A Comparative Long-Term Study of Fallback Authentication Schemes

Leona Lassak
Ruhr University Bochum
leona.lassak@rub.de

Philipp Markert
Ruhr University Bochum
philipp.markert@rub.de

Maximilian Golla
CISPA Helmholtz Center for Information Security
golla@cispa.de

Elizabeth Stobert
Carleton University
elizabeth.stobert@carleton.ca

Markus Dürmuth
Leibniz University Hannover
markus.duermuth@itsec.uni-hannover.de

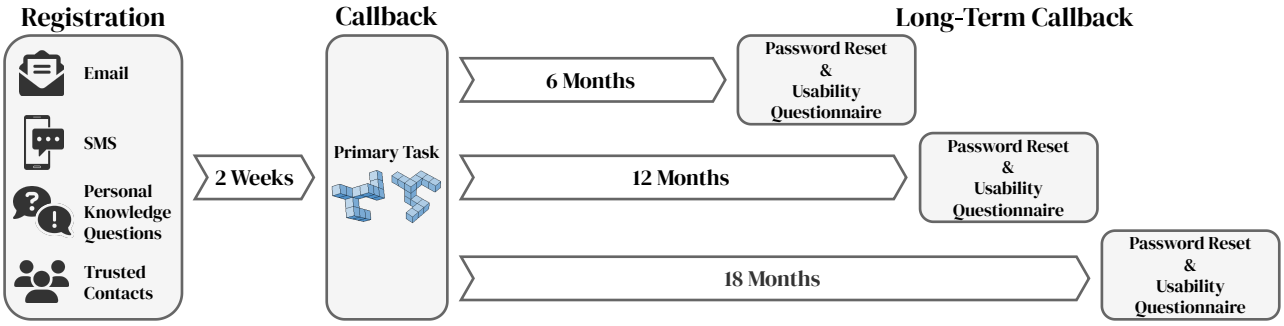


Figure 1: Structure of the conducted long-term study ($n = 97$). Participants were assigned to 1 of 4 fallback schemes (email, SMS, PKQ, trusted contacts) and a recall time (6, 12, 18 months). The 2-week callback served the purpose of minimizing dropout rates.

ABSTRACT

Fallback authentication, the process of re-establishing access to an account when the primary authenticator is unavailable, holds critical significance. Approaches range from secondary channels like email and SMS to personal knowledge questions (PKQs) and social authentication. A key difference to primary authentication is that the duration between enrollment and authentication can be much longer, typically months or years. However, few systems have been studied over extended timeframes, making it difficult to know how well these systems truly help users recover their accounts. We also lack meaningful comparisons of schemes as most prior work examined two mechanisms at most. We report the results of a long-term user study of the usability of fallback authentication over 18 months to provide a fair comparison of the four most commonly used fallback authentication methods. We show that users prefer email and SMS-based methods, while mechanisms based on PKQs and trustees lag regarding successful resets and convenience.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642889

CCS CONCEPTS

• **Security and privacy** → **Authentication**; **Usability in security and privacy**.

KEYWORDS

fallback authentication, email, SMS, personal knowledge questions

ACM Reference Format:
Leona Lassak, Philipp Markert, Maximilian Golla, Elizabeth Stobert, and Markus Dürmuth. 2024. A Comparative Long-Term Study of Fallback Authentication Schemes. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3613904.3642889

1 INTRODUCTION

Fallback authentication (also called account recovery, backup, emergency, or recovery authentication) is the mechanism for restoring access to an account if the primary authenticator becomes unavailable. It plays a central role in real-world account management, i.e., a study by Bonneau et al. showed that almost all account owners of the surveyed sample needed to reset their primary authenticator at least once [13]. Administrators have to deal with forgotten passwords and lost security tokens regularly [41, 75, 82], highlighting the need for usable mechanisms. Manual account recovery is usually a last resort as it comes at the highest cost for providers [23].

Fallback authentication creates another means by which an account can be accessed, so its security requirements are equivalent to

those of primary authentication systems. Even if the primary mechanism is secure, a weak fallback can compromise the security of an account. For example, attackers have abused the password reset function to gain access to lucrative cryptocurrency wallets [24, 63].

The most common fallback authentication mechanism is an emailed reset link sent to the user [53]. By clicking on the link, the user is directed to a page where a new password can be set. Other approaches require a reset code sent via SMS or answer previously set personal knowledge questions (PKQs). A different approach to fallback authentication is social authentication: in these schemes, peers of the account owner help to prove the owners' legitimacy, e.g., by providing codes they receive to their emails, which the account owner must collect and provide to complete the reset.

Several papers have studied the usability of fallback authentication schemes, but very few have examined more than a single scheme, making meaningful comparisons difficult [46, 77]. Additionally, research has usually focused on short periods of time [46, 81]. However, long spans between registration and reset are a central challenge of fallback authentication, as it can take months, if not years, until a reset is needed. Bonneau et al. [13] studied the usage of fallback authentication in the wild and found a nearly linear relation between the time passed and the share of fallback claims. After approximately 150 days (4.9 months), 30 % of the analyzed accounts initiated a reset, while 50 % did after 330 days (10.9 months), and 70 % after 540 days (17.8 months). Based on those observations, the following three research questions arise.

Given identical conditions and different realistic recall times:

RQ1 How do different fallback authentication schemes perform in terms of successful resets?

RQ2 How long do resets take for each scheme?

RQ3 How do users assess the schemes' usability and what issues arise?

To answer these questions, we conducted a long-term user study with 97 participants comparing the usability of four common reset schemes: (1) email, (2) SMS, (3) personal knowledge questions, and (4) trusted contacts. The structure of the study is shown in Figure 1. We used a between-subjects design with each participant using one of the schemes and had them reset their password after 6, 12, or 18 months. To provide a realistic study setting, we disguised the study as a test analyzing changes in spatial reasoning ability over time.

This work extends a *work in progress* report from 2019 titled, *"A Comparative Long-Term Study of Fallback Authentication"* [56] that outlines the study protocol (see Section 3.1) and focuses on preliminary results from our pilot study (see Section 3.5).

Our study showed that email resets were the most usable, as all participants successfully reset their passwords, and none reported any major issues. Similarly, most participants who used SMS resets did not report any problems and described the system as convenient. However, a few participants were unable to reset their password as they could not access the code sent to them. Finally, fallback authentication based on PKQs and designated trustees had the worst usability. Users had trouble remembering the answers to their PKQs, and successful resets in the designated trustee groups took a prohibitively long time if they were successful at all. Based on the results, we outline considerations service providers should make when providing fallback options to their users to allow for successful and convenient resets even after months or years.

2 BACKGROUND

Fallback authentication is used when the primary authenticator is unavailable, such as when a password is forgotten or an account is compromised. Even though fallback authentication is often the last resort before losing account access, most research was conducted between 2005 and 2017, and it has received little attention from the research community since then [13]. In fact, fallback authentication is considered a problematic issue even in the latest authentication schemes like Web Authentication (passkeys and FIDO2) [10, 51]. Often, resets are based on clicking a link in an email, which creates chains of trust and domino effects, causing problems for email providers and services or users that cannot use out-of-band communication like email or SMS [18, 50, 53]. Some research has evaluated fallback mechanisms on a high level. Maqbali et al. suggested a framework for systematically evaluating fallback authentication schemes from a security and usability standpoint [3]. AlHusain et al. conducted an extensive literature review with 70 articles but concluded that there is a lack of frameworks allowing proper comparison of fallback mechanisms [5].

2.1 The Fallback Setting

As fallback authentication is considered a last resort, it is not intended for daily usage but rather must be functional over long periods of time. Moreover, a fallback is expected to always work as no other option is left, and the danger of losing access is a stressful experience for users. These factors pose different requirements for the fallback authentication mechanism than there are for regular primary authentication schemes like passwords:

(1) *Long-Term:* The time between enrollment and authentication is almost always longer for fallback authentication. In contrast to passwords, knowledge-based fallback authentication suffers from poor memorability or outdated contact information that services often try to counter by prompting users to confirm their recovery details regularly.

(2) *Reliability:* As a last resort method, there is no other backup in place, underlining the need to register multiple recovery options and offer alternatives. In contrast to primary authentication, a failed fallback authentication can result in an unrecoverable state or create the need to contact a helpdesk and provide so-called *soft factors* to regain access, which is an error-prone and costly process for both the end user and the service provider.

(3) *Authentication Time:* Fallback authentication is intended to be a relatively infrequent action, thus, the required time for authentication can be longer than for primary authentication. Of course, there is a limit to what users are willing to endure and go through, and this may be correlated with the value of the account they are trying to recover.

A combination of those aspects is expressed by the *success rate*, i.e., the percentage of users that are able to recover their account, which can be used to benchmark different schemes. Additional protection mechanisms like CAPTCHAs or throttling (limiting the number of failed attempts), and obstacles such as temporal lockouts and the strictness of string verifications can further impact usability. The following section provides a comprehensive overview of common fallback authentication schemes.

2.2 Secondary Channel

One of the most common techniques is to use a secondary channel. For schemes that operate this way, the requirement is that the fallback is set up while the user still has access to the account.

Email. Email is by far the most common secondary channel and is used by over 90% of popular websites [53]. If access to the account is lost, the account recovery can be initiated using only the account name. The service provider sends an email containing reset information (e.g., a link, reset code, or even the password itself) to the registered email address. Clicking on the link or typing the code on the recovery website allows the user to set a new password. Strict rate-limiting is typically used alongside these mechanisms to prevent replay attacks, with codes and hyperlinks only being valid once or for a short time frame. Little research has explicitly focused on the usability of email as a fallback mechanism. Over the years, several studies have pointed out the various threats that come with the approach. Still, an extensive analysis by Li et al. in 2018 found that over 80% of popular websites employ no additional measures to prevent account access if an attacker has compromised the victim's email account, making it a single point of failure [53]. Others have shown that trusting the email ecosystem can be dangerous [78] as it usually mandates the proper configuration of security extensions and support of modern email authentication (SPF, DKIM, DMARC) and encryption technologies (TLS). Maqbali et al. [54] manually coded 50 popular English websites to identify potential issues with the emails, finding that many have poor instructions, email headers leak confidential information, and issues with spam filters.

SMS. Using text messaging (or Short Message Service (SMS)) as a secondary channel is very similar to email-based recovery, but instead of an email address, a phone number is linked to the account. An SMS is sent for account recovery, usually containing a reset code or, less commonly, links or temporary passwords. This approach can be less efficient than email-based authentication since users might need to type access codes manually. Additionally, this scheme requires possession of the phone, hence, a user whose phone is out of reach cannot use SMS for account recovery. In 2015, Bonneau et al. [13] did an extensive analysis of the memorability and security of Personal Knowledge Questions (PKQs) (see Section 2.3) with a Google dataset. Briefly, they compared PKQs' account recovery success rates (53%) to the success rates of SMS and email, revealing that the SMS-based scheme showed the highest recovery success (81%), followed by email as a close second (75%). Since their work focused on PKQs, Bonneau et al. provided no further insights into potential reasons for these differences. Our study extends this work by testing all methods in a controlled and comparable environment, reporting in-depth results about details, e.g., authentication timings, user perceptions, usability ratings, and reasons for potential errors. Other research studied the SMS-based approach's security from a theoretical perspective, pointing out concerns with network coverage in rural areas and SMS not being an inherently secure channel that can be spoofed [3]. The NIST also discourages the use of SMS as a *second factor for primary authentication* [30], yet it remains widely used for fallback authentication [53].

2.3 Knowledge-Based Authentication

Knowledge-based authentication describes a class of mechanisms that rely on something the user knows, i.e., known personal information like preferences or secrets.

Personal Knowledge Questions. The most common knowledge-based fallback authentication are "*cognitive passwords*," introduced by Zviran and Haga [94] in 1990. Nowadays they are known as security or personal knowledge questions (*PKQs*). They test the legitimacy of the user by asking them to answer questions with set responses about past experiences or demographic information. Typically, the questions are selected from a predefined list during account registration. Some services allow users to create security questions themselves as well. For account reset, the questions need to be answered whereby a certain variation may be allowed to tolerate different spellings. The initial research by Zviran and Haga [94] demonstrated a higher recall rate compared to a conventional password and found a low recall rate by even closely related persons. However, this conclusion originates from a time before social media and easily searchable online information. Newer studies [46, 67, 71, 76] show that many of the answers to PKQs are indirectly posted on the internet and that this approach cannot provide the initially claimed level of security.

This was further confirmed by Golla et al., who analyzed answers to 4 million PKQs from a leaked data set, concluding that the security level is low overall [27]. Just et al. proposed a framework for a systemic evaluation of security and memorability aspects of PKQs [45]. Others studied the general perception and creation behavior of PKQs concluding that users are mostly honest in their answers and disregard security in favor of memorability [13, 60]. The same studies also showed that the usability, in particular, the memorability of answers is concerningly low with 18% being unable to recall answers after only 20 days [46] and 40% after one year [13]. This is particularly concerning as long-term availability is a key requirement of fallback authentication. In an attempt to address the social media-induced security concerns with regular PKQs, a number of studies have explored the feasibility of PKQs about autobiographical information based on phone usage and sensor data [4, 34, 92]. Others explored using geographical information to generate dynamic PKQs [1, 35] or used nudging and memorization techniques to enhance memorability and security [7].

Recovery Codes, Keys, and Phrases. A relatively small number of services like Apple, Microsoft, and ProtonMail offer recovery code-based fallback authentication. While still having access to the account, the user receives an up to 28 character long recovery code that can be used to regain access in case the password has been forgotten. All companies recommend to "print this out and keep it in a safe place or take a picture of it," in order not to lose it [9].

2.4 Social Authentication

Social authentication describes mechanisms that rely on "who you know," i.e., information about one's social graph. Several variations of social authentication exist. Alomar et al. summarized those techniques in an extensive literature review [6]. In the following, we focus on techniques that have found real-world application.

Trust-based Techniques. The most prominent trust-based technique are *designated trustees*, first proposed in 2006 by Brainard et al. [15]. For setup, the user selects several contacts while still having access to the account. For account recovery, the trustees receive reset codes, and to regain access, the user needs to present a subset of the codes. Schechter et al. studied the same idea, but used email addresses to contact the designated trustees [77]. Their study suggested that the scheme is less efficient than other mechanisms [77]. Still, the new approach had a high success rate, with 17 out of 19 participants being able to complete the recovery process.

In October 2011, Facebook introduced *Trusted Friends* [19] which allowed users to select trusted contacts from active Facebook friends *after* the access was lost. Each of those contacts received instructions on how to obtain a reset code, three of which the user had to provide to regain access. However, the feature proved vulnerable to attacks that utilized recently added fake friends under attacker control [29, 43]. Thus, alternatively, Facebook introduced *Trusted Contacts* in May 2013 [20], which allowed selecting trustees only prior to recovery. This feature is discontinued since July 2022 as well [21]. Apple offers a trustee-based fallback called *Account Recovery Contact* [8] since September 2021. They recommend adding "someone you trust" like friends or family members who own an Apple device and can be easily reached either in person or via phone.

Research has explored further trust-based techniques such as using secondary information like PINs or biometrics to increase security or implicitly inferring users' trust relationships instead if users choosing trustees themselves [80, 91]. Guo et al. explored the usability and acceptability of video-call-based social authentication, finding major contextual influences of mood, location, and trust [31]. Stavova et al. compared trustee-based authentication to backup codes, finding that for higher-value accounts (i.e., online banking) trusted party recovery was preferred over codes [81].

Knowledge-based Social Authentication. A second group of social authentication schemes requires users to answer questions about their social environment, which ideally only the legitimate owner knows. The most common knowledge-based approach leverages photo-based information. Yardi et al. [90] first proposed the idea in 2008, basing a prototype implementation on Facebook. The system uses the social graph and other information like photos with tags of the shown persons to authenticate users. This is done by presenting photos from the user's database and asking questions, for example, the names of the photographed persons or the date the photo was taken. In 2011, Facebook adopted the idea to provide an additional barrier in case a suspicious login is detected. The fundamental concept is that an attacker, even if they manage to acquire the account password, would be unable to answer questions correctly as they do not possess knowledge of the associated social graph.

Several works have raised security concerns ranging from close friends being able to answer the questions as well [49] to automated attacks exploiting face recognition techniques [69, 93]. To enhance the resilience of photo-based methods against automated attacks, Polakis et al. [68] developed a countermeasure that reduces image quality to outperform recognition algorithms while the user's ability to recognize it is retained. Jain et al. [42], suggested utilizing other forms of social knowledge by creating challenge questions based on three elements representing the social graph.

2.5 Supplementary Security Mechanisms

Extending the above, research has proposed numerous alternative fallback mechanisms. Claiming better memorability for graphical compared to textual information, some suggest different versions of graphical or interface-based fallback techniques [32, 33, 38, 59], user-interface individualizations [47], self-assembling protocols [39], or behavioral biometrics [88].

Help Desk. As a last resort, some services offer help desks for those struggling with fallback mechanisms [23]. However, employing support personnel and maintaining help desks is costly [41, 73] and *soft factors* used for authentication are prone to *targeted* attacks [40]. Common soft factors are name, address, date of birth, parts of registered credit card numbers (Microsoft), account usage details like the account registration date (Google), or, "contacts you've recently sent emails to." Parkin et al. showed that users prefer self-service online password resets over help desk interactions despite a 4:1 ratio of failed-to-successful account recoveries [64].

Browser Fingerprinting. Browser fingerprints consist of details about the user's browser, location, and device configurations (i.e., IP address, language settings, screen resolution, hashes of browser plugins). On every website visit, fingerprints are compared to previous sessions, assuming an attacker cannot precisely mimic real sessions. Despite browser fingerprints being considered short-lived (only stable for 3 to 6 weeks [70]), with limited utility for fallback authentication, in 2018 Google disclosed using them to drive authentication decisions [62]. They recommend users to "use a device where [they]'ve signed in before" and "choose a familiar Wi-Fi network, such as at home or work" when resetting passwords [12].

Proactive Measures. Services employ various tools to improve the success rate of fallback authentications proactively. *Up-2-date checks*, deployed by eBay, GitHub, and Yahoo, prompt users with "Don't get locked out! Review your account recovery info." to confirm that the stored account recovery information, i.e., phone number or email address, is still correct. Others utilize opportune moments, such as account security checkups, asking to register additional *alternative recovery options*, like a recovery phone number, alternative email address, or security key, "in case you accidentally get locked out" (Google). Finally, *notifications* for account security-related updates like password changes or registrations of new recovery options create awareness and help users notice potential account compromise and regain access [57].

2.6 Long-Term Studies

Bonneau et al. compared several recovery schemes [14], including email, SMS, designated trustees, and PKQs but only synthesized individual analyses [15, 76, 85]. A subsequent comparative work by Bonneau et al. [13] demonstrated a higher recovery rate for SMS (81%), and email (75%) than for PKQs (61%). However, they disregarded the time between account creation and recovery claim. Raponi et al. compared whether websites adapted password management and fallback policies in a long-term evaluation but did not run a user study [72]. Research on the usability of fallback mechanisms mostly considered only one scheme at a time and covering durations of 3 months or less [4, 32, 38, 46, 92] or at most half a year [76].

Table 1: The considered fallback authentication schemes as well as the security assumption they rely on.

Scheme	Description	Security Assumption
Email	Click on reset link sent to registered email account	Secrecy of the channel and access to the email account
SMS	Provide reset code sent via SMS to a registered phone number	Secrecy of the channel and access to the phone
PKQ	Answer security questions referring to personal knowledge	Difficulty to answer the questions (targeted and trawling attacks)
Designated Trustees	Provide reset codes sent to registered trusted contacts	Ability of trusted contacts to only share the reset code with the user

3 METHOD

This study aims to provide the missing comparison of fallback authentication schemes' effectiveness and usability over realistic recall times of 6 to 18 months. We explain the protocol of the long-term study, the selected schemes and their implementation details, the recruitment process, and the participants' demographics. We also provide details on the primary task (mental rotation test), closing with limitations and ethical implications of the research.

3.1 Study Protocol

The study consisted of three stages: *registration*, a *short-term callback* after 2 weeks, and a *long-term callback* after 6, 12, or 18 months where participants used their assigned fallback authentication scheme. We limited the study to participation via desktop devices to provide identical and ideal conditions for each reset scheme. For example, typing is usually more cumbersome on mobile devices, which could affect certain schemes negatively. Importantly, the study was disguised as a measurement of long-term performance trends in a mental rotation test (MRT), justifying the need to log into the MRT website multiple times during the study [79, 86]. The full survey instrument can be found in Appendix A.

Stage 1: Registration After consenting, participants created an account on our study website using an email address and a password. Participants were then assigned round-robin to one of four fallback authentication schemes (see Section 3.2) and one of the three *callback times*—6 months, 12 months, or 18 months—which defined the time span between the second and third stage. The time spans were chosen following findings from Bonneau et al. [13], who measured that 33%, 50%, and 75% of the analyzed sample had started an account recovery after the mentioned periods. Some groups (SMS, trustee, and PKQ—see Section 3.2 for refernece) had to provide further details, which we explained by saying that the long duration of the study might make fallback authentication necessary. After the registration, participants completed five initial mental rotation tests. A demographic questionnaire (**S1-D1–S1-D4**) and an honesty question (**S1-H**) concluded the first stage.

Stage 2: Two-Week Callback After two weeks all participants were emailed to return and complete another mental rotation test. They had to log in using their email address and password combination but could also reset their password using the respective fallback authentication mechanism. This stage was included to remind participants about the study, select participants who will be more likely to return after an extended time, and give further incentives to follow through the entire study. Additionally, it gave us another data point after two weeks.

Stage 3: 6/12/18-Month Callback Depending on their condition, participants were emailed to return after 6, 12, or 18 months. When logging in, we enforced a password reset claiming internal re-configurations to be the reason. With this approach, we could measure how many participants correctly remembered their password and, more importantly, how many successfully completed the fallback authentication. At this point, we constantly monitored if participants initiated the reset but struggled to complete it. If this was the case, we manually emailed them a link to reset their password to ensure that we also collected results from unsuccessful fallback procedures.

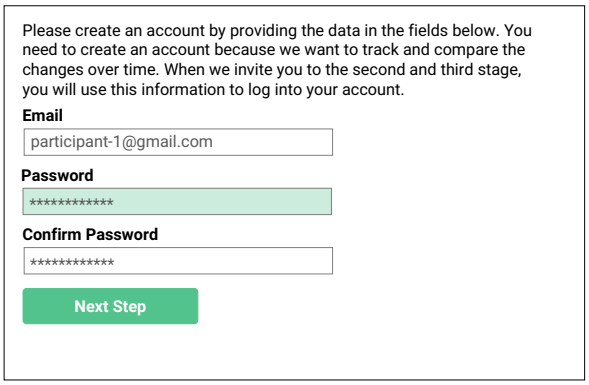
After resetting the password, participants logged in and completed the primary task a third time before we debriefed them about the real purpose of our study. No participants withdrew from the study after the debriefing. Participants who did not complete the study were debriefed via email. After the debriefing, we asked participants to complete a usability questionnaire regarding the reset process, consisting of a set of tailored questions for each scheme and the system usability scale (SUS), including an attention check, as a metric for direct comparison [17]. Finally, participants again answered the demographic questions from Stage 1 to detect any changes before we asked them about potential dishonesty (**S3-H**). We emphasized that indicating dishonesty would only exclude their data from the analysis but not affect their payment.

3.2 Selected Schemes & Implementations

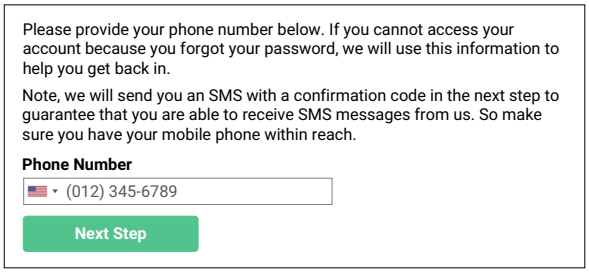
From the different real-world implementations, we selected four fallback authentication schemes (cf. Table 1) to test in our study. Figure 2 shows screenshots of the different enrollments.

3.2.1 Email. This scheme is often easy to implement as users provide an email address during account registration anyway, which can then also be used for fallback authentication. In our study, we did just that. Like all participants, the email reset group had provided their email address during account setup and was thus not asked for further information. Updating the email address was possible at any time during the study. For recovery, participants provided their email address and received an email containing a unique link directing them to the password reset page (see Figure 6 in Appendix B.1).

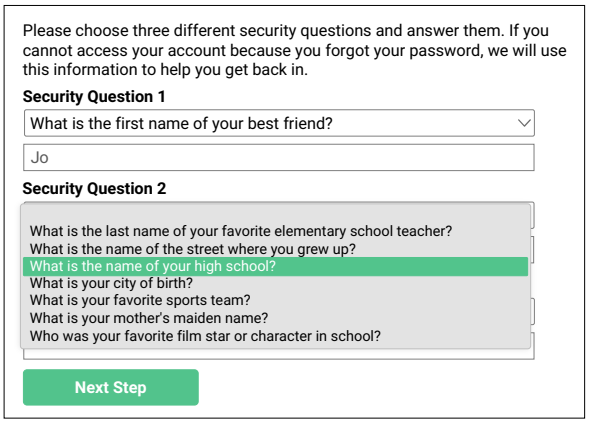
3.2.2 SMS. During account setup, we asked for the user's mobile phone number, explicitly stating account recovery as the reason. To confirm that participants can access the phone number, we asked them to input a code we sent to the provided number. The same confirmation process took place if the phone number was changed, which was possible at any time during the study. To reset the password, we first asked participants for their email address



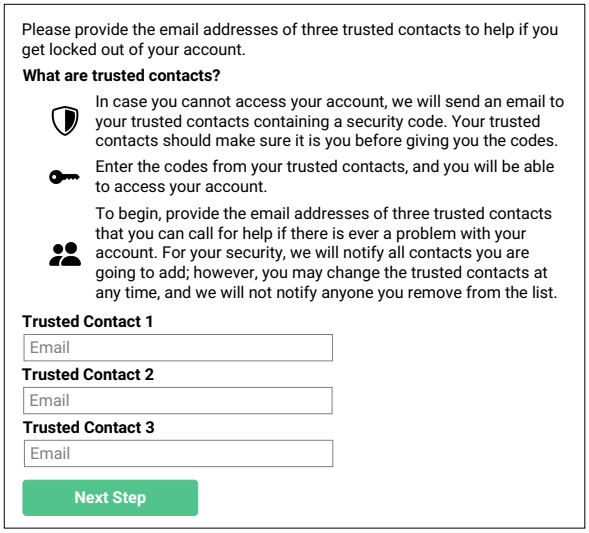
(a) Email/Signup Form



(c) SMS



(b) Personal Knowledge Questions (PKQs)



(d) Designated Trustees

Figure 2: Screenshots of the fallback setup pages displaying the respective information for the different fallback schemes. The form of the email scheme shown in Figure 2a was the standard form that all participants had to complete to create an account.

and then redirected them to a form where they had to provide a six-digit reset code, which we sent to the linked phone number. On this page, participants were also able to initiate re-sending the SMS. The SMS was written in line with best practice [26]. Figure 7a in Appendix B.2 shows the exact wording. We did not disclose the phone number during the reset for privacy reasons.

3.2.3 Personal Knowledge Questions. For the reset scheme based on personal knowledge questions (PKQs), participants had to select and answer three questions. Our set of PKQs consisted of 4 "classical" questions that have been used for a long time but are known to be insecure and easy to guess [27, 71],

- "What is your mother's maiden name?"
- "What is your city of birth?"
- "What is your favorite sports team?"
- "What is the name of your high school?"

and 4 questions with reportedly better security properties [13]:

- "What is the name of the street where you grew up?"
- "What is the first name of your best friend?"
- "Who was your favorite film star or character in school?"
- "What is the last name of your favorite elementary school teacher?"

For account recovery, two out of the three registered questions were randomly selected (see Figure 11 in Appendix B.4), which participants both had to answer correctly. When matching the originally set answers to the given ones, we ignored capitalization following practice by Apple, PayPal, and eBay and removed spaces like Apple and eBay do. In line with all of those services, we did not allow any edit distance.

3.2.4 Designated Trustees. We designed our designated trustee scheme as a variation of Schechter et al.'s approach [77] and the implementation of Apple [8]. During account creation, participants were asked to provide email addresses of three contacts, again explicitly stating recovery as the reason. All trusted contacts received an email informing them about their role (see Figure 8 in Appendix B.3), which we also used to check the existence of the email addresses. Updating the list of trustees was possible at any time. We emphasized that we do not inform trustees about being removed to prevent reluctance to adjust the list due to social concerns.

For recovery, participants had to provide their email and were then presented with a form to submit the reset codes (see Figure 10a in Appendix B.3). For a successful recovery, two out of three codes were required. Due to privacy reasons, we did not directly display

the email addresses of the trusted contacts. Instead, we offered to reveal the list by providing the email address of one trusted contact correctly (see Figures 10b and 10c in Appendix B.3).

If a reset was initiated, the three trustees received an email with a six-digit reset code, and instructions to relay the code to the owner of the account. As part of these instructions, we provided the participant's email address and explicitly told trustees only to pass the code once they verified the participant's identity. Figure 9 in Appendix B.3 shows the email's exact wording. Directly sending the reset codes to the trustees is different from the initial proposal by Schechter et al. [77]. They required the designated trustees to complete several steps before obtaining a code, among others, a pledge, to minimize the risk of an account takeover. As we wanted to minimize the risk of trustees not completing such a multi-step protocol, we decided on a simplified version, which is also more in line with the implementation by Apple [8], where the reset codes are shown in the trustees' iOS settings.

3.3 Recruitment and Demographics

We recruited participants using different channels, including mailing lists at the university, as well as websites and social media groups where researchers who are looking for participants can post their surveys. We were unable to use services like Amazon Mechanical Turk and Prolific for recruitment as we needed to collect data like email addresses and phone numbers. Overall, 201 participants completed Stage 1 of which 142 participants returned to complete Stage 2. A total of 105 participants completed the third long-term stage, of which 8 failed attention checks and were removed (**S3-AC**). The final number was $n = 97$. Since the three stages all differed in their duration, and we wanted incentives for returning to the long-term stage, we paid different compensations: In Stage 1 participants received $1.80 for an average of 3.5 minutes. The second stage took 1.5 minutes and was compensated with $0.90, whereas the final, long-term stage took 6 minutes and was compensated with $3.60.

Table 2 shows the demographics. Our sample included a slight majority of female-identifying (56%) and non-technical (53%) participants. Our participants were mostly younger, with 85% aged between 18 and 34, and were relatively educated, 32% with a Bachelor's or Master's, respectively. This comes as no surprise considering the described recruitment channels.

3.4 Primary Task

To disguise the real purpose of our study, we used a Mental Rotation Test [79, 86] as the "primary task." The layout and exact wording of the study's implementation can be seen in Figure 3. The purpose of the primary task was to distract the participants from the real purpose of the study and to increase the ecological validity of the authentication task. Framing the long-term nature of the study as being a study of cognitive ability over time allowed us to justify the length of the commitment without revealing our interest in the authentication step. The MRT is also a strong cognitive distractor and should suitably prevent participants from remaining focused on the authentication task.

Table 2: Demographics.

	Female		Male		Other		Total	
	No.	%	No.	%	No.	%	No.	%
Age	54	56	37	38	6	6	97	100
18–24	22	23	10	10	3	3	35	36
25–34	25	26	20	21	2	2	47	49
35–44	3	3	3	3	1	1	7	7
45–54	1	1	4	4	0	0	5	5
55–64	3	3	0	0	0	0	3	3
Education	54	56	37	38	6	6	97	100
High School	12	12	7	7	0	0	19	20
Some College	3	3	1	1	1	1	5	5
Associate's	6	6	1	1	0	0	7	7
Bachelor's	20	21	9	9	2	2	31	32
Master's	12	12	17	18	2	2	31	32
Doctorate	1	1	1	1	0	0	2	2
Prefer not to say	0	0	1	1	1	1	2	2
Background	54	56	37	38	6	6	97	100
Technical	14	14	26	27	5	5	45	46
Non-Technical	40	41	10	10	1	1	51	53
Prefer not to say	0	0	1	1	0	0	1	1

3.5 Pilot Study

Our pilot study intended to minimize the risk of technical issues during the main study and ensure questions were understood as intended. We recruited 74 students from our university, of which 44 completed all three stages. As testing our implementation was the primary purpose of this study, we reduced the time span between the first and second stage to 1 week, as well as the period between stages two and three to 3 weeks.

Most importantly, we identified the need to reduce the use of fictional email addresses for the designated trustees scheme. Thus, we decided to send information emails to the trustees after enrollment to check the existence of the addresses. If we received an "Undelivered Mail Returned to Sender" error, we marked the respective address and asked the participant for a new one at the beginning of the second stage. At this point, we highlighted the importance of providing valid trusted contacts for account recovery and the long time span between the second and third stage. This is different from Schechter et al. [77], who did not send emails to trustees during enrollment but also did not face the described problem to the same extent, since participants in Schechter et al.'s study used their actual Microsoft accounts. Similarly, Apple's account recovery contacts [8] are selected from a user's contact list and must be associated with an Apple ID. In contrast, our participants created an account for the study to which they presumably do not assign a high value and are thus more likely to provide fictional email addresses. This limitation, which we extend in Section 3.6, is shared by all studies with a similar methodology.

3.6 Limitations

This study aims to compare fallback schemes after realistic reset times given identical conditions for all resets, yet some confining

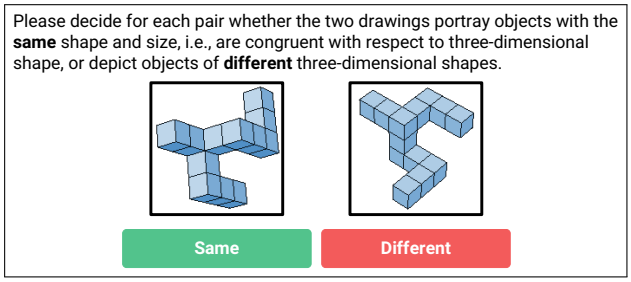


Figure 3: Example of the Mental Rotation Test (MRT), which is used as a distractor task in the study.

aspects exist. First, participants needed to create an account for the study, and we assume the perceived value of the account to be comparatively low. This could have negatively impacted the reset rates of the trusted contacts and PKQ scheme if participants stated made-up email addresses or random answers, which prevented them from resetting the password. As described earlier, we added a checkup as a countermeasure to minimize the consequences for the trusted group. Nevertheless, dealing with incorrect information and accounts not being as important as others is a problem that other studies with a similar methodology must manage, as well as regular service providers.

Additionally, we acknowledge that results could have been influenced by biases regarding how questions are formulated and participants' tendency provide socially desirable answers. Finally, due to the recruitment channels, our participants were mostly younger and more educated—results could differ for more diverse recruitment samples.

3.7 Ethical Considerations

Our study received clearance from our institutional review board. We took a number of steps to minimize the risk of ethical harm to participants. Although we concealed the true focus of the work until the end of the study, the authentication steps were always visible to participants. To clarify the true purpose of the study, participants were debriefed during the final session. Participants who did not return for the last stage received the debriefing via email. Finally, participants were educated about the data collected in this study and that it was stored and processed per the General Data Protection Regulation (GDPR). We also took care of the data entrusted to us during the study. All personally identifiable information (PII), such as participants' email addresses, phone numbers, or trusted contacts, was deleted after the study.

4 RESULTS

Below, we present the results of our user study, including results from the reset processes and more general usability aspects. First, we address research question **RQ1**, i.e., participants' ability to reset the password with the fallback scheme after the assigned recall time. Afterward, we focus on the time spans needed for the resets for each scheme and recall time (**RQ2**). We close by answering **RQ3**, i.e., how participants perceive the usability of the reset processes and show which issues arise for the schemes after 6, 12, or 18 months. Table 3 details the results for each combination of the independent variable's reset scheme and recall time. When referencing participants, e.g., for quotes, we use an identifier composed of the abbreviated treatment name (*EM*, *SMS*, *PKQ*, *DT*), the recall time (*6*, *12*, *18*), and the ID within the reset group.

4.1 RQ1: Successful Password Resets

The most important measure of usability for a fallback authentication mechanism is whether users are able to successfully access their accounts. Regarding the schemes, *email* was the leading option: all 21 participants (100%) successfully reset their passwords. This is followed by the *SMS* group, where 24 of 26 resets (92%) were completed. For the *designated trustees* scheme, 24 of 29 (83%) participants were able to reset their password. Participants who reset their password using *personal knowledge questions* had by far the lowest success rate, with only 12 of 21 (57%) being successful. The results of Fisher's exact test ($p < .001$) indicated a significant difference in the number of successful password resets between the reset schemes. Using a posthoc test, Bonferroni corrected for multiple comparisons, we observed significant differences between the email (100%) and PKQ (57%) scheme ($p < .01$) as well as SMS (92%) and PKQ ($p < .05$).

For the callback times, reset rates after 6 and 12 months are very similar, 80% and 79%, respectively. Reset success after 18 months is slightly higher (87%), yet, Fisher's exact test ($p = .681$) did not indicate any significance here.

When looking into *why* resets fail, we find that one of the two failures in the SMS group was due to the participant not residing in the US at the time of the reset and not having service:

> "*I don't have service to that phone number in this country, as i am studying abroad*" (SMS-6-P9)

The other participant described having a new SIM card altogether. While similar situations can occur at any time, chances for the latter may increase over time.

Issues reported by the trusted contacts participants include being unable to remember who they selected as a trustee, trustees not responding, or not having access to their email accounts:

> "*One of my trusted contacts' email account was not active anymore*" (DT-12-P26)

Following the pilot study, where many participants provided non-existent email addresses, we added a check to account for this issue (see Section 3.5). After the main study, we can conclude that this approach was beneficial: We caught three errors, two typing errors, and one participant, who initially provided a random email string, corrected it to a valid email address after being prompted. All three successfully reset their password at the beginning of the third stage.

Of the people who failed to regain access using PKQs, two could not remember the exact spelling of their answer (e.g., "St." vs. "Street") and seven users said they failed to remember one or both of their answers entirely. One participant described how their coping strategy failed:

> "*I thought I wrote the answers down somewhere, but I couldn't find them*" (PKQ-18-P4)

As an exploratory posthoc analysis, we examined two sub-groups for the designated trustees scheme: *dependent* and *autonomous*.

Table 3: The rate of successful resets, the median reset time, and the SUS scores for the four fallback schemes. We also depict two sub-groups for the designated trustees scheme based on whether participants were *dependent* on others for their reset or could complete it *autonomously*.

	Callback Time											
	6 Months			12 Months			18 Months			Combined		
Email	100%	47 s	76	100%	26 s	79	100%	44 s	90	100%	31 s	80
SMS	82%	35 s	65	71%	56 s	80	100%	59 s	84	92%	52 s	74
PKQ	57%	20 s	55	67%	31 s	70	63%	31 s	65	57%	30 s	63
Trustees	82%	116 s	64	78%	826 s	62	89%	85 s	53	83%	111 s	60
Combined	80%	55 s	65	79%	61 s	75	87%	54 s	78	–	–	–
Trustees Dependent	67%	22 min	51	60%	117 min	65	67%	257 min	38	64%	27 min	53
Trustees Autonomous	88%	105 s	70	100%	145 s	59	100%	75 s	60	94%	97 s	64

These two groups are defined by the number of trusted contacts that are email accounts owned by the participants (see Question **S3-DT1**). If 2 or 3 contacts were actually the participant's email accounts, the reset could be performed *autonomously*. In the other case, the participant was *dependent* on others. For the ratio of successful/failed, we observe a stark discrepancy between the dependent and autonomous sub-groups of the designated trustees scheme. While roughly every third participant (4/11; 36%) who had to rely on actual trusted contacts failed the password reset, only 1 of 18 participants (6%) of those who stated 2 or 3 own email accounts did. Further investigation of the latter failure revealed that 2 of the 3 email accounts were non-existent and flagged by our system as "Undelivered Mail Returned to Sender." Disregarding our instruction, the participant provided them again when prompted for new trusted contacts in Stage 2.

4.2 RQ2: Password Reset Times

In addition to knowing *if* participants were able to reset their passwords, we also wanted to know how long it took them to complete the process. For this analysis, we measured the time span from initiating a reset to successfully setting a new password. We removed extreme outliers from the collected reset times using Tukey fences with $k = 3$, i.e., values greater than 3 times the interquartile range. Figure 4 shows the results for each of the four reset schemes and the two sub-groups of the designated trustees.

Resets from the PKQ scheme were the fastest, with 30 s as the median. Note that while reset times were low, we previously saw that significantly fewer participants in the PKQ group could reset their passwords at all. Email resets were comparably fast with a median of 31 s. The SMS scheme ranked third in reset times ($Mdn = 52\,s$). Lastly, participants who used the designated trustees scheme spent the longest time resetting their passwords, with a median reset time of 111 s. However, an interquartile range of 247 s highlights that some participants in this group took substantially longer than others. Using a Kruskal-Wallis H test, as the data was not normally distributed, we saw that there was a significant difference in the reset time between the schemes, $\chi^2(3) = 34.53, p < .001$. The Bonferroni-corrected posthoc Dunn's test indicated that the reset times of email, SMS, and PKQ are all significantly shorter than those of the designated trustees scheme ($p < .001^{**}$).

We also separately compared the fallback schemes for each callback time. The Kruskal-Wallis H test indicated that there is a significant difference after 6 ($\chi^2(3) = 11.8, p = .008$), 12 ($\chi^2(3) = 9.78, p = .021$), and 18 months ($\chi^2(3) = 10.06, p = .018$). For each of them, the posthoc Dunn's test using a Bonferroni corrected α of 0.0083 indicated that the PKQ and trustee resets are significantly different. Taking just the callback time as the independent variable, a Kruskal-Wallis H test indicated that there is a non-significant difference between the reset times, $\chi^2(2) = 0.51, p = .775$. Both analyses suggest that the callback time does not influence the time needed to reset the password.

In the PKQ group, where resets were the fastest, 9 of the 12 participants got their answers correct on the first try. Of the remainder, two took a second attempt, and one took three tries due to different spelling or typos. As reset times were low, we initially suspected participants digitally saving answers (i.e., in a password manager), but found that all answers were typed, refuting this assumption.

The situation for the email-based resets is similar, yet outliers are more notable. One participant named the delivery of the email as a cause for the delay:

> "*The mail took 30 sec. longer than expected.*" (EM-6-P12)

For the SMS scheme, an interquartile range of 22 s suggests that the reset experience was consistent for most participants, which is further underlined by the fact that no one mistyped their code, and only two participants requested more than one reset SMS. Question **S3-SMS3** asked if participants usually have their phone within reach when surfing the web to understand if the accessibility of the phone posed a hurdle during the reset. It did not, at least for our comparably young population (see Section 3.6), with 92% saying they "often" or "always" have their phone within reach when surfing the web.

Resets for the designated trustees group took significantly longer, which is reasonable, considering that participants had to get in touch with others to reset their passwords. One participant whose reset took more than a day described the situation as follows:

> "*My contacts and I weren't online simultaneously such that collecting all codes took rather long.*" (DT-18-P1)

To better understand the reset process for this scheme, we investigated how participants got in contact with their designated trustees.

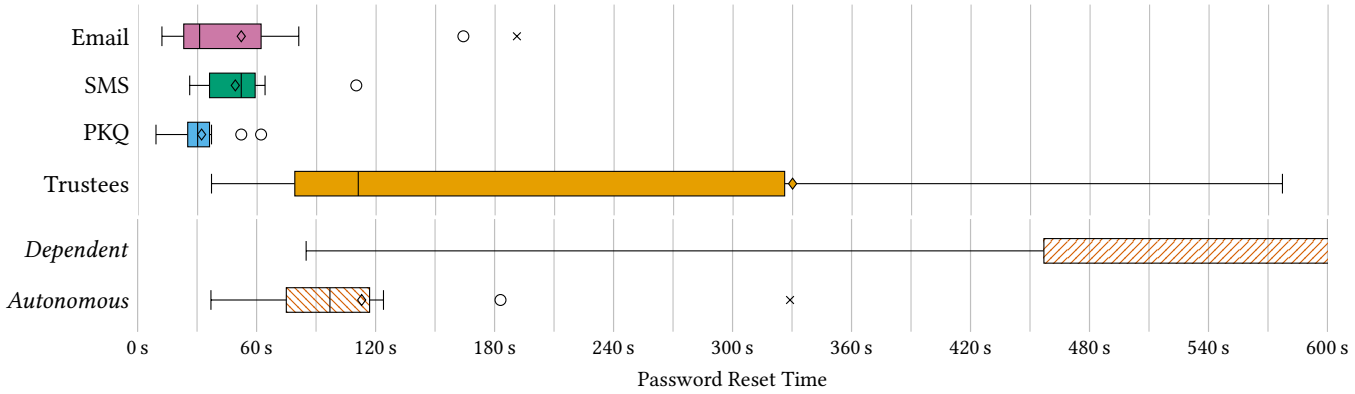


Figure 4: Password reset times for the different fallback schemes. For the designated trustees scheme, we show two additional plots where we separated participants into those who used the scheme in a *dependent* or *autonomous* way. For the sake of clarity, we limited the x-axis to 600 s. The median and average for the dependent sub-group are 27 min and 62 min, respectively.

Of the 11 participants who provided actual contacts, most (6) interacted with them using an instant messenger. Fewer participants (4) sent an email, and 3 participants met their trusted contact in person. The least participants called their trusted contacts (2) or sent an SMS (1). While the popularity of instant messengers comes as little surprise as they depict an efficient way to communicate the reset codes, it must be noted that their confidentiality and authenticity are not guaranteed. Certain messengers like Signal or WhatsApp do provide end-to-end encryption and mutual authentication, yet those mechanisms are often poorly understood by users [22, 37, 87, 89]. Phone calls, in turn, can be spoofed, further simplified by the advances in artificial intelligence [16, 83], and the insecurity of SMS or email as communication channels has also been proven repeatedly [44, 55]. Hence, in-person meetings depict the highest security level possible, as the trusted contacts can be sure that the code is only shared with the account owner. For this reason, Schechter et al.'s original proposal also prompted trustees and account owners to get in touch physically [77].

Again, we performed an exploratory analysis of the differences between the dependent and autonomous sub-groups of the designated trustees scheme. In stark contrast, even to the combined designated trustees scheme, the average reset for the dependent group took 62 minutes (3734 s); the median reset time was 1,602 s or 27 minutes. As seen in Figure 4, participants' resets were substantially faster if they could complete the protocol autonomously, averaging at only 113 s ($Mdn = 97\,s$). The results of a Mann-Whitney U test also indicate that these differences are significant, $z = 2.6308$, $p = .009$. Hence, in addition to increasing the success rate of resets, this unintended deviation from the protocol utilized by some participants also decreases reset times:

> "*Since all the accounts belonged to me, it was alright (if not tedious) but I think if I used accounts belonging to other people I wouldn't have been able to log in [...] I would've had to contact at least two people asking them to check their email and wait for a response, which, knowing the kinds of people I'd use as a 'trusted contact,' would've taken at least a day.*" (DT-12-P25)

4.3 RQ3: Perceived Usability

We used the System Usability Scale (SUS) [17] for a standardized and comparable assessment of the four reset schemes. The email-based reset ranked the highest, with an average score of 80 ($Mdn = 83$). Based on the adjective rating by Bangor et al. [11], which provides an informative description of SUS scores, most participants ranked the email scheme as "excellent." The SMS scheme's usability, on the other hand, was assessed as "good" by participants with an average SUS score of 74 ($Mdn = 83$) — ranking the second highest among all schemes. An average score of 63 ($Mdn = 60$) put the PKQ's usability between "OK" and "good", with some participants' scores even ranging to "poor." The overall greatest range in scores was seen for the designated trustees scheme. While the average SUS score of 60 ($Mdn = 58$) was only marginally lower than that of the PKQ scheme, some participants rated the system as "worst imaginable." A Kruskal-Wallis test indicated that there is a significant difference in the SUS scores of the different schemes, $\chi^2(3) = 18.07, p < .001$. Using a posthoc Dunn's test with a Bonferroni corrected α of 0.0083, we were able to observe significant differences between the following schemes: email/PKQ, email/trustees, and SMS/trustees.

The separate analysis of the subgroups of the designated trustees scheme who used it in a dependent or autonomous way was in line with the findings about the success rate of resets and the required time, yet the differences in SUS scores were not as substantial: While the average rating of the dependent group was 53 ($Mdn = 54$), it increased to only 64 ($Mdn = 65$) for the autonomous group.

An explanation for the lower SMS ratings compared to the email scheme may be the lower success rate and more effort required to reset as a code needs to be copied instead of clicking a link. Similarly, for the PKQ schemes. Examining the 12 participants who were able to successfully reset their password, the average score was 80 ($Mdn = 84$), making it comparable to the email scheme. In contrast, the average across the nine participants who failed to reset their password was only 43 ($Mdn = 43$). In the group of trusted contacts, low scores were not solely given by participants who failed the reset: Those who did fail gave an average SUS score of 44 ($Mdn = 48$), but the score for those who successfully reset their password is only slightly higher at 63 ($Mdn = 65$).

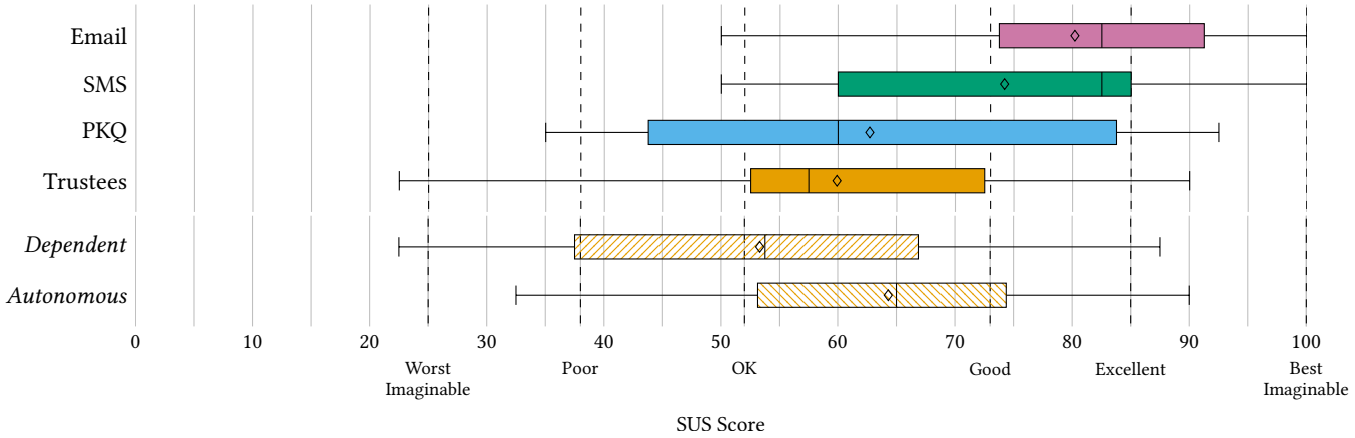


Figure 5: Scores of the System Usability Scale (SUS) for the different fallback authentication schemes. For the designated trustees scheme, we show two additional plots where we separated participants into those who used the scheme in a *dependent* or *autonomous* way. To provide additional context, we added the adjective ratings from Bangor et al. [11].

Kruskal-Wallis H tests for each callback time indicated that there is a significant difference in the SUS scores after 18 months ($\chi^2(3) = 14.29, p < .003$). A posthoc Dunn's test using a Bonferroni corrected α of 0.0083 indicated that the SUS scores of email (90) and trustee (53), as well as SMS (84) and trustee (53), are significantly different. This highlights that the longer the callback time is, the worse the perceived usability of the trusted contacts scheme becomes. This is reasonable, as two of the described issues, not remembering the contacts and not having access to their email account, may become more likely over time.

4.4 Results Summary

Below, we summarize our findings for each scheme following the three research questions. We also contextualize the results to provide a ranking.

4.4.1 Email. The email scheme showed the best overall usability. All participants in this group were able to successfully reset their passwords, the SUS scores were the highest out of all four schemes, and reset times were also not significantly longer than those of the other schemes. Generally, participants were very positive and did not report any negative aspects about this type of reset.

SMS. Compared to email, SMS-based password resets showed only marginally worse results in our study. This fallback scheme ranked second in all measured categories, and in each case, the differences from the best ranking scheme were not statistically significant. Only two participants encountered problems in this condition: (1) one participant changed their phone number without updating the number associated with the account and (2) the SMS could not be transmitted due to technical reasons.

Personal Knowledge Questions. The results for the PKQs were mixed, but the overall usability evaluation was rather negative. While the average successful reset took only 32 seconds, which was the quickest across all schemes, the number of participants who could not complete the reset at all was by far the highest. SUS scores were also low, especially among those participants who could not recall their answers and failed the reset. We did not observe any notable usability or success-rate differences between the traditional set of questions, which are often easy to guess [27, 71] and the more modern ones with slightly better security properties [13] (see Section 3.2.3).

Additionally, we highlight that some users employ strategies to compensate for usability issues and security concerns with the scheme. Websites should be aware that some users intentionally provide random answers or answer untruthfully, attempting to increase security, also observed by related work [13]. Others do not want to share the personal information that the PKQs ask them for. While some users might take note of their random answers (for example by storing it in a password manager [2] or writing it down in a notebook [65]), enabling them to reset their passwords later on, some may forget their random answer, making the reset impossible, as happened to one of our participants:

> "*I couldn't remember the answer for my favorite teacher, because I somehow forgot all my elementary teachers. [...] I might have even given some wrong answers because I do not trust sensible information on other websites. The exception is on online banking websites.*" (PKQ-12-P15)

Designated Trustees. Overall, this reset scheme ranked last as it had the most drawbacks. However, the results varied based on whether participants provided actual contacts or just used multiple of their own email accounts. If the scheme was used in the intended way (*dependent*), reset rates were among the lowest, reset times were tremendously higher than for the other schemes, and the system's usability was perceived as low. In the other case (*autonomous* usage), the reset procedure is more similar to the email-based scheme, and usability ratings were more similar to the other schemes. These findings are influenced by the implementation, which in our case required users to provide two of three reset codes. Other configurations may perform differently, e.g., requiring only one reset code would potentially decrease the effort.

5 DISCUSSION

Next, we discuss the takeaways and give recommendations.

5.1 Email as the Magic Bullet?

Throughout our analysis, resets based on email proved to be the most favorable option. Even after 18 months, all participants successfully reset their passwords in a reasonable time and did not report any usability issues. There are multiple ways to explain this positively outstanding result.

Apart from the very few steps required and the short (but not shortest) reset time, it could be argued that an email reset has the highest familiarity. By now, email has essentially become a digital identity [53], and fallback authentication mechanisms that rely on it only strengthen that association. Most people use emails frequently and may thus be very accustomed to the idea that this is how to cope with account issues. Still, it needs to be taken into account that the email-based scheme might not be deployable in every case, most prominently for the recovery of the primary email account. Moreover, the over-reliance on email creates a single point of failure [53]. While the login to an email account can theoretically be secured using two-factor authentication or passwordless schemes like FIDO2 and passkeys, which are not susceptible to most online attacks, those mechanisms are used infrequently [48, 62, 66]. Additionally, email has little protection against intimate attackers, like a partner or family members [25, 36, 58, 84]. This allows for trivial password resets to all linked accounts, not only locking the original owner out but also allowing account access for those with malicious intent. Therefore, email resets should not be recommended without restrictions, despite their many advantages, as certain aspects may confine applicability. Interestingly, in contrast to our results, Bonneau et al.'s work [13] reported SMS having a slightly higher recovery success rate than email. In the context of only slightly worse usability ratings for SMS compared to email in our study, this may indicate that both email and SMS are generally similarly favorable.

5.2 Another Nail in the Coffin

As many others have proven before, we can confirm that the memorability and perceived usability for PKQs is low, especially compared to the alternatives, and for realistic callback times in the range of months and years. This holds for both the traditional set of questions [27, 71] and the more secure ones [13] (see Section 3.2.3). Moreover, people sometimes provide random answers intentionally, either for privacy reasons or because they falsely believe that this increases security [13]. Despite extensive multi-year research efforts to improve the approach of PKQs by using dynamic [4, 34, 92], location-based [1, 35], or simply "harder to know" [13] question types, they proved to be an unsuitable approach once again. As there are neither arguments from the users' perspective [13, 46] nor the security side [27, 46, 67, 71, 76], we strongly recommend services cease the use of PKQs.

5.3 Cheating as a Solution?

Our findings were very diverse for the social authentication approach (designated trustees), where participants had to provide two of three reset codes. Participants who correctly followed the instructions and provided actual contacts described the usability as poor. Participants who "cheated" by stating their own email addresses ("*autonomous*" usage) rated the usability as tolerable.

From a security standpoint, cheating on the system also sabotages its security. The security of the scheme is grounded in the trustees' ability to check the legitimacy of the person requesting the reset codes [77]. However, when using the "autonomous" way, the security is essentially reduced to multiple email-based resets. While we fully acknowledge the aforementioned security shortcomings of email, we can also not disregard that it is the de-facto standard, and we lack compelling alternatives. Hence, requesting users to provide multiple email addresses owned by them that need to be accessed for account recovery only slightly decreases the usability and authentication times. At the same time, it could be an easily deployable improvement that reduces the risk of one email account being a single point of failure. Of course, this requires the password of those email accounts to be different and may only marginally increase the security against intimate attackers [25, 36, 58, 84].

5.4 Users Find Their Shortcuts

As advocated for by many, and laying at the core of usable security, systems must be usable to provide the intended security [74, 75]. Our participants circumventing the trusted contacts systems is a prime example of this behavior and once more stresses the importance of designing for usability and, at best, designing secure systems without users having to play a role in the "securing." Moreover, the security of the trusted contacts scheme relies on the authenticity and confidentiality of the communication channel. This may enable certain attackers to intercept [22, 37, 44, 87, 89] or spoof the reset process [16, 55, 83]. To prevent this, the original proposal by Schechter et al. prompts trustees and account owners to get in touch physically [77], which is very demanding, assuming that account parties could reside in different cities, states, or even countries.

In the case of trusted contacts, getting in touch with others can even add a level of social anxiety or concern, considering that one has to tell others that something potentially embarrassing like losing access to their account has happened. Relying on others might also not be viable for people who do not know "enough" people for whom they have email addresses or trust closely.

5.5 Resets in a Passwordless Future

With ongoing efforts to eliminate the password overall, one may argue that password resets will become obsolete eventually. Still, the necessity of a backup authentication mechanism will remain a topic of utter importance [10, 51]. In fact, the lack of a standardized fallback solution is considered one of the biggest obstacles when it comes to modern passwordless authentication based on FIDO2 and passkeys [52]. The FIDO Alliance recommends (purchasing), registering, and safely storing a second authenticator in case access to the primary gets lost [28]. However, they themselves acknowledge that this is just a quick fix recommended in default of better alternatives. Microsoft's alternative is a so-called Temporary Access Pass (TAP), which is a time-limited single-use passcode, comparable to the 6-digit security code that we sent via SMS. Users are asked to enter their TAP (e.g., received via SMS) when they register their first passwordless authenticator. While the scheme is

intended to be used during account creation, Microsoft states that "this method can also be used for easy recovery when the user has lost or forgotten their authentication factor" [61]. These examples indicate that the findings of our study will likely remain relevant even with the progression of FIDO2 passwordless authentication.

5.6 Recommendations

We want to close with recommendations based on the study's result for the four analyzed fallback schemes and the three callback times we considered:

Email and SMS Recommendable options should sustain usability criteria for frictionless resets, reducing cost-intensive manual reviews. Schemes based on email and SMS meet these requirements—being sufficiently reliable (see Section 4.1) and perceived as usable (see Section 4.3). Trusted contacts could fulfill the requirements to some extent if used in a modified, autonomous way. The original protocol showed multiple drawbacks, similar to Personal Knowledge Questions, which are also not recommended.

Authentication Time is Less Relevant For authentication, login times are an important criterion. Our results indicate that login times are generally comparable between fallback schemes, and the differences only partially influence the overall usability of a system (see Section 4.2). Thus, services can, to a certain extent, sacrifice quick recovery in favor of availability and security. Using hybrid systems might be an option, such as letting users provide multiple fallback emails to which reset codes are sent.

Provide & Encourage Multiple Reset Options No fallback method is perfect and universally accessible to everyone. Thus, services should always provide multiple reset options for users to choose from. Users should be made aware of the different security levels available, ideally promoting schemes that offer a usability and security level similar to email and SMS. Additionally, services should promote multiple enrollments to ensure continuous account access in case one reset mechanism fails (see Section 4.1). As resets lay months or years in the future, users likely cannot foresee situations requiring a reset when creating an account. Thus, services should inform users that registering multiple methods increases the chances of regaining access.

Ensure Up-To-Date Information Services should regularly remind users to review their reset information. Our study showed that information like email addresses or trusted contacts are subject to change even after only 1.5 years (see Section 4.1). Real-world accounts are usually held much longer than that, increasing the potential for changes. Regularly ensuring information is up-to-date increases the ability to complete resets, even after years. Moreover, since email addresses and phone numbers are subject to reassignment, it ensures that only the intended person can perform a reset.

ACKNOWLEDGMENTS

This research was supported by the research training group "Human Centered Systems Security" sponsored by the state of North Rhine-Westphalia and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CASA – 390781972.

REFERENCES

[1] Alaadin Addas, Amirali Salehi-Abari, and Julie Thorpe. 2019. Geographical Security Questions for Fallback Authentication. In *International Conference on Privacy, Security and Trust (PST '19)*. IEEE, Fredericton, New Brunswick, Canada, 1–6.

[2] AgileBits, Inc. 2023. Create Unique Answers to Security Questions. https://support.1password.com/generate-security-questions/, as of March 16, 2024.

[3] Fatma Al-Maqbali and Chris Mitchell. 2018. Web Password Recovery: A Necessary Evil?. In *Future Technologies Conference (FTC '18)*. Springer, Vancouver, British Columbia, Canada, 324–341.

[4] Yusuf Albayram and Mohammad Maifi Khan. 2016. Evaluating Smartphone-Based Dynamic Security Questions for Fallback Authentication: A Field Study. *Human-Centric Computing Information Sciences* 6, 16 (Dec. 2016).

[5] Reem AlHusain and Ali Alkhalifah. 2021. Evaluating Fallback Authentication Research: A Systematic Literature Review. *Computers & Security* 111, C (Dec. 2021).

[6] Noura Alomar, Mansour Alsaleh, and Abdulrahman Alarifi. 2017. Social Authentication Applications, Attacks, Defense Strategies and Future Research Directions: A Systematic Review. *IEEE Communications Surveys & Tutorials* 19, 2 (Jan. 2017), 1080–1111.

[7] Armin Anvari, Lei Pan, and Xi Zheng. 2016. Generating Security Questions for Better Protection of User Privacy. *International Journal of Computers and Applications* 42, 4 (2016), 329–350.

[8] Apple, Inc. 2021. Help a Friend or Family Member as Their Account Recovery Contact. https://support.apple.com/en-us/HT212515, as of March 16, 2024.

[9] Apple, Inc. 2021. How to Generate a Recovery Key. https://support.apple.com/en-us/HT208072, as of March 16, 2024.

[10] Sunpreet S. Arora, Saikrishna Badrinarayanan, Srinivasan Raghuraman, Maliheh Shirvanian, Kim Wagner, and Gaven Watson. 2022. Avoiding Lock Outs: Proactive FIDO Account Recovery using Managerless Group Signatures. *Cryptology ePrint Archive* 2022/1555 (Nov. 2022), 1–46.

[11] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4, 3 (May 2009), 114–123.

[12] Saikat Basu. 2022. How to Fix It When You're Locked Out of Your Gmail Account. https://www.lifewire.com/fix-it-when-locked-out-of-gmail-account-5220812, as of March 16, 2024.

[13] Joseph Bonneau, Elie Bursztein, Ilan Caron, Rob Jackson, and Mike Williamson. 2015. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In *International World Wide Web Conference (WWW '15)*. ACM, Florence, Italy, 141–150.

[14] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symposium on Security and Privacy (SP '12)*. IEEE, San Jose, California, USA, 553–567.

[15] John Brainard, Ari Juels, Ronald L. Rivest, Michael Szydlo, and Moti Yung. 2006. Fourth-Factor Authentication: Somebody You Know. In *ACM Conference on Computer and Communications Security (CCS '06)*. ACM, Alexandria, Virginia, USA, 168–178.

[16] Thomas Brewster. 2021. Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/, as of March 16, 2024.

[17] John Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Bernard Weerdmeester, and Ian Lyall McClelland (Eds.). CRC Press, London, United Kingdom, Chapter 21, 189–194.

[18] Elie Bursztein, Borbala Benko, Daniel Margolis, Tadek Pietraszek, Andy Archer, Allan Aquino, Andreas Pitsillidis, and Stefan Savage. 2014. Handcrafted Fraud and Extortion: Manual Account Hijacking in the Wild. In *Internet Measurement Conference (IMC '14)*. IEEE, Vancouver, British Columbia, Canada, 347–358.

[19] Facebook Security. 2011. Facebook: Introducing Trusted Friends. https://www.facebook.com/notes/facebook-security/national-cybersecurity-awareness-month-updates/10150335022240766/, as of March 16, 2024.

[20] Facebook Security. 2013. Facebook: Introducing Trusted Contacts. https://www.facebook.com/notes/facebook-security/introducing-trusted-contacts/10151362774980766/, as of March 16, 2024.

[21] Facebook Security. 2022. Trusted Contacts Is No Longer Supported. https://www.facebook.com/help/119897751441086, as of March 16, 2024.

[22] Matthias Fassl and Katharina Krombholz. 2023. Why I Can't Authenticate — Understanding the Low Adoption of Authentication Ceremonies with Autoethnography. In *ACM Conference on Human Factors in Computing Systems (CHI '23)*. ACM, Hamburg, Germany, 72:1–72:15.

[23] Schubert Foo, Siu Cheung Hui, Peng Chor Leong, and Shigong Liu. 2000. An Integrated Help Desk Support for Customer Services Over the World Wide Web – A Case Study. *Computers in Industry* 41, 2 (March 2000), 129–145.

[24] Lorenzo Franceschi-Bicchierai. 2023. Hackers Are Breaking Into AT&T Email Accounts to Steal Cryptocurrency. https://techcrunch.com/2023/04/26/hackers-are-breaking-into-att-email-accounts-to-steal-cryptocurrency/, as of March 16, 2024.

[25] Diana Freed, Sam Havron, Emily Tseng, Andrea Gallardo, Rahul Chatterjee, Thomas Ristenpart, and Nicola Dell. 2019. "Is my phone hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence. In *ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '19)*. ACM, Austin, Texas, USA, 202:1–202:31.

[26] Nethanel Gelernter, Senia Kalma, Bar Magnezi, and Hen Porcilan. 2017. The Password Reset MitM Attack. In *IEEE Symposium on Security and Privacy (SP '17)*. IEEE, San Francisco, California, USA, 251–267.

[27] Maximilian Golla and Markus Dürmuth. 2015. Analyzing 4 Million Real-World Personal Knowledge Questions (Short Paper). In *International Conference on Passwords (PASSWORDS '15)*. Springer, Cambridge, United Kingdom, 39–44.

[28] Hidehito Gomi, Bill Leddy, and Dean H. Saxe. 2019. Recommended Account Recovery Practices for FIDO Relying Parties. https://media.fidoalliance.org/wp-content/uploads/2019/02/FIDO_Account_Recovery_Best_Practices-1.pdf, as of March 16, 2024.

[29] Neil Zhenqiang Gong and Di Wang. 2014. On the Security of Trustee-Based Social Authentications. *IEEE Transactions on Information Forensics and Security* 9, 8 (Aug. 2014), 1251–1263.

[30] Paul A. Grassi, James L. Fenton, and William E. Burr. 2017. Digital Identity Guidelines – Authentication and Lifecycle Management: NIST Special Publication 800-63B.

[31] Cheng Guo, Brianne Campbell, Apu Kapadia, Michael K. Reiter, and Kelly Caine. 2021. Effect of Mood, Location, Trust, and Presence of Others on Video-Based Social Authentication. In *USENIX Security Symposium (SSYM '21)*. USENIX, Virtual Conference, 1–18.

[32] Joon Kuy Han, Xiaojun Bi, Hyoungshick Kim, and Simon S. Woo. 2020. PassTag: A Graphical-Textual Hybrid Fallback Authentication System. In *ACM Asia Conference on Computer and Communications Security (ASIA CCS '20)*. ACM, Taipei, Taiwan, 60–72.

[33] Alina Hang, Alexander De Luca, and Heinrich Hussmann. 2014. Using Icon Arrangement for Fallback Authentication on Smartphones. In *ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, Toronto, Ontario, Canada, 2467–2472.

[34] Alina Hang, Alexander De Luca, and Heinrich Hussmann. 2015. I Know What You Did Last Week! Do You? Dynamic Security Questions for Fallback Authentication on Smartphones. In *ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, Seoul, Republic of Korea, 1383–1392.

[35] Alina Hang, Alexander De Luca, Matthew Smith, Michael Richter, and Heinrich Hussmann. 2015. Where Have You Been? Using Location-Based Security Questions for Fallback Authentication. In *Symposium on Usable Privacy and Security (SOUPS '15)*. USENIX, Ottawa, Canada, 169–183.

[36] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2019. Clinical Computer Security for Victims of Intimate Partner Violence. In *USENIX Security Symposium (SSYM '19)*. USENIX, Santa Clara, California, USA, 105–122.

[37] Amir Herzberg, Hemi Leibowitz, Kent Seamons, Elham Vaziripour, Justin Wu, and Daniel Zappala. 2021. Secure Messaging Authentication Ceremonies Are Broken. *IEEE Security & Privacy* 19, 2 (March 2021), 39–47.

[38] Amir Herzberg and Ronen Margulies. 2016. My Authentication Album: Adaptive Images-Based Login Mechanism. In *International Conference on ICT Systems Security and Privacy Protection (IFIP SEC '16)*. IFIP, Heraklion, Greece, 315–326.

[39] Brad Hill. 2017. Moving Account Recovery beyond Email and the 'Secret' Question. In *USENIX Enigma Conference (Enigma '17)*. USENIX, Oakland, California, USA.

[40] Mat Honan. 2012. How Apple and Amazon Security Flaws Led to My Epic Hacking. http://www.wired.com/2012/08/apple-amazon-mat-honan-hacking/, as of March 16, 2024.

[41] Philip G. Inglesant and M. Angela Sasse. 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. In *ACM Conference on Human Factors in Computing Systems (CHI '10)*. ACM, Atlanta, Georgia, USA, 383–392.

[42] Sakshi Jain, Juan Lang, Neil Zhenqiang Gong, Dawn Song, Sreya Basuroy, and Prateek Mittal. 2015. New Directions in Social Authentication. In *Workshop on Usable Security (USEC '15)*. ISOC, San Diego, California, USA.

[43] Ashar Javed, David Bletgen, Florian Kohlar, Markus Dürmuth, and Jörg Schwenk. 2014. Secure Fallback Authentication and the Trusted Friend Attack. In *International Distributed Computing Systems Workshops (ICDCSW '14)*. IEEE, Madrid, Spain, 22–28.

[44] Roger Piqueras Jover. 2020. Security Analysis of SMS as a Second Factor of Authentication. *ACM Queue* 18, 4 (Aug. 2020), 37–60.

[45] Mike Just. 2004. Designing and Evaluating Challenge-question Systems. *IEEE Security & Privacy* 2, 5 (Oct. 2004), 32–39.

[46] Mike Just and David Aspinall. 2009. Personal Choice and Challenge Questions: A Security and Usability Assessment. In *Symposium on Usable Privacy and Security (SOUPS '09)*. ACM, Mountain View, California, USA, 8:1–8:11.

[47] Nader Abdel Karim, Zarina Shukur, and Al-Banna Abedal-Kareem. 2020. Uipa: User Authentication Method Based on User Interface Preferences for Account Recovery Process. *Journal of Information Security and Applications* 52 (June 2020), 102466.

[48] Markus Keil, Philipp Markert, and Markus Dürmuth. 2022. "It's Just a Lot of Prerequisites": A User Perception and Usability Analysis of the German ID Card as a FIDO2 Authenticator. In *European Symposium on Usable Security (EuroUSEC '22)*. ACM, Karlsruhe, Germany, 172–188.

[49] Hyoungshick Kim, John Tang, and Ross Anderson. 2012. Social Authentication: Harder Than It Looks. In *Financial Cryptography and Data Security (FC '12)*. Springer, Kralendijk, Bonaire, 1–15.

[50] Lydia Kraus, Mária Švidroňová, and Elizabeth Stobert. 2021. How Do Users Chain Email Accounts Together?. In *International Conference on ICT Systems Security and Privacy Protection (IFIP SEC '21)*. Springer, Oslo, Norway, 416–429.

[51] Johannes Kunke, Stephan Wiefling, Markus Ullmann, and Luigi Lo Iacono. 2021. Evaluation of Account Recovery Strategies with FIDO2-based Passwordless Authentication. In *Open Identity Summit (ODI '21)*. GI, Copenhagen, Denmark, 59–70.

[52] Leona Lassak, Elleen Pan, Blase Ur, and Maximilian Golla. 2024. Why Aren't We Using Passkeys? Obstacles Companies Face Deploying FIDO2 Passwordless Authentication. In *USENIX Security Symposium (SSYM '24)*. USENIX, Philadelphia, Pennsylvania, USA.

[53] Yue Li, Haining Wang, and Kun Sun. 2018. Email as a Master Key: Analyzing Account Recovery in the Wild. In *IEEE Conference on Computer Communications (INFOCOM '18)*. IEEE, Honolulu, Hawaii, USA, 1646–1654.

[54] Fatma Al Maqbali and Chris J Mitchell. 2018. Email-based Password Recovery - Risking or Rescuing Users?. In *International Carnahan Conference on Security Technology (ICCST '18)*. IEEE, Montreal, Quebec, Canada, 1–5.

[55] Philipp Markert, Florian Farke, and Markus Dürmuth. 2019. View The Email to Get Hacked: Attacking SMS-Based Two-Factor Authentication. In *Who Are You?! Adventures in Authentication Workshop (WAY '19)*. USENIX, Santa Clara, California, USA, 1–6.

[56] Philipp Markert, Maximilian Golla, Elizabeth Stobert, and Markus Dürmuth. 2019. Work in Progress: A Comparative Long-Term Study of Fallback Authentication. In *Workshop on Usable Security and Privacy (USEC '19)*. ISOC, San Diego, California, USA.

[57] Philipp Markert, Leona Lassak, Maximilian Golla, and Markus Dürmuth. 2023. Understanding Users' Interaction with Login Notifications. *CoRR* abs/2212.07316 (June 2023), 1–26.

[58] Tara Matthews, Kathleen O'Leary, Anna Turner, Manya Sleeper, Jill Palzkill Woelfer, Martin Shelton, Cori Manthorne, Elizabeth F. Churchill, and Sunny Consolvo. 2017. Stories from Survivors: Privacy & Security Practices when Coping with Intimate Partner Abuse. In *ACM Conference on Human Factors in Computing Systems (CHI '17)*. ACM, Denver, Colorado, USA, 2189–2201.

[59] Nicholas Micallef and Nalin Asanka Gamagedara Arachchilage. 2017. A Gamified Approach to Improve Users' Memorability of Fall-back. In *Who Are You?! Adventures in Authentication Workshop (WAY '17)*. USENIX, Santa Clara, California, USA.

[60] Nicholas Micallef and Nalin Asanka Gamagedara Arachchilage. 2021. Understanding Users' Perceptions to Improve Fallback Authentication. *Personal and Ubiquitous Computing* 25, 5 (May 2021), 897–910.

[61] Microsoft, Corporation. 2023. Configure Temporary Access Pass to Register Passwordless Authentication Methods. https://learn.microsoft.com/en-us/entra/identity/authentication/howto-authentication-temporary-access-pass, as of March 16, 2024.

[62] Grzergor Milka. 2018. Anatomy of Account Takeover. In *USENIX Enigma Conference (Enigma '18)*. USENIX, Santa Clara, California, USA.

[63] Federal Bureau of Investigation. 2022. Criminals Increasing SIM Swap Schemes to Steal Millions of Dollars from US Public. https://www.ic3.gov/Media/Y2022/PSA220208, as of March 16, 2024.

[64] Simon Parkin, Samy Driss, Kat Krol, and M. Angela Sasse. 2015. Assessing the User Experience of Password Reset Policies in a University. In *International Conference on Passwords (PASSWORDS '15)*. Springer, Cambridge, United Kingdom, 21–38.

[65] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2019. Why People (Don't) Use Password Managers Effectively. In *Symposium on Usable Privacy and Security (SOUPS '19)*. USENIX, Santa Clara, California, USA, 319–338.

[66] Thanasis Petsas, Giorgos Tsirantonakis, Elias Athanasopoulos, and Sotiris Ioannidis. 2015. Two-Factor Authentication: Is the World Ready? Quantifying 2FA Adoption. In *European Workshop on System Security (EuroSec '15)*. ACM, Bordeaux, France, 4:1–4:7.

[67] Jamie L. Pinchot and Karen L. Paullet. 2012. What's in Your Profile? Mapping Facebook Profile Data to Personal Security Questions. *Issues in Information Systems* 13, 1 (March 2012), 284–293.

[68] Iasonas Polakis, Panagiotis Ilia, Federico Maggi, Marco Lancini, Georgios Kontaxis, Stefano Zanero, Sotiris Ioannidis, and Angelos D. Keromytis. 2014. Faces in the Distorting Mirror: Revisiting Photo-Based Social Authentication. In *ACM*

Conference on Computer and Communications Security (CCS '14). ACM, Scottsdale, Arizona, USA, 501–512.

[69] Iasonas Polakis, Marco Lancini, Georgios Kontaxis, Federico Maggi, Sotiris Ioannidis, Angelos D. Keromytis, and Stefano Zanero. 2012. All Your Face Are Belong to Us: Breaking Facebook's Social Authentication. In *Annual Computer Security Applications Conference (ACSAC '12)*. ACM, Orlando, Florida, USA, 399–408.

[70] Gaston Pugliese, Christian Riess, Freya Gassmann, and Zinaida Benenson. 2020. Long-Term Observation on Browser Fingerprinting: Users' Trackability and Perspective. In *Privacy Enhancing Technologies Symposium (PETS '20)*. Sciendo, Virtual Conference, 558–577.

[71] Ariel Rabkin. 2008. Personal Knowledge Questions for Fallback Authentication: Security Questions in the Era of Facebook. In *Symposium on Usable Privacy and Security (SOUPS '08)*. ACM, Pittsburgh, Pennsylvania, USA, 13–23.

[72] Simone Raponi and Roberto Di Pietro. 2020. A Longitudinal Study on Web-Sites Password Management (in)Security: Evidence and Remedies. *IEEE Access* 8 (March 2020), 52075–52090.

[73] M. Angela Sasse and Ivan Flechais. 2005. *Usable Security: Why Do We Need It? How Do We Get It?* (1 ed.). O'Reilly and Associates, Sebastopol, California, USA, Chapter 2, 13–30.

[74] M. Angela Sasse, Matthew Smith, Cormac Herley, Heather Lipford, and Kami Vaniea. 2016. Debunking Security-Usability Tradeoff Myths. *IEEE Security & Privacy* 14, 5 (Oct. 2016), 33–39.

[75] M. Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. 2014. The Great Authentication Fatigue – And How to Overcome It. In *International Conference on Cross-Cultural Design (CCD '14)*. Springer, Heraklion, Crete, Greece, 228–239.

[76] Stuart Schechter, A. J. Bernheim Brush, and Serge Egelman. 2009. It's No Secret. Measuring the Security and Reliability of Authentication via "Secret" Questions. In *IEEE Symposium on Security and Privacy (SP '09)*. IEEE, Oakland, California, USA, 375–390.

[77] Stuart Schechter, Serge Egelman, and Robert W. Reeder. 2009. It's Not What You Know, But Who You Know: A Social Approach to Last-Resort Authentication. In *ACM Conference on Human Factors in Computing Systems (CHI '09)*. ACM, Boston, Massachusetts, USA, 1983–1992.

[78] Kaiwen Shen, Chuhan Wang, Minglei Guo, Xiaofeng Zheng, Chaoyi Lu, Baojun Liu, Yuxuan Zhao, Shuang Hao, Haixin Duan, Qingfeng Pan, and Min Yang. 2021. Weak Links in Authentication Chains: A Large-scale Analysis of Email Sender Spoofing Attacks. In *USENIX Security Symposium (SSYM '21)*. USENIX, Virtual Conference, 3201–3217.

[79] Roger N. Shepard and Jacqueline Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171, 3972 (Feb. 1971), 701–703.

[80] Bijan Soleymani and Muthucumaru Maheswaran. 2009. Social Authentication Protocol for Mobile Phones. In *IEEE International Conference on Computational Science and Engineering (CSE '09)*. IEEE, Vancouver, British Columbia, Canada, 436–441.

[81] Vlasta Stavova, Vashek Matyas, and Mike Just. 2016. Codes v. People: A Comparative Usability Study of Two Password Recovery Mechanisms. In *International Conference on Information Security Theory and Practice (WISTP '16)*. Springer, Heraklion, Greece, 33–50.

[82] Elizabeth Stobert and Robert Biddle. 2014. The Password Life Cycle: User Behaviour in Managing Passwords. In *Symposium on Usable Privacy and Security (SOUPS '14)*. USENIX, Menlo Park, California, USA, 243–255.

[83] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. 2021. Synthesizing Obama: Learning Lip Sync from Audio. *ACM Transactions on Graphics* 36, 4 (July 2021), 95:1–95:13.

[84] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. 2022. Care Infrastructures for Digital Security in Intimate Partner Violence. In *ACM Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New Orleans, Louisiana, USA, 123:1–123:20.

[85] Timothy W. van der Horst and Kent E. Seamons. 2007. Simple Authentication for the Web. In *Conference on Security and Privacy in Communication Networks (SecureComm '07)*. IEEE, Nice, France, 473–482.

[86] Steven G. Vandenberg and Allan R. Kuse. 1978. Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47, 2 (Oct. 1978), 599–604.

[87] Elham Vaziripour, Justin Wu, Mark O'Neill, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. 2017. Is that you, Alice? A Usability Study of the Authentication Ceremony of Secure Messaging Applications. In *Symposium on Usable Privacy and Security (SOUPS '17)*. USENIX, Santa Clara, California, USA, 29–47.

[88] Ahmed Wahab, Daqing Hou, Stephanie Schuckers, and A. Barbir. 2019. Utilizing Keystroke Dynamics as Additional Security Measure to Protect Account Recovery Mechanism. In *International Conference on Information Systems Security and Privacy (ICISSP '21)*. USENIX, Virtual Conference, 33–42.

[89] Justin Wu, Cyrus Gattrell, Devon Howard, Jake Tyler, Elham Vaziripour, Daniel Zappala, and Kent Seamons. 2019. "Something Isn't Secure, but I'm Not Sure How That Translates Into a Problem": Promoting Autonomy by Designing for Understanding in Signal. In *Symposium on Usable Privacy and Security (SOUPS '19)*. USENIX, Santa Clara, California, USA, 137–153.

[90] Sarita Yardi, Nick Feamster, and Amy Bruckman. 2008. Photo-Based Authentication Using Social Networks. In *Workshop on Online Social Networks (WOSN '08)*. ACM, Seattle, Washington, USA, 55–60.

[91] John D. Yesberg and Mark S. Anderson. 1996. Quantitative Authentication and Vouching. *Computers & Security* 15, 7 (1996), 633–645.

[92] Peng Zhao, Kaigui Bian, Tong Zhao, Xintong Song, Jung-Min Park, Xiaoming Li, Fan Ye, and Wei Yan. 2017. Understanding Smartphone Sensor and App Data for Enhancing the Security of Secret Questions. *IEEE Transactions on Mobile Computing* 16, 2 (Feb. 2017), 552–565.

[93] Wei Zhou, XiaoWei Yuan, Wenjun Chai, and Hui Ma. 2019. Deep Learning Based Attack On Social Authentication System. In *IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC '19)*. IEEE, Chengdu, China, 982–986.

[94] Moshe Zviran and William J. Haga. 1990. User Authentication by Cognitive Passwords: An Empirical Assessment. In *Jerusalem Conference on Information Technology (JCIT '90)*. IEEE, Jerusalem, Israel, 137–144.

A SURVEY INSTRUMENT

Stage 1: Enrollment

Participants solved 5x mental rotation tests.

Demographic Information

To improve the quality of our research, we kindly ask you to provide some demographic information in the form below.

S1-D1 What is your age range?
○ 18-24 ○ 25-34 ○ 35-44 ○ 45-54 ○ 55-64 ○ 65-74 ○ 75 or older
○ Prefer not to say

S1-D2 Which of these best describes your current gender identity?
○ Woman ○ Men ○ Non-binary ○ Prefer to self-describe: _____
○ Prefer not to say

S1-D3 What is the highest level of education you have completed?
○ Some high school ○ High school ○ Some college
○ Trade, technical, or vocational training ○ Associate's degree
○ Bachelor's degree ○ Master's degree ○ Professional degree ○ Doctorate
○ Prefer not to say

S1-D4 Which of the following best describes your educational background or job field?
○ I have an education in, or work in, the field of computer science, computer engineering or IT
○ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT
○ Prefer not to say

One More Thing

S1-H Please indicate if you have honestly participated in this survey and followed instructions completely. You will not be penalized/rejected for indicating 'No' but your data may not be included in the analysis:
○ Yes, I participated honestly ○ No, I did not participate honestly

Stage 2: Recall

Participants solved 5x mental rotation tests.

One More Thing

S2-H Please indicate if you have honestly participated in this survey and followed instructions completely. You will not be penalized/rejected for indicating 'No' but your data may not be included in the analysis:
○ Yes, I participated honestly ○ No, I did not participate honestly

Stage 3: Long-Term

Debriefing

Participants were debriefed and told about the actual purpose of the study.

Questionnaire

Please provide some information about your reset process.

*For participants who used the **email** scheme.*

S3-EM1 Which problems did you encounter during the reset process? (select all that apply):
☐ I no longer have access to the registered email account
☐ I didn't receive a reset email ☐ The reset link in the email was invalid
☐ I couldn't set a new password on the reset page ☐ Other: _____

S3-EM2 Use the text field below to give a detailed description of any problems you had during the reset process:
Answer: _____

*For participants who used the **SMS** scheme.*

S3-SMS1 Which problems did you encounter during the reset process? (select all that apply):
☐ I no longer have access to the registered phone number
☐ I didn't receive a reset SMS ☐ The SMS reset code was invalid
☐ I couldn't set a new password on the reset page ☐ Other: _____

S3-SMS2 Use the text field below to give a detailed description of any problems you had during the reset process:
Answer: _____

S3-SMS3 I have my cell phone within reach when surfing the web:
○ Never ○ Rarely ○ Sometimes ○ Often ○ Always

*For participants who used the **personal knowledge question** scheme.*

S3-PKQ1 Which problems did you encounter during the reset process? (select all that apply):
☐ I couldn't remember the answers to my reset questions
☐ I couldn't set a new password on the reset page ☐ Other: _____

S3-PKQ2 Use the text field below to give a detailed description of any problems you had during the reset process:
Answer: _____

*For participants who used the **designated trustees** scheme.*

S3-DT1 How many of the email addresses that you stated as trusted contacts belong to you?
○ 0 ○ 1 ○ 2 ○ 3

*If participants selected "0", "1", or "2" in response to **S3-DT1**.*

S3-DT2 How did you get in touch with your trusted contacts?
☐ Phone call ☐ SMS ☐ Email ☐ Instant messenger ☐ In person
☐ Other: _____

S3-DT3 Which problems did you encounter during the reset process? (select all that apply):
☐ I couldn't remember all of my trusted contacts
☐ My trusted contacts didn't receive an email
☐ My trusted contacts didn't have access to their email account
☐ My trusted contacts didn't respond
☐ The provided reset codes were invalid
☐ I couldn't set a new password on the reset page ☐ Other: _____

S3-DT4 Use the text field below to give a detailed description of any problems you had during the reset process:
Answer: _____

From now on, there was no difference between the treatments.

For the assessment of *[treatment]*, please select your agreement/disagreement with the following statements.

S3-SUS1 I think that I would like to use *treatment* frequently:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS2 I found *treatment* unnecessarily complex:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS3 I thought *treatment* was easy to use:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS4 I think that I would need the support of a technical person to be able to use *treatment*:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS5 I found the various functions in *treatment* were well integrated:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS6 I thought there was too much inconsistency in *treatment*:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-AC Select "Agree" as the answer to this question:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS7 I would imagine that most people would learn to use *treatment* very quickly:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS8 I found *treatment* very cumbersome to use:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS9 I felt very confident using *treatment*:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

S3-SUS10 I needed to learn a lot of things before I could get going with *treatment*:
○ Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly disagree

Demographic Information

To improve the quality of our research, we kindly ask you to provide some demographic information in the form below.

S3-D1 What is your age range?
○ 18-24 ○ 25-34 ○ 35-44 ○ 45-54 ○ 55-64 ○ 65-74 ○ 75 or older
○ Prefer not to say

S3-D2 Which of these best describes your current gender identity?
○ Woman ○ Men ○ Non-binary ○ Prefer to self-describe: _____
○ Prefer not to say

S3-D3 What is the highest level of education you have completed?
○ Some high school ○ High school ○ Some college
○ Trade, technical, or vocational training ○ Associate's degree
○ Bachelor's degree ○ Master's degree ○ Professional degree ○ Doctorate
○ Prefer not to say

S3-D4 Which of the following best describes your educational background or job field?
○ I have an education in, or work in, the field of computer science, computer engineering or IT
○ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT
○ Prefer not to say

One More Thing

S3-H Please indicate if you have honestly participated in this survey and followed instructions completely. You will not be penalized/rejected for indicating "No" but your data may not be included in the analysis:
○ Yes, I participated honestly ○ No, I did not participate honestly

B ILLUSTRATIONS FROM THE RESET PROCESSES

B.1 Email

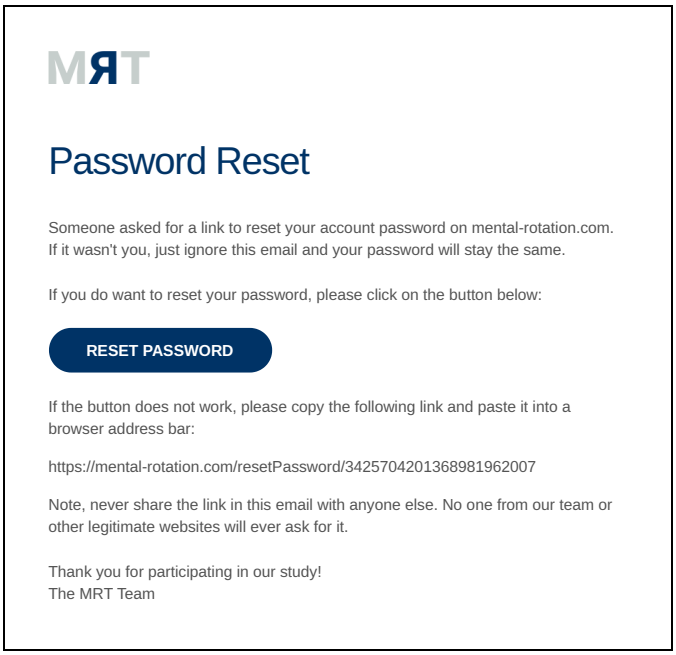
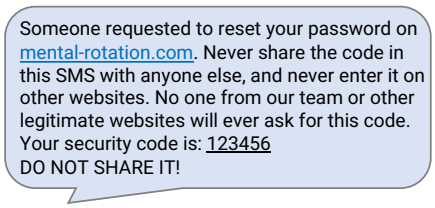
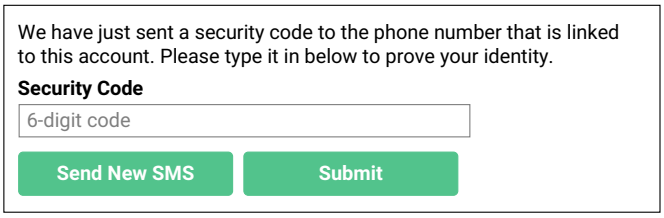


Figure 6: The reset email we sent to the participants of the email reset group.

B.2 SMS



(a) The SMS we sent containing the security code.



(b) Form to provide the security code received via SMS.

Figure 7: Message and interface shown to participants in the SMS group.

B.3 Designated Trustees

MRT

Hi,

We are researchers conducting a study about changes in spatial reasoning ability over time. The person who uses the email address participant-1@gmail.com is participating in this study and selected you as a trusted contact. As a trusted contact you will help our participant to get back into their account for the study in case the password needs to be reset.

With this email, we are just informing you about this role. You do not have to do anything at all. However, if you want to learn more about our study, feel free to visit our website https://mental-rotation.com.

Thank you!
The MRT Team

Figure 8: Email we sent to trusted contacts informing them about their role. Additionally, this email was used to confirm the provided email address exists.

MRT

Password Reset

The person who uses the email address participant-1@gmail.com is participating in our study about changes in spatial reasoning ability and wants to reset the password for the account used in the study.

You are receiving this email because we use "Trusted Contacts" for the password reset. That means our participants state the email addresses of trusted contacts and each of those contacts receives a security code when a password reset is initiated. By providing these codes, the participant confirms the identity and is able to set a new password.

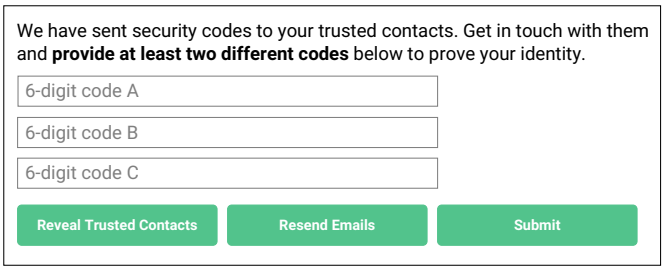
Therefore, we kindly ask you to provide the following code in case the person reaches out to you: **123456**

To prevent any malicious activity, please ensure that the person who receives the code has access to the email address participant-1@gmail.com.

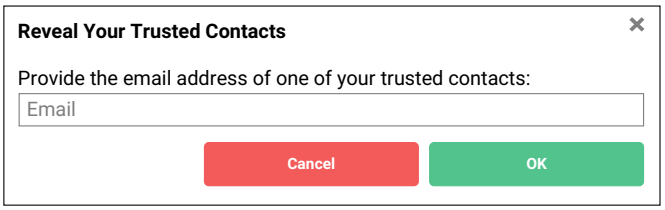
If you want to learn more about our study, feel free to visit our website https://mental-rotation.com.

Thank you for your help!
The MRT Team

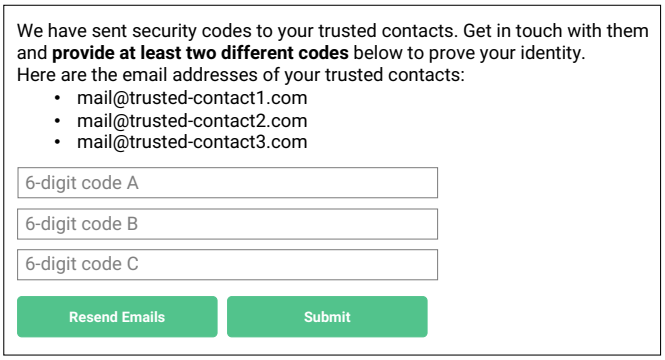
Figure 9: The email we sent to the trusted contacts containing the reset code.



(a) Form to provide reset codes received from trusted contacts.



(b) Option to reveal email addresses of trusted contacts.



(c) Form with trusted contacts revealed.

Figure 10: Interfaces shown to participants in the trustee group.

B.4 Personal Knowledge Questions

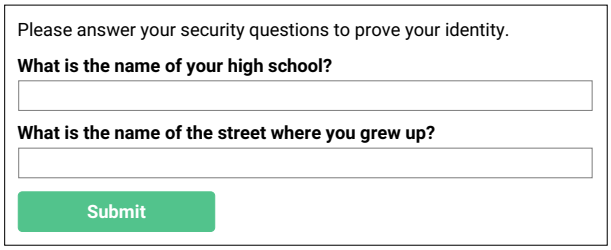


Figure 11: The form shown to the participants to answer the personal knowledge questions.