2.2). The player must shoot these barriers down by identifying the weak characteristics of the passwords they represent (e.g., the password contains an uppercase letter at the beginning, which over 70% of users do). The study found that after an average of 3 and a half minutes of playing, participants' identification of the tested password characteristics improved from an average of 60.65% to 82.96%. However, the study only tested six characteristics, focusing on the identification of password strength through the placement of uppercase letters, digits, and special characters at either the beginning or end of passwords. This omits password characteristics such as their susceptibility to dictionary attacks. The game also does little to explain why these characteristics are insecure beyond the identification they are found in a large percentage of users' passwords (Tupsamudre et al., 2018).

Chapter 3

Previous Work

This chapter provides a brief overview of the work done on this iteration of the project from the previous year. This includes an overview of the game, its design, implementation, evaluation methodology, and results. In addition to outlining the progress made, this chapter also discusses the issues identified in the previous iteration and offers suggestions for improvements to be implemented in the updated version of the game.

3.1 Overview

Passwords are a ubiquitous aspect of modern digital life, and the average person is frequently reminded to create secure passwords, often guided by heuristics such as password strength meters. However, these meters usually calculate strength based on password length and the variety of character types used (LUDS). They often fail to consider risks such as the predictability of placing numbers, special characters, and uppercase letters in common positions, as well as the usage of dictionary words. These render passwords susceptible to brute-force attacks and dictionary attacks, respectively.

While the serious games discussed in section 2.2.2 identify many of these risks, they do not cover the full range of these characteristics and do not explain why they are insecure with reference to the attack model. ”Let’s Get Cracking” addresses this by positioning players in the role of the adversary. From this unique perspective, players better understand the attack model and the tools used for cracking passwords. Players are encouraged to think creatively about how to maximise the number of accounts they can crack. Rather than being explicitly told the rules that define weak passwords, players discover these characteristics organically through trial and error, refining their strategies as they observe which passwords are easier to crack. This process naturally leads to an implicit learning experience, where players discern the characteristics of weak passwords and gain insights on how to strengthen their own.

3.1.1 Gamification

The game simulates the adversarial position by giving players simplified versions of tools used by real password crackers. This includes an interface which allows players to

input a character and/or word sequence and then simulate the brute-force generation and matching of hashes with those from a leaked password database. This input sequence can be constructed using numbers, letters, special characters, and dictionary words.

The game challenges players to crack as many accounts as possible within a set time limit. Players input potential word/character combinations, generate the hashes corresponding to all possible passwords that conform to this input sequence, and are then shown matches from the password database. Generating hashes uses up ”time remaining” which is a limited resource. The amount of time consumed is calculated proportionally to the number of hash combinations from the input.

This system encourages player freedom and experimentation, as recommended in section 2.2.1.3. It has a large number of gameplay systems, providing constantly varied gameplay to engage the player. In experimenting with these different systems, players receive immediate feedback on the number of accounts they have cracked with the hashes they have generated (section 2.2.1.2). The learning outcome of the game is thus embedded in the game's goal, which is an important part of balancing the educational and entertainment parts of the game, as outlined in section 2.2.1.1. By observing how different input sequences affect the player's score and time, the player can learn the patterns which crack the most accounts in the least amount of time and, therefore, the characteristics of these weak passwords.

To avoid overwhelming the player, new password cracking tools are unlocked progressively across five different levels. These include custom-inputs, dictionary attacks, and dictionary customisation. The difficulty of each level also increases through the use of increasingly strict password requirements (section 2.2.1.4). These requirements specify the minimum length and character types that an input combination must meet. The increasing password complexity, along with the addition of password-cracking tools, helps balance the game experience and encourages players to effectively utilise and experiment with these newly introduced features.

3.2 Design and Implementation

The game was developed using Godot 4.0, a free and open-source game engine that supports both 2D and 3D projects. The game code was written in GDScript, Godot's purpose-built programming language, while data structures such as passwords and dialogue were formatted using Python. Due to space limitations, some of the visualisations will be presented in the following chapter, where comparisons to the new user interface can be made more effectively.

3.2.1 Tutorial and Dialogue

The game includes a tutorial to familiarise players with cyber-security concepts such as hashes, their properties, how they are used in password storage, and the process involved in cracking them. The dialogue, delivered by a character named ”CiPH3R”, is accompanied by relevant visuals (Figure 3.1).

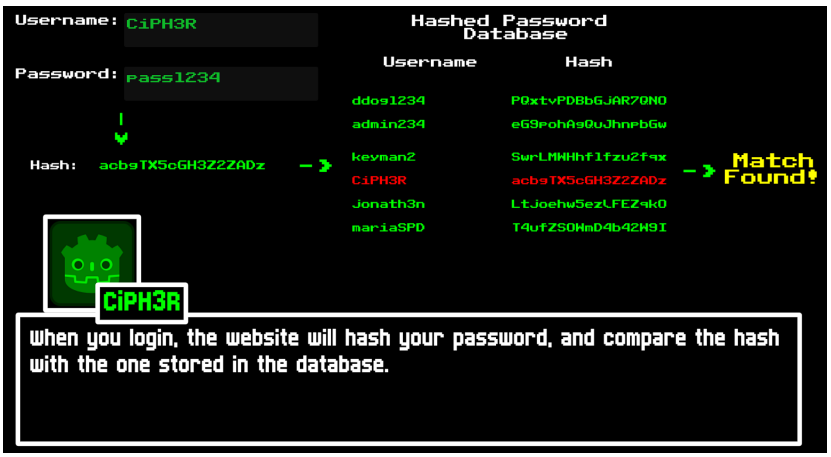


Figure 3.1: Screenshot from the original game's tutorial.

Following the initial tutorial, CiPH3R provides brief dialogue at the beginning of each level. In the first level, they introduce the user to the interface and guide them in inputting specific combinations. In later levels, the dialogue becomes less interactive, simply explaining the level's password requirements, any new password-cracking tools, how to use them, and providing relevant tips.

3.2.2 Results Screen

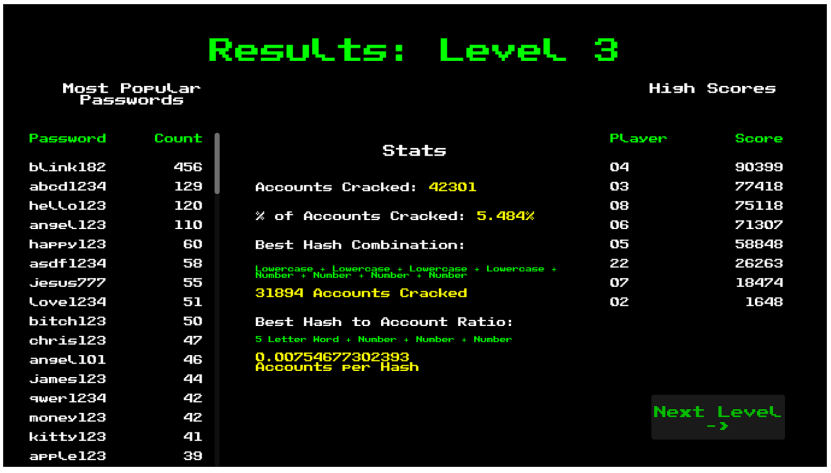


Figure 3.2: Results page from original game.

After each level, a results screen presents detailed statistics on the passwords the player has cracked, allowing them to analyse their performance and refine their strategies (Figure 3.2).

These statistics included the most frequently cracked passwords, the total number of cracked accounts, the percentage of all accounts cracked, the best input sequence that led to the highest score increase, and the input sequence with the highest efficiency (cracked accounts per hash generated). These insights allow players to evaluate the effectiveness of their attack strategies. The results screen also included a leaderboard

which compares the player's performance with others, adding a competitive element that motivates them to improve their skills and experiment with different approaches (section 2.2.1.2).

3.2.3 Simulating Password Cracking

The game simulates password cracking by performing character-type matching on a set of plaintext passwords sourced from the RockYou data breach (section 2.1.4). The RockYou website did not enforce any password complexity restrictions, and because many users perceived it as a site of low security importance, they often chose less secure passwords. Consequently, any complexity present in user-created passwords typically resulted from individual choice rather than mandated security policies. This is an important distinction, as user-chosen complexity differs from the complexity users add when forced to comply with LUDS policies (Wheeler, 2016). Despite this, analysis from Tatlı (2015) indicate that the RockYou passwords still follow similar underlying composition patterns. For example, even when users add complexity, they frequently do so in limited ways, such as appending a single number or symbol to a base word, rather than employing more diverse character combinations.

To implement the password cracking simulation, each individual password in the dataset is checked for character-type matches using a linear-time algorithm. Although this approach can be slow, the game conceals this delay by displaying cracked passwords as they are found. This is computationally intensive, leading to suboptimal performance on machines with limited processing power, and prevented the game from being hosted or distributed online.

3.3 Evaluation

The game's evaluation involved having participants play the game and complete a two-part questionnaire to assess their understanding of password security. The first part was answered before gameplay, while the second was completed afterward. Many of the questions were repeated in both parts to measure changes in participants' responses and determine the game's impact on their knowledge.

The questions asked included short-form text answers to gauge their prior knowledge and confidence of password security, questions where they would be asked to rate and compare the strength of passwords, and questions to reflect on their experience and effectiveness of the game.

3.3.1 Problems with Password Comparison

Participants rated the strength of 12 individual passwords on a five-point scale and compared the strength of 12 password pairs on a seven-point scale. This approach, inspired by Ur et al. (2016), measured how participants' perceptions of password security changed after playing the game. The passwords chosen test many of the same characteristics evaluated in PASDJO (section 2.2.5), GAP (section 2.2.6), and the aforementioned Ur et al. (2016) study.

To quantify results, password strength ratings were compared with *zxcvbn*'s estimates. However, *zxcvbn* was not designed for direct pairwise comparisons, sometimes producing misleading results. For example, it assigned equal strength to "appleton16" vs. "appletonqy" and "scotland1" vs. "1scotland," despite meaningful security differences. This occurs because *zxcvbn* relies on previously cracked password data rather than complexity analysis.

Despite *zxcvbn's* limitations, these comparisons remain valid since the primary focus is on how players' perceptions shift rather than absolute password strength.

3.4 Summary of Results

The study concluded with 22 participants, and overall reactions were highly positive. Participants praised the game's presentation, including its music, sound effects, and visuals. However, despite these positive responses, the evaluation also uncovered unexpected issues that negatively impacted the player experience.

3.4.1 Unexpected Observations from Gameplay

The study revealed several key areas where the game could be improved. Participants often favored complex input sequences over simpler single-character-type sequences, likely due to the influence of modern password policies. Players tended to overlook the number of hashes generated and the time it took, possibly due to the cluttered UI and insufficient emphasis on hash quantities. They also underutilised the word options menu and the "Any Character" button, likely due to unclear explanations and poor placement within the interface. Players may have also skipped the information as they would get tired of the dialogue and skip through important sections.

3.4.2 Summary of Important Statistics

The questionnaire responses indicate that the game effectively improved participants' understanding of password security concepts.

Participants' confidence in identifying weak passwords increased from an average of 3.36 to 4.09 out of 5 after playing the game. The percentage of participants who correctly recognised that password strength plays a role in protecting against large-scale breaches rose from 50% to 95.5%.

While participants' estimates of "how many guesses do you think your password needs to be able to withstand for it to be considered secure" improved after playing the game, only one participant answered with a numeric value within the range recommended by Ur et al. (2016) of 10^{14}–10^{20} guesses. This result may be due to several factors, including unclear question phrasing, the game's focus on short input sequences with low hash counts, the cluttered UI that made it difficult for players to notice the "Number of Hashes" quantity, and the inherent difficulty of comprehending numbers of this scale.

After playing the game, participants showed a significant improvement in their understanding of password security, with understanding of brute-force and dictionary attacks

increasing from 54.5% to 90.1% and knowledge of hashes rising from 50% to 95.5%.

3.4.2.1 Password Strength Evaluation

Participants' ability to assess password strength improved, with a reduced tendency to overestimate the strength of weak passwords. Compared to the results from the PASDJO study, players were able to more accurately identify weak passwords to a degree that would take the average PASDJO player 9 rounds of playing to achieve. However, very few participants in the PASDJO study were willing to play that many rounds (Seitz and Hussmann (2017)).

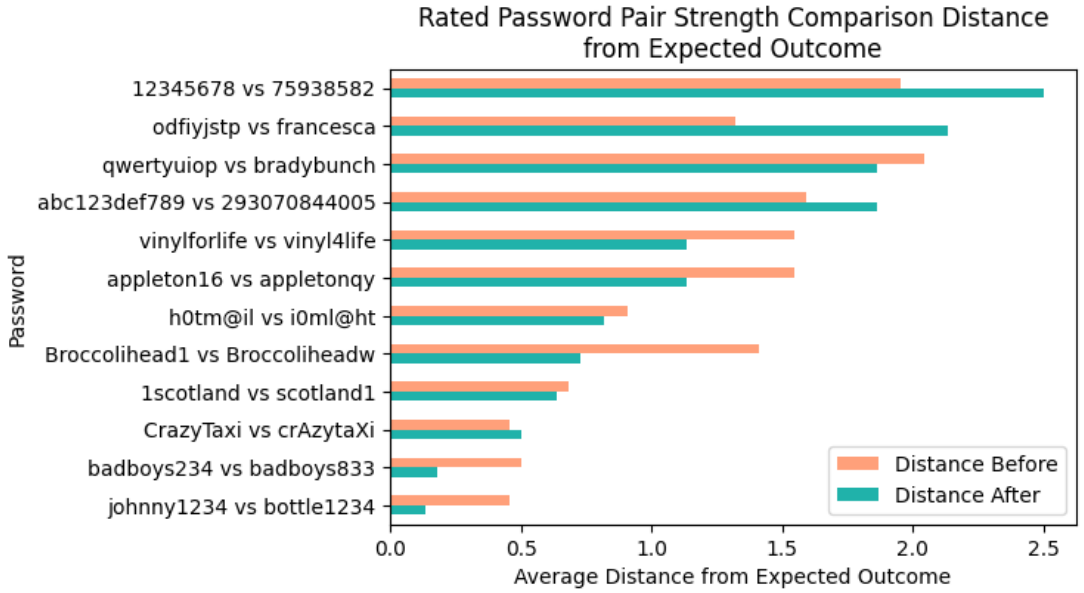


Figure 3.3: Change in average distance of participants' pair strength comparison ratings

Figure 3.3 shows that the game helped participants identify weaknesses in predictable character placement (e.g., "1scotland vs scotland1") and the use of numbers instead of letters (e.g., "appleton16 vs appletonqy"). However, players struggled to recognise the relative weakness of keyboard patterns (e.g., "12345678 vs 75938582") and the weakness of using common names over random letters (e.g. "francesca vs odfiyjstp"). This difficulty can be attributed to players not acknowledging the "Hash Number" counter and a lack of experimentation with "Word Options", the game's dictionary modifier.

3.5 Improvements

The evaluation highlighted some shortcomings. To improve the game, the new version addresses the following issues.

3.5.1 Improved User Interface

The evaluation revealed that many players struggled to make accurate judgments about password strength, largely because they missed or overlooked key information. Critical details like the ”Number of Hashes” and ”Time to Generate” often went unnoticed, and essential UI elements like ”Any Character” and ”Word Options” were easily overlooked due to the cluttered user interface. Consequently, players misjudged the strength of passwords, underestimated the weaknesses of keyboard patterns and common names, failed to understand the impact of different character types, and struggled to grasp how many guesses a password needs to withstand to be considered secure.

The main issue was that the game attempted to present all relevant information on a single screen, resulting in a cramped and overwhelming layout. To address this, the user interface should be redesigned to separate content into clear, manageable sections. This would reduce visual clutter and help players better understand and utilise the available tools without feeling overwhelmed.

3.5.2 More Engaging Dialogue and Tutorial

To address the issue of players skipping through dialogue, several strategies should be implemented to enhance engagement and retain attention. One approach is to make the dialogue box more dynamic and interactive by incorporating movement, text effects, or animations that capture the player’s focus. The dialogue should also be streamlined and made more concise. The game should also include more mandatory interactions to prevent players from skipping through the dialogue without engaging with the content.

3.5.3 Preventing Unwanted Progression

A common problem would occur when players would use up all of their in-game time by generating hashes for an unsuccessful combination, ending the level with a poor score, and still being able to progress. To prevent this, the goal of the game should shift to ensure players have cracked a certain number of accounts before they can progress to the next level. Combined with additional tools to support struggling players, this change would encourage more strategic thinking rather than allowing them to move on to more challenging levels despite a poor performance.

3.5.4 Help Struggling Players

Since the game relies on implicit learning based on players’ prior knowledge of password combination patterns, some may struggle to identify the right combinations. To help these players improve, the game should include support mechanisms. One approach is to introduce an explicit hint system, where CiPH3R offers guidance for struggling players. The game could also feature more subtle, implicit gameplay cues that nudge players toward more promising combinations without directly giving away the answer.

Chapter 4

Design and Implementation

Building off the strong foundation of the previous work, an improved version of ”Let’s Get Cracking” was created to address its shortcomings and further enhance its positive aspects.

4.1 User Interface

The results from last year highlighted several issues with the user interface (3.5.1). It lacked clear sections and was overly cluttered, making it difficult for players to find critical information and buttons. This was addressed in the new version of the game through a complete overhaul of the user interface.

4.1.1 Figma Mock-Up

Instead of the previous iterative approach, where the UI was designed gradually as new features were added, a comprehensive mock-up of the final design was created in Figma, a web-based design tool. This resulted in a more cohesive and structured design process from the outset.

A key structural change in the new interface was the separation of different phases of the gameplay loop into distinct sections, which are displayed on different pages. This contrasts with the previous design, where all information was displayed simultaneously, potentially overwhelming the player (Figure 4.1). Compartmentalising the gameplay phases enhances clarity, reduces cognitive load, and allows players to focus on the relevant information for each stage, improving the overall user experience.

The sections designed in Figma included the main screen, the inputting interface, the hash-matching screen, the upgrades shop, and the results screen.

4.1.2 Nielsen’s Heuristics

Nielsen’s heuristics are a set of usability guidelines that help improve user experience and interface design (Nielsen (1994)). The interface was designed with the following

heuristics in mind:

- **Visibility of System Status** – The compartmentalised interface clearly indicates the user's current state. Responsive UI elements and animations provide immediate feedback.
- **User Control and Freedom** – Players can freely navigate between states using back and close buttons, and inputs can be cleared using the backspace and clear buttons for ease of use.
- **Consistency and Standards** – All interactive buttons are green. Light green buttons are consistently used for navigation between states and pages, while dark green buttons are reserved for inputting combinations. This maintains a clear and predictable interaction model.
- **Error Prevention** – The requirements panel continuously displays missing password requirements to prevent invalid inputs. It also restricts overly long input combinations, prevents duplicate entries, and provides clear warnings before submitting a combination that exceeds the total available time (section 2.2.1.3).
- **Recognition Rather Than Recall** – The interface incorporates iconography, allowing users to quickly recognise their functions without needing to recall details from memory.
- **Flexibility and Efficiency of Use** – The game supports shortcuts such as the physical backspace key, which functions identically to the backspace button in the input page. Players can also advance dialogue using the Enter key, the Spacebar, or by clicking the dialogue box, offering multiple interaction options.
- **Aesthetic and Minimalist Design** – The game's consistent aesthetic provides clarity. Compartmentalisation of different UI elements across different screens prevents players from being overwhelmed by irrelevant information based on the current game state.
- **Help Users Recognise, Diagnose, and Recover from Errors** – The requirements/warning panel explicitly informs users why their input is invalid, aiding in error resolution.
- **Help and Documentation** – The game's dialogue system serves as a tutorial, guiding users on how to navigate the interface. Once viewed, players can replay the tutorial or access hints as needed.

4.1.3 Implementation in Godot

While some changes were made in the final version of the design, the implementation of the user interface in Godot remains largely similar to that of the Figma mock-up. The final in-game designs for the main page, input page, and matching page, can be seen in Figures 4.2, 4.3, and 4.4 respectively. A comparison between the Figma and implemented versions of the upgrades shop (introduced in section 4.2.2) is shown in Figure 4.6.

Using Godot's animation system, the interface is able to smoothly transition from each of the game's states, improving user feel and gameplay flow. In particular the design is more fluid and responsive than the previous one, with buttons and panels scaling dynamically based on interactions, improving feedback given to the user as they play. For example, when inputting a combination, the "Number of Hashes" text grows and shrinks in proportion to the number it is displaying. This visual cue directs the player's attention to the increasing number, something the previous game's interface failed to highlight effectively.

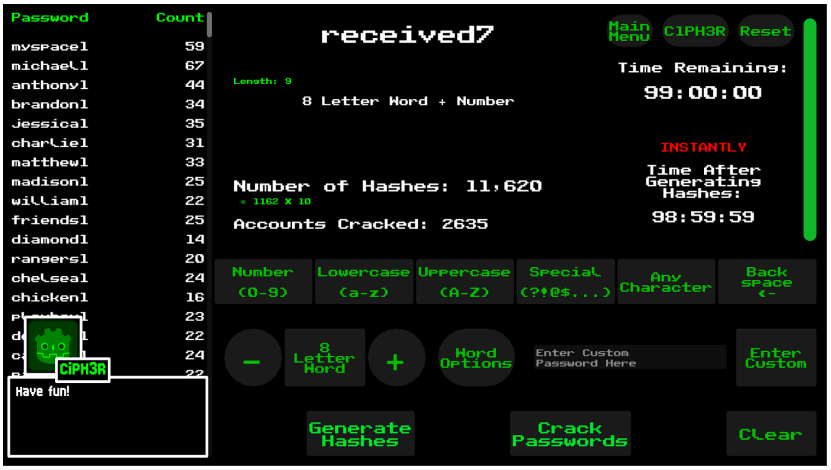


Figure 4.1: Cluttered user interface of main page of previous version.

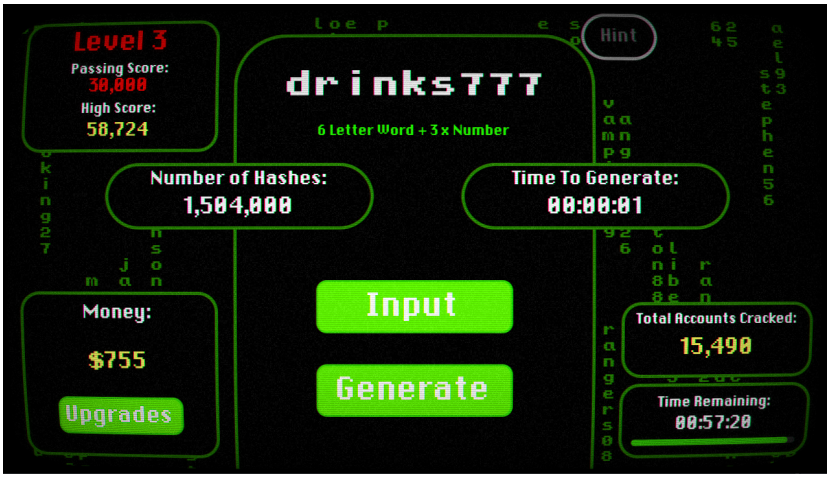


Figure 4.2: New UI - Main page

4.2 New Gameplay Features

4.2.1 Passing Score

As discussed in section 3.5.3, the goal of the game has been changed such that players must achieve a minimum passing score before progressing. If they run out of time without reaching this threshold, they are required to replay the level.

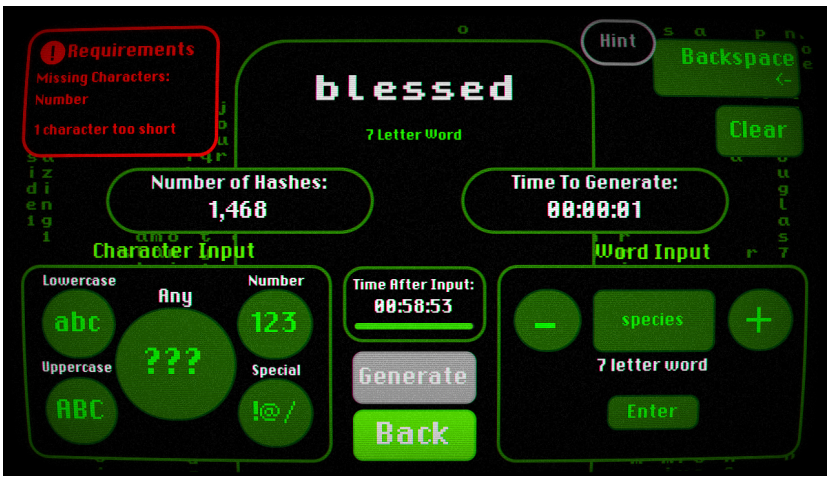


Figure 4.3: New UI - Input page with all tools unlocked and Requirements Panel warning input does not meet requirements.

Figure 4.4: New User Interface - Password Cracking Page

The purpose of this is to address several of the issues raised in section 3.4.1. One key problem is that players could accidentally progress to the next level by inputting an excessively long hash combination, which would completely deplete their in-game time. The passing score ensures that players have successfully applied the intended strategy of targeting weak passwords, reinforcing the game's educational objectives.

It also means that players will spend less time on a given level. In the previous version of the game, players had to remain on the level until their time limit expired before they could access the results page. Due to the fear of using longer hash combinations, this led to players spending excessive amounts of time on levels, disrupting the intended pacing of the game.

This passing threshold also allows for the inclusion of longer time limits, as players no longer feel pressured to remain on a level indefinitely. By extending the time limit to an hour for all levels, compared to the previous game's five-minute limit, players have more freedom to experiment with longer hash combinations without excessive time pressure, an issue with the previous game which resulted in poor password strength evaluation

by the participants. The ability to retry levels also further encourages experimentation, allowing players to refine their strategies.

4.2.2 Upgrades Shop

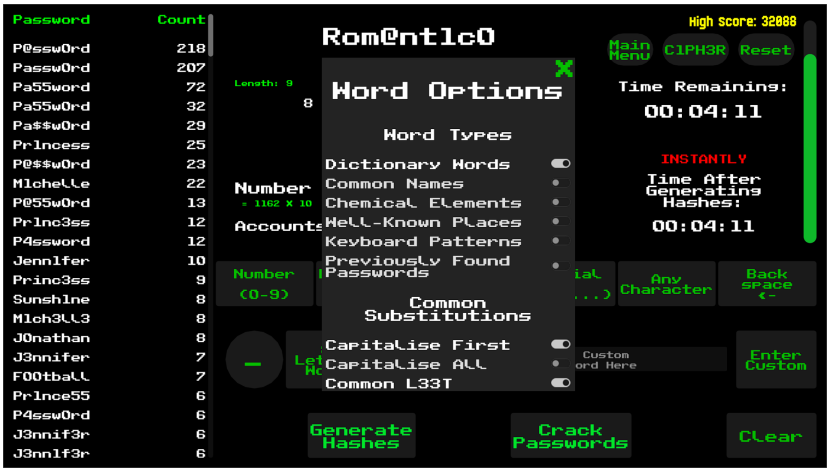
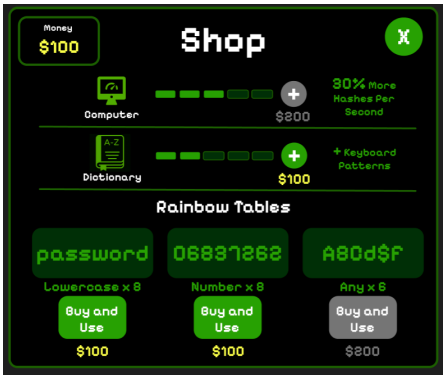


Figure 4.5: Previous version: word options menu



(a) Figma Version

(b) Implemented Version

Figure 4.6: Comparison of figma and implemented versions of the upgrades shop.

In addition to being rewarded with a score, players also earn currency through a new in-game economy system. After completing the first level, players receive $1 for every 1,000 accounts they successfully crack, with a variable conversion rate depending on the total number of accounts in a given level. This currency can then be spent in the upgrades shop, where players can enhance their computer's processing power, expand their word dictionary, and purchase Rainbow Tables (Figure 4.6b).

Along with the level's high score, the monetary reward for cracking passwords encourages players to play beyond the passing score, as upgrades are permanent and will help them in future levels. This improves the game's sense of reward (section 2.2.1.2), and the added depth to player gameplay improves player control (section 2.2.1.3).

The upgrade's shop replaces the "Word Options" menu from the previous version of the game (Figure 4.5). It was shown participants did not engage with this feature in the previous study, resulting in overestimating the strength of passwords with names, places, keyboard patterns, uppercase variations, and common substitutions.

Converting the menu into an upgrades shop achieves several goals. Because dictionary upgrades now cost money, players cannot activate all options simultaneously, unlike in the previous version where there was no such restriction. These upgrades are permanent, giving players a sense of progression and ownership, a contrast to the original menu which unlocked everything from the start. Furthermore, player frustration with settings resetting between levels has been addressed. Consequently, interacting with the shop becomes necessary in later levels, as upgrades like uppercase variations, common substitutions, and computer speed are extremely helpful for meeting stricter password requirements. This is reinforced through dialogue and hints that explain how to use the shop and its benefits.

4.2.3 Rainbow Tables

Rainbow Tables are purchasable items in the shop. Like their real-world counterparts, they are pre-generated sets of hashes. In gameplay, they allow players to match hashes from specific input combinations without consuming in-game time. The shop displays three different rainbow tables at any time, each for a specific input combination. Players can purchase a table to crack passwords matching that combination without using any in-game time.

Beyond their primary function, Rainbow Tables also educate players by providing examples of effective input combinations, as suggested in section 3.5.4. The shop displays Rainbow Tables with input combinations chosen randomly from the top 30 most effective combinations for the current level. As all tables are priced equally, it is up to the player to choose which table they believe would be the most useful, adding strategic depth to the game. The game also encourages players through dialogue and hints to analyse patterns within these tables and apply them to their own strategies. For instance, players unaware that many passwords with uppercase letters tend to begin with them can learn this trend from the Rainbow Tables and adjust their input combinations accordingly.

4.3 Improved Results Page

The results page has been overhauled in this version of the game. While the information displayed is mostly the same, it has been compartmentalised and made more digestible through the game's new design language (Figure 4.7).

The "Best Hash Combination" section has been expanded to show all of the player's input combinations, sorted by the number of accounts cracked. This allows players to compare their best-performing combinations with their worst, allowing them to observe patterns in their inputs and identify areas where their strategies can be improved.

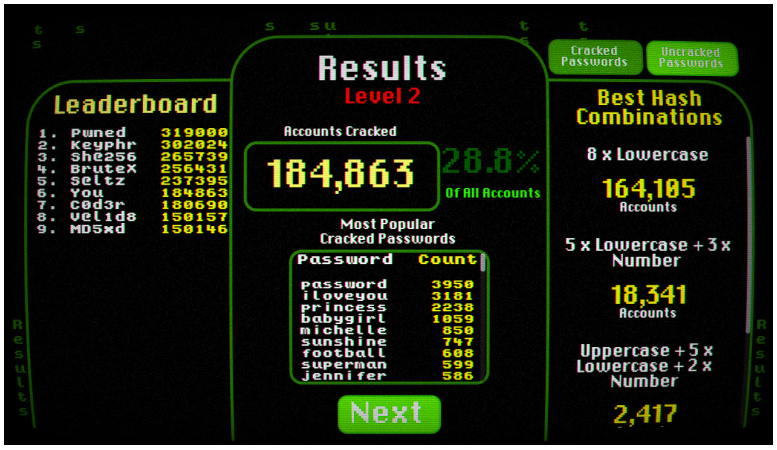


Figure 4.7: New Results Screen

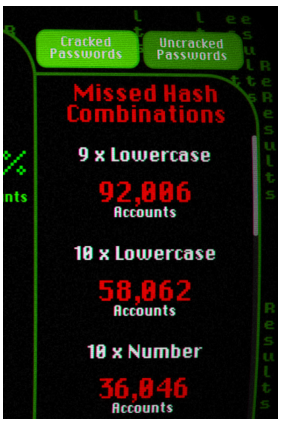


Figure 4.8: Uncracked Passwords

The leaderboard has also returned from the previous version of the game as this competitive aspect of the game was hugely encouraging for players. Unfortunately, due to the fact the game is hosted online, it was not possible to implement a live leaderboard. The leaderboard now emulates the original experience by using fictitious usernames and scores.

A key new feature is the toggle at the top of the section, which, when clicked, reveals the best input combinations that the player missed (Figure 4.8). This provides valuable feedback on overlooked patterns, highlighting strategies players may have missed, such as simpler approaches they ignored in favor of more complex ones in the previous version of the game. This feature is particularly helpful for players who may be struggling, as it steers them toward more effective strategies, which in turn, helps them understand the characteristics of weak passwords (section 3.5.4).

4.4 Improved Tutorial and Dialogue

As suggested in section 3.5.2, the game's tutorial has been significantly improved through greater interactivity, dynamic visuals, and the improved dialogue system. The dialogue box is now more dynamic and interactive, addressing previous player feedback. The text has been enlarged for better readability, and the dialogue box is no longer static; it dynamically resizes and moves around the screen to emphasise key UI elements. Furthermore, important words in the dialogue now pulse to highlight their significance and reinforce key concepts.

As with the previous version of the game, there is an introductory tutorial which explains the concept of hashes, password storage, and brute-force attacks, followed by a brief dialogue tutorial at the beginning of each level to introduce players to the interface and any new password tools.

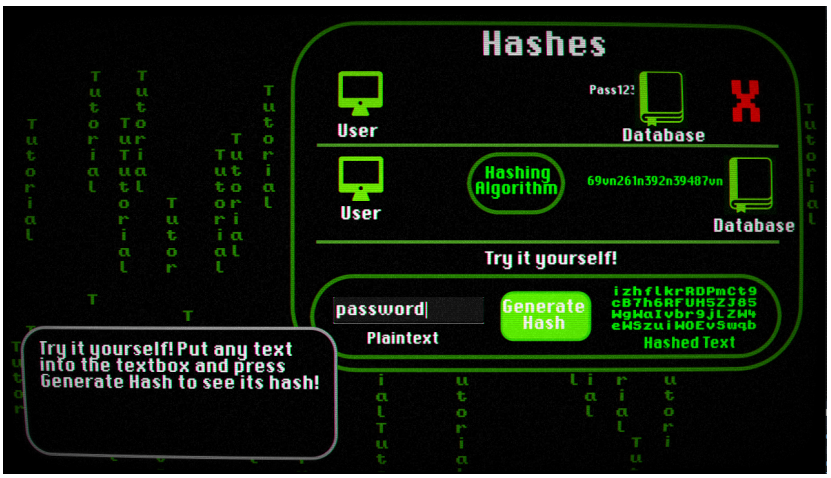


Figure 4.9: New Tutorial: Hash explanation with interactive hash converter.

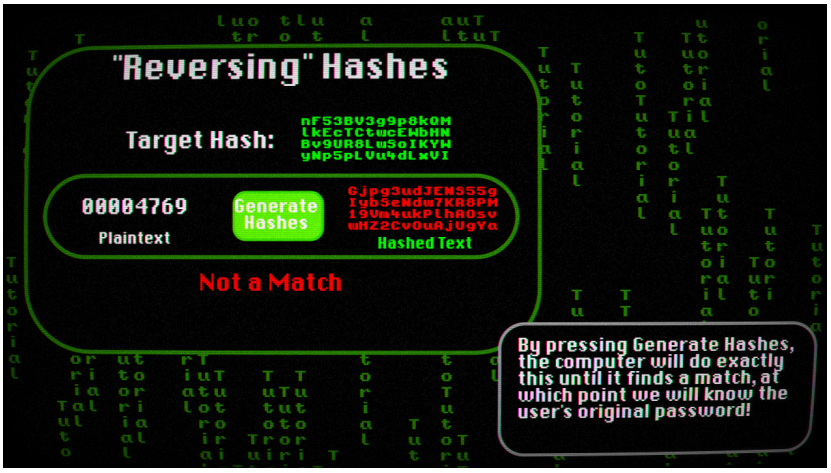


Figure 4.10: New Tutorial: Brute-Force explanation with simulation of brute force attack.

4.4.1 Introductory Tutorial

The concept of hashes is explained in the introductory tutorial using a semantic wave teaching approach. The semantic wave refers to the pedagogical process of simplifying complex concepts before systematically building towards a deeper understanding (Maton, 2013). This method involves initiating instruction with an abstract analogy, such as comparing hashes to fingerprints, and subsequently introducing more technical details.

This explanation is reinforced by an animation demonstrating how passwords are stored in a database. The visualisation first shows an incorrect approach, where a password is saved as plaintext, and then contrasts it with the correct method: converting the password into a hash before storing it.

The tutorial then introduces key properties of hashes, including irreversibility and consistency. To illustrate consistency, the tutorial provides an interactive text box where players can convert their inputted text into a hash. The dialogue encourages them to observe how the same input always produces the same hash, while minor modifications result in entirely different hashes (Figure 4.9).

Next, the tutorial explains that despite their irreversibility, hashes are vulnerable to brute-force attacks that exploit common patterns. An interactive brute-force simulation illustrates this: pressing the 'Generate Hashes' button initiates an animation that hashes and tests every possible eight-digit password. The animation concludes when a match is found, displaying the original password and demonstrating how brute-force attacks reverse hashes by systematically testing all possible inputs (Figure 4.10). The previous version of the tutorial lacked a visual analogy for brute-force attacks, making it harder for players to grasp how attackers systematically guess passwords. This approach makes the concept more tangible and intuitive, reinforcing learning through participation rather than reading (section 2.2.1.3).

The animations, dynamic text box, and interactive elements make the tutorial much more engaging then the one created for the previous game. In particular, the interactive elements prevent players from mindlessly clicking through the dialogue, and forces them to engage with the material.

4.4.2 Level Tutorials

Similarly to the previous game, completing the introductory tutorial takes the player to the first level of the game. They are taught step-by-step the process of cracking passwords with interactive checkpoints to reinforce their understanding of each stage. The dynamic dialogue box moves around the screen to highlight key UI elements, improving player engagement and focus.

After guiding the player through two cycles of the cracking process, the player is informed of the minimum passing score, leaving them to experiment freely, applying what they know about password patterns to optimise their approach.

Each subsequent level features a brief tutorial introducing new concepts, such as password requirements, the upgrades shop, rainbow tables, and dictionary attacks. Table D.1 outlines this progression. Unlike the previous game, these subsequent in-level tutorials are accompanied by more interactivity checkpoints, where the player must engage with the new mechanics. These include requiring players to open the 'Uncracked Passwords' page on the results screen and to input a specific dictionary attack, ensuring they know how to do so (a problem many participants had in the previous study).

4.4.3 Hint System

As suggested in section 3.5.4, to assist struggling players, the game features an optional hint system. At any point, the player can press the Hint button at the top of the screen, and CiPH3R will appear to provide a brief hint about a specific aspect of the game. The game features 15 hints in total, with each level having three exclusive hints, tailored to the level's particular new tools or password requirements. Examples of hints include tips on the placement of certain characters in passwords (e.g. uppercase at the start or numbers at the end), hints for using rainbow tables (such as recognising patterns and drawing inspiration from them), and the importance of upgrading the dictionary to successfully use a dictionary attack on levels with stricter password requirements.

The hint system is designed to provide targeted assistance, helping players to overcome the passing score without feeling frustrated, while still encouraging them to think critically and solve challenges on their own. Using hints is not penalised, making the game more inclusive and enjoyable for players of varying skill levels (section 2.2.1.4).

4.5 Settings and Accessibility

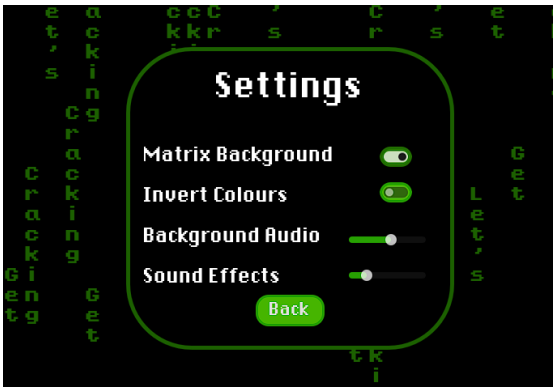


Figure 4.11: Settings Page

The game's title screen features a settings menu that allows players to customise various options for their comfort and accessibility. One of the settings is a toggle for the Matrix-style background animation. While this effect adds to the game's aesthetic, it may be distracting for some players or cause readability issues. Disabling it provides a simpler, more static background for better focus. Another important accessibility feature is a colour inversion toggle. This setting is particularly useful for players who have difficulty reading text on a dark background, such as individuals with light sensitivity, contrast sensitivity, or certain visual impairments. These customisation options enhance the overall accessibility and player experience, allowing a wider audience to engage with the game comfortably.

4.6 Improved Performance

The previous version's inefficient password cracking algorithm hindered online distribution due to high performance requirements (section 3.2.3). To address this a new approach was implemented. Passwords are now stored in a tree data structure, where each branch represents a different input character type. To find passwords matching an input combination, the tree is traversed, with the destination node containing the passwords and counts. This reduces the time complexity from linear time to $O(m)$, where m is the length of the password. For inputs using the "Any" character input, passwords are retrieved from multiple branches using a depth-first search algorithm.

Overall, the system is significantly faster at the expense of higher memory requirements for the tree data structure. Table 4.1 compares the time to retrieve passwords in both versions of the game. This comparison is somewhat misleading, as the previous version displayed passwords as they were found, whereas the new version traverses the tree

structure before displaying any passwords. The new system greatly outperforms the old system except when required to search through multiple branches such as with ”Any” character inputs. This is not an issue as inputs with multiple branches are discouraged as they use up lots of in-game time. Regardless, the improved search times significantly enhance the game’s pacing and reduce the delay between player action and feedback, improving game flow (section 2.2.1.2).

Password Format	Old Input Time (s)	New Input Time (s)
7 x Lowercase + Number	2.488	0.193
12 x Number	3.280	0.040
5 x Any	1.441	0.257
7 letter word + Number	1.332	0.226
8 x Any	1.817	2.962

Table 4.1: Comparison of time to retrieve passwords in both versions of the game.

4.7 Playtesting

As with the previous study, before starting the evaluation, the game was played by a handful of playtesters. These sessions were instrumental in identifying and rectifying various bugs and in further polishing the gameplay experience.

The most significant insight gained from playtesting was the need for careful game balancing. This balancing process involved fine-tuning the pricing of upgrades, adjusting the game’s reward system, and revising the criteria for passing levels. This proved to be a complex undertaking due to the diversity of player strategies and the uneven distribution of passwords in certain branches of the tree data structure. Some players were able to achieve passing scores with relative ease by discovering effective input combinations, while others experienced much more difficulty.

Although the game already includes features to assist struggling players, adjustments were made to lower the overall difficulty. The goal was to prevent frustration and encourage players to fully play the game and unlock all the tools and upgrades. The leaderboard and monetary reward still encourage players to exceed the minimum score, thereby accommodating those who seek a greater challenge.

Chapter 5

Evaluation and Results

This chapter presents the evaluation of the improved version of ”Let’s Get Cracking”, which involved a study with 22 participants. This chapter will first outline the methodology, and then proceed to present and analyse the data collected from the participants.

5.1 Methodology

5.1.1 Questionnaire

Participants’ understanding of password security was assessed through a questionnaire completed in two parts: one before and one after playing the game. All of the questions asked in the first half of the questionnaire are asked again in the second half, to observe how playing the game affected their answers.

The questions in the first half of the questionnaire were designed to assess participants’ confidence and understanding of password security. They aimed to gauge participants’ perception of the strength of their own passwords, their awareness of the role password strength plays in protecting against large-scale breaches, and their knowledge of common password security concepts such as hashing and attack methods. The goal was to identify any gaps in their understanding and to establish a baseline for evaluating the effectiveness of the game in improving their knowledge of password security.

In addition to observing how participants’ answers to the previous questions changed after playing the game, the second half of the questionnaire focuses on evaluating the user’s experience with the game, particularly the new features and improvements. Participants are asked about their progress in the game, the time spent, and their overall impressions. The questions assess the effectiveness of the game’s new features, such as the hint system, improved tutorial, and the redesigned user interface. Participants were also encouraged to share any difficulties they encountered, areas for improvement, and technical issues. This aims to gather feedback on the game experience as a whole and identify potential areas for further extensions.

5.1.2 Password Strength Rating

Participants were asked to rate the strength of 25 passwords, rather than the approach of the previous study of having participants rate the strength of 12 individual passwords and compare 12 password pairs. This method allows for a much more comprehensive comparison, as each password can be evaluated against every other one, significantly increasing the number of possible combinations. This provides a richer dataset for analysis. The passwords chosen test a variety of characteristics, including those studied in Ur et al. (2016) and Seitz and Hussmann (2017). Table C.1 outlines each password and the specific characteristic it aims to test.

5.2 Questionnaire Responses

All participants participated anonymously, and no personal or demographic information was collected. However, the study was advertised primarily through University of Edinburgh Informatics channels (e.g., email and flyers), likely biasing the participant demographics towards Informatics students and staff, who may have greater prior knowledge of password security than the general population.

A notable factor affecting response quality was the change in study environment. The previous study involved in-person participation on a singular device under supervision, while this study's online distribution, while increasing reach, potentially led to lower quality questionnaire data, as indicated by more "joke" responses which had to be removed from the results, and considerable differences in gameplay duration.

5.2.1 Confidence Questions

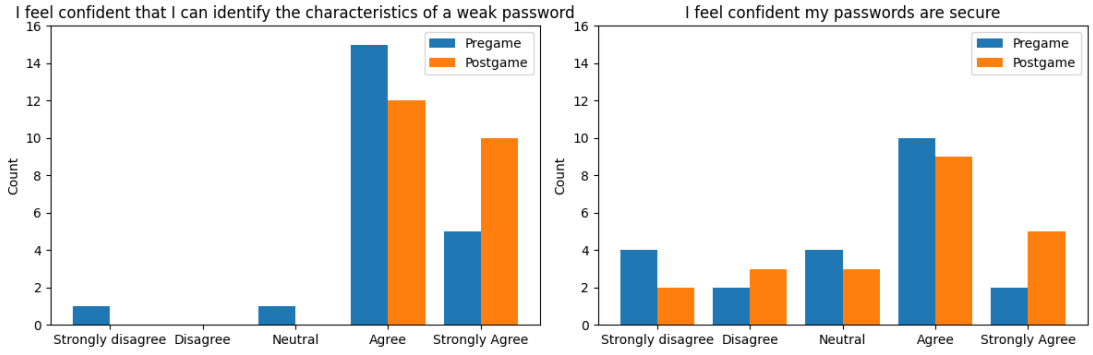


Figure 5.1: How playing the game affected confidence in identifying weak password characteristics and personal password security.

Participants were asked questions regarding their confidence in identifying weak password characteristics and the security of their own passwords (Figure 5.1).

Regarding the ability to identify weak password characteristics, last year saw an increase in average confidence from 3.36 to 4.09 out of 5. This year, the average confidence increased from an already high rating of 4.05 to 4.45. This suggests that the game

was effective at enhancing players' understanding of what makes a password weak, reinforcing the game's educational impact.

When it comes to confidence in the security of their own passwords, the average rating increased from 3.18 to 3.55. This coincides with the result in section 5.2.6 that 73% of participants reported the game helped them identify a weakness in their password practices. These results suggest that the game not only improved players' theoretical knowledge but also prompted them to reassess and strengthen their own password habits.

5.2.2 In the event of a large-scale password data breach, would the strength of your password play a role in its ability to remain secure?

Before playing the game, 36.4% of participants selected "Yes", 45.5% responded with "Maybe", and 18.2% selected "No". After playing the game, 77.3% responded with "Yes", 13.6% with "Maybe", and 9.1% with "No". In comparison, last year's version of the game showed even more pronounced improvements. Initially, 50.0% of participants selected "Yes", 40.9% selected "Maybe", and 9.1% selected "No". After playing the game, an impressive 95.5% of participants responded with "Yes", while only one participant (4.5%) selected "No". While both versions of the game successfully increased awareness of password strength's importance in data breaches, the slightly less dramatic improvement in the current version may suggest that participants entered the study with a lower baseline understanding compared to the previous year.

5.2.3 Briefly describe how you think attackers gain access to passwords after a large-scale password data breach

The expected answer to this question would mention the matching of generated hashes created through brute-force attacks that target common patterns in passwords, including dictionary attacks and/or rainbow tables.

Before playing the game, 54.5% of participants correctly identified at least one or a combination of password-cracking practices. After playing the game, this figure rose to 77.2%. This is a smaller improvement compared to last year's study, where after playing the game 90.1% of participants correctly identified one of these practices. However, both results indicate a positive impact of the game on participants' knowledge of password-cracking techniques.

5.2.4 What is a hash in the context of computer security and cryptography?

A correct response to this question would identify a hash as a unique string generated from input data, in this case, a password, through an irreversible hash function.

Before playing the game, 59.1% of participants were able to correctly identify a hash. After playing the game, this number increased to 86.7%. Last year, the results were

even more pronounced: before playing, only 50% of participants identified a hash as an encrypted string, but after playing, 95.5% provided the correct definition. Thus, while both studies demonstrate the game's effectiveness in teaching the concept of a hash, the previous year's results show a greater improvement

5.2.5 Approximately how many guesses do you think your password needs to be able to withstand for it to be considered secure

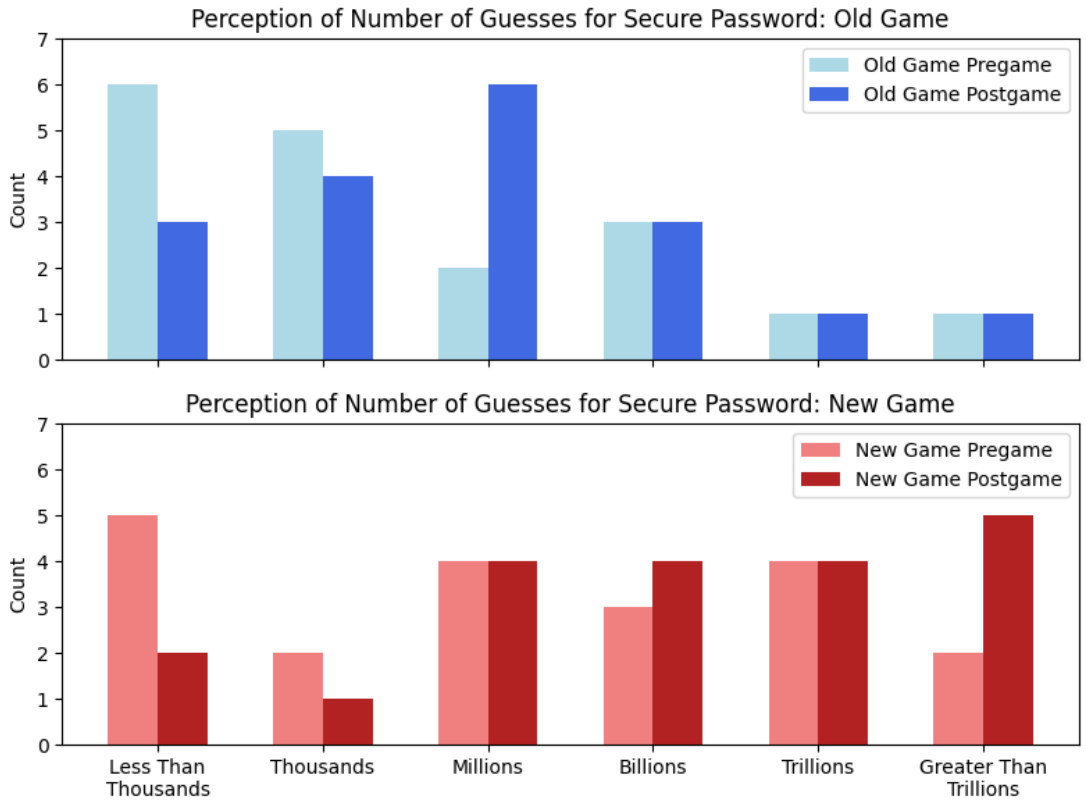


Figure 5.2: Comparison of estimate of number of guesses a secure password should withstand.

According to Ur et al. (2016), secure passwords should be able to withstand approximately 10^{14} to 10^{20} guesses to be considered secure, depending on the speed of the hashing function utilised.

Figure 5.2 compares the responses to this question with those from last year's study. Answers have been categorised into the nearest major order of magnitude.

Before playing the game, 31.8% of participants answered with a value in the thousands, and after playing the game, this decreased to 13.6%. This is a better result than in the previous study, where after playing the game, 31.8% of participants still answered with a value in the thousands or lower. For context, in the study completed by Ur et al. (2016) with 165 participants, 67% of participants estimated a value in the thousands or lower. Despite the likely demographic bias towards informatics students with pre-existing

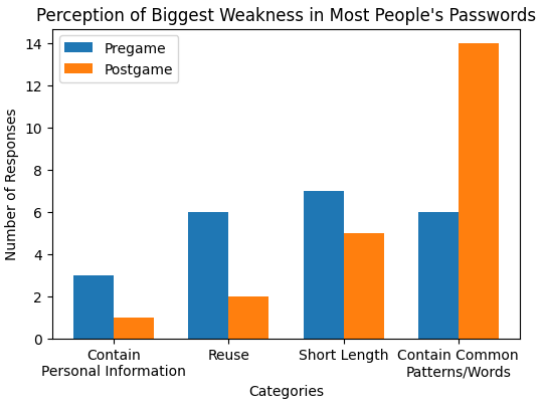


Figure 5.3: How playing the game affected participant's perception of the biggest weakness in passwords.

knowledge, the new game version significantly improved participant understanding, showing a substantial reduction in incorrect estimates compared to both the baseline and the previous study.

Furthermore, after playing, participants from this year's study selected significantly larger values. While last year only 22.7% of participants chose values in the billions or higher, this year, 59.1% selected estimates in the billions or greater. This indicates the game's UI improvements and greater emphasis on "Number of Hashes" made a significant impact on participants' practical understanding of password security scales. Although few participants reached the values recommended by Ur et al. (2016), these numbers are likely too abstract for easy comprehension. This could potentially be addressed in the future by simplifying the hash count further to the nearest order of magnitude for better comprehension (e.g., 8 billion instead of 8,127,454,186).

5.2.6 What do you think is the biggest weakness in most people's passwords?

Responses to this question were categorised into four groups: containing personal information, reusing passwords across multiple accounts, short length, and containing memorable patterns or words. Figure 5.3 displays the shift in responses before and after gameplay.

After playing the game, participants overwhelmingly identified the use of common patterns or words as the primary password weakness. While this aligns with the game's core message, it's important to acknowledge that password length and reuse are also critical security factors. The results suggest the game could more effectively emphasise the significance of these vulnerabilities.

5.2.7 Gameplay Statistics

As the game was distributed online, participants were asked to self report the amount of time they spent on the game and the level they reached.

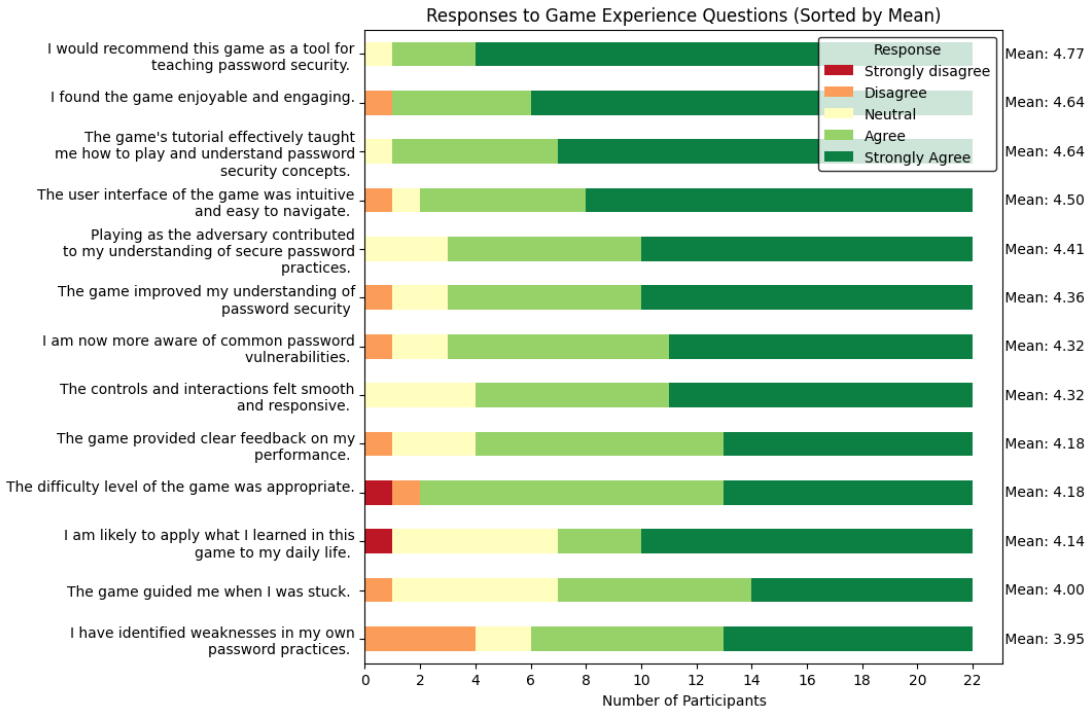


Figure 5.4: Questionnaire responses regarding game experience.

59.1% of participants beat the game, 22.7% of participants stopped at Level 5, 9.1% of participants stopped at Level 4, and 4.5% at both Level 3 and Level 2. In last year's study all participants completed the game, although it is worth noting there were no passing requirements and players could freely progress to the next level regardless of their performance.

This response may explain differences in results between the game versions. The previous in-person study likely pressured participants to complete the game. In contrast, the online distribution may have reduced this sense of obligation.

5.2.8 Game Experience Evaluation

Overall, results from this section of the questionnaire are extremely positive. For each question, a quantitative mean was calculated by assigning a numerical value to each response with "Strongly Disagree" as 1 and "Strongly Agree" as 5.

5.2.9 Comparison to Previous Study

Participants highly recommended the game as a tool for teaching password security, giving it a mean rating of 4.77. This is comparable to last year's rating of 4.81, reinforcing the game's effectiveness as a teaching tool.

Participants found the game highly enjoyable and engaging, giving it a mean rating of 4.64 out of 5. This is an improvement over last year's already positive mean rating of 4.41, indicating the game's success in teaching players in an entertaining way.

Participants felt that taking on the role of the adversary contributed to their understanding of secure password practices, giving this statement a mean rating of 4.41, compared to last year's average rating of 4.14. This suggests that the adversarial perspective continues to be an effective method for teaching password security.

5.2.10 Evaluating New Features

Participants strongly agreed that the game's user interface was intuitive and easy to navigate, giving it a mean rating of 4.50. They also rated the smoothness and responsiveness of the controls and interactions highly with a mean of 4.32. This indicates that the improvements made to the user interface and player interactions had a positive impact on the overall gameplay experience.

The game's tutorial effectively taught participants how to play and understand password security concepts, receiving a rating of 4.64. Participants agreed that the game provided helpful guidance when they were stuck, giving this statement a rating of 4.00. This indicates that the improved tutorial and hint system were successful in assisting players. The large number of neutral responses to the latter (27.2%) may also suggest that some players never encountered issues requiring hints.

Players agreed that the difficulty level of the game was appropriate and that the game provided clear feedback, rating both statements with a mean of 4.18. This suggests that the game remained appropriately challenging while ensuring players received the necessary feedback to guide their progress.

5.2.11 Identifying Password Vulnerabilities

Players agreed that the game helped them identify weaknesses in their own password practices, with an average rating of 3.95. While this is a subjective response, influenced by the varying strength of personal passwords, it still reflects a strong result. Last year, when asked in a yes or no format, 86% of participants selected ”Yes,” which is comparable to the 72.7% of participants who selected ”Agree” or ”Strongly Agree” this year.

Participants agreed that the game improved their understanding of password security and are now more aware of common password vulnerabilities, rating the statements 4.36 and 4.32 respectively. Participants are also likely to implement what they learned in their daily life, rating this statement a 4.14. This is a higher result than last year's rating of 3.91. This confirms the game's effectiveness in both educating about password security and influencing real-world behavior.

Overall, these results demonstrate that the game continues to effectively raise awareness about password security, and encourages participants to apply the knowledge gained to improve their personal practices.

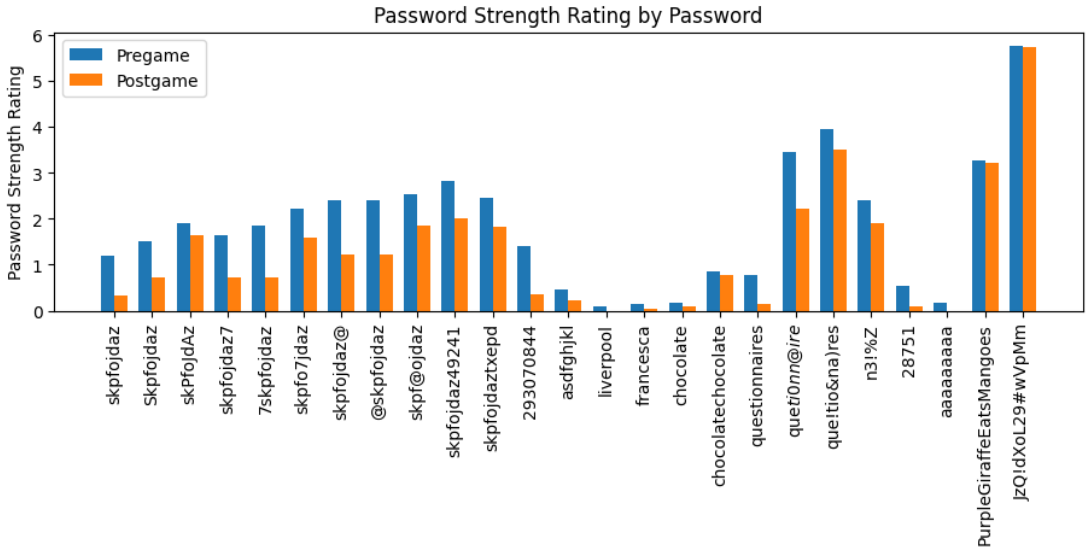


Figure 5.5: How the game affected participants' perception of strength for 25 passwords

5.2.12 Password Strength Ratings

Participants were asked to rate the strength of 25 passwords on a 7 point scale from Very Weak to Very Strong. Their responses were quantified on a scale from 0-6 where 0 is Very Weak and 6 is Very Strong. Figure 5.5 shows how the strength ratings of each password changed before and after playing the game.

One thing to note is that the average strength for all passwords decreased overall. Before playing the game, the mean strength for all passwords was 1.86, which decreased to 1.29 after playing. This indicates that, after playing, participants were more critical of the strength of all passwords.

There are some positive results, such as the decrease in the strength perception of passwords like 'questionnaires,' 'que$ti0nn@ire$,' and '293070844,' which indicates that the game successfully demonstrated the dangers of using common dictionary words, relying on common substitutions, and using only digits. There are a few results which indicate areas for potential improvement. For instance, although the strength of the passphrase 'PurpleGiraffeEatsMangoes' did not decrease significantly after playing the game, its rating remains relatively low for its complexity. This suggests that the game did not adequately convey the strength of long passphrases, likely due to players' reliance on dictionary attacks.

Similarly, the password 'n3!%Z' was rated to have a disproportionately high strength despite being only 5 characters long. This indicates that the game may not have emphasised how password length contributes to strength and the importance of experimenting with the 'Any Character' button to observe how easily a 5-character password can be cracked, even if it appears complex.

Players rated 'skpfojdaz49241' and 'skpfojdaztxepd' with similar strength ratings after playing the game. This suggests the game did not help them recognise that, while the former includes mixed characters, the latter offers more entropy due to a higher

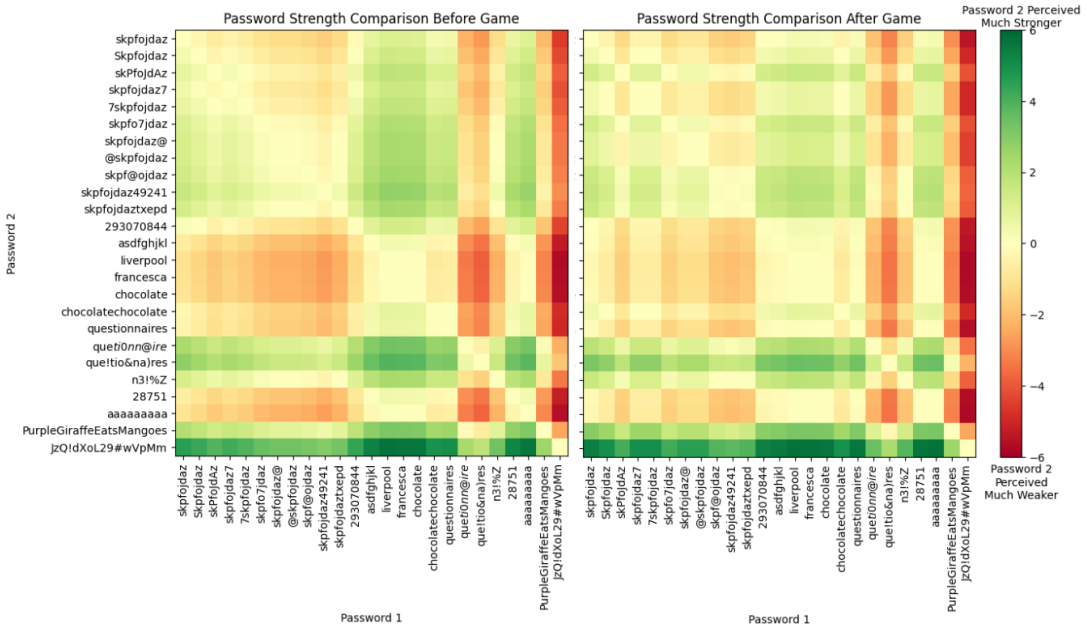


Figure 5.6: Heatmap of password pair strength before and after gameplay.

combination of lowercase letters versus numbers. The rating is also surprisingly low for a random 14-character string, which would take a computer centuries to crack, according to *zxcvbn* ((Wheeler, 2016)).

To compare the strength of the passwords relative to each other, two heatmaps were created, one for before and one for after playing the game. A green tile indicates that the password on the y-axis was perceived as stronger than the one on the x-axis, while a red tile indicates the opposite (Figure 5.6).

Comparing the before and after results reveals some interesting behavior. To observe these changes more clearly, Figure 5.7 shows the difference between these two heatmaps, allowing us to observe how the relative strength comparison between passwords was affected by the game. Blue tiles indicate that a password's relative strength increased compared to others, while red tiles indicate a decrease in relative strength.

One caveat of this visualisation is that passwords which were already perceived as very weak (including 'asdfghjkl,' 'liverpool,' 'francesca,' 'chocolate,' '28761,' 'aaaaaaaaa') were still deemed very weak after playing the game, as they could not be rated any lower. In comparison to other passwords, which on average experienced a decrease in strength, these passwords' rows on the y-axis appear blue, as their relative strength increased. This is not a true increase in strength but rather an artifact of the rating system, where the weakest passwords were capped at the lowest rating and thus appear stronger relative to the other passwords that saw a decrease.

What is promising to see is that, regarding the changes made to 'skpfojdaz', common alterations such as capitalising the first letter, adding a number or special character to the beginning or end, were perceived as relatively weaker after the game. In contrast, less predictable alterations, such as using mixed case or placing the number or special character in the middle of the password, were considered stronger. This indicates

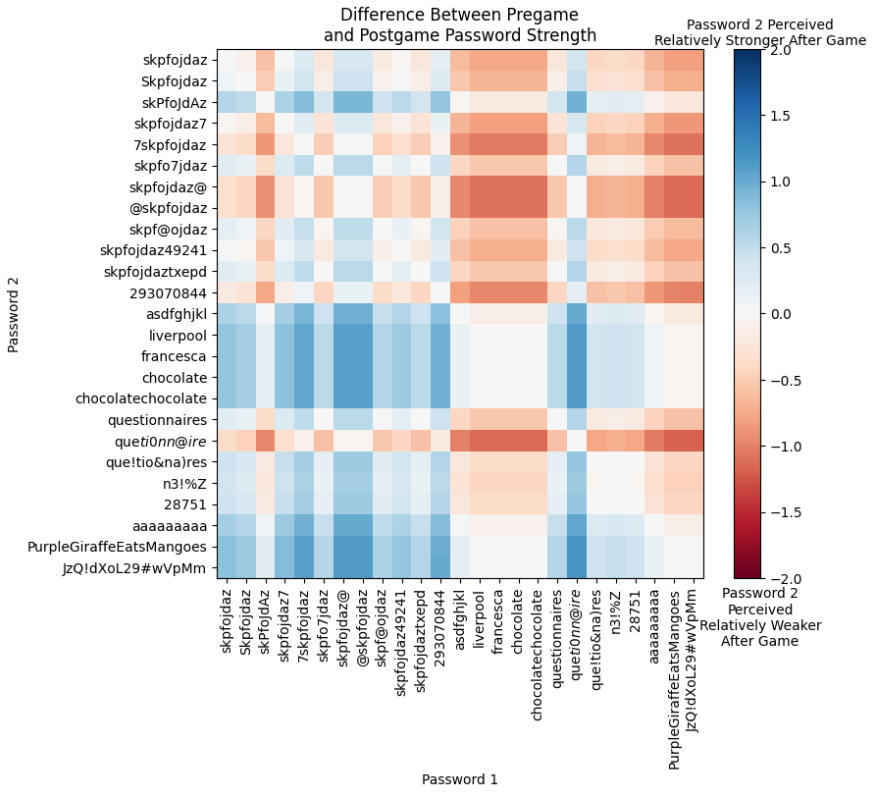


Figure 5.7: Heatmap showing difference of pregame heatmap and postgame heatmap.

that the game successfully conveyed the importance of making more complex and unpredictable changes to passwords in order to enhance their security.

It is also worth noting how the alteration of 'questionnaires' using common substitutions, 'que$ti0nn@ire$,' was rated relatively weaker compared to the same password with random substitutions, 'que!tio&na)res.' This indicates that the game helped players understand that relying on predictable substitutions-like replacing 's' with '$' or 'o' with '0'-does not significantly strengthen a password. Instead, more random and unpredictable substitutions, as demonstrated in the latter password, offer much higher security, reinforcing the game's effectiveness in teaching better password creation strategies.

5.2.13 Game Issues and Comments

Participants were asked open-ended questions about technical difficulties they encountered and suggested potential improvements. Notably, many issues raised in the previous study are no longer mentioned, including concerns about dialogue (text size and length), level time duration, unclear input validation, and the cramped user interface. This demonstrates the effectiveness of the implemented improvements in addressing key usability and design challenges identified in the earlier version of "Let's Get Cracking".

5.2.13.1 Balancing

Some participants mentioned that the game's balancing could be improved. There were a range of comments as to which levels should be made easier or harder, including some conflicting answers, showcasing how hard it is to balance a game like this for the audience. Some participants stated that rainbow tables were far too useful, sometimes providing a large proportion of the score to pass a level.

5.2.13.2 User Interface

Although the interface was praised by the majority of participants, there were a few comments on how it could be improved. Participants complained the information was scattered across the different pages, and that repeatedly pressing input every time got tiring. A potential solution would be to make the input page the default and relocate the shop and other features to a secondary screen.

5.2.13.3 History

Although the game prevents you from inputting a combination you have already tried, and the results screen shows you a list of your best combinations, some participants wished there was a feature to see your historic inputs. This would be particularly useful for struggling players as they may not remember which inputs were most successful upon retrying a level.

5.2.13.4 Game Freezing or Inability to Progress

Four participants encountered game-breaking issues, including freezes and animation bugs that obscured a crucial button, preventing further progress. While refreshing the page generally resolved these problems and allowed users to return to their level via the level select, the absence of an in-game prompt for this workaround negatively impacted the user experience.

5.2.13.5 Positive Comments

When asked if they had anything else to share about their experience with the game, all 14 participants who chose to respond praised the game, including its music, user interface, shop, and visuals. They also suggested it be promoted to schools or distributed more broadly as a valuable password education resource.

Chapter 6

Conclusion

The original ”Let’s Get Cracking” successfully demonstrated the effectiveness of an adversarial approach to teaching password security. Based on feedback from the previous study, the game was improved in several key areas, including the user interface, dialogue, tutorial, player guidance, and gameplay depth.

This study confirmed the game’s continued effectiveness in improving participants’ understanding of password security. Evidence for this includes increased confidence in identifying weak password characteristics, greater awareness of password-cracking techniques, and improved self-awareness regarding potential weaknesses in their own passwords. Notably, participants’ perception of the number of guesses required to crack a secure password increased significantly compared to the previous year, likely due to the new UI enhancements that improved information processing. The game showed similar or improved results in many key areas including its overall effectiveness as a password security teaching tool and the value of its adversarial perspective. Participants also praised the game’s new and improved elements including the enhanced UI, player guidance, and interactions.

Despite the challenges of online distribution, including varied completion rates and potentially lower data quality, the findings of this study strongly indicate that the improved version of ”Let’s Get Cracking” remains a highly effective tool for teaching password security.

6.1 Future Extensions and Improvements

6.1.1 Improved Balancing

The evaluation highlighted the need for further playtesting regarding game balancing. To maintain accessibility while accommodating players seeking a greater challenge, levels could implement a tiered system: bronze for passing scores, silver and gold for cracking even more passwords. This would allow players to choose their level of challenge.

6.1.2 Improvements to Rainbow Tables

To improve player experience with rainbow tables, their implementation should be refined. Instead of random input combination selections, levels could feature curated tables that demonstrate specific attack types that were overlooked, like 'Any' inputs or unique dictionary combinations (e.g., 8-character keyboard combinations). Players also expressed dissatisfaction with unwanted rainbow tables, as purchasing them was the only way to remove them. A 'reroll' feature, with a small fee, would provide a more flexible solution.

6.1.3 Real-Time Leaderboard

Although the fictional usernames and scores in the leaderboards were convincing, with some participants inquiring about their authenticity post-game, future iterations could feature a server with real-time player scores. This would enhance player engagement and competition.

6.1.4 Categorised Password Matching

A challenge within the game's dictionary upgrade system lies in the lack of clear feedback regarding the tangible benefits of dictionary enhancements. Players may struggle to visually correlate dictionary upgrades with a noticeable improvement in password cracking. This may result in player's overestimating the strength of passwords containing words from those dictionary upgrades. To address this issue, the game could benefit from more robust tools for categorising and visualising matched hashes. Specifically, implementing a dynamic filtering system within the matched hashes screen would allow players to toggle between different dictionary types (e.g., common words, names, keyboard patterns) and observe the corresponding passwords successfully cracked by each category. This feature would provide a direct, visual representation of how each dictionary upgrade contributes to overall cracking efficiency, enhancing the educational impact of the game.

6.1.5 Input Log

Future iterations of the game should include a log of the player's past inputs, accessible within the level. While the current system prevents duplicate inputs and displays top-performing combinations in the results page, there is no such feature within a level. This would allow players to analyse their past inputs, even when retrying a level, allowing them to refine their strategies.

Bibliography

Per Backlund and Maurice Hendrix. Educational games - are they worth the effort? a literature survey of the effectiveness of serious games. In *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, pages 1–8, 2013. doi: 10.1109/VS-GAMES.2013.6624226.

Ivan L. Beale, Pamela M. Kato, Veronica M. Marin-Bowling, Nicole Guthrie, and Steve W. Cole. Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41(3):263–270, 2007. ISSN 1054-139X. doi: https://doi.org/10.1016/j.jadohealth.2007.04.006. URL https://www.sciencedirect.com/science/article/pii/S1054139X07001759.

Polona Caserman, Katrin Hoffmann, Philipp Müller, Marcel Schaub, Katharina Straßburg, Josef Wiemeyer, Regina Bruder, and Stefan Göbel. Quality criteria for serious games: Serious part, game part, and balance. *JMIR Serious Games*, 8(3):e19037, Jul 2020. ISSN 2291-9279. doi: 10.2196/19037. URL http://games.jmir.org/2020/3/e19037/.

Otávio Filho, Barbara Maia, Amabily Reis, Vitor Santos, Thiago Felix, Pedro Cotrim, Pedro Veloso, and Angela Silva. Assessment of the potential application of the game plague inc. as a playful tool for teaching science and biology. *Revista Contemporânea*, 3:15127–15139, 09 2023. doi: 10.56083/RCV3N9-088.

Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pages 523–537, 2012. doi: 10.1109/SP.2012.38.

Fedwa Laamarti, Mohamad Eid, and Abdulmotaleb El Saddik. An overview of serious games. *Int. J. Comput. Games Technol.*, 2014, jan 2014. ISSN 1687-7047. doi: 10.1155/2014/358152. URL https://doi.org/10.1155/2014/358152.

Karl Maton. Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education*, 24:8–22, 04 2013. doi: 10.1016/j.linged.2012.11.005.

Jakob Nielsen. 10 usability heuristics for user interface design. *Nielsen Norman Group*, 1994. URL

https://www.nngroup.com/articles/ten-usability-heuristics/. Accessed: 2024-04-26.

Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why people (don't) use password managers effectively. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 319–338, Santa Clara, CA, August 2019. USENIX Association. ISBN 978-1-939133-05-2. URL https://www.usenix.org/conference/soups2019/presentation/pearman.

Prabaharan Poornachandran, M. Nithun, Soumajit Pal, Aravind Ashok, and Aravind Ajayan. Password reuse behavior: How massive online data breaches impacts personal data in web. In H. S. Saini, Rishi Sayal, and Sandeep Singh Rawat, editors, *Innovations in Computer Science and Engineering*, pages 199–210, Singapore, 2016. Springer Singapore. ISBN 978-981-10-0419-3.

George E. Raptis, Christina Katsini, Andrew Jian-lan Cen, Nalin Asanka Gamagedara Arachchilage, and Lennart E. Nacke. Better, funner, stronger: A gameful approach to nudge people into making less predictable graphical password choices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445658. URL https://doi.org/10.1145/3411764.3445658.

L. A. Robinson, I. J. Turner, and M. J. Sweet. The use of gamification in the teaching of disease epidemics and pandemics. *FEMS Microbiology Letters*, 365(11):fny111, 2018. doi: 10.1093/femsle/fny111. URL https://doi.org/10.1093/femsle/fny111.

Tobias Seitz and Heinrich Hussmann. Pasdjo: Quantifying password strength perceptions with an online game. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, OzCHI '17, page 117–125, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450353793. doi: 10.1145/3152771.3152784. URL https://doi.org/10.1145/3152771.3152784.

Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Designing password policies for strength and usability. *ACM Trans. Inf. Syst. Secur.*, 18(4), May 2016. ISSN 1094-9224. doi: 10.1145/2891411. URL https://doi.org/10.1145/2891411.

Wayne Summers and Edward Bosworth. Password policy: The good, the bad, and the ugly. pages 1–6, 01 2004.

Penelope Sweetser and Peta Wyeth. Gameflow: A model for evaluating player enjoyment in games. *Computers in Entertainment*, 3:3, 07 2005. doi: 10.1145/1077246.1077253.

Emin Tatlı. Cracking more password hashes with patterns. *IEEE Transactions on Information Forensics and Security*, 10:1–1, 08 2015. doi: 10.1109/TIFS.2015.2422259.

Harshal Tupsamudre, Rahul Wasnik, Shubhankar Biswas, Sankalp Pandit, Sukanya

Vaddepalli, Aishwarya Shinde, C. J. Gokul, Vijayanand Banahatti, and Sachin Lodha. Gap: A game for improving awareness about passwords. In Stefan Göbel, Augusto Garcia-Agundez, Thomas Tregel, Minhua Ma, Jannicke Baalsrud Hauge, Manuel Oliveira, Tim Marsh, and Polona Caserman, editors, *Serious Games*, pages 66–78, Cham, 2018. Springer International Publishing. ISBN 978-3-030-02762-9.

Blase Ur, Jonathan Bees, Sean M. Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Do users' perceptions of password security match reality? In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3748–3760, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. doi: 10.1145/2858036.2858546. URL https://doi.org/10.1145/2858036.2858546.

Daniel Lowe Wheeler. zxcvbn: Low-Budget password strength estimation. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 157–173, Austin, TX, August 2016. USENIX Association. ISBN 978-1-931971-32-4. URL https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentatio

Appendix A

Participants' Information Sheet

Participant Information Sheet

Project title:	Let's Get Cracking: Leveraging Gameplay from an Adversarial Perspective to Teach Password Security Concepts
Principal investigator:	Borislav Ikonomov
Researcher collecting data:	Dylan Lins Brasiliense Drucker
Funder (if applicable):	

This study was certified according to the Informatics Research Ethics Process, reference number 987681. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

For this undergraduate research project, the student leading the research is Dylan Lins Brasiliense Drucker. The principal supervisor for this project is Borislav Ikonomov.

What is the purpose of the study?

The purpose of this study is to design and evaluate the effectiveness of a game which educates users about offline password database attacks and characteristics of strong/weak passwords. Participants will be asked to play the game and answer two questionnaires, one before and one after the game. The questionnaires include questions to gauge user's existing knowledge of password security, their perception of password strength characteristics, and their thoughts on the game experience. The study will observe how playing the game affected user's responses between the two questionnaires and evaluate the effectiveness of the game.

Why have I been asked to take part?

The target group of participants for this research project are participants with varying degrees of familiarity regarding offline password database attacks.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until you complete the second questionnaire, without giving a reason.

After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

You will be asked to play the game for up to 20 minutes and answer two questionnaires (one before and one after playing the game) which should take no more than 10 minutes per questionnaire. If you are playing the game on the researcher's device, your gameplay may be recorded. The questionnaire will include questions evaluating your existing knowledge of offline password database attacks, your perceived strength of different password characteristics, and your thoughts on the game and its effectiveness as a teaching tool.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

Are there any benefits associated with taking part?

No.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher Dylan Lins Brasiliense Drucker.

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Dylan Lins Brasiliense Drucker (s2077148@ed.ac.uk), or the principal investigator Borislav Ikonomov (borislav.ikonomov@ed.ac.uk). If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Dylan Lins Brasiliense Drucker (s2077148@ed.ac.uk).

General information.

For general information about how we use your data, go to: edin.ac/privacy-research

Appendix B

Participants' Consent Form

Participant number:_____

Participant Consent Form

Project title:	Let's Get Cracking: Leveraging Gameplay from an Adversarial Perspective to Teach Password Security Concepts
Principal investigator:	Borislav Ikonomov
Researcher collecting data:	Dylan Lins Brasiliense Drucker
Funder (if applicable):	

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1.	I agree to having my gameplay recorded (Only required if playing the game on the researcher's device).	☐ **Yes**	☐ **No**
2.	I allow my data to be used in future ethically approved research.	☐ **Yes**	☐ **No**
3.	I agree to take part in this study.	☐ **Yes**	☐ **No**

Name of person giving consent	Date dd/mm/yy	Signature
_____	_____	_____
Name of person taking consent	Date dd/mm/yy	Signature
_____	_____	_____

THE UNIVERSITY of EDINBURGH informatics

Appendix C

Questionnaire

Let's Get Cracking (Year 2) - User Study

Thank you for your interest in participating in my user study!

You will be asked to answer a questionnaire and play the game. This should take approximately **30-40 minutes.**

The game is hosted online and can be played in-browser, but if you prefer you may also download a copy of the game if you are on Windows.

The questionnaire is split into two parts. The first part should be answered before playing the game, and the second should be answered after.

First, please read the participant information sheet below.

* Indicates required question

Participant Information Sheet

Project Title:

Let’s Get Cracking: Leveraging Gameplay from an Adversarial Perspective to Teach Password Security Concepts

Principal Investigator:

Borislav Ikonomov

Researcher Collecting Data:

Dylan Lins Brasiliense Drucker

This study was certified according to the Informatics Research Ethics Process, reference number 987681. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

For this undergraduate research project, the student leading the research is Dylan Lins Brasiliense Drucker. The principal supervisor for this project is Borislav Ikonomov.

What is the purpose of the study?

The purpose of this study is to design and evaluate the effectiveness of a game which educates users about offline password database attacks and characteristics of strong/weak passwords. Participants will be asked to play the game and answer two questionnaires, one before and one after the game. The questionnaires include questions to gauge user's existing knowledge of password security, their perception of password strength characteristics, and their thoughts on the game experience. The study will observe how playing the game affected user's responses between the two questionnaires and evaluate the effectiveness of the game.

Why have I been asked to take part?

The target group of participants for this research project are participants with varying degrees of familiarity regarding offline password database attacks.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until you complete the second questionnaire, without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove

individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

You will be asked to play the game for up to 20 minutes and answer two questionnaires (one before and one after playing the game) which should take no more than 10 minutes per questionnaire. If you are playing the game on the researcher's device, your gameplay may be recorded. The questionnaire will include questions evaluating your existing knowledge of offline password database attacks, your perceived strength of different password characteristics, and your thoughts on the game and its effectiveness as a teaching tool.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

Are there any benefits associated with taking part?

No.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher Dylan Lins Brasiliense Drucker.

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be

sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Dylan Lins Brasiliense Drucker (s2077148@ed.ac.uk), or the principal investigator Borislav Ikonomov (borislav.ikonomov@ed.ac.uk). If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Dylan Lins Brasiliense Drucker (s2077148@ed.ac.uk).

General information.

For general information about how we use your data, go to: edin.ac/privacy-research

1. **By proceeding with the study, I agree to all of the following statements:** *

I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.

My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.

I consent to my anonymised data being used in academic publications and presentations.

I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Tick all that apply.

- [ ] I allow my data to be used in future ethically approved research.
- [ ] I agree to take part in this study.

2. If you are playing on the researcher's device, do you agree to having your gameplay recorded? *

Mark only one oval.

- ◯ I am not playing on the researcher's device
- ◯ Yes
- ◯ No

Please answer the following section BEFORE playing the game.

3. Please rate how strongly do you agree with the following statements. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I feel confident that I can identify the characteristics of a weak password	◯	◯	◯	◯	◯
I feel confident my passwords are secure	◯	◯	◯	◯	◯

4. In the event of a large-scale password data breach, would the strength of your password play a role in its ability to remain secure? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

5. Briefly describe how you think attackers gain access to passwords after a large-scale password data breach *

_____

6. What is a hash in the context of computer security and cryptography? (It is ok if you do not know) *

_____

7. Approximately how many guesses do you think your password needs to be able to withstand for it to be considered secure (in the context of a password database leak)? *

_____

8. What do you think is the biggest weakness in most people's passwords? *

_____

9. Evaluate the strength of the following passwords. *

Mark only one oval per row.

	Very Weak	Weak	Somewhat Weak	Neutral	Somewhat Strong	Strong
skpfojdaz						
Skpfojdaz						
skPfoJdAz						
skpfojdaz7						
7skpfojdaz						
skpfo7jdaz						
skpfojdaz@						
@skpfojdaz						
skpf@ojdaz						
skpfojdaz49241						
skpfojdaztxepd						
293070844						
asdfghjkl						
liverpool						
francesca						
chocolate						
chocolatechocolate						
questionnaires						
que$ti0nn@ire$						
que!tio&na)res						
n3!%Z						
28751						

aaaaaaaaa						
PurpleGiraffeEatsMangoes						
JzQ!dXoL29#wVpMm						

Please now play the game! The game has 5 levels, but you are not required to finish the game. After playing the game, return to this survey and press "Next".

The game can be played at the following link: https://s2077148.itch.io/lets-get-cracking

You can play the game in browser by pressing the "RUN GAME" button. Alternatively, you may download and unzip the executable at the bottom of the page (improved performance but requires a Windows machine).

While the game is functional on mobile, I highly reccomend using a laptop/desktop/DICE machine as the performance on mobile can be slow.

Please answer the following section AFTER playing the game

10. Which level of the game did you reach *

Mark only one oval.

- Level 1
- Level 2
- Level 3
- Level 4
- Level 5
- I finished the game

11. Approximately how long did you spend playing the game? *

Mark only one oval.

- < 10 minutes
- 10-20 minutes
- 20-30 minutes
- 30-40 minutes
- > 40 minutes

12. In the event of a large-scale password data breach, would the strength of your password play a role in its ability to remain secure? *

Mark only one oval.

- Yes
- No
- Maybe

For any of the short-answer questions that you have already answered before playing the game, if your answer remains unchanged, simply write 'same.'

13. Briefly describe how you think attackers gain access to passwords after a large-scale password data breach *

_____

14. What is a hash in the context of computer security and cryptography? (It is ok if you do not know) *

_____

15. Approximately how many guesses do you think your password needs to be able to withstand for it to be considered secure (in the context of a password database leak)? *

_____

16. What do you think is the biggest weakness in most people's passwords? *

17. Please rate how strongly do you agree with the following statements. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I feel confident that I can identify the characteristics of a weak password	⬭	⬭	⬭	⬭	⬭
I feel confident my passwords are secure	⬭	⬭	⬭	⬭	⬭
The game improved my understanding of password security	⬭	⬭	⬭	⬭	⬭
I have identified weaknesses in my own password practices.	⬭	⬭	⬭	⬭	⬭
The user interface of the game was intuitive and easy to navigate.	⬭	⬭	⬭	⬭	⬭
The game's tutorial effectively taught me how to play and understand password security concepts.	⬭	⬭	⬭	⬭	⬭

I found the game enjoyable and engaging.					
I am likely to apply what I learned in this game to my daily life.					
I would recommend this game as a tool for teaching password security.					
The game guided me when I was stuck.					
The difficulty level of the game was appropriate.					
The game provided clear feedback on my performance.					
The controls and interactions felt smooth and responsive.					
I am now more aware of common password vulnerabilities.					

Playing as the adversary contributed to my understanding of secure password practices.					

18. Evaluate the strength of the following passwords. *

Mark only one oval per row.

	Very Weak	Weak	Somewhat Weak	Neutral	Somewhat Strong	Strong
skpfojdaz						
Skpfojdaz						
skPfoJdAz						
skpfojdaz7						
7skpfojdaz						
skpfo7jdaz						
skpfojdaz@						
@skpfojdaz						
skpf@ojdaz						
skpfojdaz49241						
skpfojdaztxepd						
293070844						
asdfghjkl						
liverpool						
francesca						
chocolate						
chocolatechocolate						
questionnaires						
que$ti0nn@ire$						
que!tio&na)res						
n3!%Z						
28751						

aaaaaaaaa						
PurpleGiraffeEatsMangoes						
JzQ!dXoL29#wVpMm						

19. Were there any parts of the game that you found confusing or difficult to understand?

20. What aspects of the game could be improved upon?

21. Please describe any bugs or technical issues you encountered with the game

22. Is there anything else you would like to share about your experience with the game?

C.1 Additional Figures

C.2 Password Strength Rating

Password	Characteristic Tested
skpfojdaz	Lowercase random characters
Skpfojdaz	First letter capitalised
skPfoJdAz	Mixed case
skpfojdaz7	Single digit at the end
7skpfojdaz	Single digit at the start
skpfo7jdaz	Single digit in the middle
skpfojdaz@	Single symbol at the end
@skpfojdaz	Single symbol at the start
skpf@ojdaz	Single symbol in the middle
skpfojdaz49241	Appended numbers
skpfojdaztxepd	Appended lowercase characters
293070844	Numeric-only
asdfghjkl	Common keyboard pattern
liverpool	Common place
francesca	Common name
chocolate	Common dictionary word
chocolatechocolate	Repeated word
questionnaires	Long dictionary word
que$ti0nn@ire$	Long dictionary word with common substitutions
que!tio&na)res	Randomised substitutions
n3!%Z	Short, mixed complexity
28751	Short numeric-only
aaaaaaaaa	Repeated single character
PurpleGiraffeEatsMangoes	Passphrase-style
JzQ!dXoL29#wVpMm	Long, mixed complexity

Table C.1: Passwords and their tested characteristics

Appendix D

Let's Get Cracking Levels

D.1 Level Progression

Level	New Mechanic(s)	Additional Requirements	Passing Score
1	Character Input	None	300,000
2	Upgrades Shop (Speed Upgrades, Rainbow Tables)	Min Length: 8	100,000
3	Word Input, Dictionary Upgrades	1 lowercase, 1 number	50,000
4	None	1 uppercase	30,000
5	None	1 special character	1,000

Table D.1: Game Levels, Mechanics, Requirements, and Passing Scores