Exploring Usability and Security Perceptions of Panic Passwords Against Coercion

Christina Nissen[0009−0005−7912−9249] and Oksana Kulyk[0000−0003−4218−1658]

IT University of Copenhagen, Copenhagen, Denmark
{chfn,okku}@itu.dk

Abstract. Digital accounts are often preferred targets of adversaries. An underrated threat is coercion, where adversaries force users to reveal their credentials. *Panic passwords* have been suggested to counter this human-targeted threat, enabling users to apply a credential that looks legitimate but triggers a covert security response. We conduct the first evaluation of a panic password-based authentication scheme in non-coercive use via two empirical studies (total $N = 368$). The scheme's dictionary keywords, that is, credentials belonging to a particular theme, show recall rates up to 94% after two weeks and are perceived as a secure protection. However, the responses also reveal gaps in the participants' understanding of the scheme and concerns about their predictability. Our results show that while the keyword-based scheme can be valuable, efforts in user education and application of these keywords is required to ensure effective usage in normal authentication.

Keywords: Coercion · Authentication · Panic Password · Perceptions

1 Introduction

Protecting digital accounts is crucial, as many contain sensitive information such as financial or health records. This information makes digital accounts an attractive target for attackers. One human-targeted strategy is coercion, where adversaries force users to authenticate in presence of the coercer or share their access credentials with the coercer, and thereby gain access. A 2024 survey finds that 43% of individuals have felt pressured to share their accounts with their partners [26]. This emphasises the need for authentication systems that protect users from this threat. However, traditional authentication systems do not provide such protection.

A number of proposals have been made to ensure coercion-resistant authentication. One promising approach for such authentication is the concept of *panic passwords*, which are an alternative set of credentials that trigger a coercion alarm followed by hidden countermeasures. Different variants of panic passwords have been proposed in contexts such as online accounts, ATMs, and electronic voting [9, 10, 15, 36]. In particular, the keyword-based scheme proposed by Stefanov and Atallah [36] aims to reduce the burden on the user in a stressful situation by separating the credentials into a regular password, which can be stored,

and a so-called *keyword* which has to be remembered without being stored, specifically designed to protect against coercion. This keyword is selected from a *dictionary*, an easily defined set of words within a given theme, one of which is assigned to the users. To authenticate, the users enter either their keyword in a normal authentication situation or another word from the same dictionary if they are being coerced. For additional security, the users are furthermore required to enter their regular password, which works in the same way as in traditional password-based authentication. While such a keyword-based scheme has a potential advantage in terms of cognitive load on the users, it has not been empirically evaluated yet.

To address this gap, we present the first human-centred investigation of the keyword-based scheme, focusing on *memorability*, *understandability*, and *perceived security and usability*. In our study, we consider the authentication process using the keyword-based scheme in a non-coercive scenario. In particular, we investigate two design choices central to the scheme: the choice of dictionary used for keywords and the method in which the keyword is generated to the user, namely, whether it is generated by the system or chosen by the user themselves. We consider both of these choices to be critical for the overall design of the scheme. The choice of the dictionary is critical in ensuring that the users have the capability to remember their keywords and properly apply them during authentication in both non-coercive and coercive situations. The choice of the password generation method can have further influence on both the memorability and security of the keyword. With the investigation of these choices, we therefore aim to answer the following research questions:

RQ1 How does the keyword-based scheme perform in terms of memorability, understandability and perceived security?
RQ2 How can the evaluated design choices (i.e., choice of dictionary and assignment method of the keyword) influence this performance?

To answer these questions, we run two online studies (Study 1 with $N = 213$ and Study 2 with $N = 155$) as between-subject experiments, with one study per either of the two design choices outlined above. Both studies were conducted in two stages, two weeks apart, to measure long-term memorability of the keywords. Across both studies, participants generally recalled keywords accurately and found them understandable, regardless of the dictionary or generation method. While dictionary choice had no significant effect, system-generated keywords were recalled more reliably than user-generated ones. Overall, usability and security perceptions were positive, but 6–20% of participants failed to recall their keyword after two weeks, and 30–46% struggled to understand what input to enter under coercion. Some participants also raised concerns about memorability and vulnerability to brute-force attacks due to small dictionary sizes. These results suggest that the keyword-based scheme shows practical potential but requires further research to improve memorability and clarity of security guarantees. Based on our results, we provide a set of recommendations for practical implementation of the keyword-based scheme.

2 Background and Related Work

Coercion-Resistant Authentication Traditional authentication systems provide no protection for digital accounts when users are forced to authenticate. To address this gap, several coercion-resistant mechanisms have been proposed [4,9, 14,20,22,39]. Several of them rely on *biometrics* [12,14]. Although these methods avoid knowledge-based secrets, they raise privacy concerns due to the sensitivity of biometric data. Another direction uses *implicit learning* to embed a secret in the user's memory without conscious awareness [4]. The user cannot voluntarily disclose the secret because they are unaware of it. However, this technique raises ethical considerations and transparency issues, as the learning process is hidden and users may be unaware of what information they carry. One line of work suggests different *multimodal approaches* to signal coercion [20]. These include biometrics (e.g., voice), a single-button press, or a specific sequence of actions, illustrating distinct input modalities rather than offering users multiple simultaneous options. However, the work assumes a time-limited adversary, making it less suitable for contexts, such as abusive relationships, where an adversary knows the victim. Of particular relevance to our study is the *fake credentials* approach, which is a type of knowledge-based authentication, allowing users to enter an alternative credential under coercion. Fake-credential approaches all rely on the assumption that the coercer is absent when the user receives the fake credential, and they can be grouped into three variants:

1. *One valid credential with all other inputs treated as alternative credentials:* This approach appears in coercion-resistant voting systems such as JCJ [22]. Clark et al. [9] also proposed the P-complement scheme, as well as the graphical 5-click scheme. However, since many inputs act as alternative credentials, the system cannot distinguish accidental inputs from intentional signals of coercion, which undermines usability and limits feasibility.
2. *One valid credential and one alternative credential:* This approach is deployed in the real-world disc-encryption tools TrueCrypt and VeraCrypt [19, 39]. However, if plausible deniability fails, the coercer can simply demand that people reveal both credentials. Similarly, Clark et al. [9] proposed the 2P and 2P-time-lock schemes.
3. *One valid credential and a set of alternative credentials:* This variant allows multiple, yet a limited amount of alternative credentials and has been suggested in prior work [9, 10, 15, 36]. These proposals, further, treat any input outside the set of alternative credentials and the true credential as invalid, resulting in a standard error message. In this study, we focus on this third variant, commonly known as panic-password schemes.

Panic Passwords Clark and Hengartner [9] were the first to formalise the concept of panic passwords. They introduced *5-dictionary*, in which users have five words from a dictionary, and any reordering of these words signals coercion, functioning as the set of panic passwords. Stefanov and Atallah [36] proposed a similar scheme, where users have a single keyword from a dictionary in addition to a regular password. Thus, users first enter a regular password, followed

by a keyword. While users are allowed to store the regular password (e.g., in password managers), they must be able to remember their keyword, as storing it could expose them to a coercer. An overview of the keyword-based scheme is provided in [29]. Beyond general-purpose schemes, a few context-specific panic password mechanisms have also been explored [10, 15]. Hameed et al. [15] introduced SafePass, a panic-PIN scheme for ATMs in which any change to the fourth digit of the PIN is interpreted as a coercion signal, while all other incorrect PINs are treated as invalid input. Clark et al. [10] proposed *Selections*, an Internet voting scheme implementing the concept of panic passwords. Although these authentication schemes implementing panic passwords offer a promising direction for coercion-resistant authentication, their feasibility has not been empirically evaluated. We address this gap by conducting an empirical study of the keyword-based scheme based on the proposal by Stefanov et al. [36].

Human factors in Knowledge-Based Authentication *Memorability* is a central challenge of knowledge-based authentication, and previous research shows that humans struggle with remembering random or complex passwords, graphical patterns, and multiple credentials across accounts [8,30,37]. Passphrases have been proposed to improve memorability [35,48]. However, user-selected and system-assigned passphrases both show poor long-term memorability and are associated with security misconceptions [35, 48]. Research in cognitive psychology and authentication find that humans recall user-generated items more easily compared to assigned ones [2, 7, 11, 25]. However, humans also tend to choose short and weak passwords, reused on multiple platforms [5, 46]. While allowing humans to select their own passwords could improve memorability, it also risks generating predictable choices that coercers might guess. Therefore, we investigate how the generation method in the keyword-based scheme affects memorability and security. Another factor is humans' *mental models* that strongly influence security behaviour, interpretation of risks, and decision-making during authentication tasks [13, 18]. Humans may adopt insecure practices, misunderstand feedback, or misuse security mechanisms if they hold inaccurate mental models. Previous studies have explored mental models in password reuse [46], implicit authentication relying on behavioural traits [24], and home computer security [45]. These demonstrate that humans' perceptions can differ substantially from technical realities. Compared to conventional authentication, the keyword-based scheme poses unique challenges. Unlike standard passwords, keywords should not be stored, are short meaningful words, and are used with systems that suppress certain error feedback to avoid alerting attackers. These differences may conflict with humans' mental models, making it critical to evaluate the practical viability of such a coercion-resistant solution. In addition to mental models, human's *perceptions of security* strongly influence adoption, as even technically robust mechanisms fail if they are perceived as unnecessary, confusing, or burdensome [3, 16, 30, 40]. Convenience often takes priority over security when mechanisms disrupt routines or appear complex [16, 30], and adoption is further shaped by perceived risks, trust in the system, and clarity of how mechanisms function [3, 40]. Mechanisms designed for special situations, such as

panic passwords to resist coercion, may face additional adoption barriers due to low perceived threat likelihood, usability challenges, and limited user awareness. These findings suggest that understanding how humans perceive security is essential for the adoption of a novel authentication mechanism. Therefore, we examine the perceived security of the keyword-based scheme.

3 Study 1: Dictionary Choice

3.1 Methodology

The goal of our first study was to assess the viability of an authentication scheme based on panic passwords for practical implementation in non-coercive situations. We chose the scheme by Stefanov and Atallah [36] because it requires fewer cognitive skills from users, needing only to recall a single keyword from a dictionary, compared to five words in the scheme proposed in [9]. Furthermore, the solutions in [10, 15] is limited to ATMs or internet voting, which is why we did not consider it for our evaluation.

Design of the scheme In order to evaluate the keyword-based scheme proposed in [36], we had to consider two design choices: the choice of the dictionary for the keywords and whether users are assigned a keyword by the system or select one themselves. We decided to focus on the dictionary in this first study because it is a defining element of the scheme, making it different from traditional authentication approaches. Users must recall a dictionary keyword at every attempt of non-coercive authentication, and this recall is a prerequisite for the scheme to function well in a real-world implementation. Thus, if the dictionary is not usable, the scheme cannot be meaningfully deployed. Consequently, this study also served as a feasibility evaluation, establishing whether the keyword-based scheme was usable in practice before exploring additional design choices. Below, we elaborate on the choice of dictionaries in this study.

Choice of dictionary We considered the requirements set by the authors of the original proposal, namely, (1) the dictionary must be well defined, meaning that users must be able to pick a random word and be sure that it is in the dictionary by only knowing the topic of it, and (2) keywords should have a high *typo quality*, that is, a large distance between keywords so that single typos cannot lead to mistakenly entering a panic password. We decided to use *animals*, *countries* and *colours* as our dictionaries. The selected dictionaries are potentially well-defined, satisfying (1); it is likely easy for users to come up with a panic password that they know is in the dictionary. To assess requirement (2), we computed pairwise edit distances within each dictionary[1]. We found three country pairs and 21 animal pairs with an edit distance of 1, and none among the colours. While there are some exceptions (e.g., "Ireland" vs. "Iceland"), these

[1] We used the 193 UN-recognised countries [41], a 571-animal list provided from GitHub [27], and the 11 basic colour terms described in the Berlin–Kay tradition and summarised by the International Society for Knowledge Organization (ISKO) [21] to compute distances.

words are few enough so that they can be removed from the dictionary. Furthermore, these dictionaries were relatively easy to assess due to their somewhat finite size, whereas broader dictionaries, such as those of verbs, would have been difficult to evaluate for high-typo quality. Therefore, while we acknowledge that there are other dictionaries that might be good, or perhaps even better, candidates for the scheme, we decided to use these three dictionaries.

Study Design We conducted a two-part online survey study with a between-subjects design[2]. The goal of the study was to compare the performance in terms of *memorability* and *understandability*. Furthermore, the goal was to understand users' perception of the security and usability of the keyword-based scheme. We chose between-subject design to avoid reducing recall accuracy due to confusion or increased cognitive load if participants had tried all dictionaries, which could otherwise also impact understandability due to learning effects. To identify potential issues, we conducted a test session with a researcher present to observe a user and a pilot study with 30 participants.

Study Procedure Our study consisted of two surveys spaced two weeks apart (see [29] for full questionnaire) [3]. We chose this interval following a prior study on password memorability in a normal authentication context [48]. Importantly, we evaluated the keyword-based scheme under typical, non-coercive usage. Thus, a two-week period is sufficient to assess how well participants remember their keywords in everyday authentication.

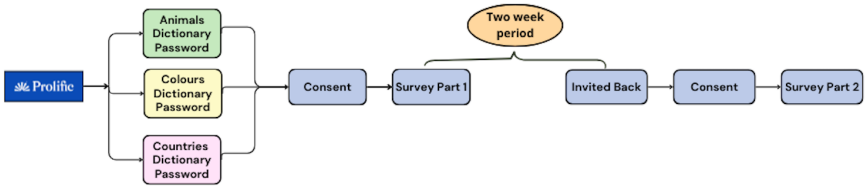

Fig. 1: Overview of study procedure

Survey: Part 1 Participants provided consent and were introduced to the keyword-based scheme with an example(see [29]). Keywords were called "thematic passwords" to provide a semantic cue. Due to platform limitations, passwords could not be dynamically generated per participant and were therefore fixed: the regular password was *Rm2VRGNrWp*, and keywords were *elephant*, *norway*, and *purple*. We embedded the assigned keywords as images to prevent copying, and the back button was disabled to prevent revisiting the keyword. To investigate *understandability*, participants then answered three multiple-choice questions, presenting different login scenarios in which they had to select the correct input. These scenarios capture the different types of inputs the system

[2] Due to extensive length, the full surveys, demographics, codebook, and statistical results for both studies are accessible in [29].

[3] Our surveys included additional questions that were not analysed in this paper and evaluated separately, see [28]

can receive, each producing a distinct system response: (1) entering their assigned keyword in a normal login situation, which results in a successful login; (2) entering a panic password in a coercive situation, which triggers a silent alarm; and (3) entering any unrelated word to the dictionary (e.g., 'pizza'), which produces an error message in both coercive and normal login situations. To assess *memorability*, participants then first entered their regular password into a free-text field, followed by their keyword. We asked participants to enter their regular password as well to imitate a real-world scenario but omitted the evaluation from the paper for the sake of brevity. We followed the methodology of a similar study [48], giving the participants five attempts. Subsequently, the participants completed the IPIP-NEO 120 personality inventory as a cognitive distractor (∼10 minutes). Here, we also included an attention check for quality control to follow Prolific recommendations [32]. Finally, they answered demographic questions and re-entered their regular password and keyword.

Survey: Part 2 We re-contacted the participants two weeks after the first survey. After providing informed consent, we gave them five attempts to enter the regular password and keyword. Afterwards, we asked them whether they had saved their keyword despite being instructed to refrain from doing so. Afterwards, we briefly explained the keyword-based scheme to them. This was provided after their password inputs to not affect the recall performance. Further, we presented them with the same three scenarios as in the first survey part to assess the understandability. Finally, we presented them with multiple-choice questions about how confident they were that the keyword-based scheme could prevent forced login, security compared to traditional authentication systems, and their confidence about memorising multiple keywords.

Variables Our independent variable was the *dictionary*. Our dependent variables were defined as follows. *Memorability* was measured as (a) correct keyword entry on the first attempt at three time points - (1) shortly after learning, (2) after a short distraction, and (3) after two weeks - and (b) *perfect recall*, defined as correct entry at all three points. *Understandability* was measured as (a) the number of correct answers per participant (range 0–3) and (b) the binary correctness of each scenario question (0 = incorrect, 1 = correct). Each survey provided one time point.

Recruitment and Ethics We recruited 300 participants via Prolific, randomly assigning 100 to each dictionary. A chi-square power analysis (df = 2, α = 0.05, Cohen's w = 0.3) indicated 172 participants were needed for 80% power. Anticipating some drop-out based on a reported 77% return rate in a similar study [48], we oversampled to ensure sufficient follow-up participation. Participants were UK or US residents, fluent in English, and received £2.6 for the first survey and £1.2 for the follow-up.

While our institution requires no formal ethical approval, we followed their guidelines. We had participants sign a consent form outlining the study, provided them with withdrawal options, and described our data handling (see consent form in [29]). Data were collected on a GDPR-compliant platform, and all identifying information was removed.

Data Analysis We used the statistical software R 4.3.0 (R Studio) to analyse our data. We included participants who completed both surveys, passed the attention check, and reported that they did not store their keyword. From 300 participants, we excluded seven participants for storing the keyword, 21 for failing the attention check, and 59 for not completing both surveys.

Memorability Keywords were coded as correct or incorrect, where an input had to match the letter sequence, case-insensitive (e.g., 'Elephant' = 'elephant'), reflecting the preprocessing a system can perform. Memorability on the first recall attempt was assessed using chi-square or Fisher's exact tests on counts of correct keywords per dictionary at each time point. We also conducted a logistic regression for *perfect recall*. This test was conducted to aggregate the repeated measures from the same participants.[4] We conducted a descriptive analysis of errors, distinguishing (1) panic password inputs, which are other dictionary words that would trigger a coercion response (e.g., "lion"), and (2) inputs causing failure and an error message (e.g., "mango2"). We analyse panic password errors both in terms of the proportion of all errors and the number of participants with at least one such error across dictionaries and time points.

Understandability For the number of correct scenario answers per participant (range 0–3) per time point, we used a Kruskal-Wallis test due to violation of the normality (Shapiro-Wilk $p = 9.688e - 14$) for ANOVA. For binary correctness of each scenario question (0 = incorrect, 1 = correct), we applied a generalised linear mixed-effects model (GLMM) with dictionary as a fixed effect and participants as a random effect, excluding time point and its interaction due to non-significance. This test accounted for repeated measures from the same participants.

Security and Usability Perceptions For the multiple-choice questions, we conducted a descriptive analysis with frequencies of responses grouped by the three dictionaries.

3.2 Results

We report results from 213 participants, with 74 assigned to the animals, 67 to the countries, and 72 to the colours conditions. Among participants, 132 were women, 80 were men, and one was non-binary. The biggest age group was 35-44 years, with 58 participants (see [29] for full demographics).

Memorability Table 1 summarises the rate of participants assigned to each dictionary correctly recalling their keyword on the first attempt at the three time points, as well as the number of participants with perfect recall. While the vast majority manage to recall their keyword correctly at the first two time points, 6% to 15% of participants fail to do so after the two-week period.

[4] A generalised linear mixed model (GLMM) with random effects could not be used due to non-normality.

Table 1: First-attempt keyword recall rates by dictionary and time point.

	Shortly after learning	**After distraction**	**After two weeks**	**Perfect recall**
Countries	65/67 (97%)	65/67 (97%)	57/67 (85%)	57/67 (85%)
Colours	69/72 (96%)	70/72 (97%)	68/72 (94%)	64/72 (89%)
Animals	73/74 (99%)	71/74 (96%)	65/74 (88%)	64/74 (86%)
Total	207/213 (97%)	206/213 (97%)	190/213 (89%)	185/213 (87%)

No statistical differences between the dictionaries are detected at any time point (testing via chi-square and Fisher's exact tests [5], p-values > 0.05) [6]. Furthermore, *perfect recall* reveal no significant differences between the dictionaries (using logistic regression, $p > 0.05$).

Errors Table 2 gives and overview of the participants' panic password errors, i.e., entering another dictionary word (before recalling the keyword or failing to recall it after five attempts). Overall, 27 participants entered a panic password. Across dictionaries and time points, panic password errors account for 55.6% of errors. By dictionary type, they represent 38.5%, 57.1%, and 77.8% of errors for the animal, country, and colour dictionaries, respectively (see table in [29]).

Table 2: Number of participants making a panic password error.

Dictionary	**Shortly after learning**	**After distraction**	**After two weeks**
Animals	1/74 (1%)	2/74 (3%)	4/74 (5%)
Colours	3/72 (4%)	1/72 (1%)	4/72 (6%)
Countries	2/67 (3%)	2/67 (3%)	8/67 (12%)
Total	6/213 (3%)	5/213 (2%)	16/213 (8%)

Understandability Mean scores for the number of correct answers per participant range from 2.04 to 2.48 out of 3 across time points for the dictionaries (see Table 3). The Kruskal-Wallis tests reveal no significant difference between the dictionaries for each time point (p-values > 0.05).

Table 3: Mean understanding scores (0–3) by dictionary and survey round.

Dictionary	**Part 1 Mean (SD)**	**Part 2 Mean (SD)**
Animals	2.22 (0.815)	2.04 (0.801)
Countries	2.48 (0.704)	2.27 (0.750)
Colours	2.28 (0.843)	2.18 (0.775)

Participants' understanding of inputs outside the dictionary is highest (88%), while their understanding of coercive scenario is lowest (57%) (see Table 4). Scores decline across survey parts for coercive and non-coercive questions but increase for inputs outside the dictionary.

The GLMM indicates that using the countries dictionary significantly increases the odds of a correct response compared to the baseline dictionary, which is the animals dictionary (OR = 1.67, 95% CI [1.07, 2.33], p = 0.026), while the

[5] Fisher's exact tests were used in place of chi-square tests for comparisons with small counts.

[6] The detailed information about all the statistical tests performed is included in [29]

Table 4: Rates of correct answers per understandability question across survey rounds.

Understandability Question	Part 1	Part 2
Non-coercive scenario	160/213 (75%)	152/213 (71%)
Coercive scenario	150/213 (70%)	121/213 (57%)
Inputs outside dictionary	184/213 (86%)	187/213 (88%)

colours dictionary does not differ significantly from the baseline (OR = 1.23, 95% CI [0.80, 1.65], p = 0.351), as shown in Table 5. However, the post-hoc pairwise comparison with the Tukey method, adjusting for multiple comparisons, shows no significant differences between the dictionaries [7].

Table 5: GLMM predicting correctness on understandability questions across dictionaries (random intercepts for participants; baseline = *animals*). Participant intercept variance = 0.765. Significance: *** $p < 0.001$, * $p < 0.05$.

Predictor	Estimate	Std. Error	z value	Pr(>|z|)	Odds Ratio (OR)	95% CI
Intercept	1.0362	0.1554	6.667	<0.001***	2.82	[2.01, 3.97]
Countries	0.5131	0.2303	2.228	0.026*	1.67	[1.07, 2.33]
Colours	0.2057	0.2203	0.934	0.351	1.23	[0.80, 1.65]

Security and Usability Perceptions Most participants (154 of 213) are moderately confident or higher that keywords can prevent access under forced login, with colours showing slightly more very high confidence responses (17) than animals (9) or countries (5) (see figures in [29]). Furthermore, most participants (154 of 213) are moderately or very confident that keywords can prevent access under forced login, with colours eliciting more very high confidence responses (17) than animals (9) or countries (5). Moreover, keywords are generally perceived as more secure than traditional passwords (146 out of 213), where slightly more participants rate animals (14) and colours (17) as much more secure than countries (5). Regarding usability perceptions, confidence in remembering multiple keywords is low with 123 out of 213 participants feeling not confident or slightly confident. However, more participants report moderate confidence for colours (21) than for animals (14) or countries (9).

4 Study 2: Password Generation Method

4.1 Methodology

The goal of our second study was to assess the effect of the keyword generation method by comparing user-selected and system-assigned keywords. While user-generated passwords often are insecure [17, 42, 46], they may be easier to remember, which is critical since keywords cannot be written down to ensure coercion resistance. Building on Study 1, we focused on the colours dictionary

[7] The detailed information about all the statistical tests performed is included in [29]

to isolate the effect of the generation method, as this dictionary showed a slight memorability advantage after two weeks.

Study Design We conducted a two-part online survey with a between-subjects design, assigning participants to either a user-generated or system-generated keyword. As in Study 1, we compared memorability, understandability, and perceived security and usability. All participants were assigned the same regular password as in Study 1, while the system-generated group was assigned the keyword *purple* and the user-generated group chose a colour keyword themselves. Due to platform limitations, the system-assigned keyword was fixed, as in Study 1. We chose a between-subjects design for the same reasons as in Study 1. We also ran a pilot with 20 Prolific participants, which led us to present open-ended questions about usability and security before the corresponding multiple-choice questions to reduce priming effects.

Study Procedure The study procedure was largely consistent with that of Study 1 (see Section 3.1). Differences were as follows: (1) participants in the user-generated group chose a keyword themselves from the colours dictionary; (2) participants only had one attempt to recall the keyword due to high recall rates in Study 1; and (3) participants completed several additional multiple-choice questions on security and usability, along with two open-ended questions about security concerns related to this type of authentication method and their preferred keyword generation method.

Variables The measures for *memorability* and *understandability* were almost identical to those in Study 1. However, we only measured understandability as the number of correct answers per participant (range 0–3) in this study.

Recruitment and Ethics As in study 1, we recruited participants via Prolific and randomly assigned them to a keyword generation method. Inclusion criteria and ethical procedures were the same as in Study 1, and participants of Study 1 were furthermore excluded. Based on a power analysis and expected attrition, we recruited 200 participants. Participants received £3 and £1.20 for the first and the second survey, respectively.

Data Analysis From 200 participants, we excluded eight participants for storing the keyword, five for failing to choose a colour as their keyword, and 32 for not completing both surveys. We followed the same analysis approach as Study 1. However, we only used a Kruskal-Wallis test for understandability, as the GLMM for binary correctness of each scenario question could not be used due to unmet model assumptions. Similarly, normality assumptions for ANOVA were violated (Shapiro-Wilk $p = 7.741e{-}11$). Open-ended responses were analysed inductively: one researcher developed a codebook per question from 50 responses, a second researcher independently coded the same set, discrepancies were resolved through discussion, and one researcher coded the remaining responses (see codebook in [29]). To evaluate the security of user-generated keywords, we analysed the frequencies of the chosen colours.

4.2 Results

We report results from 155 participants, where 84 were assigned to a system-generated keyword and 71 to a user-generated keyword. Among participants, 69 were women, and 86 were men. The biggest age group was 25-34 years, with 38 participants (See [29] for full demographics.)

Blue is the most user-chosen colour as a keyword, with 38% (27 of 71) of participants choosing it (see figure in [29]). The colours green, purple, yellow, pink, and red have frequencies between 7 and 10% (5-10 of 71 participants choosing each colour). The renaming colours have been chosen fewer than five times.

Memorability Across all time points, recall rates range between 99% and 82% (see Table 6). Furthermore, while recall rates decline after two weeks, the system-generated keywords have a higher percentage of successful recalls (93%) compared to user-generated (82%).

Table 6: Keyword recall rates by generation method and time point.

	Shortly After Learning	**After Distraction**	**After Two Weeks**	**Perfect Recall**
User-generated	69/71 (97%)	70/71 (99%)	58/71 (82%)	56/71 (79%)
System-generated	82/84 (98%)	82/84 (98%)	78/84 (93%)	78/84 (93%)
Total	151/155 (97%)	152/155 (98%)	136/155 (88%)	134/155 (86%)

Chi-square and Fisher's exact tests indicate no statistical difference between the generation method at any time point for keywords (all p-values > 0.05). The logistic regression of perfect recall indicates a significant effect of the generation method on the recall rate ($p = 0.02$) (See Table 7). Participants with a system-generated keyword have a 93% probability of a perfect recall compared to 79% for those with a user-generated one.

Table 7: Logistic regression on perfect recall of keywords. Baseline: *System-generated*.

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**OR**	**95% CI**
(Intercept)	2.5649	0.4237	6.054	<0.001	13.0	[6.17 – 33.46]
User-generated keyword	-1.2476	0.5138	-2.428	0.015	0.29	[0.10 – 0.75]

Errors Table 8 provides the number of participants entering a panic password as an error. Overall, 15 participants enter a panic password at different time points, with 11 of these errors occurring after two weeks. Across generation methods and time points, panic password errors account for 57.7% of all errors. Moreover, panic password errors represent 50% of all errors for system-generated keywords and 62.5% for user-generated passwords (see table in [29]).

Understandability Mean scores for the number of correct answers per participant range from 2.08 to 2.34 of 3 across the two surveys (see Table 9). While the score declines for user-generated keywords between the surveys (2.34 to 2.17), the opposite pattern shows for system-generated keywords (2.08 to 2.13.) Our participants' understanding of inputs outside the dictionary is highest (88%), while understanding of coercive password use is lowest (54%) (see Table 10). Scores decline or remain stable across survey parts for coercive and non-coercive

Table 8: Number of participants making a panic password error.

Generation method	Shortly after learning	After distraction	After two weeks
User-generated	1/71 (1%)	1/71 (1%)	8/71 (11%)
System-generated	1/84 (1%)	1/84 (1%)	3/84 (4%)
Total	2/155 (1%)	2/155 (1%)	11/155 (7%)

Table 9: Mean understanding scores (0–3 correct answers) by generation method and survey round.

Generation Method	Part 1 Mean (SD)	Part 2 Mean (SD)
System-generated	2.08 (0.895)	2.13 (0.902)
User-generated	2.34 (0.827)	2.17 (0.910)

questions but increase for inputs outside the dictionary. The Kruskal-Wallis tests reveal no statistically significant difference between the generation method for understandability for each time point (all p-values > 0.05) [8]

Security and Usability Perceptions Overall, 61 out of 155 participants are confident or very confident that keywords can prevent access under forced login. More participants assigned a system-generated keyword (30) report being confident compared to those with a user-generated keyword (17)(see figures in [29]). Further, 88 out of 155 believe that attackers are unlikely to detect the use of a keyword, and 97 out of 155 consider the use of keywords as more secure compared to traditional passwords. Regarding usability, most participants consider keywords perfectly easy to remember and very likely to be recalled after one week, with 93 and 102 participants reporting each outcome, respectively (see figures in [29]). However, confidence in remembering multiple keywords across accounts is lower, with 91 out of 155 participants reporting no or slight confidence. Overall, 100 out of 155 participants are generally positive about using the scheme beyond the study, and 105 out of 155 agree or strongly agree with feeling more confident using user-generated keywords.

Open-Ended Questions Across all responses, *guessability* is the most dominant security concern, with 118 mentions. Participants fear that keywords will be too guessable by other people. In particular, they point out that there is little variety of choices in colours and that attackers can do brute force or social engineering. One participant says: *"Someone close to me could guess based on what colours they see me wear..."* In contrast, 36 mention that they have *no concerns*. Furthermore, *usability* is another concern, with 23 mentions. This is not strictly security-related, but participants still mention the possibility of forgetting keywords, writing them down to remember, or mistakenly triggering a silent alarm. Finally, 22 mentions belong to the *other* category. Amongst these, participants compare keywords to other security measures (e.g., 2FA), and raise concerns about the human risks of forced logins. A majority of 111 participants

[8] The detailed information about all the statistical tests performed is included in [29]

Table 10: Rates of correct answers per question across survey rounds

Understandability Question	Part 1	Part 2
Non-coercive scenario	115/155 (74%)	114/155 (74%)
Coercive scenario	101/155 (65%)	83/155 (54%)
Inputs outside dictionary	125/155 (81%)	136/155 (88%)

prefer user-generated keywords. The most dominant reason is *memorability*, with 94 mentions, as participants consider it easier to remember a self-chosen keyword. One participant says: *"Choose my own, so I can remember it."* Other reasons for preference towards a user-generated keyword include security (15), other (11), and usability (6). System-generated keywords are preferred by 30 participants. The most dominant reason is *security*, with 17 mentions, as participants consider them harder to guess, less predictable, and more resistant to brute-force attacks. Other reasons include usability (5 mentions), memorability (5 mentions), and other (5 mentions). A smaller number of participants have either no preference (10) or mixed opinions with no clear answer (4).

5 Discussion and Conclusion

Memorability Despite promising recall rates for the keyword-based scheme, we observe recall errors, indicating that mistakes are unavoidable. Consistent with this, prior work shows that nearly all account owners eventually require credential resets [6], and research on other types of authentication, such as paraphrases, also demonstrates failure in recall after a certain time [35,48]. Therefore, we recommend for systems implementing the concept of panic passwords to offer a recovery mechanism in the user interface that cannot be invoked by the coercer, such as in-person visits. The aspect of recall errors is even more relevant if the keyword-based scheme is implemented large-scale for multiple accounts, as managing several credentials can exacerbate errors due to memory interference [8], which aligns with our participants' reported low confidence in recalling multiple keywords. In particular, the high prevalence of panic-password errors in this non-coercive situation (55.6% and 57.7%) illustrates a practical issue with the keyword-based scheme, as entering a panic password would trigger the system's covert countermeasures, which in sensitive domains such as military systems could escalate unnecessarily. This shows that memorability is not just a usability concern but a key factor for operational viability. Thus, we recommend to only deploy panic password schemes for accounts with elevated coercion risks and significant potential consequences. The recall errors further suggest that some participants may have formed incorrect mental models of the panic-password mechanism. In particular, participants may have assumed that any dictionary word can be safely entered. Prior research has shown that incorrect mental models can influence user's security behaviour during authentication tasks [13,18,24,47]. Avoiding panic-password errors that arise from incorrect mental models is therefore critical to prevent false alarms in real-world deployment. Therefore, we recommend to educate users about the consequences of triggering a false alarms, emphasising that panic passwords will not produce visible feedback to avoid alerting a coercer. Additionally, the user interface should provide a clear prompt above the password input fields advising users not to enter arbitrary inputs if they cannot recall their credentials. In particular, it should clearly explain when error messages are displayed or intentionally suppressed, helping users understand the system's covert behaviour. Still, future work should examine whether

these panic password errors stem from misunderstandings of the scheme, experimental behaviour, or gaps in instructions. Interestingly, system-generated keywords being easier to recall contrasts with prior research that self-generated items typically are easier to remember [2, 7, 11, 25]. This also conflicts with our participants' perceptions, as most perceived user-generated keywords easier to recall. Overconfidence may explain the discrepancy, as participants may have invested less effort in memorising self-chosen keywords. Still, it remains an open question for future research to explore.

Understandability Our results suggest that some participants understood the concept and answered all scenarios correctly, while others struggled with only one correct answer (means = 2.04–2.48; SDs = 0.70–0.91). Furthermore, they indicate that difficulties are not tied to the two key design aspects (dictionary or generation method) but rather to the keyword-based scheme in itself. Participants understand the non-coercive scenario and inputs outside the dictionary scenario, as they resemble familiar password systems. Yet, they find the coercive scenario harder to understand (down to 54% in Study 2), as it differs from their mental models of how authentication works. This emphasises a problematic aspect of understanding the scheme, i.e., if users misunderstand how to avoid coercion, they might enter a word outside the dictionary, which would trigger an error that could expose their attempt at deception, or they might use their real keyword, giving the coercer what they want. Thus, we recommend training users in the concept of panic passwords to ensure they know how to act, even under coercion. Future work should investigate how to improve users' understanding of the keyword-based scheme to align their mental models, as these strongly affect their security behaviour and decision-making during authentication [13, 24].

Security perceptions Participants' generally positive attitudes toward the scheme suggest a promising baseline for user acceptance in real-world deployment, which is a key factor in the adoption of new security mechanisms [31, 50]. Nevertheless, their concerns about brute-force attacks are valid in certain scenarios, particularly when a coercer forces a user to reveal their regular password or otherwise gains access to it. In such cases, the attacker could attempt a brute-force attack on the associated keywords. The risk is especially high if the coercer knows the victim personally and the keywords are user-generated (e.g., guessing a favourite colour), particularly when combined with a small dictionary of likely choices. This is supported by participants' tendency to select the same colours (blue accounting for 38%), indicating that common choices are easily predictable. Consequently, system-generated keywords provide a clear security advantage over user-generated ones. Taken together with system-generated keywords having better memorability, we recommend for the keyword-based scheme to deploy system-generated passwords to ensure security and increase memorability. Still, a brute-force attack of the keyword would be infeasible without access to the regular password, provided it is sufficiently secure. This suggests a potential mismatch in participants' mental models: they may not fully understand that the regular password, rather than the keyword, protects against brute-force attacks, while the keyword's purpose is to provide protection under coercion. Thus, we

recommend for the keyword-based scheme that users should be clearly informed that the regular password prevents brute-force attacks. Furthermore, given the importance of the regular password and the known challenges associated with knowledge-based authentication [38, 43, 49], alternative authentication methods, such as biometric authentication [34] or password-less authentication [23, 33], could strengthen security in the keyword-based scheme without overburdening users. Still, it is unclear how the concept of panic passwords can be transferred to other authentication methods that future work should investigate. Somewhat paradoxically, participants felt it was unlikely that a coercer would detect their use of a panic password. This may reflect overconfidence in their own abilities, consistent with prior research showing that people tend to overestimate their security skills relative to their actual performance [1, 44]. However, it remains an open question whether this confidence holds under realistic conditions.

Limitations and Future Work Our study comes with several limitations. The study was not longitudinal; thus, it remains unclear how keywords are remembered over longer periods. As we relied on self-reports about storage of the keyword, we cannot verify their accuracy, which may have influenced recall. Attrition over the two-week period further complicates the results, as missing participants may have forgotten their keywords. We examined only three dictionaries. Other dictionaries (e.g., plants) may offer different usability or perceived-security benefits and should be explored. Study 2 assessed keyword-generation methods only within the colours dictionary, so it remains an open question whether the results are generalise to other dictionaries. We also studied the dictionary and generation method separately, leaving potential interactions unexamined. Moreover, the system-generated keywords in both studies were assigned, so fully random system-generated keywords remain untested; in particular, it remains an open question whether e.g. assigning more obscure colours would perform differently to a relatively common colour (purple) that was assigned in our studies. We used only one distinct word per dictionary and did not examine cases where small typos could transform a keyword into a panic password. Participants did not interact with a functioning authentication system implementing the keyword-based scheme. In addition, our study results relies heavily on self-reporting in a survey rather than actual behaviour. Future studies should therefore employ prototypes and usability testing of users actual behaviour to validate our results. We also did not simulate coercive situations, leaving an open question about how the scheme performs under stress. Another direction of future work is a more thorough security analysis of the scheme, including the quantification of the keyword space for chosen dictionaries.

Conclusion We conducted two empirical studies ($N = 368$) to evaluate the design of the keyword-based authentication scheme in non-coercive usage. While the scheme shows promising memorability, users struggle to understand the scheme, particularly in coercive situations. A majority of the errors made by users are panic passwords, which would trigger a covert security response in real-world usage. Taken together, these results highlight a clear need to educate users about the scheme to ensure effective usage. Furthermore, system-generated

keywords have better memorability while also having the security advantage of being less predictable for coercers to guess. Therefore, we recommend implementing system-generated keywords. Future work should investigate how to improve users' understanding of such a coercion-resistant solution.

Acknowledgments. The work reported in this paper was supported by Independent Research Fund Denmark, grant number 10.46540/2101-00006B

References

1. Ament, C.: The ubiquitous security expert: Overconfidence in information security. In: ICIS 2017 Proceedings. vol. 2 (2017), https://aisel.aisnet.org/icis2017/Security/Presentations/2
2. Benney, K.S., Henkel, L.A.: The role of free choice in memory for past decisions. Memory **14**(8), 1001–1011 (Nov 2006). https://doi.org/10.1080/09658210601046163
3. Bhattacharya, S., Panyam, S., Deshmukh, G., Gatala, S., Vemoori, V., Seth, D.: Integrating user experience and acceptance in authentication: A synthesis of technology acceptance model and user-centered design principles. International Journal of Computer Trends and Technology **72**(4), 15–23 (2024). https://doi.org/10.14445/22312803/IJCTT-V72I4P102
4. Bojinov, H., Sanchez, D., Reber, P., Boneh, D., Lincoln, P.: Neuroscience meets cryptography: crypto primitives secure against rubber hose attacks. Commun. ACM **57**(5), 110–118 (May 2014). https://doi.org/10.1145/2594445
5. Bonneau, J.: The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords . In: 2012 IEEE Symposium on Security and Privacy. pp. 538–552. IEEE Computer Society, Los Alamitos, CA, USA (2012). https://doi.org/10.1109/SP.2012.49
6. Bonneau, J., Herley, C., van Oorschot, P.C., Stajano, F.: Passwords and the evolution of imperfect authentication. Commun. ACM **58**(7), 78–87 (Jun 2015). https://doi.org/10.1145/2699390
7. Cheng, S., Ding, Z., Chen, C., Sun, W., Jiang, T., Liu, X., Zhang, M.: The effect of choice on memory: The role of theta oscillations. Psychophysiology **60**(12), e14390 (Dec 2023). https://doi.org/10.1111/psyp.14390, epub 2023 Jul 16
8. Chiasson, S., Forget, A., Stobert, E., van Oorschot, P.C., Biddle, R.: Multiple password interference in text passwords and click-based graphical passwords. In: Proceedings of the 16th ACM Conference on Computer and Communications Security. p. 500–511. CCS '09 (2009). https://doi.org/10.1145/1653662.1653722
9. Clark, J., Hengartner, U.: Panic passwords: Authenticating under duress. In: Proceedings of the 3rd Conference on Hot Topics in Security. HOTSEC'08, USENIX Association, USA (2008)
10. Clark, J., Hengartner, U.: Selections: Internet voting with over-the-shoulder coercion-resistance. In: Danezis, G. (ed.) Financial Cryptography and Data Security. pp. 47–61. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27576-0_4
11. Clark, M., Bond, T., Omer, M., Johnson, C., Snow, G.L., Seamons, K.: Password memorability: What matters most? In: Proceedings of the Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025). Seattle, WA, USA (August 2025), https://www.usenix.org/system/files/soups2025_poster33_abstract-clark.pdf

12. Dietz, F., Mecke, L., Riesner, D., Alt, F.: Delusio - plausible deniability for face recognition. Proc. ACM Hum.-Comput. Interact. **8**(MHCI) (Sep 2024). https://doi.org/10.1145/3676494
13. Fassl, M., Krombholz, K.: Why i can't authenticate — understanding the low adoption of authentication ceremonies with autoethnography. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3581508
14. Gupta, P., Ding, X., Gao, D.: Coercion resistance in authentication responsibility shifting. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. p. 97–98. ASIACCS '12 (2012). https://doi.org/10.1145/2414456.2414512
15. Hameed, S., Hussain, S., Ali, S.: Safepass: Authentication under duress for atm transactions. In: Conference Proceedings - 2013 2nd National Conference on Information Assurance, NCIA 2013. pp. 1–5 (12 2013). https://doi.org/10.1109/NCIA.2013.6725317
16. Hanamsagar, A., Woo, S.S., Kanich, C., Mirkovic, J.: Leveraging semantic transformation to investigate password habits and their causes. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 1–12. CHI '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3174144
17. Haque, S.T., Wright, M., Scielzo, S.: A study of user password strategy for multiple accounts. In: Proceedings of the Third ACM Conference on Data and Application Security and Privacy. p. 173–176. CODASPY '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2435349.2435373, https://doi-org.ep.ituproxy.kb.dk/10.1145/2435349.2435373
18. Herbert, F., Becker, S., Schaewitz, L., Hielscher, J., Kowalewski, M., Sasse, A., Acar, Y., Dürmuth, M.: A world full of privacy and security (mis)conceptions? findings of a representative survey in 12 countries. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3581410
19. IDRIX / AM Crypto: Veracrypt – free open source disk encryption. https://www.veracrypt.jp/en/Home.html (2025), accessed: 2025-11-18
20. Imanda, H., Rasmussen, K.: Nakula: Coercion resistant data storage against time-limited adversary. In: Proceedings of the 18th International Conference on Availability, Reliability and Security. ARES '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3600160.3600175
21. International Society for Knowledge Organization (ISKO): Colour naming: Berlin–kay tradition. https://www.isko.org/cyclo/colour.htm, accessed: 2025-11-21
22. Juels, A., Catalano, D., Jakobsson, M.: Coercion-resistant electronic elections. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society. pp. 61–70. ACM (2005)
23. Kepkowski, M., Machulak, M., Wood, I., Kaafar, D.: Challenges with passwordless fido2 in an enterprise setting: A usability study (2023), https://arxiv.org/abs/2308.08096
24. Koushki, M.M., Obada-Obieh, B., Huh, J.H., Beznosov, K.: On smartphone users' difficulty with understanding implicit authentication. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3411764.3445386

25. Lennartsson, M.: Evaluating the Memorability of Different Password Creation Strategies: A Systematic Literature Review. Bachelor's thesis (15 credits), University of Skövde, School of Informatics (June 2019), https://www.diva-portal.org/smash/record.jsf?pid=diva2:1321294, available from DiVA – Digitala Vetenskapliga Arkivet
26. Malwarebytes Labs: What's mine is yours: How couples share an all-access pass to their digital lives (Jun 18 2024), https://www.malwarebytes.com/wp-content/uploads/sites/2/2024/06/Whats_Mine_is_Yours_Malwarebytes_Report_2024.pdf, accessed: 2025-08-22
27. Nickelza: Animal names list (571 items). https://gist.github.com/atduskgreg/3cf8ef48cb0d29cf151bedad81553a54 (2024), accessed: 2025-11-21
28. Nissen, C., Kulyk, O.: The password you hope you never use: Use cases for duress authentication. In: Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing. pp. 435–439 (2025)
29. Nissen, C., Kulyk, O.: Exploring usability and security perceptions of panic passwords against coercion: Supplementary materials. https://anonymous.4open.science/r/Exploring-Usability-and-Security-Perceptions-of-Panic-Passwords-Against-Coercion-B480/README.md (2026)
30. Oesch, S., Ruoti, S., Simmons, J., Gautam, A.: "it basically started using me:" an observational study of password manager usage. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491102.3517534
31. von Preuschen, A., Benda, C., Schuhmacher, M.C., Zimmermann, V.: Fear, fun or none: A qualitative quest towards unlocking cybersecurity attitudes. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. CHI '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3706598.3713538
32. Prolific: Prolific's attention and comprehension check policy (April 2025), https://researcher-help.prolific.com/en/article/fb63bb, accessed: 2025-05-15
33. Rivera-Dourado, M., Pérez-Jove, R., Pazos, A., Vázquez-Naya, J.: Usability of passwordless authentication in wi-fi networks: A comparative study of passkeys and passwords in captive portals (2026), https://arxiv.org/abs/2603.25290
34. Sasse, M.A., Steves, M., Krol, K., Chisnell, D.: The Great Authentication Fatigue – And How to Overcome It. In: Cross-Cultural Design. pp. 228 – 239. CCD '14 (2014). https://doi.org/10.1007/978-3-319-07308-8_23
35. Shay, R., Kelley, P.G., Komanduri, S., Mazurek, M.L., Ur, B., Vidas, T., Bauer, L., Christin, N., 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. SOUPS '12 (2012). https://doi.org/10.1145/2335356.2335366
36. Stefanov, E., Atallah, M.: Duress detection for authentication attacks against multiple administrators. In: Proceedings of the 2010 ACM Workshop on Insider Threats. p. 37–46. Insider Threats '10 (2010). https://doi.org/10.1145/1866886.1866895
37. Stobert, E., Biddle, R.: Memory retrieval and graphical passwords. In: Proceedings of the Ninth Symposium on Usable Privacy and Security. SOUPS '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2501604.2501619
38. Stobert, E., Biddle, R.: The Password Life Cycle. ACM Transactions on Privacy and Security (TOPS) **21**(3) (2018). https://doi.org/10.1145/3183341

39. TrueCrypt Foundation: Truecrypt 7.1a: Free open source disk encryption software (2012), https://www.truecrypt71a.com/, accessed: 2025-11-18
40. Ul Haque, E., Khan, M.M.H., Fahim, M.A.A.: The nuanced nature of trust and privacy control adoption in the context of google. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3581387
41. United Nations: Growth in un membership. https://www.un.org/en/about-us/growth-in-un-membership (2025), accessed: 2025-11-21
42. Ur, B., Noma, F., Bees, J., Segreti, S.M., Shay, R., Bauer, L., Christin, N., Cranor, L.F.: "i added '!' at the end to make it secure": observing password creation in the lab. In: Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security. p. 123–140. SOUPS '15, USENIX Association, USA (2015)
43. Veras, R., Collins, C., Thorpe, J.: On the Semantic Patterns of Passwords and their Security Impact. In: Proceedings of the Network and Distributed System Security Symposium. NDSS '14, Internet Society (2014)
44. Wang, J., Li, Y., Rao, H.R.: Overconfidence in phishing email detection. Journal of the Association for Information Systems **17**(11) (2016). https://doi.org/10.17705/1jais.00442
45. Wash, R., Rader, E.: Too much knowledge? security beliefs and protective behaviors among united states internet users. In: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015). pp. 309–325. USENIX Association, Ottawa (Jul 2015), https://www.usenix.org/conference/soups2015/proceedings/presentation/wash
46. Wash, R., Rader, E., Berman, R., Wellmer, Z.: Understanding password choices: How frequently entered passwords are re-used across websites. In: Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS 2016). USENIX Association, Denver, CO, USA (2016), https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-wash.pdf
47. Wolf, F., Kuber, R., Aviv, A.J.: "pretty close to a must-have": Balancing usability desire and security concern in biometric adoption. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. p. 1–12. CHI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300381
48. Wu, X., Munyendo, C.W., Cosic, E., Flynn, G.A., Legault, O., Aviv, A.J.: User perceptions of five-word passwords. In: Proceedings of the 38th Annual Computer Security Applications Conference. p. 605–618. ACSAC '22 (2022). https://doi.org/10.1145/3564625.3567981
49. Zhang, Y., Monrose, F., Reiter, M.K.: The security of modern password expiration: an algorithmic framework and empirical analysis. In: ACM Conference on Computer and Communications Security. pp. 176 – 186. ACM conference on computer and communications security (2010). https://doi.org/10.1145/1866307.1866328
50. Zou, Y., Roundy, K., Tamersoy, A., Shintre, S., Roturier, J., Schaub, F.: Examining the adoption and abandonment of security, privacy, and identity theft protection practices. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. p. 1–15. CHI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376570