A Usability Study of Five Two-Factor Authentication Methods

Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, Kent Seamons
Brigham Young University

Abstract

Two-factor authentication (2FA) defends against account compromise. An account secured with 2FA typically requires an individual to authenticate using something they know—typically a password—as well as something they have, such as a cell phone or hardware token. Many 2FA methods in widespread use today have not been subjected to adequate usability testing. Furthermore, previous 2FA usability research is difficult to compare due to widely-varying contexts across different studies. We conducted a two-week, between-subjects usability study of five common 2FA methods with 72 participants, collecting both quantitative and qualitative data. Participants logged into a simulated banking website nearly every day using 2FA and completed an assigned task. Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA to provide more security for their sensitive online accounts. We also conducted a within-subjects laboratory study with 30 participants to assess the general usability of the setup procedure for the five methods. While a few participants experienced difficulty setting up a hardware token and a one-time password, in general, users found the methods easy to set up.

1 Introduction

Passwords are the most widespread form of user authentication on the web today [9]. Although many password-replacement schemes have been proposed, none of them compete with the deployability and usability of passwords [8]. Recently, large service providers, including Google, Facebook, and Microsoft, have deployed an optional 2FA layer as part of their authentication processes to defend against account compromise. Two-factor authentication requires users to present two of the following types of authentication factors:

1. Something they *know* (traditionally a password)
2. Something they *have* (such as a phone or hardware token)
3. Something they *are* (referring to biometrics, such as a fingerprint)

Several 2FA methods are in use. Methods such as SMS, TOTP (time-based one-time password), and hardware code generators (such as the RSA SecurID) require the user to enter a single-use code in addition to their password. These codes are either sent to the user via a separate channel or are generated on the fly by the user's device. In commercial and government settings, smart cards are a commonly used second factor, requiring the user to insert an ID badge into a card reader attached to their computer. Online banking systems, particularly in the UK, frequently use variants of hardware code generators and card readers in their 2FA implementations. Companies including Google, Dropbox, and Github have deployed USB hardware tokens (aka security keys), such as YubiKey, internally [18].

Two-factor authentication provides a strong defense against account compromise. The number of recent password database leaks [2] underscores the risk of account compromise. Because users tend to reuse the same username and password across multiple sites [11], password leaks from a single site can lead to a chain-reaction of account compromises as attackers access other accounts with the same credentials [15]. Even if an attacker steals or guesses a user's password, the attacker must compromise the user's phone or steal a physical token to gain access to the account. Thus it is significantly more difficult for a remote attacker to compromise an account protected by a second authentication factor.

Despite the attractive security benefits of 2FA, its impact on the user experience remains unclear. Previous studies on

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

2FA have produced results which may appear contradictory. While one set of studies [14] [16] [17] [25] concludes that 2FA is completely unusable, others [13] [18] find that some 2FA methods are actually very usable.

It is difficult to draw general conclusions from these prior surveys and studies because of widely-differing conditions. These confounding factors make it very difficult to determine how the different methods compare in terms of usability.

We conducted a two-week, between-subjects usability study of five common 2FA methods with 72 participants, collecting both quantitative and qualitative data. Participants logged into a simulated banking website nearly every day using 2FA and completed an assigned task. Having all the participants experience 2FA within the context of a single application reduces the confounding factors that are usually present when comparing the results of different 2FA methods across usability studies. Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA to provide more security for their sensitive online accounts.

We purposely ignored setup issues during our initial study to not bias participants toward the day-to-day usability of one of the factors based on a poor setup experience. However, the promising results from the two-week study leave open the question about whether encouraging results for a given factor are incomplete if there is an associated usability hurdle to set up that factor. To gain insight into this question, we conducted a within-subjects laboratory study of the setup process for the five 2FA methods. While a few participants experienced difficulty setting up a hardware token and a one-time password, in general, users found the methods easy to set up.

2 Related Work

Previous research has explored the usability of 2FA methods through lab studies and surveys.

2.1 Lab Studies

Ace et al. [4] studied the setup and login of four of Google's 2FA methods. They found that participants experienced many failures and found Google's 2FA system hard to use. The order of preference of the four systems reported in their study exactly match the preference ordering of those four systems in our study, but setup results differ significantly. Our differing results may be explained in part because they measured 2FA setup and login with the same participants in a single study. Also, Google changed the setup instructions between their setup study and ours, which may account for our more positive setup results.

Weir et al. [25] compared the usability of three hardware code generator under evaluation by a bank in the UK. Users preferred the system that was most convenient over systems with stronger security. Weir et al. [26] also conducted a lab study of three authentication systems, including SMS and hardware code generator based two-factor systems. Participants were most successful using the SMS-based system.

Lang et al. [18] report on Google's internal deployment of security keys to their employees. They report a long-term reduction in the number of authentication-related support tickets after deploying the hardware keys. Further, they demonstrate a significant reduction in overall authentication time compared to other one-time code based methods.

Das et al. [12] performed two studies measuring both the usability and the acceptability of using the YubiKey (a type of FIDO U2F compliant hardware token) as a second factor in securing a Google account. Employing a think-aloud protocol, they made some recommendations to Yubico (the manufacturer of the YubiKey) based on common points of confusion. After one year, they repeated the study with a new group of users and found that although many of the previous usability concerns had been addressed, many users still did not see much benefit in using the YubiKey. Das et al. postulated that this lack of acceptability was due partly to the lack of awareness of the risks mitigated through using the YubiKey.

Reynolds et al. [21] describe two usability studies of YubiKeys. The study found many usability concerns with the setup process of the YubiKey but found that day-to-day usability was significantly higher. Similar to our study, participants used the YubiKey for several weeks, although we studied the YubiKey in conjunction with several other 2FA methods.

2.2 Surveys

Krol et al. [17] conducted interviews with 21 individuals who used two-factor authentication as part of the login process for several UK banks. Participants used a variety of two-factor methods, including card readers, hardware code generators, SMS, phone calls, and smartphone apps that generated single-use codes. Participants particularly disliked hardware code generators; in fact, a few individuals changed banks because of the difficulty of using the tokens. De Cristofaro et al. [13] conducted a Mechanical Turk survey of participants with experience using hardware code generators, one-time codes via SMS and email, and smartphone code generator apps. They found that email or SMS messages were the most commonly used second factor for financial or personal sites, and hardware tokens were the most common for work. Each of the methods received SUS (System Usability Scale) scores in the 'A' range.

Duo is a commercial 2FA product that supports second-factor authentication using a smartphone, phone calls, U2F, and several other methods. Weidman and Grossklags [24] studied the transition from a token-based 2FA system to Duo for employees through a survey at Pennsylvania State University. They found that employees preferred the prior

token-based system to using the Duo app. Some employees' preference was influenced by their dissatisfaction with being required to use personal devices for work. Colnago et al. [10] conducted a large-scale survey of faculty and students at Carnegie Mellon University during a campus-wide deployment of the Duo 2FA system. The results showed that many participants in the survey recognized the security benefits of using 2FA. They also identified usability issues with the deployment of Duo. Differences in perceived usability between users that *voluntarily* adopted 2FA and those that were *required* to adopt 2FA were fairly small, and many participants that were required to use 2FA reported it to be easier to use than they expected.

3 Five 2FA Methods

We compare five common 2FA methods: SMS, TOTP, pre-generated-codes, push, and U2F security keys. The differences between our study and the prior work is that we study all five methods in the context of a single simulated web application to reduce the potential for confounding factors and to be able to measure the time to authenticate using each method. We also separate setup and daily use. We are also the first study to include pre-generated codes. This section describes each method and its security properties.

3.1 SMS

One of the most widely deployed 2FA methods is SMS. The user is sent a one-time verification code (usually six digits) through a text message to their mobile phone. The broad deployment is partly because most consumers already own a mobile phone capable of receiving text messages—99% of Americans according to a recent Pew study [3]. Potential usability problems may include delayed delivery, lack of cellular service (such as in a foreign country or remote location), and miscopying the code from phone to computer.

SMS-based authentication is vulnerable at several stages. Mobile networks do not encrypt messages while in transit, allowing attackers to conduct man-in-the-middle attacks. Of particular concern, is the well-documented SIM-swapping attack [5, 20] Also, the server (or relying party) must securely store the one-time code while the SMS message is sent, received by the user, and entered back into the site for verification. The code could be salted and hashed to prevent casual theft, but a determined attacker could easily conduct a brute force attack on a stolen hashed code given the relatively small number of codes. Attackers may also steal SMS codes through targeted phishing attacks. Some ways to mitigate these threats are to invalidate a code after a short time window and limit the number of failed attempts to log in with a code.

3.2 TOTP

To set up TOTP, the user first synchronizes a secret key generated by the provider to their smartphone, usually by scanning a QR code. The app generates a verification code by combining the secret with a truncated timestamp, hashing the value, and truncating the result to derive the verification code (as with SMS, usually 6 or 7 digits long). The server verifies the user-supplied code using the same method. The advantage of using a TOTP code generator app is that after syncing the secret to the phone, the user does not need to rely on a cellular provider to deliver the one-time codes—eliminating both a potential attack surface and a problem with usability. However, if an attacker steals the TOTP secret from the server or the phone, then the attacker may be able to impersonate the user.

Each code is valid for a set time interval, usually only 30 seconds, after which a new code must be generated. The smartphone and the server must both have a clock that is reasonably in sync. A server accepts tokens for the current 30-second window and the 30-second window just before and after the current window to account for clock drift. Crucially, this means that users may have as little as 30 seconds to enter the code because codes can be generated anytime during the 30-second interval. As with SMS, the verification codes still must be manually keyed in by the user, leaving additional room for user error. According to Pew [3], 77% of Americans own a smartphone, meaning that TOTP is not as broadly deployable to all customer bases as SMS.

TOTP requires a shared secret key between the server and the user's mobile device. This secret must be stored securely, but a one-way hashing mechanism is not useful since the secret is an input to the code-generation and verification process. On the server side, the shared secret could be encrypted using the user's password to prevent casual theft. Assuming secure storage of the shared secret on both the client and server, TOTP has a significant advantage over SMS codes—it does not rely on the insecure mobile network for delivery of the code, thus eliminating an entire attack surface.

3.3 Pre-generated Codes

Pre-generated codes are often a backup 2FA authentication method in case the user is unable to access their primary 2FA method. Implementation is straightforward: the service provider generates a list of verification codes and has the user print or write the codes down. The length of the list itself is variable, and the codes are usually around 8 digits long. The codes may be used in any order and must be kept secure by both the server and the user to prevent theft. Because these codes are usually longer than the codes sent through SMS or generated with TOTP, there is additional room for user error when entering the codes. Furthermore, the user must be careful not to lose the medium on which they recorded the

codes.

Printed codes are usually used as a backup authentication mechanism, and must be stored on the server for long periods. Even applying the hashing mechanism discussed for SMS codes, the non-expiring nature of the codes would make them vulnerable to an offline brute-force attack. Although more technically complex to implement, one mitigation against a brute-force attack would be to hash the backup code with the user's password. On the user's side, the printed codes must be stored securely using traditional physical security measures. An open question is how users would prefer to store such backup codes—do users prefer to keep the codes on their person for convenience (perhaps storing the codes in a wallet or purse), or would they prefer to take more stringent security measures to safeguard the codes.

3.4 Push

In the push method, the user receives a push notification on their smartphone that allows the user to either "Approve" or "Deny" a login attempt. Push authentication requires Internet access. Google supports this technique (through their "Google prompt"), and it is also available through commercial applications such as Authy OneTouch and DUO Mobile. The advantage of this method is that there is less chance of user error since there are no numbers to copy off a phone screen correctly. We hypothesize that not having to type in numbers, as required by other 2FA methods, is both faster and perceived as more usable by participants.

Push authentication does not explicitly require the storage of a secret key; however, the server must ensure that the push notifications are sent to the correct device, suggesting that some form of two-way verification of the client and server must take place. Additionally, communication between the user's device and the server must be kept secure, such as through the use of TLS. The most prominent push-based authentication methods are proprietary, making it difficult to verify the exact security measures in place and requires implicitly trusting a third party. Push-based authentication has not yet been well-studied by the security community.

3.5 U2F Security Keys

Originally developed through a collaboration with Google and Yubico, and now sponsored by the FIDO (Fast IDentity Online) Alliance, Universal 2nd Factor (U2F) is an open standard for authentication using a USB hardware device. To authenticate with a security key, the user must connect the device to their computer and activate the device when prompted by the website.

The U2F standard was designed to be highly secure while still boasting good usability [18]. In contrast to the other four 2FA methods described above, the U2F standard itself is designed to prevent phishing attacks and provide more security and privacy protections than other forms of 2FA. U2F authentication requires that the server store a public key that the user generates at registration time—the secret key never leaves the U2F device. The main risk is that a user might lose their U2F device—but device loss is also a risk with the other four 2FA methods.

4 Two-week Study Methodology

We conducted an IRB-approved, two-week study of 2FA at Brigham Young University (BYU). The research objective of the study was to compare the usability of the five common 2FA methods described in Section 3.

4.1 Study Design

A total of 72 participants were divided into 6 groups of 12 participants each. Five of the groups were assigned to a specific 2FA method, and the final group was a control group that used only passwords with no second factor. Each participant initially met with a study coordinator to create an account on the study website. During this meeting, the participant was given a list of 12 tasks to complete on the study website over the next two-week period (with no more than one task completed per day). As part of completing each task, each participant would log in to the study website using their assigned authentication mechanism. After the two weeks, participants returned for an exit interview with a study coordinator. Using a combination of authentication event timing data, survey responses, and qualitative data gathered from the exit interviews, we compared the usability of the various authentication methods under test and made some observations and recommendations based on this data.

4.2 Banking Website

Our test scenario was that of a participant needing to log in to an online banking interface and complete a task, such as transferring money between accounts or paying a bill online. To support this scenario, we built a simulated online banking interface. supported authentication through either a password alone or a password plus one of the five 2FA methods described previously.

4.3 Recruitment

We recruited the 72 participants using flyers posted throughout the BYU campus. Prospective subjects were informed that they would need daily access to an Internet-connected computer with Google Chrome. We required Chrome because it is the only major browser that supports U2F security keys by default. To be eligible for the study, potential participants

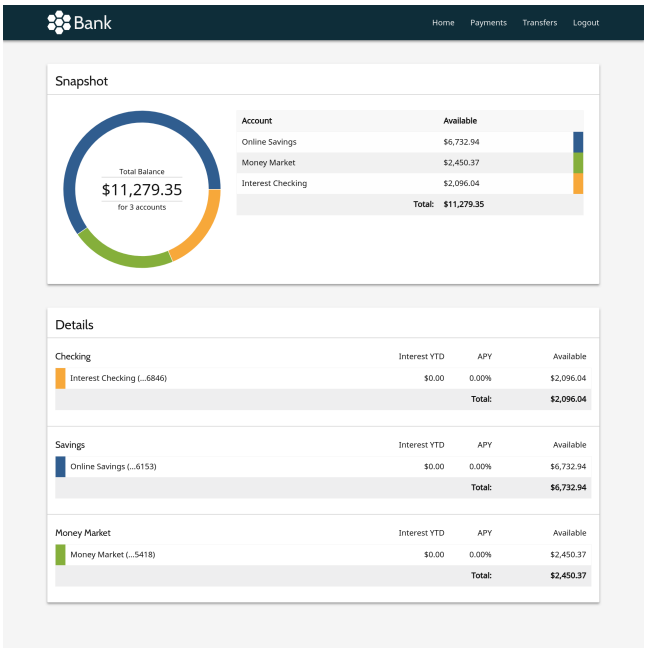

Figure 1: Example of the banking interface we constructed for our study

filled out a short survey to report whether they owned an Android or iOS smartphone, or if they owned a phone able to receive text messages.

Participants were then randomly assigned to a study group. One participant did not own a smartphone and was randomly assigned to a study group that did not require the use of a smartphone. Once a group reached 12 participants, we removed it from the pool of potential groups to which a participant could be assigned.

4.4 Demographics

We had a slightly higher number of female participants (38; 55%) as compared to male participants (31; 45%) in our study. Participants were largely young adults: 18–19 years (3; 4%), 20–29 (61; 88%), and 30–39 (5; 7%). Over two-thirds of the participants (49; 71%) had completed some college but had not yet completed a degree. Participants self-reported their level of computer expertise: far above average (13; 19%), somewhat above average (28; 41%), average (25; 36%), and somewhat below average (3; 4%).

4.5 Setup and Initial Meeting

Participants scheduled an initial appointment to meet with a study coordinator. During the initial meeting, the study coordinator assisted them in setting up an account on the online banking interface. We allowed participants to choose their username and password, with the only restriction being that the password had to be at least eight characters long.

If the participant belonged to one of the study groups using a second-factor method, the coordinator would also help them configure 2FA on their account for the study website. Depending on the study group, this included helping the participant install any necessary apps (Authy for push, Google Authenticator for TOTP), verifying their phone number, issuing the participant a U2F device (the YubiKey NEO), or printing the backup codes. Finally, the study coordinator assisted the participant in completing the first listed task during the initial meeting, leaving the participant with 11 tasks to complete on their own.

For this study, we purposely chose to focus only on the day-to-day use of 2FA methods and not confound those results with any negative issues arising from the usability of the setup process. Recent papers have studied 2FA setup of YubiKeys [12, 21], for instance, and argue that researchers should examine setup and day-to-day use independently. If day-to-day use is acceptable and promising to users, this can motivate more energy to address problematic setup procedures.

4.6 Two-week Task Completion Period

Over the next two weeks, participants were asked to complete no more than one task per day in the order given on their task list. To complete each task, the participant would need to visit our online banking website and log in with their previously selected username and password. Except for the control group using only a username-password pair, the participant would also authenticate using their assigned second-factor method for each login. After logging in, the participant would go to either the "Payments" or "Transfers" page and complete the banking component of the task. The purpose of having participants complete the banking-related task after logging in (as opposed to merely having the individual log in and do nothing) was to encourage the user to act more naturally during the login process and make the simulation more realistic—most real-world users do not authenticate for amusement; instead authentication is a means to an end.

4.7 Exit Interview

Participants reported back for an exit interview with a study coordinator after the two weeks. The coordinator first had the participant take a brief survey to gather a small amount of demographic data. Participants also completed a SUS (System Usability Scale) assessment of the website as a whole, and for the authentication method they had used during the study. Following this, the coordinator conducted a semi-structured interview with the participant to gather additional information about how the participant felt about the website overall as well as the login process. In particular, we asked participants questions about their overall online security posture to better

Table 1: Repeated measures correlation (rmcorr) between amount of time participating in study versus amount of time to authenticate.

2FA Method	p-value	r	df	95% confidence interval
SMS	0.280	-0.097	124	(-0.269, 0.081)
TOTP	0.586	-0.049	122	(-0.225, 0.129)
Push	0.029	-0.204	113	(-0.374, -0.020)
U2F	<0.003	-0.269	118	(-0.429, -0.093)
Codes	0.426	-0.076	110	(-0.260, 0.113)

understand their background and feelings about online security. With the consent of each participant, we recorded the audio of each interview. Two coders listened to the recordings and coded each interview, discussing each response until reaching agreement. Common themes identified from the recordings are discussed in section 5.2.

4.8 Compensation

Participants were compensated a maximum of 25 USD after their participation in the study according to a tiered compensation structure based on the total number of tasks completed through the banking interface.

5 Two-week Study Results

5.1 Quantitative Results

5.1.1 Timing Data

We measured both the time for the password login and the time for the 2FA on the server side based on events sent from the client. Password timing began when the page initially loaded and ended when the user submitted a password. 2FA timing began when the 2FA prompt was loaded and ended when the 2FA was verified (or rejected). We recorded timestamps on the server since each client may have a slightly different clock. By comparing adjacent timestamp events, we were able to compute the overall login time. It is possible that users spent time obtaining their 2FA device before accessing the login page, which is not accounted for in the timing data.

Individual Learnability We computed the correlation between the amount of time an individual had been in the study and the amount of time it took them to authenticate. We used the repeated measures correlation (rmcorr) technique described by Bakdash and Marusich [6] to estimate the common regression slope for each 2FA method in our study. We hypothesized that participants would get faster over time as they became more familiar with the 2FA method. We found statistically significant ($p < 0.05$) support for this hypothesis for both push notifications and U2F security keys (see Table 1).

Table 2: Authentication Time (seconds), Summary Statistics

Authentication Method	Q1	Median	Mean	Q3
Codes	11.3	17.2	28.0	25.4
Push	8.4	11.8	16.1	17.6
SMS	13.0	16.6	18.5	22.1
TOTP	10.7	15.1	23.9	23.3
U2F	4.5	9.1	13.0	16.3

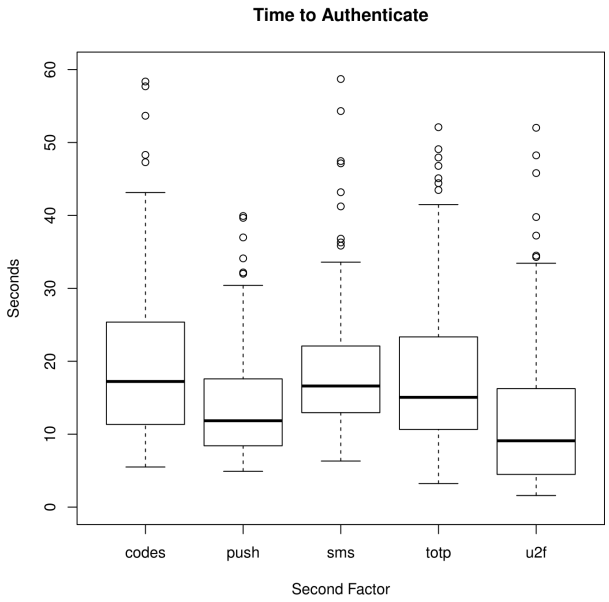

Figure 2: Time to authenticate for five 2FA methods

Comparison of 2FA Authentication Times We applied a Kruskal-Wallis one-way analysis of variance and found there was a significant difference ($p < 0.001$, $\alpha = 0.05$) in the median authentication time between the methods. We did not include the time that it took the user to enter their password; the observed authentication times reported here include only the time to get through the second-factor authentication step. The security key (U2F) devices had the fastest median authentication time, followed by push notifications. These timing results are summarized in Table 2 and Figure 2.

5.1.2 Usability Survey Rankings

We administered two SUS surveys to participants at the beginning of each exit interview session. The first survey addressed

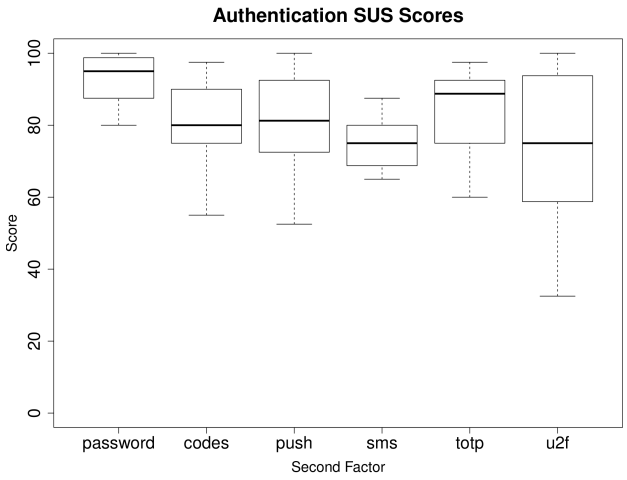

Figure 3: SUS scores for five 2FA methods.

the usability of the banking website as a whole, and the second addressed only the usability of the login system. The purpose of administering two surveys was to determine how large an impact the banking website itself had on the participants' feelings about the authentication method. Additionally, we felt that participants would be more accurate with their opinions about the 2FA method if we had first given them an opportunity to both consider and express their feelings about the system as a whole; had we only given a SUS survey on the authentication method we felt participants would be more likely to (incorrectly) report their feelings about unrelated website features.

The SUS scores for the authentication methods are summarized in Figure 3 and Table 3. We performed a Kruskal-Wallis one-way analysis of variance and determined that the authentication method used was a statistically significant ($p = 0.02579$, $\alpha = 0.05$) predictor of the median SUS score for the 2FA method. We also computed the value of $\rho = 0.7576$ for Spearman's rank correlation coefficient and confirmed that there was a significant ($p < 0.001$) correlation between the overall website SUS scores and the SUS scores of the individual authentication methods. Passwords with no second factor had the highest median SUS score, with a median score of 95, followed by TOTP (via Google Authenticator) which had a median SUS score of 88.75.

5.2 Qualitative Results

5.2.1 Security and Inconvenience

We asked participants whether logging in with a second verification step felt more secure. Most participants did feel more secure, although 3 of 12 participants that used the printed backup codes did not feel like the codes added any additional security to the method.

Table 3: SUS Scores for each Method, Summary Statistics

Authentication Method	Q1	Median	Mean	Q3
Password	87.5	95.0	92.5	98.8
Codes	75.0	80.0	80.2	90.0
Push	72.5	81.3	81.0	92.5
SMS	68.8	75.0	75.0	80.0
TOTP	75.0	88.8	83.1	92.5
U2F	61.9	75.0	73.1	93.1

> **P6:** *"I felt like the codes didn't accomplish anything, because that's just more passwords—anyone could guess them."*

We also asked participants if the additional security would be worth the additional login time or inconvenience they might face when using the second-factor method. Several people (20; 29%) said the extra security was definitely worth the tradeoff, and an additional group (25; 36%) said that they would be willing to use 2FA depending on the importance of the account.

> **P25:** *"In my opinion, it may be a little obsessive for everything, but for banking it's something that I actually do want some authentication. I almost wish that it was a requirement that the bank said, oh here set [two-factor authentication] up. Because now that I think about it, I don't know how to set up 2FA with my bank. If it were an option I would definitely use 2FA."*

> **P33:** *"It was pretty quick, so that was good; I didn't feel like I had to jump through a lot of hoops. I can imagine it being nice having an extra wall of security if it's your bank information so that even if somebody else gets your password, it's not like they're going to be able to hack into your account because they don't have the [security key]."*

Some participants were particularly concerned about the centrality and importance of their email account, particularly considering the potentially large amount of sensitive data stored there. For example, one participant reported they had already turned on 2FA for their Gmail account to gain extra protection:

> **P24:** *"I use my email for everything, and so I thought it wouldn't hurt to have some extra security. The thought of someone hacking into [my account] and having everything vulnerable... better to be safe than sorry."*

Other participants (9; 13%) expressly stated that they would not be willing to use 2FA to gain additional security because the inconvenience was too high.

> **P37:** *"I don't know how much my level of convenience and my need for level of security would balance out because for me having something that is convenient and is at hand is almost more important than having something that is more secure... I know if people hack your credit cards, then the bank will take care of that and get the money back and so having that extra security makes me care less about having a second factor."*

5.2.2 Availability of Second-factor Device

Each participant in the study in one of the 2FA groups was required to use something external to their computer to login, whether it be the sheet of paper with printed codes, a YubiKey, or their phone. Many participants (24; 35%) mentioned not having their second factor immediately available to them when they needed to log in.

> **P8:** *"I don't always have my phone on me, and so if I'm doing something on the computer, I'm usually doing homework, so I actually try to keep my phone away from me."*

> **P42:** *"Honestly, once I'm home I kind of just set my phone down and forget where I put it sometimes, so that was a little bit hard ... I needed to go find my phone and pull up the app."*

5.2.3 TOTP Timeout

Although the participants using TOTP (via the Google Authenticator app) were overall very positive about their experience, 8 of 12 participants mentioned that they had problems entering the six-digit verification code before it timed out.

> **P30:** *"I have to type in these numbers so fast or else it's going to go away."*

5.2.4 Likelihood of Account Compromise

Participants expressed a wide spectrum of views on how much value they placed on their online accounts. Some participants (9; 13%) felt that they had nothing to protect and would therefore not be a target of criminals.

> **P5:** *"I guess maybe because it's that I don't have anything to protect... I'm at a stage in my life where nothing I own is that valuable and none of my information is that wanted that it makes a difference."*

Table 4: Account Compromise and Inconvenience

2FA worth the inconvenience?	Hacked	Not Hacked
Definitely	11	9
Sometimes	6	19
Never	4	5

> **P8:** *"I mean, you hear a lot about stuff being broken into; I just don't think I have anything that people would want to take from me, so I think that's why I haven't been very worried about it."*

> **P30:** *"I don't have a lot of money in my accounts right now, so if someone stole my money, that would be bad, but it's not enough that it would be the end of the world if I lost all my money— I don't feel like I'm a target for someone to steal my stuff. I can imagine in the future if I had a huge retirement fund or something then I would want that to be more secure."*

5.2.5 Prior Compromise vs. 2FA Inconvenience

We asked each participant in this study whether any of their online accounts had ever been compromised. Several participants (26; 38%) described experiences with remote attackers taking over their online accounts, and a few people (7; 10%) mentioned that someone they know has had one of their online accounts hacked. Although not directly a form of online account compromise, a few participants also mentioned experiences with financial theft from having their credit or debit card number stolen or having their bank account credentials stolen. Others mentioned having their personal information stolen as part of one or more data breach events, including the highly publicized Equifax compromise of millions of individuals' personally identifying information [7]. When asked how they noticed that their account was compromised, most participants said they received an email indicating a new login from a suspicious location. We hypothesized that participants with previous experience having an account compromised would be more likely to feel that using a second factor was worth any extra inconvenience. Using data extracted from coding the interviews (see Table 4), we used Pearson's chi-squared test with two degrees of freedom to test the dependence of these variables. Not all participants expressly talked about both of these variables; thus we analyzed only participants for which we had coded data for both variables.

We observed no statistically significant relationship between a participant's previous history with account compromise and whether they felt that two-factor authentication was worth the inconvenience ($\chi = 4.6332$, $p = 0.0098$, $\alpha = 0.05$). One limitation of this analysis is that it does not consider the exact nature of the previous account compromise (such

as whether a financial loss was involved). However, we do note that numerous individuals independently stated that using 2FA would be worth the inconvenience at least some of the time, particularly for financial accounts.

5.3 Discussion

In this section, we further highlight some of the most interesting results of our study and discuss their meaning in the context of usable 2FA.

5.3.1 Relationship between Authentication Time and Usability

Although both push-based authentication and the U2F security keys had faster median authentication times, neither of these methods received the highest median SUS score. Conversely, TOTP was the highest scoring second-factor method we tested but had a median authentication time that was slower than either push or U2F. From our exit interviews, we identified some explanations for this result. First, some participants receiving push requests through Authy did not always receive the authentication request in their notification area and instead had to open the app and approve the request manually. It was unclear whether this was a bug in the Authy or the result of notification configuration on some participants' phones. Several U2F participants using both Windows and Mac operating systems reported a variety of minor troubles getting the YubiKey to work with their computers (possibly because they plugged it in the wrong direction). However, other participants reported no problems using the YubiKey. Ultimately, participants using TOTP reported liking the relative simplicity of the Google Authenticator app. The app functioned very similarly to SMS, a 2FA method with which many participants were already familiar while not requiring them to always have cellular service.

We believe that the minor issues encountered by participants using the Authy app and the YubiKey likely explains most of the lower scores they received. That said, no authentication method we tested received a poor usability rating, suggesting that, although there is a noticeable impact on usability from requiring 2FA, the presence of 2FA itself does not doom the method as a whole to poor usability.

5.3.2 Remember Me?

A novel aspect of our study is that participants used their second factor repeatedly for two weeks instead of using it just once in a laboratory setting. We purposely did not provide a "Remember Me" option, thus requiring participants in the non-control groups to use their second factor every day. We believe that some of the usability impacts of needing a second factor could be mitigated by only requiring the second factor on new computers or after logging out. Requiring less frequent 2FA login would provide a similar level of protection against remote attackers while mostly allowing users unfettered access to their accounts. Some systems allow access for a limited amount of time (30 days, for instance) without requiring a second factor on the same machine. Participants with previous experience using such systems (typically for a university login system) made some remarks to the effect that they were never quite sure when the second factor would be required. One solution to this problem would be to have a small count-down displayed to the user telling them how many days were left until they would need to provide their second factor again to avoid the "ambush" effect described by Sasse et al. [22]. Further research needs to be done to determine the right balance of when to ask the user for the second factor again when they have already been logged in previously on the same machine.

5.3.3 Positive Feedback

Given the weak usability results of previous 2FA studies, we expected an overall poor usability response. During the exit interviews, we were surprised at the number of participants that reported an overall positive experience using 2FA. Many participants wanted to use 2FA for some of their actual online accounts but were either unaware it was an option or were unsure how to configure it.

5.3.4 Differentiating Between High-value and Low-value Accounts

Although participants generally tended to care less about the security of their social media accounts, many expressed concern about the security of their banking and financial accounts. There were mixed feelings about frequently used accounts like email accounts, however, particularly in balancing whether it would be worth using 2FA for such accounts. Participants generally agreed that they did not want to be required to use their second factor to log in to their email account from a known computer. Other participants felt they had no confidential information in their email, and that having a second factor would not be worth the extra login step. In general, the higher the perceived value of the account, the more likely the participant was to be willing to use 2FA for the account.

5.4 Limitations

Our study has several limitations. First, the participants were not asked about their prior use of 2FA. A user assigned to a second-factor they were already familiar with could bias the results. Second, the participants were university students that were younger and more technically savvy than the general population. The students are also more likely to have fewer material assets to be concerned with, as discussed in the qualitative results. Third, we deliberately chose not to have the participants setup the 2FA mechanism on their own so that a

poor setup experience would not negatively bias day-to-day usability. This decision means the day-to-day usability results could be biased more positively compared to users that will have to setup and use 2FA. Fourth, because we wanted to capture authentication timing data, we were unable to have participants use a real banking system or an existing online account; this may have altered their behavior. Fifth, participants were required to use 2FA for every authentication attempt, which may have caused them to acclimate to using 2FA more quickly than would be seen if 2FA was required only on new machines. Sixth, participants' discussions of the necessity of 2FA and online security may have been different had we mocked our website as a social media site. Finally, with only 12 participants in each study group, we may not have reached saturation in the qualitative data that was gathered. Even if we had reached saturation, the limited demographics of the study still warrant further studies with a broader population.

6 Setup Study Methodology

We purposely ignored the setup phase during our two-week study to avoid having a poor setup experience negatively bias participant's evaluation of the day-to-day usability of one of the factors. However, the promising results from the study beg the question about whether the results are incomplete and miss an important associated usability hurdle to set up that factor. To gain insight into this question, we conducted an IRB-approved, within-subjects laboratory study comparing the usability of the setup phase for the five 2FA methods. Based on our initial review of the setup process on some popular websites, we did not expect that there would be significant usability issues for setting up the five 2FA methods.

6.1 Study Design

Each participant was tasked with setting up the five 2FA methods from a desktop computer using a provided Google account. We chose to test the methods on Google because it supports all five 2FA methods and is an industry leader in supporting 2FA for its customers and employees. The setup for Google security keys has been studied previously, and improvements have been made based on those results [12, 21]. Our goal was to observe the general usability of the setup process and not focus on provider-specific details since we did not compare the setup between multiple providers.

Participants were provided with an Android phone and a YubiKey NEO for methods requiring a physical device. Testing for every possible ordering of setting up the five methods requires 120 treatments. To reduce the time and cost of our study, we created an incomplete counterbalanced measure designed to mitigate biases due to the order participants set up each of the 2FA methods. We used two five-by-five balanced Latin squares to generate 10 different orderings of the setup methods to counterbalance sequential effects caused by ordering [19]. Each of the 10 orderings was completed 3 times during the study. After each attempt to set up the second factor, participants were asked to complete the Single Ease Question (SEQ) to measure the difficulty of each task. The SEQ is a standard usability questionnaire with a single question ("Overall, how difficult or easy was the task to complete?") rated on a 7-point scale. Although it contains only one question, the SEQ has been found to perform reliably [23]. We chose SEQ to avoid survey fatigue since participants were asked to rate five different methods. We used timing data and SEQ responses to compare the setup usability for the five methods.

We posted flyers on the BYU campus to recruit 30 participants who were familiar with Google accounts and Android phones. As each participant met with a study coordinator, they first signed a consent form. Participants were compensated 10 USD at the conclusion of the study. We assigned participants to the ten Latin square orderings in a round-robin fashion.

6.2 Setup Process

The coordinator provided the participant with an Android phone, a YubiKey, and an information sheet that listed the cellphone number and lockscreen passcode. We made an audio recording of the verbal comments of each participant along with a video recording of the computer screen.

Google does not allow backup codes, push notifications, or TOTP to be set up without first setting up SMS or U2F. In order to test each method independently, we used one Google account for setting up SMS, and a separate account for the other four options. Study coordinators navigated to Google's 2FA setup page on a Chrome browser and then instructed participants in what order to set up the five second factors. The coordinators also navigated between the two Google accounts before and after the participant setup SMS. After the participant completed the task or was unable to finish the setup, the coordinator prompted the participant to complete the SEQ. Coordinators did not assist participants in setting up any of the second factors.

The following is a brief summary of each setup task.

SMS. Participants were asked to type in a phone number. Google sends a confirmation text containing a six-digit code to the provided number. The participant completes setup by entering the code on Google's webpage.

TOTP. Participants were provided with an Android phone without Google's Authenticator app installed. We wanted participants to download the app as a part of the setup because we assumed that the typical Google user would not have the app already installed on their phone. The phone was set up with the Play Store app on the home page for easy accessibility. Google instructed participants to install the app using the Play Store, and then to scan a Quick Response (QR) code shown on the webpage. After scanning the QR code, participants completed the setup by entering a six-digit code from the

Authenticator app into the webpage.

Pre-generated codes. Google autogenerates 10 backup codes upon request. Participants were not required to print or download these codes but were asked by the coordinator how they would store these codes if they were using their own Google account. Some participants shared that they would choose to take a photo of the codes using the camera on their phone, while others said they would write down the codes and keep them in a safe place. For timing data, we measured from the time the participant began the task to the time the backup codes were displayed on the screen. Even though we asked participants how they would store the backup codes, we did not include the time taken to store codes in the setup time for backup codes since the time to store the codes varies widely depending on the storage method chosen.

Push. Push notifications require that the phone is signed in to the user's Google account. The phone provided to participants was already signed in, based on the assumption that the typical Google user would already be signed in to their Google account on their phone. When a phone is online, has screen locking enabled, and is connected to the Google account, Google sends a push notification that can be approved by unlocking the phone and tapping "Yes" on the notification.

U2F Security Key. We provided participants with a YubiKey NEO. Google directed participants to insert the security key into an open USB port, and then to tap the gold button on the key. Before the device could be recognized, participants were required to dismiss an alert from the browser asking for permission to see the U2F device's make and model. Whether or not a user allows or denies this request, the U2F device is registered and optionally given a name. Since this is optional, we excluded the time taken to name the device.

7 Setup Study Results

7.1 Timing Data

We reviewed the video screen recordings to measure the setup time for each method. Time was measured in seconds from when the participant began the setup task to the time Google notified the participant that the setup had been successful. The cases where the participant failed to complete the setup are not included in the timing analysis. Setup failure occurred twice with the U2F device and twice with the TOTP application. A summary of our results is shown in Table 5 and Figure 4.

As expected, backup codes had the fastest setup time because all that was involved was clicking the webpage button to generate the codes. However, backup codes had the longest mean authentication time in the day-to-day study, followed by push notifications and SMS messaging. While U2F devices had the fastest mean authentication time in our day-to-day study, they had the second slowest mean setup time. TOTP had the slowest mean setup time.

Table 5: Setup Time (in seconds), Summary Statistics

Authentication Method	Q1	Median	Mean	Q3
Codes	1.0	1.0	2.2	2.0
Push	16.0	23.5	27.3	33.0
SMS	27.5	32.0	34.5	40.0
TOTP	73.3	84.0	109.6	120.0
U2F	31.8	44.0	57.8	67.8

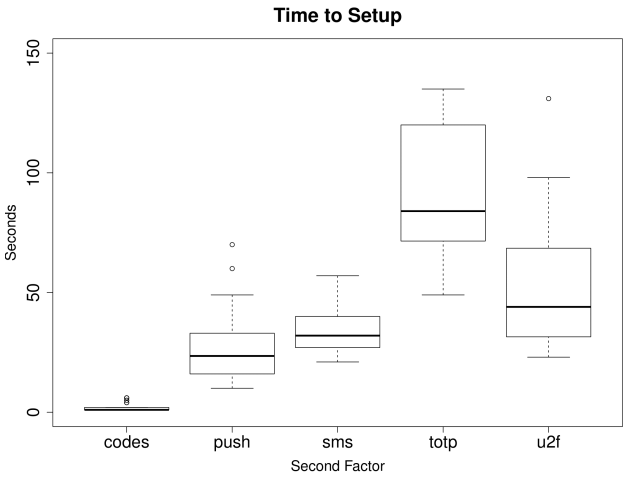

Figure 4: Setup time for five 2FA methods.

7.2 SEQ Scores

Participants answered the SEQ after completing (or being unable to finish) each 2FA method. Mean SEQ scores are shown in Table 6 and the distribution of all SEQ scores is shown in Figure 5. With the exception of backup codes, the ranking of best SEQ score to worst corresponds with the ranking of time to set up, i.e. the faster the setup, the higher the mean SEQ score. We were surprised that backup codes received a lower ranking since setup involved nothing more than pushing a button. We hypothesize that a participants' perceptions about the day-to-day usability of the 2FA method influenced their SEQ score even though they were instructed to rate only the usability of the setup task.

TOTP setup received the lowest mean SEQ score of the five methods. The low score is in stark contrast with the two-week study, where TOTP received the highest mean SUS score of all five 2FA methods. Users may have been more unfamiliar with setting up TOTP then they were with the setup of more common methods, such as SMS. However, once users have TOTP authentication successfully enabled, they may find it to be more usable than other 2FA methods they may have traditionally relied on.

Table 6: Mean SEQ Scores

Push	SMS	Codes	U2F	TOTP
6.7	6.2	5.9	4.7	4.5

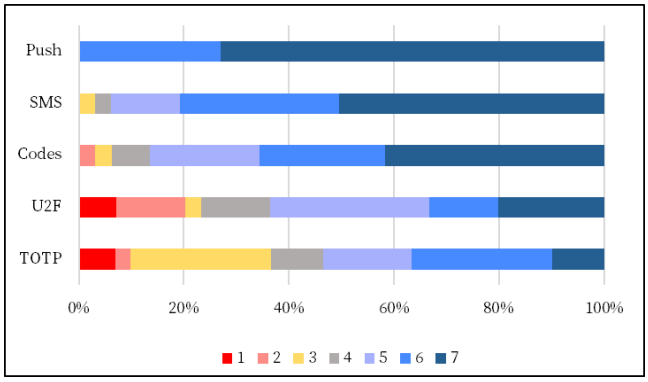

Figure 5: SEQ scores for five 2FA methods.

7.3 Discussion

Our study suggests that when 2FA setup can be implemented well, users generally find it easy to accomplish. Each of the five second factors had a mean score closer to the "easy" side than the "difficult" side. This is notable considering that study coordinators provided no assistance during setup, and many participants were required to set up second factors that were unfamiliar to them (such as the U2F device or the TOTP generator). SMS authentication is one of the most common forms of 2FA, and familiarity with using SMS as a second factor likely influenced its SEQ score.

Setup failure occurred twice with TOTP and twice with U2F. Both failures for TOTP happened when the participant immediately attempted to scan the QR code with the phone's camera, instead of downloading the Authenticator app to scan the code. An additional two participants initially tried to scan the QR code with the phone's camera but realized their mistake and successfully completed setup after downloading the app. The failures for U2F both occurred when the participant did not notice the browser alert requesting permission to see the U2F device's make and model. Google does not require the make or model to authenticate the device, so the U2F device would be registered whether or not the user allowed or denied the browser's request. However, participants who did not notice the alert at all were not able to complete the setup.

Based on our observations, we present two recommendations for reducing setup failures on Google accounts. First, users may be less likely to skip over installing the Authenticator app if the installation instructions were on a prompt separate from the QR code. Second, because the U2F browser alert occurs on many of the browsers that support U2F (including Chrome, Opera, and Firefox), 2FA-providers should notify users about the alert during the setup process. Yubico does this on their support page: "Touch the YubiKey when prompted, and if asked, allow it to see the make and model of the device" [1].

7.4 Limitations

Participants from our study were recruited at a university, and our results may not be generalizable to the general population. We tested setup on a desktop computer, and the setup experience may be different using a phone as the primary computing platform. Our timing data for backup codes did not include the time taken to store codes. Timing data and SEQ scores may have been negatively impacted by our participants' unfamiliarity with the provided phone. If participants had used a personal phone, they likely would have been able to perform tasks requiring a phone more quickly (e.g., entering the phone number, or unlocking the phone). Although our study did not focus on provider-specific details, Google's implementation of 2FA setup inevitably influenced user's perceptions.

8 Conclusion

We conducted a user study to evaluate the day-to-day usability of multiple 2FA methods by having participants log in to a simulated banking website nearly every day for two weeks and completing an assigned banking task. Having all the participants experience a 2FA method within the context of a single application reduces the confounding factors that are usually present when comparing the results of different 2FA methods across usability studies.

Participants generally gave high marks to the methods studied, and many expressed an interest in using 2FA for their sensitive online accounts. However, about one-third of the participants reported an instance of not having their second-factor device immediately available when they needed it.

There are several lessons learned from our two-week study. Participants using push notifications and U2F security keys decreased their login time as they gained experience with the method. Two-thirds of the participants using TOTP (via the Google Authenticator app) had problems entering the six-digit code before it timed out. Approximately 25% of the participants using printed backup codes did not feel like the codes added any additional security to the system—it seemed like just another password that an attacker could compromise.

We also compared the usability of the setup phase for each of the five 2FA methods. While a few participants experienced difficulty setting up U2F and TOTP as second factors, in general, users found the methods easy to set up. Together, our two studies show that well-implemented 2FA methods may be set up and used daily without major difficulty.

Acknowledgments

The authors thank the reviewers for their helpful feedback. This material is based in part on work supported by the National Science Foundation under Grant No. CNS-1816929.

References

[1] How to confirm your Yubico device is genuine with U2F. *Using Your YubiKey with Authenticator Codes : Yubico Support*.

[2] Data Breach Investigations Report, 2017.

[3] Mobile Fact Sheet, Jan 2017.

[4] Claudia Acemyan, Philip Kortum, Jeffrey Xiong, and Dan Wallach. 2fa might be secure, but it's not usable: A summative usability assessment of google's two-factor authentication (2fa) methods, 2018.

[5] Nathanael Andrews. "can i get your digits?": Illegal acquisition of wireless phone numbers for sim-swap attacks and wireless provider liability. *Northwestern Journal of Technology and Intellectual Property*, 16(2):78–106, Nov 2018.

[6] Jonathan Z Bakdash and Laura R Marusich. Repeated Measures Correlation. *Frontiers in Psychology*, 8:456, 2017.

[7] Tara Siegel Bernard, Tiffany Hsu, Nicole Perlroth, and Ron Lieber. Equifax Says Cyberattack May Have Affected 143 Million in the U.S., September 2017.

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

[9] Joseph Bonneau and Sören Preibusch. The Password Thicket: Technical and Market Failures in Human Authentication on the Web. In *The Ninth Workshop on the Economics of Information Security (WEIS)*, 2010.

[10] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. "It's not actually that horrible": Exploring Adoption of Two-Factor Authentication at a University. In *2018 CHI Conference on Human Factors in Computing Systems*, page 456. ACM, 2018.

[11] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The Tangled Web of Password Reuse. In *Network and Distributed System Security (NDSS)*, volume 14, pages 23–26, 2014.

[12] Sanchari Das, Andrew Dingman, and L Jean Camp. Why Johnny Doesn't Use Two Factor: A Two-Phase Usability Study of the FIDO U2F Security Key. In *2018 International Conference on Financial Cryptography and Data Security (FC)*, 2018.

[13] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *Workshop on Usable Security (USEC)*, 2014.

[14] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User Perceptions of Security and Usability of Single-factor and Two-factor Authentication in Automated Telephone Banking. *Computers & Security*, 30(4):208–220, 2011.

[15] Blake Ives, Kenneth R Walsh, and Helmut Schneider. The Domino Effect of Password Reuse. *Communications of the ACM*, 47(4):75–78, 2004.

[16] Mike Just and David Aspinall. On the Security and Usability of Dual Credential Authentication in UK Online Banking. In *International Conference for Internet Technology And Secured Transactions (ICITST)*, pages 259–264. IEEE, 2012.

[17] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M Angela Sasse. 'They brought in the horrible key ring thing!' Analysing the Usability of Two-Factor Authentication in UK Online Banking. *Workshop on Usable Security (USEC)*, 2015.

[18] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security Keys: Practical Cryptographic Second Factors for the Modern Web. In *International Conference on Financial Cryptography and Data Security (FC)*, pages 422–440. Springer, 2016.

[19] James R. Lewis. Pairs of latin squares to counterbalance sequential effects and pairing of conditions and stimuli. *Proceedings of the Human Factors Society Annual Meeting*, 33(18):1223–1227, 1989.

[20] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert. Sms-based one-time passwords: Attacks and defense. In Konrad Rieck, Patrick Stewin, and Jean-Pierre Seifert, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 150–159, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[21] Reynolds, Joshua and Smith, Trevor and Reese, Ken and Dickinson, Luke and Ruoti, Scott and Seamons, Kent. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018.

[22] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'Weakest Link'—A Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3):122–131, 2001.

[23] Jeff Sauro and Joseph S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1599–1608, New York, NY, USA, 2009. ACM.

[24] Jake Weidman and Jens Grossklags. I like it, but i hate it: Employee perceptions towards an institutional transition to byod second-factor authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference*, ACSAC 2017, pages 212–224, New York, NY, USA, 2017. ACM.

[25] Catherine S Weir, Gary Douglas, Martin Carruthers, and Mervyn Jack. User Perceptions of Security, Convenience and Usability for Ebanking Authentication Tokens. *Computers & Security*, 28(1):47–62, 2009.

[26] Catherine S Weir, Gary Douglas, Tim Richardson, and Mervyn Jack. Usable security: User Preferences for Authentication Methods in eBanking and the Effects of Experience. *Interacting with Computers*, 22(3):153–164, 2010.