PiXi: An Approach to Nudge Secure Password Creation

by

Shengqian Wang

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Science in Computer Science

Faculty of Science
University of Ontario Institute of Technology (Ontario Tech University)
Oshawa, Ontario, Canada

April 2023

THESIS EXAMINATION INFORMATION

Submitted by: **Shengqian Wang**

Master of Science in **Computer Science**

PiXi: An Approach to Nudge Secure Password Creation

An oral defense of this thesis took place on June 15th, 2023 in front of the following examining committee:

Examining Committee:

Research Supervisor	Julie Thorpe Faculty of Business and Information Technology
Research Co-supervisor	Amirali Salehi-Abrari Faculty of Business and Information Technology
Examining Committee Member	Patrick Hung Faculty of Business and Information Technology
Thesis Examiner	Pejman Mirza-Babaei Faculty of Business and Information Technology

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

Abstract

Passwords, a first line of defense against unauthorized access, must be secure and memorable. However, people often struggle to create secure passwords they can recall. To address this problem, we design ***P**assword **i**nspiration by e**X**ploring **i**nformation (PiXi)*, a novel approach to nudge users towards creating secure passwords. PiXi is the first of its kind that employs a password creation nudge to support users in the task of generating a unique secure password themselves. PiXi prompts users to explore unusual information right before creating a password, to shake them out of their typical habits and thought processes, and to inspire them to create unique (and therefore stronger) passwords. PiXi's design aims to create an engaging, interactive, and effective nudge to improve secure password creation. We conducted a user study ($N = 238$) to compare the efficacy of PiXi to typical password creation. Our findings indicate that PiXi's nudges do influence users' password choices such that passwords are significantly longer and more secure (less predictable and guessable).

Keywords: Passwords; Authentication; Nudging; User Studies;

Author's Declaration

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.
I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize the University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.
The research work in this thesis was performed in compliance with the regulations of the Research Ethics Board under REB Certificate number #16688.

_____ Shengqian Wang

Statement of Contributions

I hereby certify that I am the sole author of this thesis. I have used standard referencing practices outlined in the *Publication Manual of the Association for Computing Machinery (2021)* to acknowledge ideas, research techniques, or other materials that belong to other researchers. Furthermore, I hereby certify that I am the primary contributor of this thesis, including the related works, user studies, system implementation, security analysis and usability analysis. Some text of this thesis is borrowed from the paper [80], which was written in collaboration with my supervisors, Julie Thorpe and Amirali Salehi-Abrari.

Acknowledgements

Firstly, I want to say thank you to my thesis advisors, Dr. Amirali Salehi-Abari and Dr. Julie Thorpe, for your unwavering support, and extensive knowledge, and I look forward to working with you again in my upcoming PhD program.

Next, I want to say thank you to my thesis committee member, Dr. Patrick Hung, and external examiner, Dr Pejman Mirza-Babaei for your detailed work and valuable suggestions.

Finally, my deepest appreciation goes to my wife, Fangyi for her patience and unwavering support, my parents for their unconditional love and assistance, assistance from my lab members, eye doctors from LakeRidge Health and SunnyBrook Hospital, and all of my friends.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

F Frequency

PH PiXi-Hints

ASA absolute scrolling activity

TSA total scrolling activity

Kw Keyword

Kws Keywords

Pop Popular

COCA 5000 COCA 5000 top words list

Chapter 1

Introduction

Passwords have been utilized since ancient times as a method to establish one's authenticity. In the recent decades, with the remarkable advancements in internet technology, passwords have gained widespread adoption in information security, effectively safeguarding digital assets. As the internet continues to evolve, they have become an essential component for activities such as website or application logins, and accessing databases.

Despite decades of development in password authentication alternatives, the majority of websites and applications still require passwords for authentication. Unfortunately, due to time constraints, labor costs, lack of expertise, or apathy, a significant number of people reuse passwords [22] or choose simple, predictable passwords (e.g., birthdays or names). These insecure password choices do not necessarily imply users' lack of intelligence or motivation, but may simply be due to their lack of inspiration or guidance when confronted with a blank password field. Frustration can also arise from unhelpful password policy suggestions, such as "please use special characters to make your password stronger" or "make your password longer to create a strong password." Unfortunately, few solutions exist to support users with creating secure passwords in such helpless situations.

While password managers, when used with random password generators, can improve

password security [72, 73, 86], some users are not comfortable using them. Even some official organizations (e.g., governments, enterprises, etc.) do not typically recommend their use for sensitive accounts due to the fear of the password manager vault being compromised. Password manager users still require a strong master password as the key to encrypt the stored passwords in the vault. Therefore, users, regardless of employing password managers or not, still require support for creating secure and memorable passwords for (at least) these sensitive accounts.

1.1 Motivation

In recent years, the topic of nudging has been actively researched. Nudging is a promising strategy to encourage people to make better or more desirable choices. It has been applied in a variety of domains including education, ethics, social context, health, finance, energy savings, privacy, and security. Nudging studies have proved that nudges can successfully influence people's decisions by minor and inexpensive interventions. A real-life example is when you go to big shopping malls, you have to go through a long trail full of small items or candies before you check out. Those items are a kind of nudge by altering the visual arrangement of the options provided. Computer security experts have also been investigating nudges to encourage users to create more secure and memorable passwords. However, most nudges in password systems apply a one-size-fits-all approach and primarily focus on password meters [1, 50, 73], which use rigorous password standards to convince users to adjust their passwords to satisfy specific requirements. Unfortunately, many users find effective password meter designs to be annoying [73]. Other approaches suggest modifications to the initial password to make it secure [31, 37, 44]. However, these systems can be vulnerable to attacks that predict common passwords and their modifications [55]. These approaches also haven't been studied as true nudges, but rather systems that require the user

to accept the suggested modification, or request a new one. Limiting users' free choice in passwords may come with unanticipated usability problems.

To address these shortcomings, we design ***P**assword **i**nspiration by e**X**ploring **i**nformation (PiXi)*, a novel approach to nudge users towards creating secure passwords. PiXi is the first of its kind that employs a password creation nudge to support users in the task of generating a unique password themselves. PiXi prompts users to explore unusual information right before creating a password, to shake them out of their typical habits and thought processes, and to inspire them to create unique (and therefore stronger) passwords. We implemented and evaluated a web-based version of PiXi to answer our research questions: (Q1) Which nudges in PiXi are most effective, and do they influence users' password choices? (Q2) Does our PiXi system support users to create more secure passwords? (Q3) How usable is our PiXi system, and how can its usability be improved?

To investigate these research questions, we conducted a user study ($N = 238$) to evaluate the security and usability of passwords generated by users of PiXi.

1.2 Use Case

Users often face challenges when they must create strong and memorable passwords manually, especially for critical accounts like financial accounts or the master password for password managers. Often, users resort to password reuse, weak and easily guessable passwords, or even writing them down. To address this, we aim to provide a system that allows users to generate better memorable, unpredictable, and robust passwords in such situations. Our PiXi system offers a solution where users can select their preferred information to create inspire the creation of strong and memorable passwords.

1.3 Threat Model

In our threat model, we consider an attacker that is motivated to gain unauthorized access to user accounts to compromise sensitive user information. We primarily focus on offline and online attacks performed by unknown adversaries. We do not consider known adversaries that might employ attacks that rely on personal information or physical access. For offline attacks, attacks can be executed upon gaining control of the website's database. They may leverage various password guessing methods, including dictionary attacks, brute force attacks, and rainbow table attacks, to guess (a.k.a., "crack") user passwords stored in the database.

Our research objective revolves around reducing susceptibility to common password threats. To achieve this, we encourage users to create more complex and unique passwords using PiXi.

1.4 Contributions

Our contributions and findings include: (i) The design of PiXi—a novel approach to nudging users to create secure passwords. (ii) Security analysis of passwords produced with PiXi. Our study results demonstrated that PiXi successfully influences users' password choices, such that passwords are longer and more secure (less guessable) than a control group using a typical password creation process. (iii) Usability analysis of the PiXi system. Our study results indicated that PiXi shows promising usability in terms of user perception and memorability. (iv) Analysis of nudge efficacy of PiXi. Our findings identified that some nudges are more effective than others and that PiXi's combination of nudges do influence users' password choices. (v) We present a threat model identifying the vulnerability of weak passphrases containing COCA top 5000 words, urging users to avoid such

common words. However, we acknowledge that the data from Amazon MTurk is insufficient for definitive conclusions, underscoring the need for further research to enhance our understanding of passphrase security and associated threat models.

1.5 Thesis Organization

The remainder of the thesis is structured as follows: Chapter 2 summarizes related work in nudging and password creation nudges. Chapter 3 introduces the PiXi methodology, Chapter 4 shows the user studies implemented for PiXi, Chapter 5 shows the results from the user studies, and Chapter 6 is the thesis conclusion and future work, followed by the bibliography and appendices.

Chapter 2

Related Work

We describe the password security overview in Section 2.1, password attacks forms in Section 2.2, nudging using cybersecurity examples in Section 2.3, then review literature on nudges at the time of password creation in Section 2.4.

2.1 Password Security Overview

For decades, traditional alphanumeric or textual passwords have been the most widely used authentication methods. However, they come with significant drawbacks. Simple passwords are vulnerable to attacks, while complex ones are difficult to remember and often lead to reuse and writing down issues. As a result, researchers have explored various solutions [12, 66, 67, 82] to improve alphanumeric passwords or create new alternatives to replace them. Figure 2.1 displays all the existing authentication categories and some examples. Unfortunately, each of these methods has its flaws. Biometrics can be expensive and cannot be recovered if lost. Phone-based or token-based authentications require hardware support, while the password space and security of graphical passwords still require investigation [2, 54, 68, 69, 75]. Bonneau's framework [8] emphasizes that each method has its

limitations, and none can completely replace alphanumeric passwords.

Category	Schemes
Alphanumeric	
Password managers	LastPass, 1Password, Chrome/Firefox/Safari browser
Proxy	URRSA, Impostor
Federated	OpenID, Microsoft Passport, Facebook Connect
Graphical	PassPoints, DAS, BDAS, GeoPass, VIP, RouteMap, S3PAS
Protocol	FIDO/FIDO2
VR environment	3D password
Digital objects	ObPwd
Cognitive	GrIDsure, Weinshall
Paper tokens	OTPW, S/KEY
Video Password	Video-passwords
Hardware tokens	RSA SecurID, Blizzard authenticator
Phone-based	2-step authentications, Microsoft authenticator, OTP over SMS
Biometric	Finger/palm vein, iris/voice/facial recognition

Table 2.1

2.2 Password Attacks

Password attacks are often classified into two categories: online and offline attacks.

Online attacks. An online attack occurs when an adversary attempts to guess the victim's password through web applications or software portals. However, modern authentication systems can make it very difficult for the attacker as they typically freeze or lock the victim's account after a certain number of failed login attempts or honeywords have been detected during the guessing process [85]. This limited number of attempts makes it a significant challenge for the attacker to successfully guess the password.

Figure 2.1: Thresholds for online and offline attacks [30].

Offline attacks. Offline attacks occur when an adversary gains access to the back-end system and the database is exposed. In such cases, the passwords in the database are usually hashed and salted, which makes it challenging for the attacker to crack the database, then leading him to use different password cracking tools (e.g., John the Ripper [47], and Hashcat [43]), or state-of-the-art machine learning techniques [3, 36, 48].

According to Florencio et al. [30], a secure password should be capable of withstanding up to 10^6 for online guesses and 10^{14} for offline guesses, which is shown in Figure 2.1.

2.3 A Survey of Nudges

Nudging is a promising strategy to encourage people to make better or more desirable choices. Examples of nudging include framing, priming, and visualizations [15]. Thaler and Sunstein [62] defined nudges as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives". Nudging has been applied in a variety of domains including education [9], ethics [6], social context [39], health [45, 52, 77], finance [9, 13, 16, 41, 61], energy savings [14, 21], privacy [1], and security [20]. Nudging studies have proved that nudges can successfully influence people's decisions by minor and inexpensive interventions [26]. Hansen et al. [35] redefined nudges according to the two human thinking forms (System 1 & System 2) and the transparency proposed by Kahneman's framework [40, 64]. They expanded the human dual process theories which contains two modes of thinking,

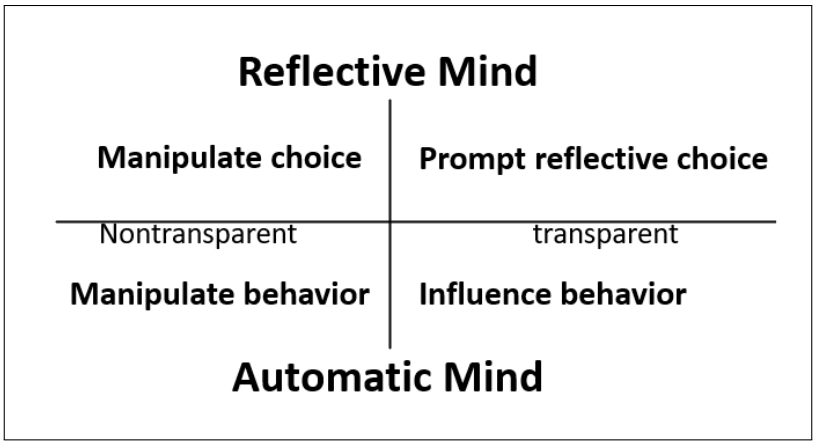


Figure 2.2: The 4 categories of nudges, adapted from Hansenand Jespersen [35].

System 1 (automatic mind) and System 2 (reflective mind) into 4 categories shown in Figure 2.2: 1) System 1, manipulating behavior (e.g., changing the order of WiFi list to promote stronger WiFi), 2) System 1, influencing behavior (e.g., painting illusions of speed bumps on streets to encourage careful driving), 3) System 2, manipulating choice (e.g., offering unattractive options to the set of choices to make the certain choice more attractive), 4) System 2, prompting reflective choice (e.g. providing real-time feedback to evaluate password strength). Humans prefer to think and make decisions via autonomic processing with minimum effort, which occasionally leads to making unwise decisions.

Throughout the years, computer security experts and administrators have been investigating nudges to encourage secure behaviors [88]. Seitz et al. [56] sought to persuade users to pick stronger passwords by providing multiple password suggestions using the decoy effect, which uses one unpleasant choice as a decoy to make the other option seem more appealing. The study's findings concluded that an effective nudge should provide only the best suggestion with more transparent and perceivable explanation. Von Zezschwitz et al. [78] tried to enhance the effectively utilized password space for Android unlock patterns by presenting background visuals and animations throughout the password formation process. The experimental study result indicated that their nudge had very limited influ-

ence when a substantial proportion of users were unaware or confused about the nudge, such as background images or animations changing, and the existence of "counter-nudges" caused by the user's own strong habits. Similarly, Briggs et al. [10] changes order of color and menu to encourage users to choose a more secure Wi-Fi. The authors indicated that the nudge effect is reduced when users have different decision-making styles, and personal preferences have significant influence on security decisions.

2.3.1 Hybrid Nudges

According to Hansen et al.'s work [34], Renaud et al. [53] defined simple nudge is defined as "a simple intervention that using system 1 (automatic mind) only". However, in many real-world circumstances, simple nudges may not handle the issues of pre-existing habitual behaviors, or strong personal preferences when it delivers messages mainly through System 1 without supporting System 2 [60]. In order to address the above issues, Zimmermann et al. [88] proposed hybrid nudge, which is defined as "a more comprehensive intervention combining a simple nudge and information provision involving both systems". Hansen et al.'s [34] examined the effectiveness of simple nudges and hybrid nudges in promoting the use of strong passwords. They reported that the simple nudge has little influence on the strength of a password, and cannot offset habitual password choice behaviors. By comparison, the hybrid nudge reduced the uncertainty, highlighted the benefits, and simplified the internal cost–benefit analysis when each intervention plays a significant role, leading to the hybrid nudge as a whole achieving an outstanding outcome.

2.3.2 Personalized Nudges

The majority of nudges we see today adhere to the "one-size-fits-all" approach and are built for the 'average user' [16, 50, 84]. While a one-size-fits-all nudge is easy to implement, it

is inefficient to protect most users away from dangerous behaviors or even have a negative impact on some users [71] when users are significantly diverse in terms of cognitive styles, educational backgrounds, expertise, or personal preferences. The first concept of personalized nudge was proposed by Sunstein el at. [59]. He argues that personalized nudge would "reduce the problems posed by one-size-fits-all approaches" if it collects enough data. Then Thaler et al. [63] introduced "choice engines", which use different kinds of heterogeneity data to generate personalized recommendations. Mills et al. [46] summarized prior research and offered the choice and delivery personalization framework. "choice personalization uses personal information to determine which choice to nudge the decision-maker towards, and delivery personalization uses personal information to determine which nudging method to apply to the decision-maker. Peer et al. [50] explored the delivery personalization in password creation processes shown in Figure 2.3, using individual differences in decision-making styles to determine the best online password nudge for each user. They found that personalized nudges provided more promising results, increasing nudges' efficacy up to four times compared to one-size-fits-all nudges.

2.4 A Survey of Password Creation Nudges

Nudging techniques have been employed, with varying degrees of success, to enhance the security of both graphical and text passwords. Throughout this review, we categorize the types of nudges employed using the framework of Caraban et al. [15]. As part of this review, we identified a number of papers previously not considered/named as nudges in other surveys (e.g., [27, 37, 42, 44, 49, 65, 79].

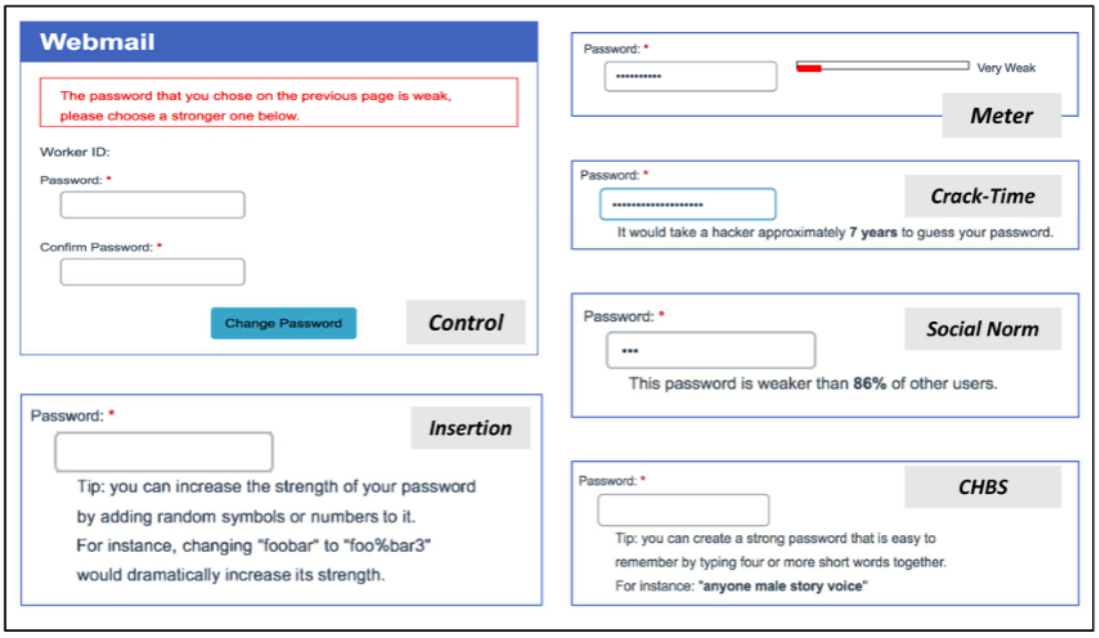


Figure 2.3: The password creation design of Peer et al. [50]. The nudges are shown in different formats based on users own preferences.

2.4.1 Graphical Passwords

Graphical passwords are a type of authentication that involve remembering picture(s) or parts of pictures instead of a word. Graphical passwords can be categorized as [7] 1) Pure-recall, requiring users to draw a secret using images or patterns (e.g., Draw-A-Secret(DAS) [38] shown in Figure 2.6, BDAS [27] shown in Figure 2.7) 2) Cued-recall(e.g., PassPoints [82] shown in Figure 2.5, CCP [19], and GeoPass [66, 70] shown in Figure 2.4) require users remember and target specific locations within an image. 3) Recognition-based (e.g., PassFaces [12], Story [23], and VIP [24]) asks users to memorize a portfolio of images during password creation, and then recognize their images from among decoys to log in. To improve graphical password security, many nudging techniques have been proposed and evaluated.

PassPoints is a widely recognized example of a graphical password [82]. Instead of using traditional text-based passwords, PassPoints employs a single image with five click-

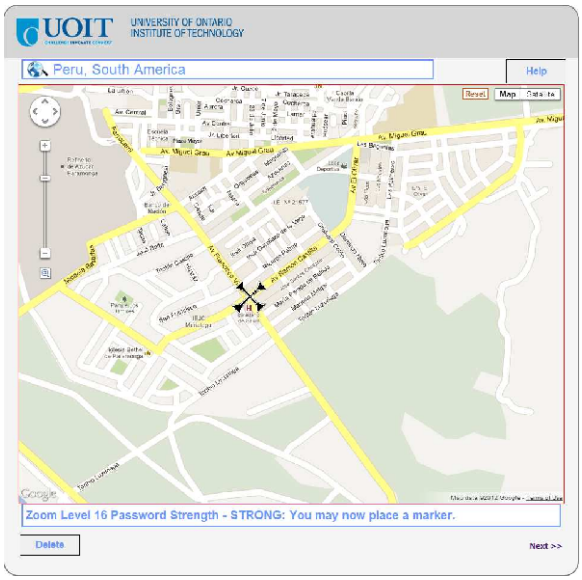


Figure 2.4: The user interface of GeoPass. Users need to select a location on a map as their passwords.

points. Users select a sequence of their preferred points to form their password during registration, and then repeat the same sequence by clicking the points in the correct order during login. This process is similar to selecting a PIN, but with the added benefit of personalized points that are more difficult for attackers to guess. Although PassPoints has some limitations, such as restricted password space due to the images and screen resolution, as well as easily identifiable hot-spot points, its benefits outweigh its drawbacks. PassPoints has served as a basis for more sophisticated graphical password systems, inspiring researchers to explore this area further.

Following this work, Chiasson et al. [19] proposed Cued Click Points (CCP), a click-based graphical password authentication scheme that adopted some features of PassPoints, Passfaces, and Story. Compared to PassPoints, Users can create passwords using points (coordinates) of a sequence of images instead of a singe image. The study results indicated that the login success rate could reach 96%, and the mean login time is around 6.0 seconds only. The questionnaire score results also reveal that participants can easily create

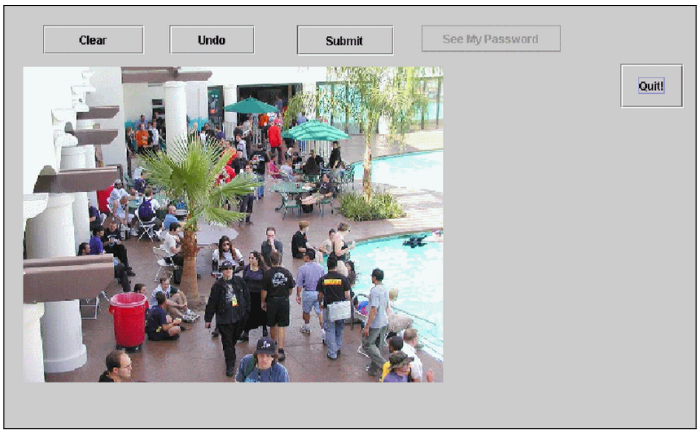


Figure 2.5: In PassPoints, users need to choose a point on an image as the password.

(8.2/10), memorize (7.2/10), and use (8.3/10) CCP. However, the study's small sample size (24 participants) and the questionable potential password space between $2^4 3$ to $2^5 4.5$ when hot-spot images and coordinates exist raise concerns.

A number of systems [42, 49, 65] based on PassPoints [82] can be considered a facilitate (hiding) and reinforce (subliminal priming) nudge, the interface is shown in 2.5. They aim to nudge users away from common patterns by presenting the background image differently at the password creation. Image presentations reveal the background image slowly so the user will perceive it differently. Since these nudges temporarily hide certain options (making them harder to reach), it can be categorized as a facilitate (hiding) nudge. As some parts of the image will be initially revealed and exposed to the user for a longer time, it also involves a reinforce (subliminal priming) nudge. Studies found simple image presentations produced different distributions in user-chosen click points, suggesting that it can significantly impact user's choice without reducing usability [49, 65]. Presenting the image such that it reveals starting from the least salient parts, ending with the most salient parts has been found to have different impacts on users password strength depends on users' cognitive style and background image [42].

Background Draw-A-Secret (BDAS) [27] may have been the first attempt to nudge

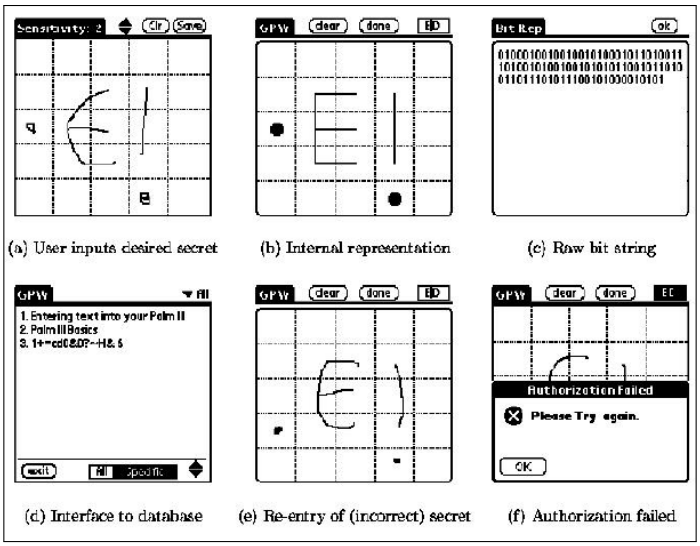


Figure 2.6: The complete steps to complete an authorization of DAS. During the registration, users need to draw shape as their passwords. When login, users need to re-draw the shapes.

users away from typical patterns during graphical password selection. It presents users with a background image, on which they need to draw their graphical password. Its background image should evoke the “salience bias", which refers to the fact that users are more likely to focus on information that is more prominent. The visual element of the background image aims to facilitate the creation of different graphical passwords than if the background image was not present. It can also be considered as a reinforce (subliminal priming) nudge. Findings indicated that BDAS did significantly increase indicators associated with password strength, (e.g., stroke count and length). Zezschwitz et al. [79] used nudging to help users create stronger patterns on Android mobile devices. They designed a variation based on background images (similar to BDAS, but using the Android pattern as the basis instead of Draw-A-Secret). While the lab-based evaluation of the system was very promising, the field study involving 496 users on MTurk platform yielded only small effects. The study results revealed that users selected a more diverse set of longer patterns when background images were present.

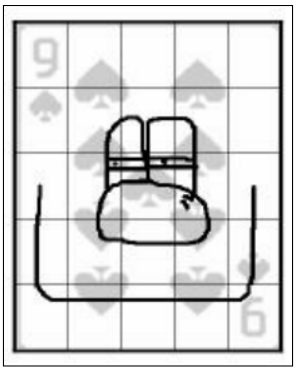


Figure 2.7: Compared to DAS, the only difference is BDAS added a background image to nudge or motivate users create unique draws.

Persuasive Cued Click Points (PCCP) [18] can be considered a facilitate (suggesting alternatives) nudge where users have to select from a point within a randomly positioned view-port (all other options are not available). If desired, they can use the shuffle button to change the view-port to another random location. The system was found to have minimal usability impact compared to other cued-recall systems; however, it does hamper a user's free choice if they wish to select a specific point — in this case the user would need to "shuffle" many times in the hopes of being able to select their desired choice. To implement PCCP as a true nudge, it would need to permit a user to override the view-port to select any point of their choosing. The researchers conducted a study to compare the performance of PCCP with CCP, PassPoints (PP), and the control condition. The study found that PCCP had similar success rates (94%, 98%, 96%) and login times (18s, 12s, 10s) to the other schemes. However, the usability of PCCP was not evaluated as the researchers did not provide any information regarding participants' preferences for PCCP compared to CCP or other systems.

Priming, a kind of implicit memory has been explored and examined as a viable approach to help users recognize system-assigned graphical passwords [25].

Then, Castelluccia et al. proposed and examined a new authentication scheme called MooneyAuth [17]. MooneyAuth using priming to train users memorize mooney images

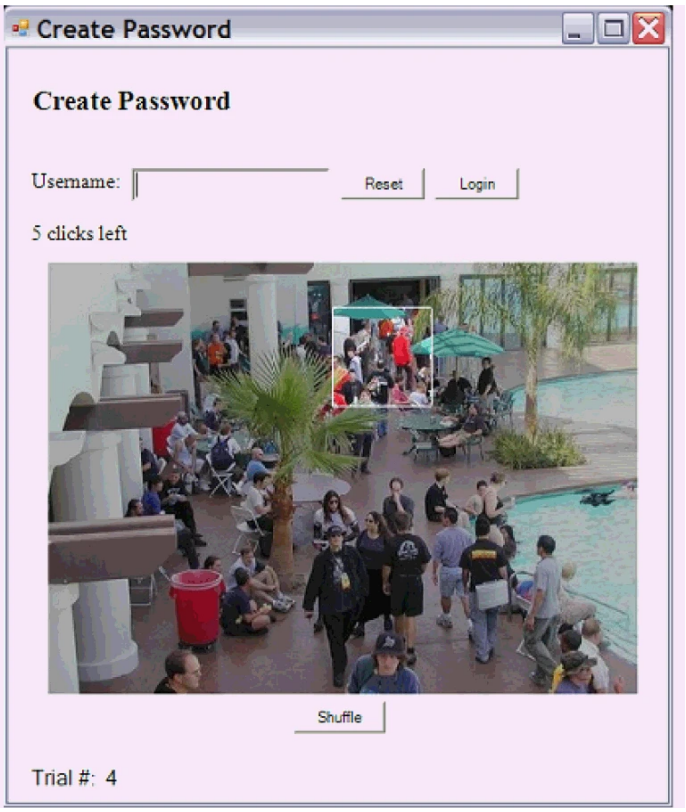


Figure 2.8: Compared to CCP, a randomly positioned view-port is used to limited users choices to encourage users to choose random points on images.

(thresholded, two-tone image showing a single object.) of system-assigned images, users need to recognize the a mooney image, and relabel it during the login process. The authors conducted an online user study with 70 participants to investigate if the participants can memorize the images in a long term period (8.5 months). The results showed that that around 75% participants can memorize mooney images and relabel the images after 264 days. The The main limitation of MooneyAuth is the complexity of mooney images and different human recognition abilities. When the original images with rich information or complex patterns, the mooney images will be very hard to recognize. Besides, elders may have trouble in recognizing those mooney images due to personal health conditions (Ex. eye disease, brain issues, and etc).

Another priming format called presentation effect has also been explored by Thorpe

et al. [65] using PassPoints as the foundation, but added simple image presentations before password creation. The study contains 3 session and recruited 34 participants from the researchers' university to examine whether different background image presentations can influence user choice. The modified PassPoints was drawing the curtain over a background image to reveal it slowly (from left-to-right or right-to-left), which were found to produce a different distributions in user-chosen click points. The results suggested that presentation effect can significantly impact user's choice (80% of users choice have been influenced) without reducing usability when the majority (80%) of participants accepted image presentation.

Katsini et al. [42] examined image presentations effect based on a cued-recall graphical authentication scheme, used for creating gesture-based passwords using different background images as cues. The authors conducted 2 studies contains 36 participants from the university, and asks participants create passwords using different gestures. The background images are varies based on conditions, they are either in normal image with rich information or Mooney image with few information. The results suggested the image complexity has a strong correlation to users passwords strength, simple images lead to stronger passwords, and users tend to use saliency points as references to create passwords.

Zezschwitz et al. [79] focused on how to use nudging to help users create strong patterns as passwords on Android mobile devices. They designed an unlock pattern concept based on background images. While the lab-based evaluation of the system was very promising, the field study involving 496 users yielded only small effects. The studying results revealed that users selected a more diverse set of longer patterns when background images were present.

Graphical passwords have been in development for several years, but the market has yet to see many viable products due to the significant cost of changing authentication schemes and concerns about their password space and security. Bonneau et al. [8] con-

ducted a comparative evaluation of all exisiting authentication schemes. The evaluation involved 25 criteria related to usability, deploy-ability. The aim was to find potential replacements for text passwords. However, the study found that each scheme had limitations and could not replace traditional passwords completely. The authors suggested that combining different schemes could provide a better solution. They also highlighted the importance of considering different criteria weights and adding more criteria to obtain unbiased evaluation results. As a result, there is a need to find a practical solution that incorporates graphical password features into text passwords.

2.4.2 Text Passwords

The most straightforward way to nudge strong password selection is to suggest a random password to the user. This is a form of facilitate (default) nudge, if implemented so the user has a choice to accept the random password or not. However, memorability is a significant problem for system-assigned random passwords [83]. Attempts have been made to help improve memorability of secure system-assigned passwords, e.g. through passphrases. Shay et al. [57] explored the possibility of using system-assigned random passphrases instead of system-assigned passwords with random characters, by encouraging users to imagine a scene that find connection among each passphrase, and re-type those passphrases as their passwords when they need to login. However, the results of the user study were not positive. The study found that the usability and memorability of system-assigned random passphrases did not improve significantly compared to system-assigned passwords with random characters. Only 48.5% of participants were able to recall the passphrases after 2-3 days, indicating that they were not significantly easier to remember.

Password Managers. Password managers are a popular technique that can help alleviate the burden of managing multiple accounts and passwords for users. These tools make it easy to handle multiple accounts, and password generators can create strong and random

passwords that fit any password policy. Even for those users who are successfully nudged to choose a random password and store it in a password manager, it is recommended to avoid using password managers for sensitive accounts (e.g., email, financial, workplace, etc.) [33]. For these reasons, finding other ways to nudge users towards creating secure passwords remains of interest.

Versipass, a password manager that integrates key elements of password managers and cued graphical passwords (ImagePassTiles), was proposed by Stobert et al. [58] in order to address the issues of password reuse and memorability. To use Versipass to generate and store passwords, users must first manually add account details for the new website, select hashing and salt values for the new password, and then Versipass generates a password based on those values. Finally, users must change the password on the target website to match the password generated by Versipass. When using Versipass to login, users must click the PMLogin bookmarklet (a component of Versipass), and then complete a task to locate the correct cue in an image. The entire process is quite complicated and time-consuming. The authors conducted a preliminary user study and cognitive walk-through to evaluate the effectiveness of Versipass. However, the results were not positive. Although there were only five participants in total, all of whom were from the same lab, many reported that the system was difficult to learn and use. The authors did not clarify where the cues and images came from or the mechanism for generating them. Additionally, there has been no follow-up study of Versipass. These findings suggest that a complex password manager may not assist users in storing and remembering passwords, but rather discourage them from using the system.

Pearman et al. conducted a study to investigate the reasons behind the low adoption and effective use of password managers and random password generators by users. The study involved on-campus interviews with 30 participants, who were recruited through both online and offline approaches using a codebook to evaluate the results. According to the

interview, 12 participants used built-in password managers, 5 used independent password managers, 9 never used a password manager, and 4 were unclear and not mentioned in the paper. The study confirmed the findings of previous research by Alkaldi and Renaud [5], which highlighted factors such as lack of awareness or knowledge, having few passwords, and security concerns as reasons for low adoption and effective use of password managers. The study also revealed new issues, including participants' lack of understanding about how passwords are stored in password managers and confusion around the purpose of certain functions (e.g., "Remember Me"), which could lead to lose trust and a reluctance to use password managers.

Zibaei et al. [72] conducted a study on the effectiveness of nudges in encouraging users to adopt randomly generated passwords in Chrome, Firefox, and Safari. The researchers found that users in Safari (61%) were more likely to accept randomly generated passwords when prompted by a better nudge browser with attractive and interesting visual effects. However, strict password policies may not be helpful in encouraging users to accept these passwords. While password managers can be helpful, users still need to create a strong and memorable master password for their password manager, which can be challenging. If the master password is lost, all other passwords stored in the password manager become inaccessible, rendering their security measures meaningless. Government of Canada [33] has suggested that passwords for sensitive accounts like Email and financial should not be stored in password managers. Therefore, users must find a way to create a strong master password that they can remember.

Password Meters. One way to nudge users towards creating stronger text passwords is through password meters, which employ a confront (creating friction) nudge to provide real-time feedback on password strength. The hope is that its feedback will nudge users to revise their passwords to be stronger, but has limited effectiveness on not "important" accounts [29]. Ur et al. [73] designed and examined 14 password meters, each of them using

different visual effects and nudge techniques. The results show that some password meters with visually appealing interfaces (half-score and third-score groups) and strict prompts can significantly enhance security against a strong adversary. When no meter was used, 46.7% of passwords were cracked, compared with 39.4% in the baseline (an insignificant improvement). However, there was a significant improvement in the half-score (26.3% cracked) and third-score groups, where 26.3% and 27.9% were cracked respectively.

Very similar to Ur et al. [73] work, Zimmermann et al. [87] conducted a comprehensive literature review and a user study using several password meters with different visual effects and nudges through Amazon MTurk to examine what factors impact password meters effectiveness. The results suggested that hybrid password meters (in combination with a feedback nudge and additional guidance) are the most effective when they encourage users to choose more secured and memorable passwords.

Vance et al. [76] designed and evaluated a password meter focused on fear (reducing the distance) nudge to help users choose stronger passwords. They conducted a user study collaborated with Socwall.com and finally 437 participants data were collected. There are 4 conditions in total, 1) Control: shows no feedback during passwords creation, 2) Static fear appeal treatment: Displays static/fixed fear appeal about password security, 3) Password Strength Meter: A standard password meter. and 4) Interactive Fear appeal: Displays static/dynamic fear appeal about password security based on users' inputs. The results indicated that interactive fear appeal treatment shown in Figure 2.9 substantially increased password security by at least 49.5 years on average, over the control, static fear appeal, and interactive password strength meter treatments.

Dupuis et al. [28] also investigated the utilization of fear to promote the adoption of secure passwords and examined the subsequent long-term emotional effects. The study involved 811 valid participants recruited through Amazon MTurk and contains 4 different conditions. Each participant was required to watch videos based on their assigned con-

dition. 1) Control condition: Participants viewed a peaceful video that did not bring any emotional changes or convey any specific messages, 2) Threat appraisal condition: Participants watched a video demonstrating the significant negative consequences of compromised passwords and the vulnerability associated with weak passwords, 3) Coping appraisal condition: Participants viewed a video offering several tips on how to enhance password security, 4) Threat + coping appraisal condition: Participants in this group first watched the Threat appraisal video, followed by the Coping appraisal video. The findings of the study indicated that fear could motivate users to select secure passwords in the short term. However, it was also observed that this approach led to long-term negative emotional effects (e.g., anxiety, worried, frightened, and scared). Consequently, future studies should explore and analyze the trade-off associated with employing fear as a motivator for password hygiene behaviors.

Egelman et al. [29] investigated the impact of password meters on users' password choices. Initially, the authors conducted a two-week laboratory study to examine the effectiveness of two password meters with different nudges. Participants were asked to change their existing passwords, and the study included three conditions: control (no password meters), existing motivator (EM, a traditional password meter that shows the strength of passwords), and peer pressure motivator (PPM, a password meter that shows the strength of passwords relative to all other users on the system, considered as social influence (enabling social comparisons) nudge). The results revealed that both password meters encouraged users to select stronger (entropy of control: 49.3 bits, EM: 60.8 bits, PPM: 64.9 bits) and longer (nine to ten characters) passwords. Subsequently, the authors conducted a follow-up study to test if various scenarios could also impact password meters' effectiveness. In the follow-up study, participants were asked to create new passwords for unimportant accounts, with or without password meters, and then log back in after two weeks. The study recruited 541 participants through Amazon MTurk. The results indicated that password meters did

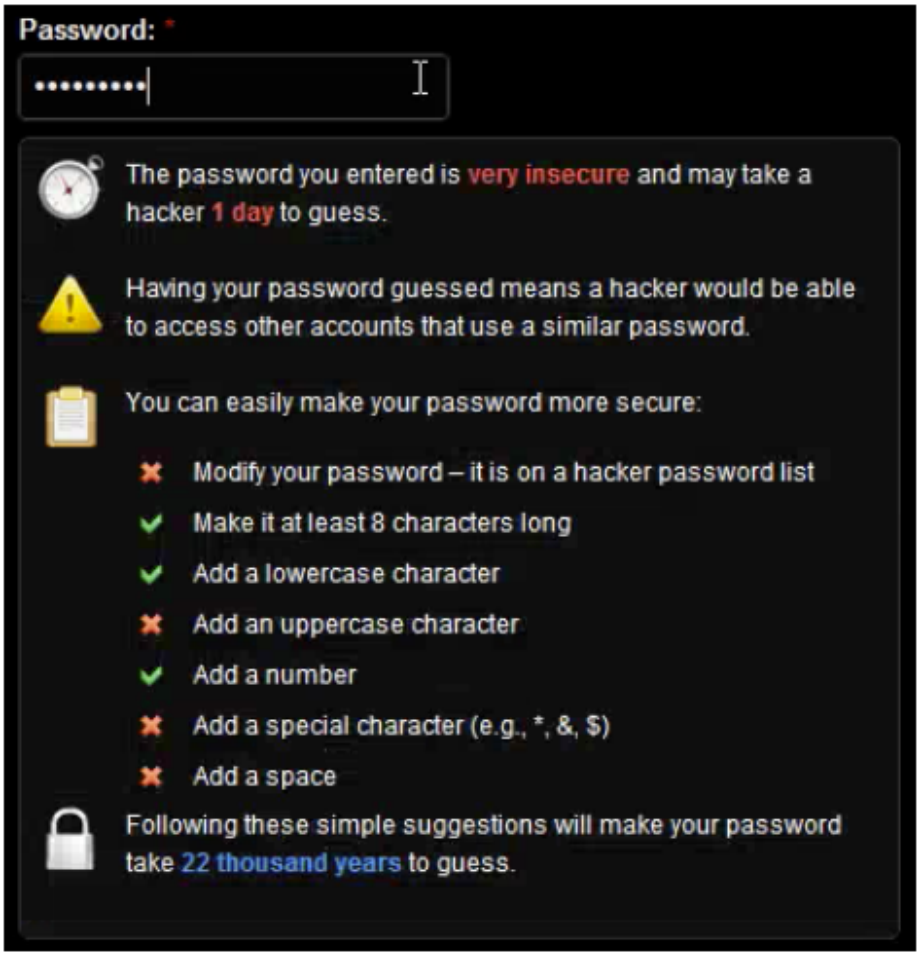


Figure 2.9: The screenshot of interactive fear appeal (fear nudge) of Vance et al.'s work [76].

not encourage users to create stronger passwords for unimportant accounts when password strength was similar across different conditions (median entropy: 41.4, median length: 8). Additionally, a significant number of participants (63.8%) tended to reuse weak passwords.

The finding indicates that strict nudge prompts can be more effective at creating friction. Unfortunately, many participants of these successful meters (38-39%) found the meter annoying. One possible reason is that password meter nudges create friction, but do not facilitate users in coming up with an acceptable secure password.

Golla et al. (2018) [32] aimed to investigate the impact of password meters with differ-

ent visualization designs on users' password choices. The authors conducted a user study with five conditions, including one control condition and four password meters utilizing different elements, and recruited 342 participants from their campus. Participants were asked to register a new account in a simulated scenario (for a university account). The study results revealed no significant difference among the password meters in the resulting password strength using ZXCVBN [81] password guesses (log_{10} = 7.94 and log_{10} = 8.00 across the five conditions). In comparison to Egelman et al.'s work [29], the authors found that the password meter "high score," which use social influence (enabling social comparisons) nudge, resulted in significantly higher password creation time (control: 38 seconds, high score: 45 seconds). The results suggested that password meters with significantly different visualizations have little to no effect on password strength and users satisfactory. Also, non-standard password meters may have an adverse effect on creation times and have reduced usability.

System-Assigned Passwords. Another way to encourage users use more secured passwords is system-assigned passwords. Compared to user chosen passwords, system-assigned passwords are often more stronger. However, most of the system-assigned passwords are almost impossible to memorize, which lead users to reuse or write down passwords. To solve the memorability issues, some researchers implemented different kinds of nudges to system-assigned passwords.

Shayetal. [57] explored the possibility of using system-assigned random passphrases instead of system-assigned passwords with random characters, by encouraging users to imagine a scene that find connection among each passphrase, and re-type those passphrases as their passwords when they need to login. However, the results of the user study were not positive. The study found that the usability and memorability of system-assigned random passphrases did not improve significantly compared to system-assigned passwords with random characters. Only 48.5% of participants were able to recall the passphrases after

2-3 days, , indicating that they were not significantly easier to recall, and system-assigned passwords are ineffective without the implementation of nudges.

Al-Ameen and colleagues [4] investigate whether using verbal cues could improve users' ability to memorize system-assigned passwords. They conducted two user studies included 52 and 54 participants, respectively, from the researchers' university. In the first study, which lasted for two sessions over one week, participants were divided into three groups and asked to memorize five keywords. The control group had no verbal cues, while the second group (TextV) was given textual verbal cues, and the third group (GraphicV) was given both verbal cues and related images (a reinforce nudge (or subliminal priming)) to the keywords. The results suggested that the TextV and GraphicV groups (shown in Figure 2.10) had significantly higher login success rates (94.23% and 96.15%, respectively) compared to the control group (61.54%). In the second study, the authors evaluated the usability of GraphicV only. The results suggested that GraphicV was sufficiently memorable for nearly all of the participants. 72% of participants had a 100% login success rate, and 94% had at least a 90% success rate. However, the authors reported that in the second study, participants performed an average of 25 login attempts.

Password Modification Systems. Other approaches employ a facilitate (suggesting alternatives) nudge that suggests modifications to the initial password to make it secure. These systems could be implemented as a nudge, where the user can dismiss the suggested modification if desired, but the following implementations only provided the option to "shuffle" to generate a new suggestion if they were unhappy with the one the system provided.

In Persuasive Text Passwords (PTP) [31], after a user creates a password, random characters are added at random positions to increase the complexity of the password. The user can then use a shuffle button to repeat this process until they are satisfied with the final password. Initial memory tests (in the same session as creation) showed promise, but password creation time was significantly increased as users often use the shuffle button (18

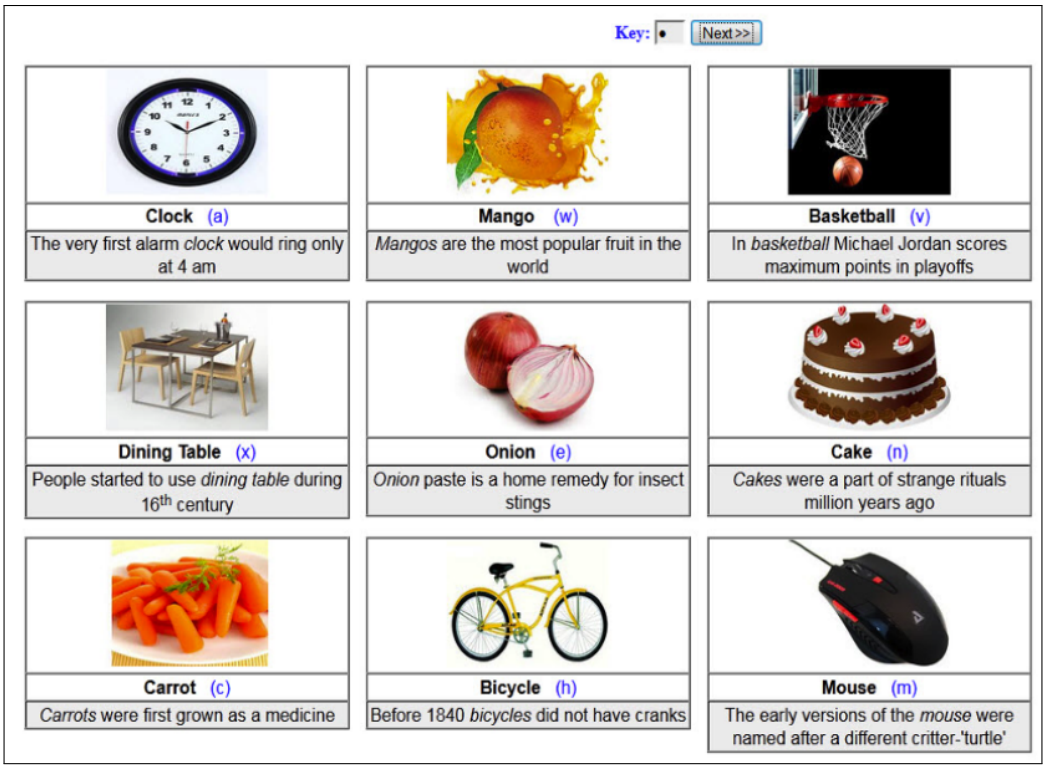


Figure 2.10: The login page of GraphicV. Each keyword is associate with a corresponding image to help users memorize the keyword.

times on average) before finding an acceptable system-suggested password. Houshmand et al. [37] proposed that the suggestions be built using probabilistic context free grammars. Both PTP and Houshmand's modification systems were found to be vulnerable to Guided Brute Force attacks [55].

MacRae et al. [44] proposed and designed PassMod, shown in Figure 2.12, a password recommending system similar to PTP, but with a different strategy to create strong password suggestions. PassMod uses semantic grammar generation algorithms to extract weak words from user input and replace them with strong words. Users can use the "New Suggestions" button to shuffle the system recommendations until the password is desirable. A small feasibility study indicates PassMod successfully strengthened participants' passwords and users could remember their suggested passwords very well (74.5% recall over 7-8 days in

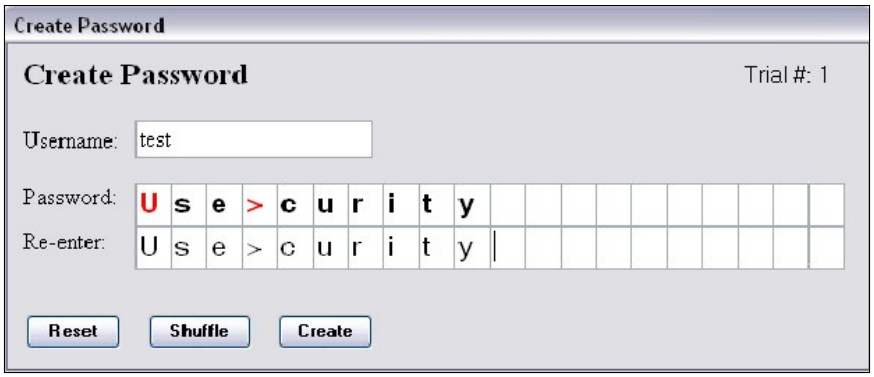


Figure 2.11: Persuasive Text Passwords (PTP) add random characters at random positions to users' chosen default passwords in order to increase the password strength. Users can click "shuffle" button to get new suggestions. However, the modified passwords are almost unable to memorize.

session#3), and most participants (71.4%) were satisfied with the system suggestions in the post questionnaire when they knew the suggestions made their passwords more secured. These results suggest that PassMod shows promise; however, the study only involved 43 participants, all of whom were students. It is difficult to compare the results with typical passwords as there was no control group, so further study is needed to determine its efficacy.

Summary. In conclusion, the existing approaches to nudge stronger passwords are either (a) default nudges to use a randomly generated password (typically employed as a nudge in password managers [86]), (b) confront (friction) nudges that aim to increase user's awareness of their chosen password's weakness, with no facilitation in coming up with a new password (e.g. password meters [74]), and (c) facilitate (suggesting alternatives) nudges that suggest modifications to a user's initially weak password to make it secure (e.g., [31, 37, 44]. Our approach with PiXi is entirely different than previous text password nudges; we aim to facilitate user's password creation without suggesting alternatives, but instead using the following set of nudges immediately prior to password creation: (i) facilitate (positioning and suggesting alternatives) to help users explore an unusual path (and set of selections) through the PiXi system, (ii) confront (throttling mindless activity) to ensure

Figure 2.12: Compared to PTP, PassMod provides more secured and memorable password suggestions based on users' inputs using chunks. If they are not satisfied with the result, they can also click "New Suggestion' to get new ones.

users consider their PiXi selections, and (iii) reinforce (subliminal priming) to make the user's PiXi selections more prominent and easily accessible at the time the user is attempting to conceive a new password. The goal of this combination of nudges is to create an engaging, interactive, and effective nudge to impact password creation.

Chapter 3

Methodology

The PiXi system aims to nudge users to create stronger passwords, by engaging them with an interactive system for information exploration (e.g., search and select a sequence of keywords) before they create their typical alphanumeric passwords. Instead of limiting user choice, PiXi exposes its users to some unusual and randomized information to shake them out of their typical password creation patterns and get them thinking about new possibilities for their passwords.

Users interact with PiXi just before password creation through:

Introduction. The introduction page (see Figure 3.1) offers a brief description of the system via a YouTube video tutorial that guides users through the step-by-step process of PiXi. It illustrates how to select a category and a keyword. A short paragraph and a simple animation are also included on the introduction page to assist users in selecting keywords. The users can bypass this page by clicking the "Next" or "X" buttons, and they can always return to it by clicking on the question icon located at the interface.

Category Selection. The category page (see Figure 3.2) contains three possible content categories for user selection: images, books, or movies. The order of categories is randomly shuffled for each user. This page contains a *facilitate (positioning) nudge* [15] as it positions

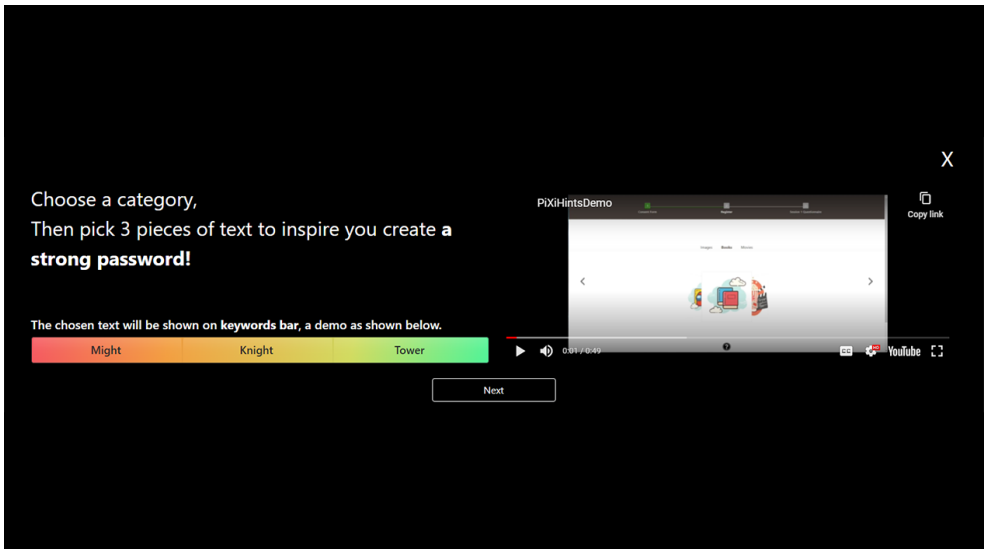


Figure 3.1: Introduction Page. The introduction page provides a video tutorial and instructions to users on how to use the system.

a category in the center more prominently to nudge the user to select it. The user still has the option to choose another category. Once a category is selected, the user is directed to an item page (see below).

Item Page. The item page contains a set of 20 randomly selected items (e.g., book covers, movie covers, or images) from the selected category.[1] A user can click on the image of interest to view the accompanying content. If she is unsatisfied with any of the results, she can continue her search using the search bar with auto-complete feature. Comparable to Google Search, the search mechanism employs a search bar to locate results containing the search terms. The main difference is that users can interact with search results (e.g., click a book cover to see the content, and pick any interesting text as keywords). Each user see different items when seeds are generated using their IDs. The algorithm for generating seeds for randomly displaying search results is shown in Appendix B. The maximum number of items per page is limited to 20 in order to maintain a neat and organized user interface.

1 In our PiXi prototype configuration, there are around 6 million possible items (all categories); 20 items are randomly selected from the pool of possible items and shown to the user, for their selected category. However, the number of possible items could be configured to be much larger.

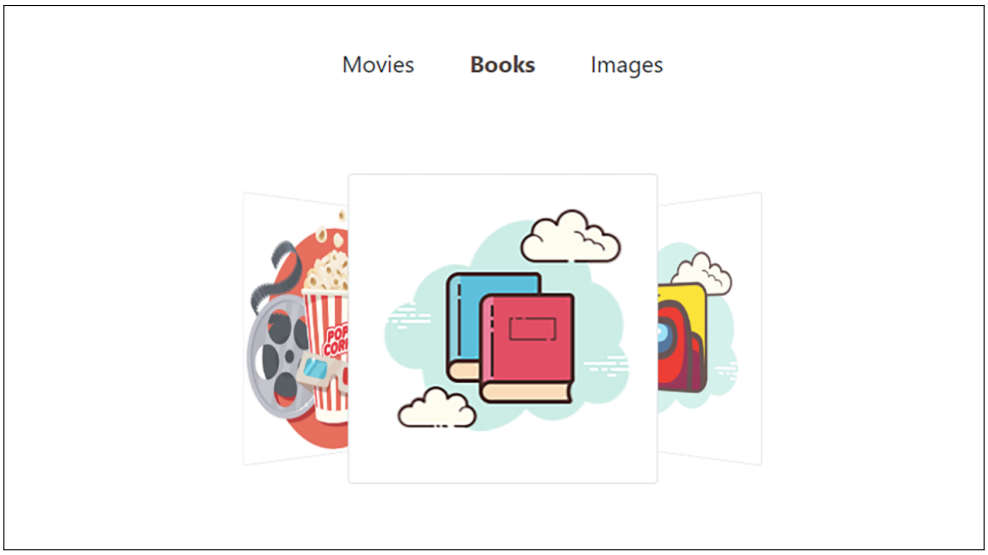


Figure 3.2: Category Page. By clicking the “Next” or “X” buttons, they will be directed to the category page, which contains three possible content categories: Books, Movies, and Images.

The first row of items, along with the search bar, is shown in Figure 3.3. The full page of each item is page is shown in Appendix A.1, Appendix A.3 and Appendix A.2. This page contains a *facilitate (suggesting alternatives) nudge* [15], by facilitating the selection from a random set of items over many others.

Keyword Selection Page. After selecting an item, the user is brought to the keyword selection page, where she must choose three keywords. For example, if a user selects “Harry Potter 4” as an item, she will be shown a random excerpt of the book (see Figure 3.4) from which she is expected to select her keywords. Once each keyword is chosen, it is shown in a bar at the top of the page. We set the maximum number of words per excerpt to 50 to avoid scrolling the page for the user, but the user can click on the shuffle button to land on another random excerpt of the book. After the selection of each keyword, the user is directed to another random excerpt containing the previously selected keyword. Suppose that the user has already selected “had” and “Herminone” as the first and second keywords. For the third keyword selection, she would be shown a random excerpt containing the word

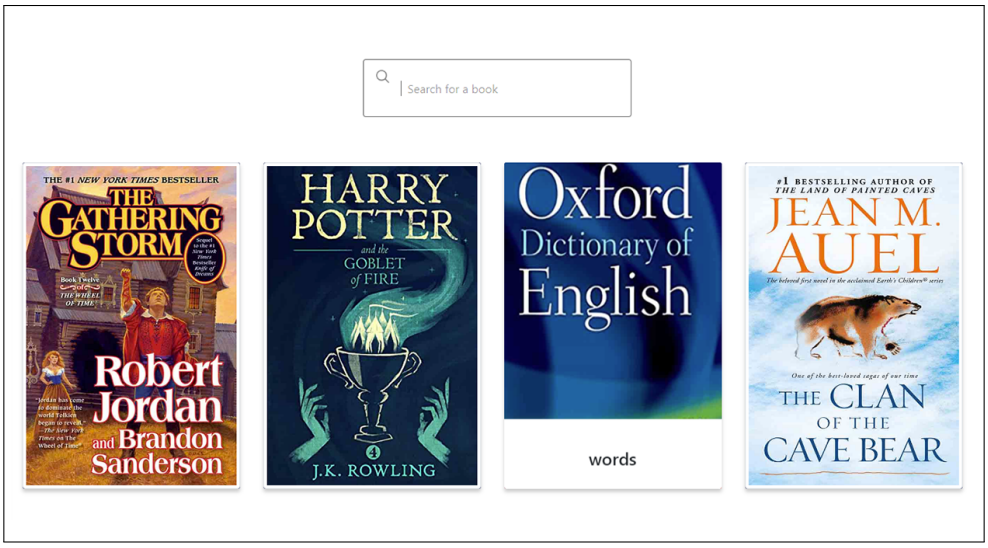


Figure 3.3: Item Page, Books. Once users select their desired category, they will be taken to the item page, which contains 20 randomly selected items.

"Herminone" (highlighted in red); see Figure 3.4 for this exact scenario. Then, she can select "apologize" (highlighted in blue) as the final third keyword.

Keyword Splash. Once three keywords are selected, the user will be shown the selected keywords in a "splash" page as shown in Figure 3.5. This page intends to employ further nudging towards selected keywords just before the password creation phase. This page has a black background with soft-white text to create a dramatic color contrast for drawing visual attention to selected keywords, and it automatically close after 3 seconds. But users can manually close it by clicking anywhere on the screen. This splash page aims to offer *a confront nudge* (throttling mindless activity) [15], to nudge users to review the content again.

Registration. PiXi adds a large display area of selected items and keywords on the left side of the typical registration input panel (see Figure 3.6). This addition serves *a reinforce nudge (or subliminal priming)* [15], as they make the image cover and keywords more prominent and easily accessible at the time the user is attempting to conceive a new password. We implement the password length requirement of at least 8 characters.

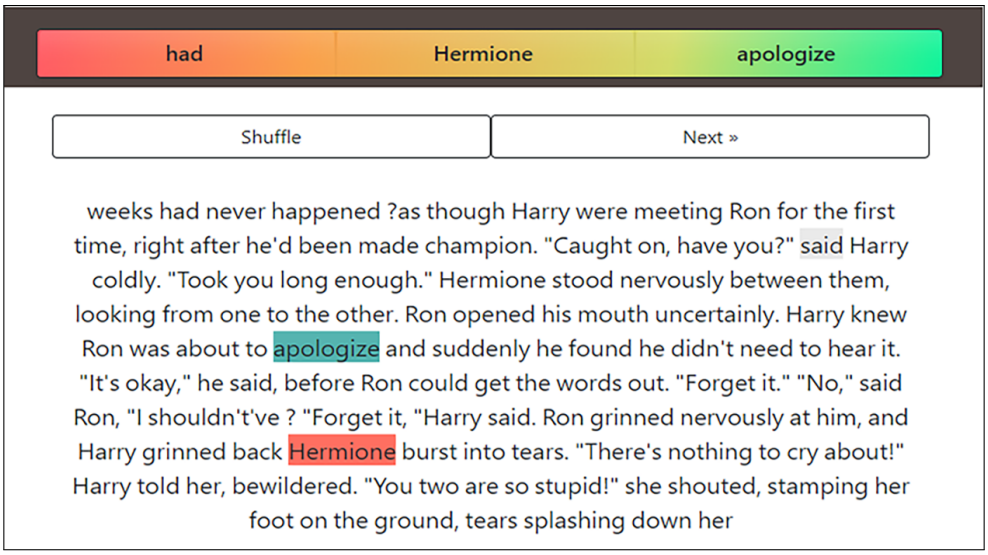


Figure 3.4: Keyword Selection Page. Selecting an item will lead users to the keyword selection page, where they choose three keywords from a random excerpt of the text of the selected item.

Login. PiXi does not modify the standard login page, and users simply need to enter their username and password to complete the login process.

An Extension: PiXi-Hints. PiXi can also be deployed as a hint for password recovery. To this end, we also have designed a PiXi extension called *PiXi-Hints (PH)*, which has all the components of PiXi but slightly differs at the login time. It requires the user to interact with PiXi just before login by inputting their keywords. This interaction intends to help users remember their passwords. In our implementation, we did not require users to recall their keywords but recorded their recall for analysis purposes. The introduction video had some minor differences for users of PiXi-Hints: they have an additional sentence that advises them to select interesting and memorable keywords. This recommendation is provided to encourage users to remember their keywords as they will need to reuse PiXi to input them again before each login.

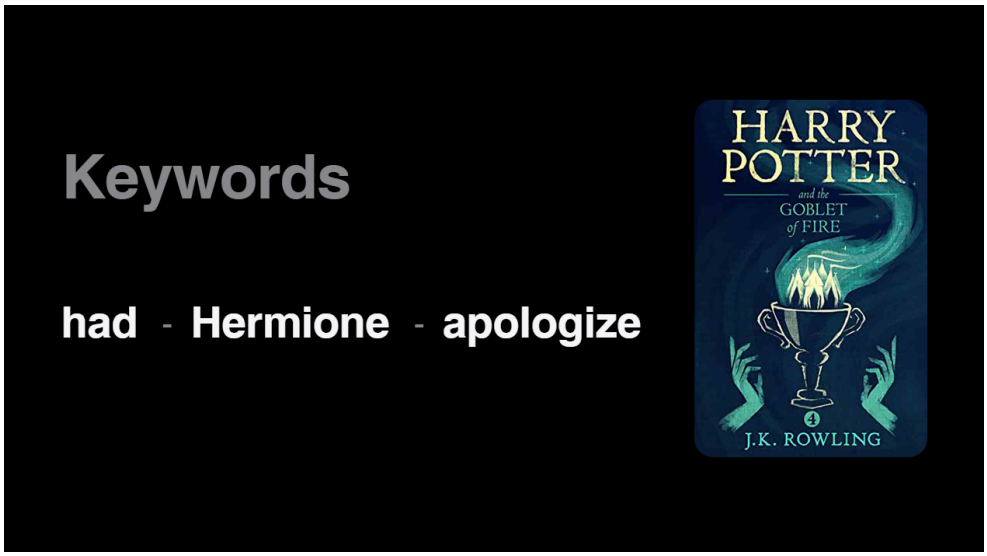


Figure 3.5: Keyword Splash Page. After selecting all three keywords, users will see the keyword splash page that displays all three chosen keywords (for three seconds) to nudge them further to selecting their passwords.

3.1 Theoretical Keywords Space

This section discuss theoretical keywords space of each category of the PiXi.

In PiXi, each category contains a significant number of items that cater to the needs of the average knowledgeable user.The theoretical keyword space is the total number of possible combinations of keywords that could be generated. However, the theoretical keyword space in each category varies due to the limitations and capacities of the APIs. By calculating this value, we can determine the level of security provided by the keyword generation algorithm.

For example, if a user chooses "Harry" as one of their keywords, and an attacker knows that PiXi uses a category of "Movies", the attacker could try other keywords in the "Movies" category such as "Potter" "Hogwarts" or "Hermione" to see if any of these combinations form the user's password. If the keyword space is too small, and if users choose keywords as their passwords directly, an attacker can potentially guess or brute-force these passwords

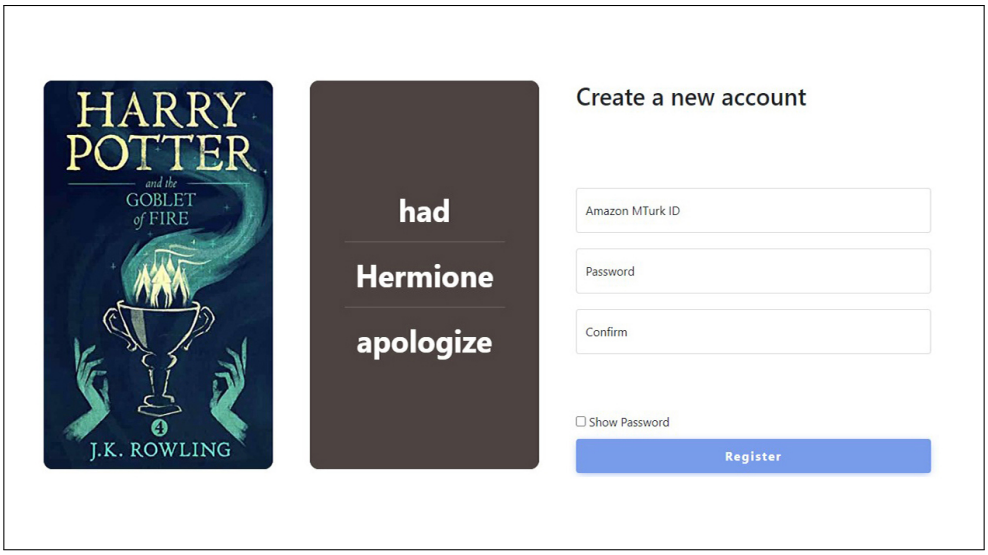


Figure 3.6: The Register Page. Finally, users will see the register page which features a large display area of the selected items and keywords on the left side of the typical registration input panel.

by trying out different combinations of keywords.

As a result, it is essential for PiXi to provide a sufficiently large keyword space in each category to ensure that generated passwords are strong and resistant to attacks.

Unique words references (Books)		
Title	**Total Words**	**Unique Words**
Bible (KJV)	857,116	13,076
Breakfast at Tiffanys	27,790	4,893
The Wheel of Time 11 - Knife of Dreams	328,633	13,844
Harry Potter 7 - Deathly Hollows	200,888	11,411
The Two Towers	156,379	19,622

Table 3.1: Unique Words Reference List

Books: The Books API provided a list of 87 best-selling books from The New York Times Best Sellers. Additional books may be added to the list in the future. It is assumed that each book contains an average of 10,000 to 12,000 unique words, excluding blank spaces,

special characters, or numbers, some examples are shown in Figure 3.1.

Movies: The Movies API contains more than 805,778 movies, and more movies are being added everyday. Some of the movies may be duplicated and some of those have very limited content/descriptions or missing covers, which are blocked by the PiXi system automatically. Each movie descriptions contains around 50 unique words.

Images: The Images API contains more than 5.1 million images, we are using tags of each image as keywords. Different from the Books or Movies section, tags can be a single word or a combination of words. According to the Unsplash API document, there are more than 5,000,000 unique tags. Each image is associated with in average 8 tags and each tag is unique.

The keyword space of the Books category is $\log_2 99.88$, it indicates that there is a large database of unique contents available, and users may interact with the system frequently to search, and browse. Similarly, the keyword space of the Movies category is $\log_2 92.72$, it suggests that the movie database may be slightly smaller than the book database, but still significant enough to attract user interactions. Lastly, the keyword space of the Images category is $\log_2 84.84$, it shows that the image database may be the smallest of the three categories, but still sizable enough to attract user interactions. Overall, it is difficult to make precise assumptions about the total number of PiXi interactions with these categories without more information. However, based on the size of the keyword spaces, it can be inferred that all three categories are likely to attract a significant amount of user interactions.

Chapter 4

User Studies

We conducted a two-Session study (N=550) on Amazon MTurk to evaluate PiXi's ability to nudge users. Our study was approved by Ontario Tech University's Research Ethics Board (#16688) on (05/27/2022).

4.1 Recruitment and Compensation

Our advertisement was made visible to all MTurk workers, but only US workers with an approval rate of 95% or above were allowed to participate. We asked users to imagine a scenario which requires them to create a password for a very important banking account. The users first reviewed and signed the consent form, then were redirected to the PiXi system. Workers who were interested in the study can view and sign the consent form before being redirected to the PiXi system from Amazon MTurk. The consent form can be found in Appendix F. 7 days later, participants who successfully completed Session #1 and passed an attention test were invited back through Amazon MTurk to take part in the Session #2 study. The Sessions were compensated at the US minimum wage at the time of the study ($7.25/hour). For Sessions 1 and 2 (resp.) with estimated completion times of 7

and 2 minutes (resp.), the participants received $0.85 and $0.35 (resp.). Participants could withdraw from the study at any time, and their data would be destroyed and not used in any future data analysis.

Conditions/Groups. Upon beginning the study, users were randomly assigned to one of three groups:

1. *Control:* Users create a password and log in as usual (without PiXi).
2. *PiXi:* Users are asked to use PiXi only prior to password creation.
3. *PiXi-Hints:* Users are asked to use PiXi immediately prior to password creation. Users are also asked to use PiXi immediately prior to login. This was intended to test (a) whether using PiXi as a password hint might help improve memorability, and (b) whether users recall their PiXi information (image and keywords).

4.2 Sessions and Tasks

Our study contains two Sessions.

Session 1. Participants were required to register an online account (the process differs based on condition/group), and then complete questionnaire #1. The system randomly assigned the participant to a condition to balance the population of each. We did implement the basic password length requirement (e.g., at least 8 characters long) as it is the basic password policy of the majority of websites. After participants create new passwords, they filled out the questionnaire. Questionnaire #1 15 questions. Question 1 to Q3 are used to collect users password behaviours. Questions 4 is the attention question to see if users pay attention to the study. Question 5 is the overall rating of the system. Question 6 - Question 10 are SUS questions to exam usability hypotheses discussed in 5.3.2. Participants who did not complete the questionnaire or quit during the Session had their responses destroyed and

excluded from our data analysis.

Session 2. 7 days later, participants who successfully completed Session #1 were invited back through Amazon MTurk to return for taking the Session #2. After successful login or three unsuccessful login attempts, the participants filled out the exit Questionnaire 2.

Storage of Data All collected data was anonymous and encrypted using SHA-512.

4.3 Data Cleaning

In order to ensure that only valid participants were included in the study, we reviewed all responses carefully. Unfortunately, we discovered several critical issues that required us to take more actions to sanitize the data.

1. *Users with multiple accounts (N=193).* Once removing the remarkably predictable passwords, one wouldn't expect to see many repeated passwords in our dataset. However, we discovered that 193 participants chose identical passwords that are uncommon in other password datasets. A likely interpretation of this finding is that these accounts belong to the same worker controlling multiple accounts; as such we have removed these accounts from our dataset.

2. *Inattentive (N=58).* We added an attention question to the Session #1 questionnaire:"Seven plus three equals eight," to confirm that participants were paying attention to the survey. Participants who did not select either "Strongly Disagree" or "Disagree" were considered as inattentive and all their data were excluded and destroyed from our data analysis.

3. *Users who did not try to create a new password (N=67).* Many participants had used weak and predictable passwords, such as their MTurk IDs, or simple number sequences like "123456789" and "25252525", in order to quickly complete the study.

	Control	PiXi	PiXi-Hints
Participants	181	185	192
Multi-Identity	76	53	64
Inattentive	15	35	8
Weakly-Committed	19	14	34
Valid Participants (Session 1)	71	83	84
Valid Participants (Session 2)	10	9	12

Table 4.1: Statistic of session completion and filtered participants across conditions.

Additionally, we noticed that a number of these participants had provided similar responses with specific patterns, such as “111111” and “121212”.

Overall, we were surprised by the initial amount of noise in our dataset. This experience has taught us the importance of careful screening when using Amazon MTurk as a research tool in the future.

4.4 Demographics

Table 4.2 presents an overview of the participant demographics for our study collected through the questionnaire in Session 1. Overall, our participants were composed of 41% female, 58% male, and 1% who preferred not to specify their gender. The majority of participants (51%) fell within the 20–30 age group, followed by the age group of 30–40 making up 32% of participants. Regarding participants’ education level, most participants (68%) had a Bachelor’s degree, followed by a Master’s degree (25%). The majority of participants in our study worked in Business (24%), Technology (21%), or Health (13%).

4.5 Limitations

Amazon MTurk Reliability Concerns

Our study has some limitations due to Amazon MTurk, which introduced a notable amount of noise in our collected data. While we did our best to fairly catch noise and remove it from our data, it is possible we couldn't catch and filter all noisy data. However, since the noise should be consistent between each group, any statistically significant finding should be reliable. Another potential limitation in our user study arises from the distinctive behavioral patterns exhibited by Amazon MTurk users, which could different from those of regular users. The context of being on the MTurk platform might introduce certain biases or motivations that are not representative of typical user behavior.

Gender	**Control**	**PiXi**	**PH**	**Language**	**Control**	**PiXi**	**PH**
Female	42.3%	39.8%	40.5%	English	98.6%	100.0%	98.8%
Male	56.3%	59%	59.5%	Other	1.4%	0.0%	1.2%
N/A	1.4%	1.2%	0.0%	N/A	0.0%	0.0%	0.0%
Age	**Control**	**PiXi**	**PH**	**Occupation**	**Control**	**PiXi**	**PH**
Under 20	0.0%	0.0%	0.0%	Engineering	7.0%	6.0%	7.1%
20-30	54.9%	50.6%	48.8%	Arts and Entmt.	1.4%	4.8%	7.1%
30-40	25.4%	27.7%	34.5%	Business	31.0%	18.1%	26.2%
40-50	11.3%	9.6%	9.5%	Comms.	4.2%	2.4%	3.6%
50-60	5.6%	6.0%	6.0%	Social services	5.6%	6.0%	2.4%
60+	2.8%	6.0%	1.2%	Education	7.0%	7.2%	8.3%
N/A	0.0%	0.0%	0.0%	Technology	14.1%	24.1%	23.8%
Education	**Control**	**PiXi**	**PiXi-Hints**	General Labour	2.8%	7.2%	1.2%
None	0.0%	0.0%	0.0%	Agriculture	1.4%	3.6%	3.6%
High School	1.4%	4.8%	8.3%	Government	2.8%	2.4%	2.4%
Bachelor's	74.6%	68.7%	63.1%	Health	18.3%	10.8%	11.9%
Master's	23.9%	24.1%	27.4%	Law	0.0%	0.0%	0.0%
PhD.	0.0%	2.4%	1.2%	Sales	2.8%	4.8%	0.0%
N/A	0.0%	0.0%	0.0%	N/A	1.4%	2.4%	2.4%

Table 4.2: The user demographics across the three conditions.

Chapter 5

Results

We begin by evaluating indicators that PiXi's nudges work in Section 5.1. We perform an extensive security analysis in Section 5.2, and usability analysis in Section 5.3.

5.1 Evaluation of Nudging Efficacy

Table 5.1 shows nudges implemented in different pages. Through various metrics, we evaluate time spent on pages (nudges) during the registration phase and the efficacy of (i) the positioning nudge in the Category Page, (ii) the suggesting alternatives nudge in the Items Page, and (iii) PiXi's overall nudge ability in the users' password.

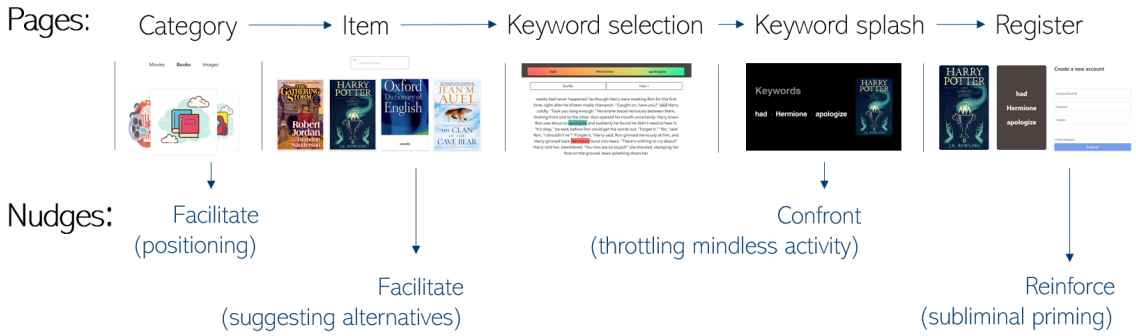


Figure 5.1: Summary of nudges implemented in different pages.

Average Time Spent on Pages (Registration Phase). In order to determine which page draw users' attention, we summarized time spent on pages in Table 5.1. We found users spent 46.49s on modal, 10.89s on category selection, 30.54s on selecting an item, 9.2s, 10.28s, 11.37s on choosing keywords respectively, 6s on keyword splash modal and 56.80s on password creation. These results indicate that users spent nearly equal amounts of time on each page, suggesting a balanced level of user engagement.

		Average Time	**Average Percentage**
Category Page	Modal	46.49s	25.6%
	Category Selection	10.89s	6.00%
Items Page	Item Selection	30.54s	16.82%
	1st Keyword	9.2s	5.07%
	2nd Keyword	10.28s	5.66%
	3rd Keyword	11.37s	6.26%
Register Page	Keyword Splash	6s	3.30%
	Password Creation	56.80s	31.29%
Total		181.57s	100%

Table 5.1: Average time spent on nudges during the registration phase, combining PiXi and PiXi-Hints.

Positioning Nudge in Category Page. Table 5.2 shows the acceptance rates of the positioning nudge for categories where one category is initially positioned in the center of the Category Page (for both PiXi and PiXi-Hints). Approximately half of the participants accepted the category positioned such that it can be selected without scrolling. There appears to be a slightly higher preference for the Images and Movies categories. It suggests that users prefer thecategories with more visual content.

Suggesting Alternative Nudge in Items Page. Table 5.2 also shows the acceptance rates of the suggested alternative nudge in item pages, where the set of 20 randomly selected

	Positioning Nudge (Category Page)	Suggesting Alternatives Nudge (Items Page)
Books	20/56 (35.71%)	40/41 (97.56%)
Movies	29/59 (49.15%)	40/55 (72.73%)
Images	30/51 (58.82%)	63/71 (88.73%)

Table 5.2: The acceptance rates of the facilitate nudges, combining PiXi and PiXi-Hints.

items initially appeared on the page for both PiXi and PiXi-Hints. Most users (72%-97%, depending on category) accepted one of the suggested items, indicating that this nudge was successful at nudging users towards exploring unique items they might not otherwise consider.

Do PiXi Nudges Influence Resulting Passwords? We aim to determine whether PiXi influenced users' password choices. The most straightforward method to measure this is to determine how many users incorporate their keywords directly in their passwords. Our findings, shown in Table 5.3, revealed that 39% of users (31% for PiXi, 46% for PiXi-Hints) incorporated at least one keyword into their passwords. We consider this metric an underestimate of the number of users who are nudged by PiXi, since users may see a relationship between their passwords and keywords that we are unable to detect (e.g., if it is indirectly related and personal in nature). Although it is likely an underestimate, it still provides evidence that a large percentage of users are influenced by the PiXi system during password creation. An emerging critical question is how these nudges have impacted the security of the chosen passwords, which we will address next.

We analyzed the frequencies of all 501 keywords for both PiXi and PiX-Hints using the COCA top 220,000 popular keywords database.[1]. We firstly calculate the average frequency of keywords in PiXi and PiXi-Hints, and the results are shown in Table 5.4. It indicates that the average frequency of keywords in PiXi-Hints is significantly lower than

1 https://www.english-corpora.org/coca/

	1 keyword	2 keywords	3 keywords	Total
PiXi	7	12	7	26/83 (31%)
PiXi-Hints	11	14	14	39/84 (46%)
Total	22	26	17	65/167 (39%)

Table 5.3: The keywords usage rate for both PiXi and PiXi-Hints, including direct and indirect use (e.g., uppercase, lowercase, or additional punctuation).

for PiXi, which means the keyword complexity in PiXi-Hints is significant higher than for PiXi. It suggesting that reinforce-priming in PiXi-Hints works well when encourage users to use more complex keywords.

	Mean	N	Std. Deviation
PiXi	691,112	248	3,906,156
PiXi-Hints	363,475	252	2,197,102
Total	524,966	501	3,160,398

Table 5.4: Mean frequencies of all 501 keywords for both PiXi and PiX-Hints.

Page Scrolling Events: We collected page scrolling data from each user, which included both the absolute scrolling activity (ASA) [2] and the total scrolling activity (TSA) [3]. This data was used to track how much scrolling was done using the mouse while the user selected items.

TSA is calculated as follows: scrolling up is recorded as $(+1 * UP\_Count)$, scrolling down is recorded as $(+1 * Down\_Count)$, and the total value is $(+1 * UP\_Count) + (+1 * Down\_Count)$. The range of TSA is $[0 - screen\_size]$.

2 ASA captures the absolute value of scrolling activity to determine the exact location of a chosen item on the screen. However, it may not accurately track users who frequently scroll up and down, and will return a value of 0 if the up and down scrolling activity is equal. The ASA function serves as an aid in precisely monitoring user behavior.

3 TSA logs all scrolling activity related to an item. A higher TSA value indicates that the user is scrolling more, which can indicate whether the user is motivated to explore further or simply scrolling out of habit.

We conducted a detailed analysis of the scrolling behavior of participants in both PiXi and PiXi-Hints systems to determine the effectiveness of different nudges, and the result in shown in Table 5.5. Our findings indicate that users who accepted suggested categories in both systems tend to scroll more (higher TSA values) when under the Movies category, while the TSA values are lower and very similar in Images and Books categories. Based on the ASA value, we found that participants did not choose the very first item but tended to choose other items in all the three categories.

	Suggested		**Independent**	
	ASA	**TSA**	**ASA**	**TSA**
Books	18	30	0	0
Movies	21	73	7	55
Images	15	24	14	47

Table 5.5: The ASA. and TSA. values of each category, combining PiXi and PiXi-Hints.

For participants who rejected suggested categories, we also found that they scroll more (higher TSA value) when under the Movies category, while the TSA values are lower in Images and almost equal to zero in Books category due to only one participant choosing it. Based on the ASA value, we found that participants did not choose the very first item but tended to choose other items in Movies and Images categories.

Our analysis suggests that the "positioning" and "Suggesting alternatives" nudges work better in the Movies and Images categories, as these categories are more attractive and well-accepted among users with different backgrounds compared to the Books category. These nudges can encourage users to explore more information and scroll through more items. Overall, our findings highlight the importance of carefully selecting and implementing nudges to improve user engagement and interaction with PiXi and PiXi-Hints systems.

However, because there are concerns about trust when using Amazon MTurk, the

suggestions we made based on our user study may not be reliable. This is especially true if many Amazon MTurk users do not prioritize keeping their keywords, passwords, and the study itself. Some users may participate in the study only to earn money, without caring about its purpose or validity. Therefore, we need to conduct more studies on different platforms to confirm whether our suggestions are correct or not.

5.2 Security Analysis

We study the security of passwords created under each condition from different perspectives including their length, ZXCVBN score, strength against online and offline attacks, hot-spots items, and the security analysis for passphrases. We do not discuss any shoulder-surfing attacks since PiXi should be the same as a traditional password system.

5.2.1 Password Length

We recorded the length of all the passwords to measure the password strength of each condition. To determine whether a condition can influence the password length, we test the following Hypothesis:

$\mathcal{H}_0$ *The distribution of password lengths is similar across PiXi, PiXi-Hints, and Control conditions.*

$\mathcal{H}_a$ *The distribution of password lengths differs between PiXi, PiXi-Hints, and Control conditions.*

The one-way ANOVA test ($df = 2$, $N = 238$) rejects the null hypothesis $\mathcal{H}_0$ ($F = 6.5, P = 0.002$) after Holm-Bonferroni correction ($\alpha'_{(1)} = 0.0167$), indicating a significant difference in password length among the three conditions with a large effect size ($\eta^2 = 0.44$). Table 5.7 shows the mean password length for each condition. The Control condition ($\mu = 9.35$) had a significantly lower password length compared to PiXi ($\mu = 10.87$) and PiXi-Hints

($\mu = 11.42$), while the mean in PiXi and PiXi-Hints are comparable. This suggests that PiXi and PiXi-Hints users tend to create longer passwords than those in the Control condition, which can offer security advantages.

Score	# Guesses X	Description
0	$1 \leq X \leq 10^3$	Too guessable: risky password.
1	$10^3 < X \leq 10^6$	Very guessable: protection from throttled online attacks.
2	$10^6 < X \leq 10^8$	Somewhat guessable: protection from unthrottled online attacks.
3	$10^8 < X \leq 10^{10}$	Safely unguessable: moderate protection from offline slow-hash scenario.
4	$X > 10^{10}$	Very unguessable: strong protection from offline slow-hash scenario.

Table 5.6: ZXCVBN password score range and descriptions [81].

5.2.2 Password Score

We use ZXCVBN (JavaScript version) [81], a widely used password meter that is easy to implement and cost-effective. Given an input password, it returns a strength score as described in Table 5.6. To determine whether a condition can influence the ZXCVBN password score, we test the following hypothesis:

$\mathcal{H}_0$ *The distribution of ZXCVBN scores is similar across PiXi, PiXi-Hints, and Control conditions.*

$\mathcal{H}_a$ *The distribution of ZXCVBN scores differs between PiXi, PiXi-Hints, and Control conditions.*

A one-way ANOVA test ($df = 2, N = 238$) revealed a significant difference in password score among the three conditions ($F = 3.868, P = 0.022$) with a medium effect size ($\eta^2 = 0.032$), leading us to reject the null hypothesis $\mathcal{H}_0$ after Holm-Bonferroni correction ($\alpha'_{(3)} =$

	Password Length	ZXCVBN Score	SUS Score
Control	9.35 ± 1.73	1.83 ± 1.04	56.60 ± 13.28
PiXi	10.87 ± 4.38	2.16 ± 1.02	54.48 ± 11.93
PiXi-Hints	11.42 ± 4.01	2.31 ± 1.17	56.68 ± 11.49

Table 5.7: The Mean ± Std. for password length, password score, and SUS score.

0.05). As shown in Table 5.7, the Control condition with an average of (μ = 1.83) has a lower password score than PiXi (μ = 2.16) and PiXi-Hints GT y (μ = 2.31). These findings suggest that passwords created through PiXi and PiXi-Hints are stronger than those created by users in the Control condition.

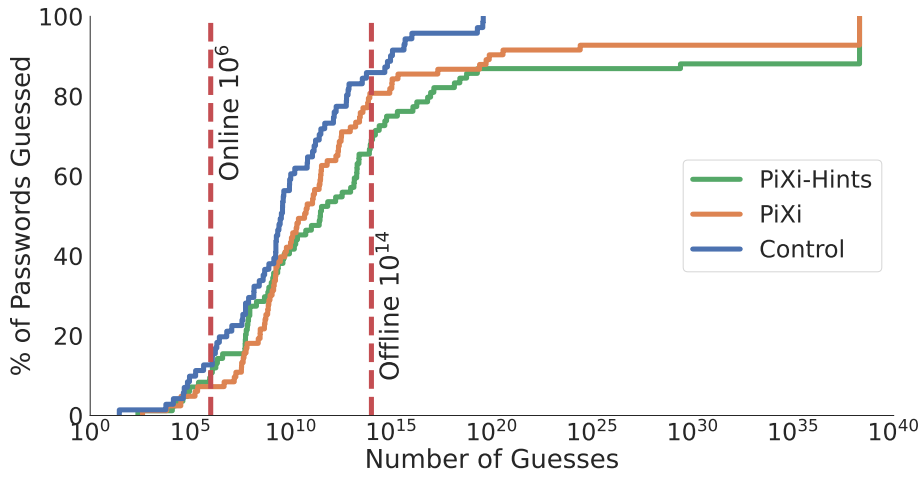


Figure 5.2: Password strength across the three conditions using CKL_PSM.

5.2.3 Password Strength

We evaluate password strength by CMU's Password Guessability Service (PGS) [74] which uses numerous state-of-the-art password cracking algorithms to calculate guessability. The use of "guessibility" as a metric of password security helps assess the vulnerability of a password to different types of attacks, particularly those involving guessing or brute-forcing

techniques. [4]. To assess password strength under online and offline attacks, we employed online and offline attack thresholds of 10^6 and 10^{14} guesses [30]. When a password can be guessed before the online (or offline) attack threshold, we call it *online-unsafe* (or *offline-unsafe*). The summary of our analyses is reported in Table 5.8 and Figure 5.3. Passwords that can withstand offline attacks in PiXi (14.4%) and PiXi-Hints (32.1%) are significantly higher than in the Control (7%) condition. Conversely, weak passwords are more common in the Control (18.3%) than in PiXi (or 10.8%) and PiXi-Hints (15.5%). We conducted a test to determine whether password strength depends on different conditions, by testing the following hypotheses:

$\mathcal{H}_0$ *The distribution of password strength measurements is similar across PiXi, PiXi-Hints, and Control conditions.*

$\mathcal{H}_a$ *The distribution of password strength measurements differs across PiXi, PiXi-Hints, and Control conditions.*

We performed a χ^2 test ($df = 4, N = 238$) to examine these hypotheses. The results in Table 5.8 showed a significant difference ($\chi^2 = 17.120$, $P = 0.004$) with a medium effect size (Cramer's $V = 0.187$) across different conditions, so we reject the null hypothesis ($\mathcal{H}_0$) after Holm-Bonferroni correction ($\alpha'_{(2)} = 0.025$). This finding further supports that PiXi and PiXi-Hints encourage users to create more unique and stronger passwords than the Control condition.

Password Uniqueness. An interesting finding from the CMU results was that when using Markov Model or Neural network approaches, a significant number of guessing results were indicated as "-5," particularly in the PiXi 51 out of 83 (or 61.4%) and PiXi-Hints 50 out of 84 (or 59.5%), while the Control condition was lower at 32 out of 71 (or 43.6%).

4 We also study them by CKL_PSM–a password strength meter based on the chunk-level PCFG model (CKL_PCFG). However, the results were quantitatively and qualitatively very similar, and the results are shown in 5.2

	Online-unsafe	Offline-unsafe	Safe
Control	18.3%	74.7%	7%
PiXi	10.8%	74.8%	14.4%
PiXi-Hints	15.5%	52.4%	32.1%

Table 5.8: Passwords guessability at the online and offline thresholds of 10^6 and 10^{14}, CMU's Password Guessability Service.

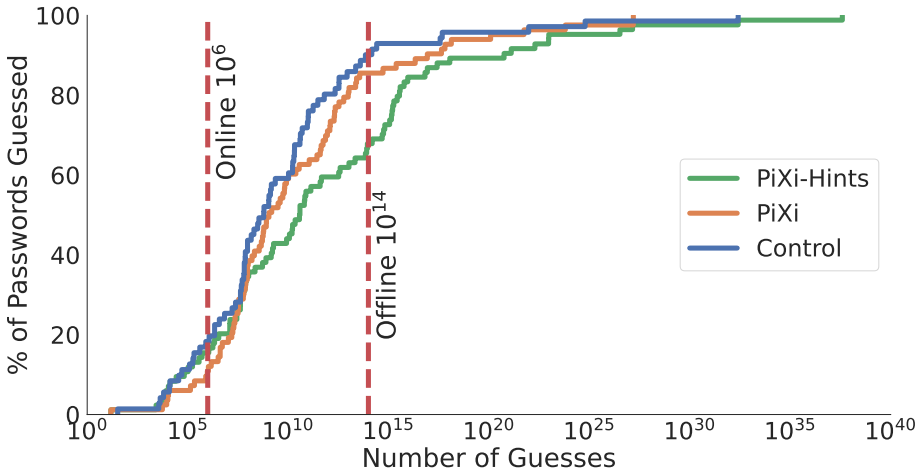


Figure 5.3: Password strength across the three conditions.

According to CMU's responses, "-5" means the password was not discovered by that approach. This finding proves that PiXi and PiXi-Hints encourage users to create more unique and stronger passwords than the Control condition.

Should Users Incorporate Keywords in Passwords? As observed in Section 5.1, many users incorporate their keywords into their passwords. Here we aim to determine the security impact of this behavior, to determine whether PiXi should encourage or prevent it. As shown in Table 5.10, for both PiXi and PiXi-Hints, the passwords using keywords had much higher length, score, and guesses than the average passwords. This suggests that users who used keywords were able to create stronger and longer passwords, and as such future versions of PiXi might encourage this behavior.

Rank	Word	Category	F.	Rank	Word	Category	F.	Rank	Word	Category	F.
1	Adam	Movies, Books	5	19	coronavirus	Movies,Images	2	37	money	Images	2
2	green	Images	5	20	data	Images	2	38	pattern	Images,Books	2
3	nature	Movies, Images	5	21	deadly	Books	2	39	pc	Images	2
4	family	Movies	4	22	down	Movies,Books	2	40	people	Images	2
5	world	Movies,Images	4	23	Drizzt	Books	2	41	person	Images	2
6	and	Movies, Books	3	24	egg	Movies,Images	2	42	phone	Images	2
7	flower	Images	3	25	escape	Movies	2	43	plant	Images	2
8	grey	Images	3	26	evening	Movies	2	44	safe	Movies,Images	2
9	happy	Images	3	27	eye	Images	2	45	scifi	Movies	2
10	human	Images	3	28	face	Images	2	46	seats	Movies,Books	2
11	story	Movies	3	29	food	Images	2	47	sky	Images	2
12	tree	Images	3	30	friends	Movies	2	48	Strange	Movies,Books	2
13	trouble	Movies,Books	3	31	girl	Images	2	49	unique	Movies	2
14	united states	Images	3	32	hand	Images	2	50	usa	Images	2
15	beach	Movies,Images	2	33	Lando	Books	2	51	Vietnam	Movies,Images	2
16	blue	Images	2	34	leaf	Images	2	52	water	Images	2
17	brown	Images	2	35	Merry	Books	2	53	weapons	Books	2
18	business	Movies	2	36	middle	Movies,Books	2				

Table 5.9: The overall most popular words (used more than once) across PiXi and PiXi-Hints.

5.2.4 Popular Selections (Hot-spots)

In this section, we examine whether certain selections (items, categories, or keywords) are “hot-spots", meaning they are frequently selected by users. Figure 5.4 displays the distribution of categories for PiXi and PiXi-Hints, with Images being the most popular, followed by Movies and Books being the least popular. This suggests that users prefer exploring visual content over textual content.

Table 5.9 illustrates the most commonly used words for both PiXi and PiXi-Hints, with the top three being “Adam,” “green,” and “nature,” each chosen 5 times. Figure 5.6 displays the most frequently used words. We did not find any significant hot-spot items or words, indicating that each user has individual content preferences.

Do Some Categories Nudge Stronger Passwords? We also investigate whether password strength depends on the nudge category (Books, Movies, or Images). Table 5.11 shows that

Figure 5.4: Category distribution for both PiXi and PiXi-Hints.

	Keywords	Length	Score	CMU Guesses
PiXi	Yes	14.15	2.81	$10^{15.45}$
	No	9.25	1.89	$10^{8.89}$
PiXi-Hints	Yes	13.05	2.51	$10^{14.37}$
	No	9.79	2.17	$10^{10.61}$

Table 5.10: Comparison of security metrics for passwords with vs. without keywords.

passwords created by users who selected Books were most resistant to online and offline attacks. Passwords created by users who selected Images have the least “safe" passwords. One possible reason for this is that keywords from the Images category tend to be less unique compared to the other categories. These results suggest that password strength differs between categories and that future PiXi implementations might avoid using the Images category.

5.2.5 Passphrases Security Analysis

During our analysis of user passwords, we found that many users were utilizing passphrases[5] instead of traditional passwords. Subsequently, we compiled the distribution of passphrases, which is summarized in Table 5.12. Overall, the result suggested that PiXi (43%) and PiXi-Hints (45%) users used more passphrases than the control condition (7%). The findings indicated that PiXi and PiXi-Hints influenced users to prefer passphrases compared to the control condition.

	Online-unsafe	**Offline-unsafe**	**Safe**
Books	7.3%	56.1%	36.6%
Movies	14.5%	56.4%	29.1%
Images	15.5%	73.2%	11.3%
Total	14.7%	66.8%	18.5%

Table 5.11: The guessability of passwords at the online and offline thresholds across three categories, combining PiXi and PiXi-Hints.

Interestingly, we observed that a significant number of users who employed keywords directly (without any modifications) tended to incorporate all the keywords into their passphrases. However, this behavior raises some security concerns. By relying solely on keywords, users put themselves at risk of generating weak and easily guessable passphrases, which can lead to predictable patterns. Attackers can leverage various techniques, such as brute-force attacks, dictionary-based attacks, or social engineering, to crack passphrases that rely solely on keywords.

Although PiXi does not store users' keywords in any format, attackers possessing knowledge of category APIs can create dictionaries using keywords from each category. If a user chooses the first three words (in sequence) from the first item of the first category

5 Passphrase: a password string contains a sequence of words (comprising at least 2 words, in any language). We only determine passphrase in English in our analysis.

	Passphrase Contains 2 Words					
	Use Kw. Directly		Use Modified Kw.		No Kw.	Total
	1 Kw.	2 Kw.	1 Kw.	2 Kws.		
Control	N/A	N/A	N/A	N/A	N/A	9/71 (12.7%)
PiXi	0%	9/83 (10.9%)	5/83 (6.0%)	2/83 (2.4%)	7/83 (8.4%)	23/83 (27.7%)
PiXi-Hints	1/84 (1.2%)	7/84 (8.3%)	0%	6/84 (7.2%)	8/84 (9.5%)	22/84 (26.2%)

	Passphrase Contains 3 Words							
	Use Kw. Directly			Use Modified Kw.			No Kw.	Total
	1 Kw.	2 Kws.	3 Kws.	1 Kw.	2 Kws.	3 Kws.		
Control	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1/71 (1.4%)
PiXi	0%	0%	7/83 (8.5%)	0%	0%	1/83 (1.2%)	5/83 (6.0%)	13/83 (15.7%)
PH	0%	0%	9/84 (10.6%)	1/84 (1.2%)	3/84 (3.6%)	0%	3/84 (3.6%)	16/84 (19.0%)

Table 5.12: Passphrase (contains 2 and 3 words) distribution across the three conditions.

to construct a passphrase, fortunately, it becomes challenging to determine the first word of each item in every category due to the unique seed for each user. The attacker's best strategy is therefore to guess sequences of keywords, where this attack would be most effective.

Do Passphrases "Popular (Pop)" in COCA[6]? By implementing the above suggestions, we can somehow mitigate the identified challenges and enhance the overall security of user passphrases in PiXi. Nevertheless, the security of user created passphrases using PiXi remains uncertain due to the absence of reliable passphrase guessers now. One possible ap-

6 The Corpus of Contemporary American English (COCA) is the largest and most representative collection of American English. It is widely used and connected to other corpora of English [51].

	COCA 5000 Permutations	COCA 5000 Permutations + characters and numbers	No Effective Passphrase Cracking Methods	Total
2 Words	(27/167) 16.2%	(5/167) 3.0%	(2/167) 1.2%	(34/167) 20.4%
3 Words	(17/167) 10.2%	(6/167) 3.6%	(3/167) 1.8%	(26/167) 15.6%

Table 5.13: Passphrase (contains 2 and 3 words) cracking rates using (COCA 5000 permutations) and (COCA 5000 Permutations + characters and numbers), combining PiXi and PiXi-Hints.

proach to assess this is by comparing each passphrase with each word in the COCA 5000[7] in the Corpus of Contemporary American English (COCA) database to investigate if those passphrases contains the most popular words and potential dictionary attacks using COCA. The specific procedure for this comparison is explained in the following section.

During our analysis, we made a significant discovery regarding the passphrases used in the three conditions. The majority of these passphrases contained 1 - 3 words from the COCA top 5000 popular words list. Specifically, in the Control condition, approximately 90% of passphrases included such words. In the PiXi condition, this figure rose to 92% (33 out of 36 passphrases), and in the PiXi-Hints condition, it further increased to 95% (36 out of 38 passphrases). It is worth noting that even though most of the popular words were indirectly used (for instance, with cases or tense changes, and adding or removing some characters), these alterations were still relatively easy to identify, categorize, and record.

To demonstrate how attackers crack passphrases containing 2 or 3 words using the COCA 5000, we combined data from PiXi and PiXi-Hints, presenting it in Table 5.13. The analysis revealed that a significant portion of passphrases can be found among the permutations of 2 or 3 words from the COCA 5000. Additionally, a notable percentage can be de-

7 An exploration of potential relationships between user keyword selections using COCA bi-grams and tri-grams lists was conducted. However, no passphrases were found within these lists. Consequently, our focus for further examination is on permutations of words from the COCA 5000.

	Offline-Unsafe	Offline-UnSafe	Safe
PiXi			
2 Words	1/167 (0.6%)	10/167 (6.0%)	0
3 Words	0	10/167 (6.0%)	0
PiXi-Hints			
2 Words	1/167 (0.6%)	15/167 (9.0%)	0
3 Words	0	7/167 (4.2%)	0

Table 5.14: The guessability of pure passphrases (2 and 3 words, without numbers, characters) using COCA 5000 permutations only at the online and offline thresholds for PiXi and PiXi-Hints.

tected in permutations of 2 or 3 words from the COCA 5000 combined with characters and numbers. On the other hand, certain passphrases showed high resistance to cracking methods, proving challenging to guess both with the help of the COCA list or password crackers. Examples of such resilient phrases include "ontariooshawaelimateuniversity" and "whitby-includingpickering," making them challenging to guess using conventional approaches.

In Table 5.14, for passphrases (denoted as p) comprised of 2 or 3 words (without special characters) from the COCA 5000, we set the guess value m as the number of permutations of 2 or 3 from 5000. The calculations for *COCA Permutations* (2 *words*) and *COCA Permutations* (3 *words*) are shown below:

$$\textit{COCA Permutations}\ (2\ \textit{words}) = \log_{10} 5000^2 = 7.397$$

$$\textit{COCA Permutations}\ (3\ \textit{words}) = \log_{10} 5000^3 = 11.09$$

By reusing existing CMU PGS guess numbers (value n), the final guess number for each passphrase becomes $min(n, m)$. This represents the "worst-case" scenario for guessing the password p, assuming the attacker chooses the best guessing method for p by chance. The final password guessibility distribution across the 3 conditions is shown in Table 5.15. We found the the number of “Safe” in PiXi and PiXi-hints has decreased (PiXi: 19.3% to

16.7%, PiXi-hints: 31.6% to 24.0%) while the number of "Offline-unSafe" has increased (PiXi: 73.5% to 76.1%, PiXi-hints: 65.5% to 57.9%). We also re-performed a χ^2 test ($df = 4, N = 238$) to examine the results, the results still showed a significant difference ($\chi^2 = 30.47$, $P < 0.001$) with a medium effect size (Cramer's $V = 0.253$) across different conditions. When a passphrase consists of 3 words from COCA 5000, attackers may opt for permutations of COCA 5000 as the primary approach to effectively crack passwords, and then utilize other password cracking tools subsequently.

	Online-Unsafe	**Offline-Unsafe**	**Safe**
Control	18.3%	74.7%	7%
PiXi	10.8%	78.4%	10.8%
PiXi-Hints	15.5%	60.5%	24.0%

Table 5.15: Updated password guesses distribution for the Control, PiXi and PiXi-Hints after adding in the possibility of a passphrase guesser based on COCA's top 5000 wordlist. (described in Section 5.2.3)

Challenges Mitigation To address the above passphrases challenges, we propose the following suggestions:

1. *Increase user awareness.* It is crucial to educate users about the importance of generating strong and unique passphrases using text or video materials, emphasizing the potential risks associated with using solely keyword-based passphrases.

2. *Encourage passphrase complexity.* Promote the use of more complex and diverse words in passphrases, discouraging users from relying solely on keywords and encouraging modifications such as combining words or adding special characters.

3. *Untraceable content generation.* In real-life products, implement the use of random seeds to provide genuine randomness in generating visual content for each user. This measure will effectively thwart attackers attempting to track users' histories.

However, it is very important to note that the data obtained from Amazon MTurk was not sufficient for drawing definitive conclusions, and the results could be subject to change when using other corpus databases and if attackers know adversarial is using a password or a passphrase. Consequently, we plan to gather more data and conduct further analysis in the future to gain a deeper understanding of passphrase security and the associated threat models.

5.3 Usability Analysis

We analyze the usability of PiXi and PiXi-Hints, according to register times 5.3.1, SUS score 5.3.2, system rating 5.3.3 , login rates, login times 5.3.4, Page using times breakdown 5.3.5 and user feedback and comments 5.3.6.

	Mean	**N**	**Std. Deviation**
Control	49.61	71	59.71
PiXi	208.63	83	281.25
PiXi-Hints	249.85	84	240.44
Total	175.74	238	236.13

Table 5.16: Mean register time across the three conditions.

5.3.1 Register Times

Once a participant complete the consent form, the system automatically start a timer to record the using time until user click the "register" button in the register page. We want to compare the register time for each condition, so we calculated the means in Table 5.16. The register time for control condition (49.61s) is significantly lower than PiXi (208.63s) and PiXi-Hints condition (249.85s), while PiXi and PiXi-Hints are similar. The average PiXi using time for account registration is around 180 seconds.

SUS Score	Grade	Adjective Rating
> 80.3	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
< 51	F	Awful

Table 5.17: General guideline on the interpretation of SUS score [11].

5.3.2 SUS Score

To measure the usability of the Control, PiXi and PiXi-Hints, we compare the System Usability Scale (SUS)— a commonly-used questionnaire to measure the usability of a system [11]. SUS consists of 10 questions with 5 options to choose from that were asked in our Session 1 questionnaire. The SUS evaluation metrics are shown in 5.17. As shown in Table 5.7, the SUS score is very close across conditions, supporting that PiXi has no noticeable usability impact. Although the SUS score is relatively low for PiXi and PiXi-Hints (comparable to Control), this indicates that although PiXi added some steps prior to password creation, that users were not bothered by these steps. As described further below, this may be due to increased user satisfaction that PiXi facilitates creating secure and memorable passwords. To compare PiXi's usability to password meters, where it was found that users were more likely to report creating a password that meets the requirements was difficult [73], we report the relative agreement to Question 8: "The password creation method in this study was easy to use." Our results indicate that PiXi (4.03 ± 0.822) and PiXi-Hints (3.95 ± 0.764) users are more likely to agree the system is easy to use than Control (3.81 ± 0.903), where 1 indicates strong disagreement and 5 strong agreement.

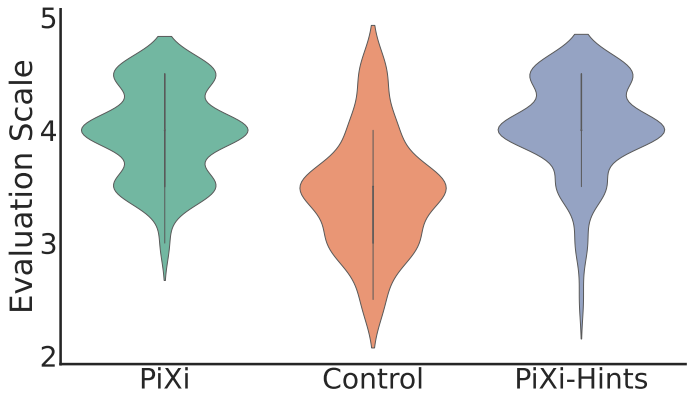


Figure 5.5: The violin plot of user overall rating distributions for three conditions. PiXi and PiXi-Hints users have a similar score distribution, with the majority of users reporting scores of 4 or higher, while Control users have scores concentrated between 3 and 4.

5.3.3 Overall System Rating

To determine the extent to which participants value each password system/process, we asked users their level of agreement with the question "I believe this password creating method helped me to choose a secure and memorable text password." Figure 5.5 gives a visual representation of the distribution of the answers, where 5 is for strongly agree, and 1 for strongly disagree. The users of PiXi or PiXi-Hints (with averages of 3.95 and 4.05) report higher levels of agreement compared to those in the Control condition (with an average of 2.9). Thus, PiXi and PiXi-Hints systems were successful at inspiring/nudging users to select secure and memorable passwords.

5.3.4 Login Rates and Times

The Memorability Test is an essential component of the PiXi system, as it helps to evaluate the effectiveness of our approach in generating memorable passwords. Unfortunately, we encountered some issues during Session 2 with the Amazon Mechanical Turk robots, which prevented us from conducting the memorability test as originally planned. However,

	Control	PiXi	PiXi-Hints
Login time	14.87 ± 7.38	27.68 ± 22.1	139.5 ± 36.08
Login success rate	7/10 (70%)	8/9 (88.9%)	10/12 (83.3%)

Table 5.18: Login data for each condition.

we were able to gather some data on participants' password recall rate, which we present in this section. We analyze our login data from Session 2 for indications of usability and memorability problems in each condition. While the MTurk return rate was low for Session 2, we believe exploring this information can still provide useful insights about system memorability. Table 5.18 shows the login success rates (over 3 login attempts) and login time. While the Control group has a higher rate of login failure, we only see this as an indication that PiXi shows promise for helping create stronger and possibly more memorable passwords, and as such further study is required for any concrete statistical analyses.

As shown in Table 5.18, PiXi-Hints with the additional hint task have higher login times compared to Control. However, surprisingly, PiXi requires a longer login time than Control, while Pixi users tended to require more than one login attempt, which increased the average login time. This issue should be analyzed in future work to determine whether it improves over successive logins or not.

5.3.5 Pages Using Times Breakdown

The using time for each page during registration process is shown in Table 5.19. We found that in the Category Page, a significant portion of time (average: 46.49 seconds) was dedicated to utilizing the Modal, where users engaged with instructional content or videos. During the Items Page interaction, users primarily invested substantial time (average: 30.54 seconds) in the process of item selection, while approximately 10 seconds were allocated to each keyword selection. In the Register Page, users devoting the majority of their time

		Average Time	Average Percentage
Category Page	Modal	$46.49s \pm 59.32s$	25.6%
	Category Selection	$10.89s \pm 7.07s$	6.00%
Items Page	Item Selection	$30.54s \pm 34.41s$	16.82%
	1st Keyword	$9.2s \pm 11.29s$	5.07%
	2nd Keyword	$10.28s \pm 10.64s$	5.66%
	3rd Keyword	$11.37s \pm 13.08s$	6.26%
Register Page	Keyword Splash	$6s$	3.30%
	Password Creation	$56.80s \pm 157.30s$	31.29%
Total		181.57s	100%

Table 5.19: Pages using time breakdown of each page.

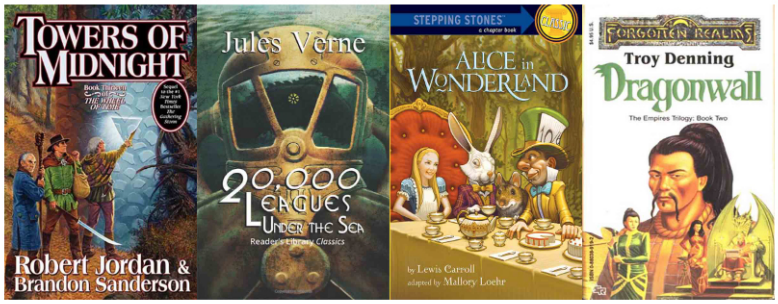
(average: 56.80 seconds) to the creation of passwords. These insights into time distribution provide valuable context for understanding user engagement and preferences within the registration process and inform a future direction on condensing the system, such as reducing keywords from 3 to 1, reducing items per page from 20 to 10, shorten modal video length, or presenting recommendations during password creations.

5.3.6 User Feedback and Comments

We did not force participants leave feedback for the both questionnaires, but we still collect and analysis all their free-form comments to investigate if they are satisfied with systems or have any suggestions to the user studies. For PiXi, we collect 3 comments, P1:“remembered part of my password but the PiXi didn’t give me all the keywords to recall all of it so I have no clue Sorry.”. P2:“I am not sure if I remember the password hardest part I guess I should of done better at is remembering the movie I think I remembered the right one but not sure but really like this method and if I use it a couple times would get better at knowing what I pick.” P3:“It was a little difficult to remember my password from

this survey but I was able to remember it after thinking about what I did in the 1st part of the study."; For PiXi-Hints, we collected 2 comments. PH1: "It was great.". PH2: "It was a interesting study to have taken part in."; For Control condition, we only collected 1 comment. C1:"Happy to participate."

Although the comments are not sufficient to do a quantitative analysis, we can still see that some participants still like PiXi (general positive), and without negative/abnormal comments. We will improve the questionnaire wording or requirements to collect more feedback from participants.



Rank: 1

Category: Books

Frequency: 5/41 (12.1951%)

Rank: 2

Category: Books

Frequency: 3/41 (7.3171%)

Rank: 3

Category: Books

Frequency: 3/41 (7.3171%)

Rank: 4

Category: Books

Frequency: 3/41 (7.3171%)

Rank: 5

Category: Books

Frequency: 3/41 (7.3171%)

Rank: 6

Category: Books

Frequency: 3/41 (7.3171%)

Rank: 7

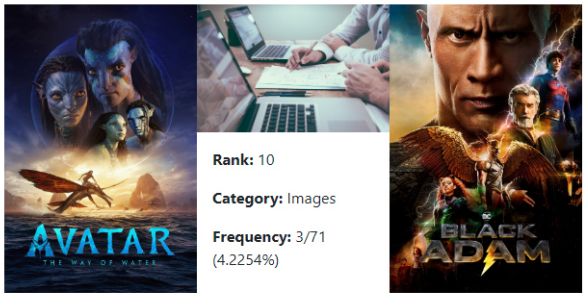
Category: Books

Frequency: 3/41 (7.3171%)

Rank: 8

Category: Books

Frequency: 3/41 (7.3171%)



Rank: 9

Category: Movies

Frequency: 3/55 (5.4545%)

Rank: 10

Category: Images

Frequency: 3/71 (4.2254%)

Rank: 11

Category: Movies

Frequency: 2/55 (3.6364%)

Figure 5.6: The most popular items (selected more than once) across PiXi and PiXi-Hints.

Chapter 6

Conclusions and Future Work

In this study, we introduced PiXi (Password Inspiration by eXploring Information), a novel approach designed to nudge users towards creating secure and memorable passwords. PiXi stands out as the first approach that employs a text password creation nudge to support users in crafting unique passwords independently. By encouraging users to explore unusual information just before password creation, PiXi aims to break free from their typical habits and thought processes, inspiring the generation of stronger and more resilient passwords.

Our study results, based on data from 238 participants, demonstrate the effectiveness of PiXi in nudging users towards secure password creation without explicitly instructing them to do so. Participants who used PiXi crafted significantly longer and more resistant passwords against password-guessing attacks. Remarkably, PiXi achieved these outcomes while maintaining a comparable overall perception to typical password creation systems. User feedback indicated that PiXi was helpful in creating more secure and memorable passwords, and unlike password meters, users found it easier to create passwords using PiXi.

Looking ahead, our future work will focus on several key areas of improvement. Firstly, PiXi presently only offers three categories due to the limited amount of time and ef-

fort, and one category, music, is still underdeveloped. We are going to add other categories in order to provide participants with more comprehensive resources and to encourage them to select a range of keywords in a number of different media, such as short videos, locations through maps, text and images in breaking news, blog posts, or audio recordings. We are also exploring potential adaptations for PiXi. One prospective transformation involves re-imagining PiXi as a password recommendation system, leveraging user-chosen keywords. It would also be interesting to study whether a shortened version of the PiXi system (e.g., involving only one keyword) could be equally effective at nudging users toward choosing secure passwords.

Secondly, PiXi is using 3 free-to-use APIs, two of which (Unsplash and TMDB) are in demo mode and have relatively low rate limits, which means they will cease to function for a period of time if an abnormally high volume of queries occurs in a short period of time. Additionally, if they are unavailable (either for maintenance or due to cyber-attacks), availability of PiXi will be greatly impacted. To overcome the aforementioned concerns, we intend to apply for production licenses for the future research, which entitles us to higher rate limits. Another option is to develop our own APIs, which may takes a long time and effort.

Thirdly, PiXi was created to motivate people to create more secure and memorable textual passwords. In other words, PiXi continues to support the traditional text password authentication framework, which has well-known issues with memorability and susceptibility. In future study, we will investigate PiXi variants that might be used in replacement of traditional text password authentication. For instance, a variation that enables participants login by clicking a sequence of chosen keywords form a category without typing passwords, another variation is modify PiXi to a password recommendation or suggestion system. PiXi takes keywords and categories from users and output different strong passwords suggestions (depends on categories) directly. Another possibility variation is that

PiXi only ask the user to select less keywords (1 or 2 words only) to reduce registration time.

Lastly, while our user study evaluated the effectiveness of PiXi and PiXi-Hints in generating secure and memorable passwords, we encountered limitations with data collection and memorability evaluation. Future studies should focus on other populations or enhanced methods to filter noise on MTurk, long-term recall rates and login times over successive logins. To address these issues, we plan to conduct another in-person user study in the future to provide more reliable data and assess the memorability of passwords generated by our proposed system.

Bibliography

[1] Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., and Wilson, S. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR) 50*, 3 (2017), 1–41.

[2] Addas, A., Salehi-Abari, A., and Thorpe, J. Geographical security questions for fallback authentication. In *International Conference on Privacy, Security and Trust (PST)* (2019).

[3] Addas, A., Thorpe, J., and Salehi-Abari, A. Towards models for quantifying the known adversary. In *Proceedings of the New Security Paradigms Workshop* (2020).

[4] Al-Ameen, M. N., Marne, S. T., Fatema, K., Wright, M., and Scielzo, S. On improving the memorability of system-assigned recognition-based passwords. *Behaviour & Information Technology* (2022), 1115–1131.

[5] Alkaldi, N., and Renaud, K. Why do people adopt, or reject, smartphone password managers?

[6] Bazerman, M. H., and Gino, F. Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual review of law and social science 8* (2012), 85–104.

[7] Biddle, R., Chiasson, S., and Van Oorschot, P. Graphical passwords: Learning from the first twelve years. *ACM Comput. Surv. 44*, 4 (2012), 1–41.

[8] Bonneau, J., Herley, C., Oorschot, P. C. v., and Stajano, F. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy* (2012).

[9] Breman, A. Give more tomorrow: Two field experiments on altruism and intertemporal choice. *Journal of Public Economics 95*, 11-12 (2011).

[10] Briggs, D., and van Moorsel, A. Nudging whom how: It proficiency, impulse control and secure behaviour. *Networks 49* (2014), 18.

[11] Brooke, J. SUS: A quick and dirty usability scale. *Usability Eval. Ind. 189* (1995).

[12] Brostoff, S., and Sasse, M. A. Are passfaces more usable than passwords? a field trial investigation. In *People and computers XIV—usability or else!* Springer, 2000.

[13] Cai, C. W. Nudging the financial market? a review of the nudge theory. *Accounting & Finance 60*, 4 (2020), 3341–3365.

[14] Camilleri, A. R., and Larrick, R. P. Metric and scale design as choice architecture tools. *Journal of Public Policy & Marketing 33*, 1 (2014), 108–125.

[15] Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2019).

[16] Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. Optimal defaults and active decisions. *The quarterly journal of economics 124*, 4 (2009), 1639–1674.

[17] Castelluccia, C., Dürmuth, M., Golla, M., and Deniz, F. Towards implicit visual memory-based authentication. In *Network and Distributed System Security Symposium (NDSS)* (2017).

[18] Chiasson, S., Stobert, E., Forget, A., Biddle, R., and Van Oorschot, P. C. Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism. *IEEE Transactions on Dependable and Secure Computing 9*, 2 (2012), 222–235.

[19] Chiasson, S., Van Oorschot, P. C., and Biddle, R. Graphical password authentication using cued click points. In *ESORICS* (2007), vol. 7.

[20] Collier, C. A. Nudge theory in information systems research a comprehensive systematic review of the literature. *Academy of Management Proceedings*, 1 (2018), 18642.

[21] Costa, D. L., and Kahn, M. E. Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*, 3 (2013), 680–702.

[22] Das, A., Bonneau, J., Caesar, M. C., Borisov, N., and Wang, X. The tangled web of password reuse. In *Network and Distributed System Security Symposium* (2014).

[23] Davis, D., Monrose, F., and Reiter, M. K. On user choice in graphical password schemes. In *USENIX Security Symposium* (2004), USENIX Association.

[24] De Angeli, A., Coutts, M., Coventry, L., Johnson, G. I., Cameron, D., and Fischer, M. H. VIP: A visual approach to user authentication. In *Advanced Visual Interfaces* (2002).

[25] Denning, T., Bowers, K., van Dijk, M., and Juels, A. Exploring implicit memory for painless password recovery. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2011).

[26] Dijksterhuis, A., Aarts, H., Bargh, J. A., and Van Knippenberg, A. On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology 36*, 5 (2000).

[27] Dunphy, P., and Yan, J. Do background images improve "draw a secret" graphical passwords? In *ACM Computer and Communications Security* (2007).

[28] Dupuis, M., Renaud, K., and Jennings, A. Fear might motivate secure password choices in the short term, but at what cost? In *HICSS* (2021).

[29] Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., and Herley, C. Does my password go up to eleven? the impact of password meters on password selection. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2013).

[30] Florêncio, D., Herley, C., and Van Oorschot, P. C. Pushing on string: The "don't care" region of password strength. *Commun. ACM 59*, 11 (2016), 66–74.

[31] Forget, A., Chiasson, S., van Oorschot, P. C., and Biddle, R. Improving text passwords through persuasion. In *Proceedings of the 4th Symposium on Usable Privacy and Security* (2008).

[32] Golla, M., Hahn, B., Selhausen, K., Hosseini, H., and Dürmuth, M. Bars, badges, and high scores: On the impact of password strength visualizations.

[33] Government of Canada. Password managers - get cyber safe. `https://www.getcybersafe.gc.ca/en/secure-your-accounts/password-managers#defn-password`. Online; Accessed: 2023-07-10.

[34] Hansen, P. G. The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation 7*, 1 (2016), 155–174.

[35] Hansen, P. G., and Jespersen, A. M. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation 4*, 1 (2013), 3–28.

[36] Hitaj, B., Gasti, P., Ateniese, G., and Perez-Cruz, F. Passgan: A deep learning approach for password guessing. In *Applied Cryptography and Network Security* (2019), R. H. Deng, V. Gauthier-Umaña, M. Ochoa, and M. Yung, Eds.

[37] Houshmand, S., and Aggarwal, S. Building better passwords using probabilistic techniques. In *Annual Computer Security Applications* (2012).

[38] Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., and Rubin, A. D. The design and analysis of graphical passwords. In *USENIX Security Symposium* (1999).

[39] Johnson, E. J., and Goldstein, D. Do defaults save lives?, 2003.

[40] Kahneman, D. Maps of bounded rationality: Psychology for behavioral economics. *American economic review 93*, 5 (2003), 1449–1475.

[41] Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. Getting to the top of mind: How reminders increase saving. *Management Science 62*, 12 (2016), 3393–3411.

[42] Katsini, C., Fidas, C., Raptis, G. E., Belk, M., Samaras, G., and Avouris, N. Influences of human cognition and visual behavior on password strength during picture password composition. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2018).

[43] Linux, K. Hashcat. https://www.kali.org/tools/hashcat/. Accessed: 2023-07-10.

[44] MacRae, B. A. Strategies and applications for creating more memorable passwords. Master's thesis, Ontario Tech University, 2016.

[45] Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences 108*, 26 (2011), 10415–10420.

[46] Mills, S. Personalized nudging. *Behavioural Public Policy 6*, 1 (2022), 150–159.

[47] Openwall. JohnTheRipper. https://www.openwall.com/john/doc/. Accessed: 2023-07-10.

[48] Parish, Z., Cushing, C., Aggarwal, S., Salehi-Abari, A., and Thorpe, J. Password guessers under a microscope: An in-depth analysis to inform deployments. *Int. J. Inf. Secur.* (2022).

[49] Parish, Z., Salehi-Abari, A., and Thorpe, J. A study on priming methods for graphical passwords. *Journal of Information Security and Applications 62* (2021), 102913.

[50] Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., and Frik, A. Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior 109* (2020), 106347.

[51] Polybius. English-Corpora: COCA. https://www.english-corpora.org/coca/. Accessed: 2023-07-10.

[52] Quigley, M. Nudging for health: On public policy and designing choice architecture. *Medical law review 21*, 4 (2013).

[53] Renaud, K., and Zimmermann, V. Nudging folks towards stronger password choices: Providing certainty is the key. *Behavioural Public Policy 3*, 2 (2019), 228–258.

[54] Salehi-Abari, A., Thorpe, J., and van Oorschot, P. On purely automated attacks and click-based graphical passwords. In *Annual Computer Security Applications Conference (ACSAC)* (2008), pp. 111–120.

[55] Schmidt, D., and Jaeger, T. Pitfalls in the automated strengthening of passwords. In *Annual Computer Security Applications* (2013).

[56] Seitz, T., von Zezschwitz, E., Meitner, S., and Hussmann, H. Influencing self-selected passwords through suggestions and the decoy effect. In *Proceedings of the 1st European Workshop on Usable Security. Internet Society, Darmstadt* (2016), vol. 2.

[57] Shay, R., Kelley, P. G., Komanduri, S., Mazurek, M. L., Ur, B., Vidas, T., Bauer, L., Christin, N., and Cranor, L. F. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (2012).

[58] Stobert, E., and Biddle, R. A password manager that doesn't remember passwords. In *Proceedings of the New Security Paradigms Workshop* (2014).

[59] Sunstein, C. R. Impersonal default rules vs. active choices vs. personalized default rules: A triptych. *Active Choices vs. Personalized Default Rules: A Triptych (May 19, 2013)* (2013).

[60] Sunstein, C. R. Nudges that fail. *Behavioural public policy 1*, 1 (2017), 4–25.

[61] Thaler, R. H., and Benartzi, S. Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of political Economy 112*, S1 (2004), 164–187.

[62] Thaler, R. H., and Sunstein, C. R. Nudge: Improving decisions about health, wealth, and happiness.

[63] Thaler, R. H., and Tucker, W. Smarter information, smarter consumers. *Harvard Business Review 91*, 1 (2013), 44–54.

[64] Thinking Kahneman, D. Fast and slow new york: Farrar. *Straus, and Giroux* (2011), 111–113.

[65] Thorpe, J., Al-Badawi, M., MacRae, B., and Salehi-Abari, A. The presentation effect on graphical passwords. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2014).

[66] Thorpe, J., MacRae, B., and Salehi-Abari, A. Usability and security evaluation of GeoPass: A geographic location-password scheme. In *Proceedings of the Symposium on Usable Privacy and Security* (2013).

[67] Thorpe, J., Salehi-Abari, A., and Burden, R. Video-passwords: Advertising while authenticating. In *New Security Paradigms Workshop* (2012), pp. 127–140.

[68] Thorpe, J., and van Oorschot, P. Graphical dictionaries and the memorable space of graphical passwords. In *USENIX Security Symposium* (2004).

[69] Thorpe, J., and van Oorschot, P. Human-Seeded attacks and exploiting Hot-Spots in graphical passwords. In *USENIX Security Symposium* (2010).

[70] Thorpe, J., and van Oorschot, P. C. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *USENIX Security Symposium* (2007).

[71] Thunström, L., Gilbert, B., and Ritten, C. J. Nudges that hurt those already hurting–distributional and unintended effects of salience nudges. *Journal of Economic Behavior & Organization 153* (2018), 267–282.

[72] Ur, B., Alfieri, F., Aung, M., Bauer, L., Christin, N., Colnago, J., Cranor, L. F., Dixon, H., Emami Naeini, P., Habib, H., Johnson, N., and Melicher, W. Design and evaluation of a data-driven password meter. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2017).

[73] Ur, B., Kelley, P. G., Komanduri, S., Lee, J., Maass, M., Mazurek, M. L., Passaro, T., Shay, R., Vidas, T., Bauer, L., Christin, N., and Cranor, L. F. How does your password measure up? The effect of strength meters on password creation. In *USENIX Security Symposium* (2012).

[74] Ur, B., Segreti, S. M., Bauer, L., Christin, N., Cranor, L. F., Komanduri, S., Kurilova, D., Mazurek, M. L., Melicher, W., and Shay, R. Measuring real-world accuracies and biases in modeling password guessability. In *USENIX Security Symposium* (2015).

[75] van Oorschot, P. C., Salehi-Abari, A., and Thorpe, J. Purely automated attacks on passpoints-style graphical passwords. *IEEE Transactions on Information Forensics and Security* (2010).

[76] Vance, A., Eargle, D., Ouimet, K., and Straub, D. Enhancing password security through interactive fear appeals: A web-based field experiment. In *System Sciences* (2013).

[77] VanEpps, E. M., Downs, J. S., and Loewenstein, G. Calorie label formats: Using numeric and traffic light calorie labels to reduce lunch calories. *Journal of Public Policy & Marketing 35*, 1 (2016), 26–36.

[78] von Zezschwitz, E., Eiband, M., Buschek, D., Oberhuber, S., De Luca, A., Alt, F., and Hussmann, H. On quantifying the effective password space of grid-based unlock gestures. In *Mobile and Ubiquitous Multimedia* (2016).

[79] von Zezschwitz, E., Eiband, M., Buschek, D., Oberhuber, S., De Luca, A., Alt, F., and Hussmann, H. On quantifying the effective password space of grid-based unlock gestures. In *Mobile and Ubiquitous Multimedia* (2016).

[80] Wang, S., Salehi-Abari, A., and Thorpe, J. Pixi: Password inspiration by exploring information. http://https://doi.org/10.48550/arXiv.2304.10728, 2023.

[81] Wheeler, D. L. ZXCVBN: Low-budget password strength estimation. In *USENIX Security Symposium* (2016).

[82] Wiedenbeck, S., Waters, J., Birget, J.-C., Brodskiy, A., and Memon, N. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies 63* (2005), 102–127.

[83] Yan, J., Blackwell, A., Anderson, R., and Grant, A. Password memorability and security: Empirical results. *IEEE Security & Privacy 2*, 5 (2004), 25–31.

[84] Yeung, K. "hypernudge": Big data as a mode of regulation by design. *Information, Communication & Society 20*, 1 (2017), 118–136.

[85] Yu, F., and Vargas Martin, M. Honeygan: Creating indistinguishable honeywords with improved generative adversarial networks. In *Security and Trust Management: International Workshop* (2023).

[86] Zibaei, S., Malapaya, D. R., Mercier, B., Salehi-Abari, A., and Thorpe, J. Do password managers nudge secure (random) passwords? In *Symposium on Usable Privacy and Security* (2022).

[87] Zimmermann, V., Marky, K., and Renaud, K. Hybrid password meters for more secure passwords – a comprehensive study of password meters including nudges and password information. *Behaviour & Information Technology* (2023), 700–743.

[88] Zimmermann, V., and Renaud, K. The nudge puzzle: Matching nudge interventions to cybersecurity decisions. *ACM Trans. Comput.-Hum. Interact. 28*, 1 (jan 2021).

Appendix A

Categories Pages

This chapter includes the full screenshot of each item page (Books, Movies, and Images).

Figure A.1: The full page of Images.

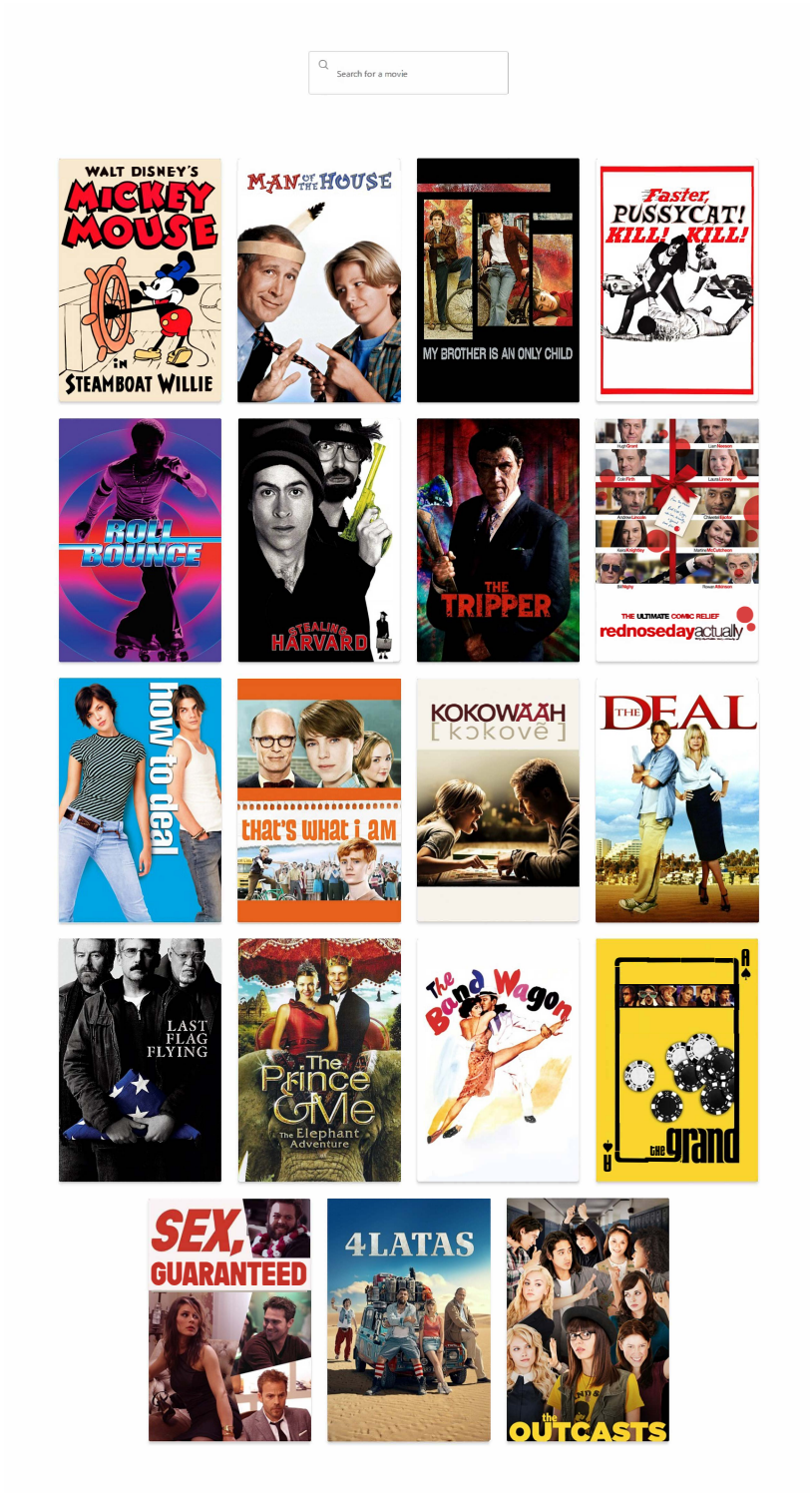


Figure A.2: The full page of Movies.

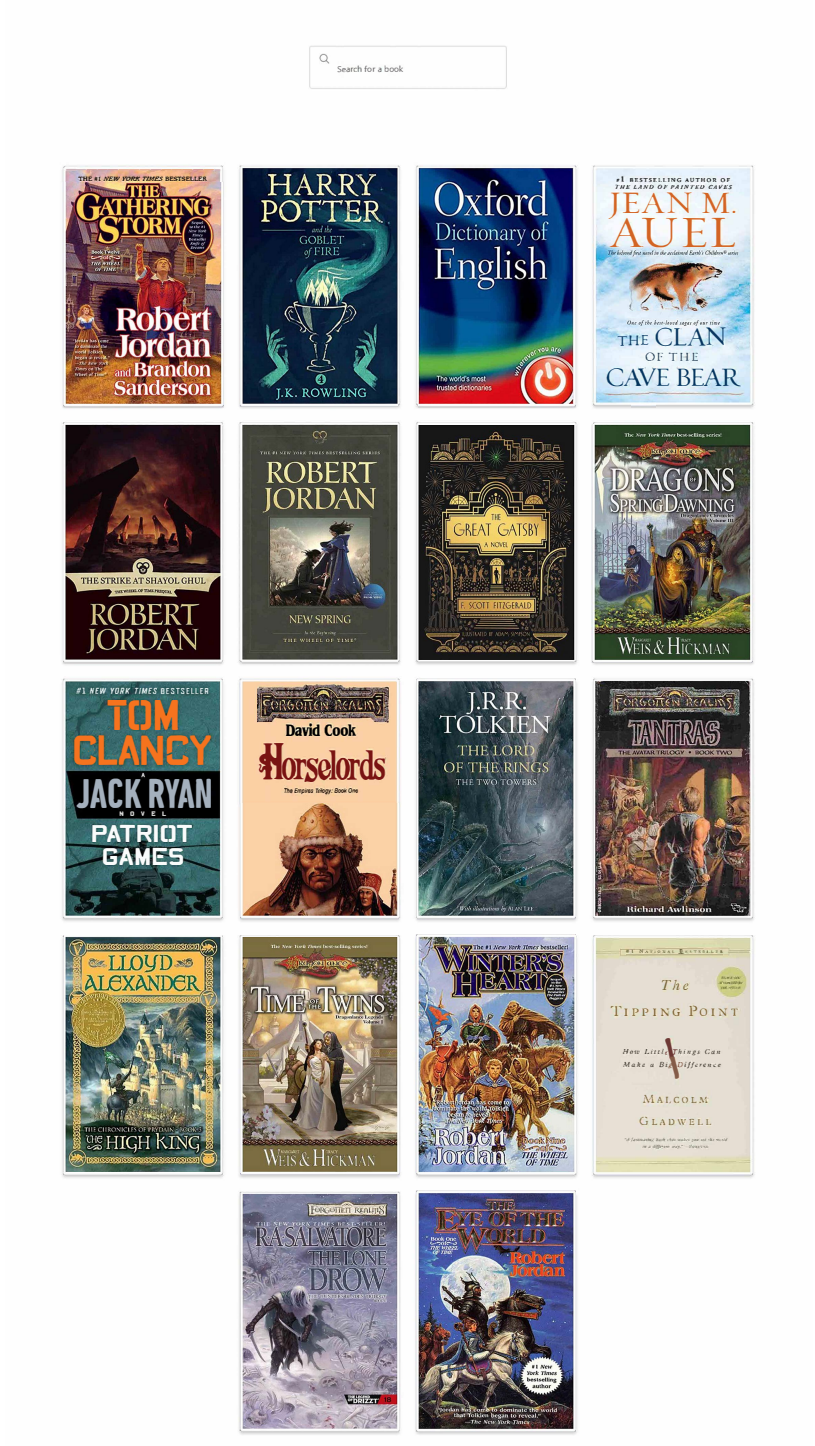


Figure A.3: The full page of Books.

Appendix B

Seed Generating Algorithms

No.	Formula	Description
1	CategorySeed = Math.seedrandom[Amazon MTurk ID + Category]	When generate a seed, the category is added to make sure the seed is different for the same user when she choose a different category.
2	CategoryKeyword = BookList[CategorySeed]	A category keyword is randomly selected from a English word list.
3	DefaultCategoryItems = APIRequest[CategoryKeyword] *20	The default 20 random items are fetched from the corresponding API.

Table B.1: Seed Generating Algorithm #1

No.	Formula	Description
1	KeywordSeed = [Amazon MTurk ID] + [Category] + [Search keyword]	A new seed is generated by adding a new element "Search keyword" when the user use the search bar.
2	SearchedCategoryItems = APIRequest[SearchSeed] *20	20 random items are fetched using the new seed from the corresponding API.

Table B.2: Seed Generating Algorithm #2

Appendix C

Book API List

1. The Gathering Storm - Chapter One
2. 20,000 Leagues Under the Sea
3. Alice in Wonderland
4. Beautiful and the Damned
5. Bible
6. Breakfast at Tiffanys
7. Clancy Tom - Patriot Games
8. Clancy Tom - Red Storm Rising
9. Clarke Arthur C - 3001 The Final Odissey
10. Cleric Quintet 1 - Canticle
11. Cleric Quintet 2 - In Sylvan Shadows
12. Cleric Quintet 3 - Night Masks
13. Crown of Fire
14. Crucible - The Trial Of Cyric The Mad
15. Dragon lance Preludes 2 vol 2 - Flint the King
16. Dragons of Autumn Twilight
17. Dragons of Spring Dawning
18. Dragons of Summer Flame
19. Dragons of Winter Night
20. Dragonwall
21. Earth's Children 01 - The Clan of the Cave Bear
22. Earth's Children 02 - The Valley of Horses
23. Earth's Children 03 - The Mammoth Hunters

24. Earth's Children 04 - Plains Of Passage
25. Earth's Children 05 - The Shelters Of Stone
26. Frank Herbert - Children of Dune
27. Frank Herbert - Dune Messiah
28. Frank Herbert - Dune
29. Hamlet
30. Harry Potter 1 - Sorcerer's Stone
31. Harry Potter 2 - Chamber of Secrets
32. Harry Potter 3 - The Prisoner Of Azkaban
33. Harry Potter 4 - The Goblet Of Fire
34. Harry Potter 5 - Order of the Phoenix
35. Harry Potter 6 - The Half Blood Prince
36. Harry Potter 7 - Deathly Hollows
37. Horselords
38. Maggie a girl of the streets
39. Other half lives
40. Pii
41. Pride and Prejudice
42. Prince of Lies
43. Robert Jordan - The Strike at Shayol Ghul
44. Robinson Crusoe
45. SCleric Quintet 4 - The Fallen Fortress
46. Shadowdale
47. Spellfire
48. Star Wars-Dark Force Rising

49. Star Wars-Heir to the Empire
50. Star Wars-The Last Command
51. Tantras
52. Test Of The Twins
53. The Age Of Innocence
54. THE Catcher In The Rye
55. The Fellowship Of The Ring
56. The Great Gatsby
57. The Legend Of Huma
58. The little Prince
59. The Lone Drow
60. The Return Of The King
61. The Second Generation
62. The Thousand Orcs
63. The Two Swords
64. The Two Towers
65. The Wheel of Time 01 - Eye of the world
66. The Wheel of Time 02 - The Great Hunt
67. The Wheel of Time 03 - The Dragon Reborn
68. The Wheel of Time 04 - The Shadow Rising
69. The Wheel of Time 05 - The Fires of Heaven
70. The Wheel of Time 06 - Lord of Chaos
71. The Wheel of Time 07 - A Crown of Swords
72. The Wheel of Time 08 - The Path of Daggers
73. The Wheel of Time 09 - Winters Heart

74. The Wheel of Time 10 - Crossroads of Twilight
75. The Wheel of Time 11 - Knife of Dreams
76. The Wheel of Time Prelude - New Spring
77. Time Of The Twins
78. To Kill a Mockingbird
79. The Tipping Point
80. War Of The Twins
81. Washington Square
82. Waterdeep
83. [Chronicles Of Prydain 1] Book of Three
84. [Chronicles Of Prydain 2] Black Cauldron
85. [Chronicles Of Prydain 3] Castle Of Llyr
86. [Chronicles Of Prydain 4] Taran Wanderer
87. [Chronicles Of Prydain 5] High King

Appendix D

Amazon MTurk Advertisement

Recruitment Material-Amazon Mechanical Turk Advertisement

We posted 2 requests (Account Registration Task) in Amazon Mechanical Turk. These advertisements are shown to qualified participants (i.e., from the United States and having an approval rate of >= 95%).

Session 1 Advertisement

Project Title: A study of online account registration methods - Session 1

Description: This study will require you to register a new account using either a new system or a traditional method. There are two sessions for this study. It will take approximately 7 mins for the 1st session, and you will receive $0.85 USD upon completion. You will be invited to come back after 7 days to login again and answer a questionnaire; this 2nd session will take approximately 1-2 mins and you will receive $0.35 USD upon completion. This study has been reviewed by Ontario Tech University Research Ethics Board (#16688) on (05/27/2022).

Session 2 Advertisement (advertised to those who completed Session 1)

Project Title: A study of online account registration methods - Session 2

Description: This is the 2nd session for the "A study of online account registration methods". Please use your login info created in Session 1 and answer a questionnaire; this 2nd session will take approximately 1-2 mins and you will receive $0.35 USD upon completion. This study has been reviewed by Ontario Tech University Research Ethics Board (#16688) on (05/27/2022).

Appendix E

Questionnaires

Session 1 questionnaire

Basic Information

A. Area of Work

   a. Architecture and engineering
   b. Arts, culture and entertainment
   c. Business, management and administration
   d. Communications
   e. Community and social services
   f. Education
   g. Science and technology
   h. Installation, repair and maintenance
   i. Farming, fishing and forestry
   j. Government
   k. Health and medicine
   l. Law and public policy
   m. Sales
   n. Prefer not to answer

B. Level of Education

   a. None, or less than secondary (high school)
   b. Secondary diploma (high school graduation)
   c. Bachelor's degree (three or more year program at a university, college, trade or technical school, or other institute)
   d. Master's degree, or professional degree

e. Doctoral level university degree (PhD)

f. Prefer not to answer

C. Gender

a. Male

b. Female

c. Prefer not to answer

D. Language

a. English

b. Other

c. Prefer not to answer

E. Age Range

a. Under 20

b. 20-30

c. 30-40

d. 40-50

e. 50-60

f. Above 60

g. Prefer not to answer

Questionnaire

1. I'm reusing passwords for my accounts (not including this system).

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

2. I'm using password manager(s) to help me manage passwords.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

3. I'm using random password generators to help me create passwords.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

4. Seven plus three = eight

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

5. I believe this password creation method helped me to create a unique password.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

6. I would use this password creation method in this study frequently.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

7. The password creation method in this study is unnecessarily complex.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

8. The password creation method in this study was easy to use.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

9. I need the support of a technical person to be able to use the password creation method in this study.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

10. The various functions in this password creation method were well integrated.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

11. There was too much inconsistency in the password creation method in this study.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

12. Most people would learn to use the password creation method in this study very quickly.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

13. The password creation method in this study is very cumbersome to use.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

14. I felt very confident using the password creation method in this study.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

15. I needed to learn a lot of things before I could get going with the password creation method in this study.

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6. (Prefer not to answer)

Session 2 questionnaire

- **I did not record my password in any format (Ex. password manager, save in a file, write it on paper) in this study.**

1 (strongly disagree) 2 (disagree) 3 (neutral) 4 (agree) 5 (strongly agree)

6 (Prefer not to answer)

Other comments (optional)

Appendix F

Consent Form

Consent Form to Participate in a Research Study

Title of Research Study: A Study of Online Account Registration Methods

Name of Principal Investigator (PI): Shengqian Wang

PI's contact email: Shengqian.Wang@ontariotechu.net

Departmental and institutional affiliation: Faculty of Business and Information Technology, Ontario Tech University.

Introduction:

You are invited to participate in a research study entitled A study of online account registration methods. Please read the information about the study presented in this form. The form includes details on study's procedures, risks and benefits that you should know before you decide if you would like to take part. You should take as much time as you need to make your decision. You should ask the Principal Investigator (PI) or study team to explain anything that you do not understand and make sure that all of your questions have been answered before signing this consent form. Before you make your decision, feel free to talk about this study with anyone you wish including your friends and family. Participation in this study is voluntary. This study has been reviewed by Ontario Tech University Research Ethics Board (#16688) on (05/27/2022).

Purpose:

This research aims to test whether a new user interface nudges users to choose secure and memorable passwords.

Procedures:

There are approx. 1000-1500 participants in this study. The research activity and data storage are occurring in Canada.

This study has 2 sessions:

Session	Study procedure	Duration of session
Session 1	1. Go to the register page, use our system to navigate through some book, movie, music, or image-related information (much like an internet search), and register an account. (2-3 mins) 2. Complete a questionnaire. (2-4mins)	4-7 mins
Session 2, 7 days after Session 1	1. Go to the login page and login. (0 - 1min) 2. Complete a questionnaire. (0 - 1 min)	1-2 mins

Please ensure that you answer the questionnaires on your own without any assistance from others. We would kindly ask you to take your time to carefully consider your answers for each of the questions.

Potential Benefits:

Aside from receiving compensation, you will not directly benefit from participating in this study. Your participation will contribute to knowledge about nudging in password systems, which may help researchers design more secure and memorable password approaches.

Potential Risks or Discomforts:

The greatest potential risk is a breach of the data we collect. Although we employ best practices for security and ensure your information will be stored with an anonymous identifier, we ask that you please do not use your existing passwords in this research study or reuse any passwords created in this study.

Our system will ask you to search and navigate through some book, movie, music, or image-related information, much like an internet search. It is unlikely, but possible, that some images or words shown by the system may contain visual or verbally uncomfortable information (e.g., covers for horror movies or violent images).

Use and Storage of Data:

The data includes demographic information and feedback (i.e., gender, age, and education level). All the data is anonymous, and the data doesn't include any personal, confidential, or valuable information.

Confidentiality:

Your MTurk ID will be kept confidential. Collected data will be anonymous and it will not include any information that reveals your identity. Your privacy shall be respected. No information about your identity will be shared or published without your permission, unless required by law. Confidentiality will be provided to the fullest extent possible by law, professional practice, and

ethical codes of conduct. Please note that confidentiality cannot be guaranteed while data is in transit over the Internet.

This research study includes the collection of demographic data which will be aggregated in an effort to protect your anonymity. Despite best efforts it is possible that your identity can be determined even when data is aggregated.

Voluntary Participation:

Your participation in this study is voluntary and you may partake in only those parts of the study in which you feel comfortable. You may also decide not to be in this study, or to be in the study now, and then change your mind later.

You may refuse to answer any question you do not want to answer, or not answer an interview question by saying, 'prefered not to answer'.

Right to Withdraw:

You can withdraw at any time simply by leaving our website and not submitting your data. If you withdraw from the research project prior to submitting your data at the end of a session, any data collected in that session will be removed. If you wish to withdraw after completing a session, you can do so within one week by contacting the researchers directly by email. You do not need to offer any reason for your withdrawal.

Compensation, Reimbursement, Incentives:

You will use your own personal computer to complete this study.

You will be paid $0.85 for completing online session 1, and $0.35 USD for online session 2 on Amazon Mechanical Turk after submitting your data.

Debriefing and Dissemination of Results:

After the study is complete, we will publish aggregated and anonymous data in a research paper. If you are interested in the study, please contact shengqian.wang@ontariotechu.net to provide you with the results of the study after the research is published. .

Participant Rights and Concerns:

Please read this consent form carefully and feel free to ask the researcher any questions that you might have about the study. If you have any questions about your rights as a participant in this study, complaints, or adverse events, please contact the Research Ethics Office at (905) 721-8668 ext. 3693 or at researchethics@ontariotechu.ca.

If you have any questions concerning the research study or experience any discomfort related to the study, please contact the researcher Shengqian Wang at shengqian.wang@ontariotechu.net.

By signing this form you do not give up any of your legal rights against the investigators, sponsor or involved institutions for compensation, nor does this form relieve the investigators, sponsor or involved institutions of their legal and professional responsibilities.

Secondary Use of Research for Future Research Purposes:

Please note, if you agree to participate (and do not withdraw from the study), your anonymous data may also be used for future research studies relating to our research on password authentication

systems, etc.

Consent to Participate:

1. *I have read the consent form and understand the study being described.*
2. *I have had an opportunity to ask questions and my questions have been answered. I am free to ask questions about the study in the future.*
3. *I freely consent to participate in the research study, understanding that I may discontinue participation at any time without penalty.*
4. *I understand the possible need for secondary research uses of my research data for future research use and provide consent for the use of my data to be used in future studies.*
5. *The study may contain graphic or textual content that may be inappropriate or offensive to some users.*
6. *Discretion while searching and exploring content on the system is advised.*

☐ **I agree**

☐ **I understand and wish to proceed**