Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key

Sanchari Das(sancdas@indiana.edu)
Andrew Dingman(adingman@indiana.edu)
L Jean Camp (ljcamp@indiana.edu)

Indiana University Bloomington

Abstract. Why do individuals choose to use (or not use) Two Factor Authentication (2FA)? We sought to answer this by implementing a two-phase study of the Yubico Security Key. We analyzed acceptability and usability of the Yubico Security Key, a 2FA hardware token implementing Fast Identity Online (FIDO). This token has notable usability attributes: tactile interaction, convenient form factor, physical resilience, and ease of use. Despite the Yubico Security Key being among best in class for usability among hardware tokens, participants in a think-aloud protocol still encountered several difficulties in usage. Based on these findings, we proposed certain design changes, some of which were adopted by Yubico. We repeated the experiment, showing that these recommendations enhanced ease of use but not necessarily acceptability. With the primary halt points mitigated, we could identify the remaining principle reasons for rejecting 2FA, like fear of losing the device and perceptions that there is no individual risk of account takeover. Our results illustrate both the importance and limits of usability on acceptability, adoption, and adherence in Two-Factor Authentication.

Keywords: Two-factor Authentication, Hardware Authentication Device, Usable Security, Adaptability.

1 Introduction

The Yubico Security Key is an implementation of Fast Identity Online (FIDO) [22] Universal Second Factor (U2F) in a USB token form. The Security Key is designed to appeal to high-touch, low-tech users who want more secure interactions and improved ease of use from their online service providers [13]. According to Brett McDowell, Executive Director of the FIDO Alliance, "We fail if FIDO is not more usable than all the other (hardware token) options you have used before" [15].

We explore the acceptability and usability of the FIDO U2F technology, in the form of the Yubico Security Key, against these goals and metrics using a think-aloud protocol. We recruited students from STEM degree programs and tested different setup instruction sets. Our goal was to identify difficulties that

might be barriers to adoption. Usability measures if individuals can complete a set of tasks with a given technology. Acceptability addresses the experience of use, including perceived risks and benefits, and impinges on user adoption.

We asked experiment participants to configure a FIDO U2F Security Key for their Gmail account. From the analysis of the participants' experiences, we developed a series of recommendations for Yubico. Yubico adopted and implemented a subset of the recommendations. A year later, after the changes were made, we repeated the experiment to evaluate the new interaction with the security keys. There was a significant increase in usability, but we could not assert any corresponding increase in acceptability.

We detail the related literary work in section 2, our experiment methods in section 3, and results in section 5. We further discuss on future recommendations in section 7 both for Yubico Security Keys in specific and Two-Factor Authentication at large. We conclude by providing an overview of the study and giving a future direction towards a better usability, acceptability, and adaptability of the FIDO security keys in section 8.

2 Related Work

Our work was primarily informed by research on usable authentication as well as influenced by research on online risk perception and risk communication. Bonneau et al. provided a framework for examining the usability of authentication technologies [4]. Before authoring this framework with Bonneau, Stajano provided a set of recommendations for any authentication system through research grounded in the Pico hardware token authentication project [23]. This earlier work demonstrated five core requirements: security, memory-less operation, scalability, loss resistance, and theft resistance.

Previous work has shown that FIDO Security Keys are easy to deploy. Lang et al. refer to the use of a Security Key as "brainless", which seems to indicate a belief that there are no halt points in Security Key adoption [14]. However, this study included neither qualitative components nor human subject experiments. In our work, we have implemented a think-aloud protocol and found numerous halt points and challenges to acceptability. A previous human-centered evaluation of 2FA also found that users perceived twice the utility from *avoiding* 2FA compared to adopting it [6]. Our results echoed this finding, with most subjects simply leaving or returning the keys. Usability of 2FA methods has been studied by Krol et al. [12], however they studied online banking customers and people often relate financial accounts as more confidential than their email accounts. The study by Krol et al. also focused on 2FA in general and approached the usability of 2FA from the steps a user has to follow in contrast to how we studied the Yubico Security Key, it's usability and acceptability.

Passwords have been heavily critiqued in academic research. Archaic recommendations such as formulaic complexity requirements of passwords and periodic password changes may be helpful, but still cannot ensure protection against password vulnerabilities. Instead, best practice guidelines are proposed such as validating newly created passwords against commonly-used or known compro-

mised passwords [18]. In *Understanding Password Choices*, Wash et al. showed that users tend to re-use passwords across sites, especially where they must enter passwords frequently [25]. Komanduri et al. found that users frequently have critical misunderstandings about what make passwords secure - with a tendency to overestimate the effects of additional complexity, while underestimating the impact of using common phrases [11].

Unusable password policies often result in insecure workarounds, but Inglesant and Sasse assess that the cost goes beyond insecurity and often negatively impacts the productivity of both individuals and their organizations. Their work indicated that human-centered design principles should influence policy creation which guides users to create suitably secure passwords in accordance with the usage context [9]. Through the password system, Abbott et al. showed that users can indeed be guided towards better password decisions without corresponding increase in cost [5]. Biddle et al. shows that though it was more acceptable to the users, it gradually resulted in decreased predictability of user password behavior [3]. The major alternative to FIDO in 2FA is time-based (TOTP) or hash-based (HOTP) one-time codes. However, neither TOTP or HOTP offers mutual authentication of both the user and the service [16, 17]. Our study was informed by a two-phase examination of Tor by Norcie et al. [19]. Norcie et al.'s study followed the same process of a think-aloud protocol implemented in our study. Their study also analyzed the design modifications made to Tor and was in turn strongly influenced by the canonical *Why Johnny Can't Encrypt*, which examined the use of Pretty Good Privacy(PGP) [27] by asking participants to complete the tasks required for adoption and use. By observing the difficulties encountered by participants, Norcie and Whitten offered design heuristics for anonymized systems and public key systems for emails respectively. Some of the recommendations, such as focusing on the importance of the setup steps prior to operation or conveying to the user why a feature exists may be generalized to authentication systems. Our target was to provide specific solutions to enhance the adaptability of the security keys.

3 Methodology

We investigated the end user experience of configuring and using the FIDO Security Key by combining a think-aloud protocol and two surveys before and after the experiment. In a think-aloud protocol the subjects narrate their actions, providing a real-time description of their decisions, choices, or motivations. The preliminary survey was online, followed by the think-aloud protocol being implemented in a university computer laboratory. There were open questions asked after the think-aloud protocol and a final survey sent via email. We then implemented the first phase of the study and made recommendations to Yubico and Google, some of which were adopted. This experiment was repeated after a year.

Participants were recruited from an undergraduate non-technical introductory security course. We recruited the participants from the same course in the consecutive year. Participants completed a preliminary background, knowledge, and skills survey to evaluate any differences in the security and computing ex-

pertise of the subject pool. These technical knowledge and skill inventories were implemented and calculated as done by Rajivan et al. [20]. The participants were required to be (i) at least 18 years old, (ii) have a personal Gmail account, (iii) own a laptop with the Chrome browser, and (iv) own a personal mobile phone.

The participants were consciously selected to have more security and computer expertise than the general population. Similar to the experiments on setting up access control rules [2], firewalls [21] and PGP [27], we chose a population likely to be successful.

A coin flip was used to randomly divide the participants into two groups. In one group, participants were directed to the official Yubico Security Key instructions. The other group was directed to the Security Key instructions provided by Google. The think-aloud protocol began by giving each participant a Yubico Security Key, as shown in Fig. 1. The participant was then asked to configure 2FA using the Yubico Security Key with their Gmail account while narrating the experience. Each participant was paired with one researcher who took notes, but did not offer additional guidance unless requested when the participant was unable to proceed without some guidance. After the task was complete, participants were asked to describe the benefits and importance of the Yubico Security Key using the seven open-ended questions below.

1. How could you test to confirm that your key is working?
2. If your key was lost or stolen, what would you do?
3. Based on your current understanding of the technology, could you use the same key with an account on another web site, or would you need to obtain an additional key?
4. Based on your current understanding, could you add a second key to your account?
5. Do you see any benefits from using the security key? Please explain.
6. Do you expect to continue to use your key after today? Why or why not?
7. How would you remove a key from your account if you decided to?

There were two goals for this closing interview. One was to explore the participant's reflection on the experience of configuring 2FA. The second was to ensure that we would not harm the participants by locking them out of their accounts. Each participant departed only after the researchers were certain that the research subjects were capable of removing 2FA without any assistance.

Fig. 1. Yubico Security Key

One month after the end of the think-aloud protocol, the subjects received a follow-up survey inquiring about their continued use of 2FA. The follow-up survey was sent over email.

3.1 Coding and Analysis

Recall that there was a preliminary survey, a think-aloud protocol, an interview after the protocol, and a survey on continued use well

after the experiment. In this section, we describe the coding and analysis of the qualitative data. The procedure was identical in both phases. We solved the few discrepancies between the codes which were discussed between the coders and the researchers.

There were two sources of qualitative data. The first source was the transcripts of the think-aloud protocol itself. These transcripts began after the Yubico Security Key was handed to the participant and ended after the enrollment into 2FA was complete. The second source consisted of the transcription of the responses to the open-ended questions from the interview immediately following the experiment.

For the enrollment task, the halt points were noted for each subject. The conversations around the halt points as well as the responses to the open-ended questions were transcribed. Three researchers trained independently in qualitative methods read through the transcripts. As standard in qualitative research, the themes were compiled into a code book. Of the recorded halt points, the cause was identified to be centered around 4 major mutually exclusive points: form factor, a setup demo, setup validation, and security valuation of the device. *Halt* points occurred when participants could not move forward alone. *Confusion* points occurred when the participants significantly slowed down due to confusion, or stopped but would have been able to continue with the registration procedure without assistance. We also noted expressions of value, where participants expressed ideas or opinions about the perceived utility of the technology, device, or installation process.

In both experiments, many participants recognized the potential value of the Security Key in theory, but not in practice for themselves. The details of the two phases are described in the following two sections followed by a discussion addressing both.

4 Experiment

4.1 Phase-I

In Phase-I, we discovered that the most significant halt point was the confusion resulting from a Yubico demonstration tool. Yubico had built a tool clearly illustrating how to register the 2FA token with a service. Participants went through the demo and believed that they had completed the installation process. No participant in the experimental group that was directed to the Yubico demo was able to realize that they needed to continue and complete the installation without researcher intervention.

Phase-I concluded with a set of recommendations about the instructions, visualization, device identification, and guidance provided to users. The details of the recommendations are described in Section 5.3. We repeated the experiment to test the efficacy of the adopted recommendations after Yubico implemented a subset of them. We also revisited the recommendations (in section 7) from Phase-I that were not implemented to determine if those changes were still needed.

As reported in Section 3, our participants for both the phases were recruited from the same class to ensure that the sample was moderately security savvy. We had 27 young scholar participants in total - 6 were between 18 and 20, 16 were between 21 and 23, 4 were between 24 and 26, and 1 was over 30 years old. There were 20 male and 7 female students, a 74% to 26% split. Every student was enrolled in at least one information or computer science class, by definition. Unfortunately, due to the nature of the recorded data we lost data of 6 participants and the results in Table 1 and Table 2 indicate reflect that from 21 participants.

Depending on the computer and security expertise questions answered by the participants before the in-lab experiment, the mean security expertise was 2.96 of 5 and the mean computing expertise was 4.34. We compared this with a general population survey of 593 participants where the results showed a mean security expertise (using the same calculation) of 1.7 and a mean computing expertise of 1.77 [10]. As a result, it is reasonable to assume that any halt points encountered by this population would also occur in a less technical and less educated population.

4.2 Phase-II

In Phase-II, as with Phase-I, the participants were students recruited from the same computer security course after a year. We had 34 young scholar participants in total - 1 was between 18 and 20, 29 were between 21 and 23, 2 were between 24 and 26, 1 was over 30 years old, and one chose not to answer. There were 26 male students, and 8 female students, a 76.4% to 23.6% split. Every student was enrolled in at least one information or computer science class, by definition. In Phase-II, the mean security expertise score was 2.95 of 5 and the mean computing score was 4.34. Again, it is reasonable to assert that our participants have more security and computing expertise than the general population. The differences in the mean values were not significant.

5 Results

5.1 Phase I Findings and Usability

In this section, we discuss about the various halt and confusion points where the participants found it difficult to register the Yubico Security Key.

Inserting the Device We were able to identify several points of confusion related to device form factor. Primarily, users experienced confusion about the correct orientation of the key due to the slim design which allows it to enter the USB port both correctly or upside-down.

Finding instructions Once the device was successfully inserted and individuals were directed to setup the device for their account, they had trouble getting started. Over half of the users navigated to their browser settings or their email settings first. The second time they encountered an instruction-centered challenge was when they had actually found the correct 'account settings' control panel. For successful setup, users were required to follow a non-linear path through the control panel, and at each page, a large array of options were of-

fered. This presented many opportunities for confusion and abandonment of setup altogether for several participants.

Illumination To activate the Security Key, either for enrollment or authentication, users must touch a capacitive button on the device. The button light would blink on insertion and at other seemingly unrelated points. Participants frequently displayed confusion over the timing of button press and the meaning of the blinking light.

Correctly identifying the device Participants in the first condition found the Yubico landing page to be difficult to understand and navigate. Despite having the original packaging for the device, participants generally were not confident about which model of Yubico Security Key they were using. This was a halt point where device identification was required to receive setup instructions. The most commonly mentioned reason for choosing a particular device was color. No subjects mentioned using the images on the button to differentiate between Security Key models.

"Try out this key" link Once subjects had determined which model of key they were using, the next challenge was finding the correct setup instructions. Without exception, participants identified a link to a demo application as the most salient option for their goal of setting up their key with a Google account.

Demo versus reality Many participants either believed they had completed the task after successfully authenticating to the demo, or repeated the enrollment and test cycle of the demo tool several times without progressing. After ten minutes of repeating the demonstration cycle, we considered subjects to have reached a halt point. As one participant noted, *"The web site is kinda confusing because I do not know what it wants me to do."*

Biometric versus touch Many participants thought the circular touch sensor was a biometric authenticator that read their fingerprints. This has both positive and negative implications. On the positive side, this indicated awareness that interaction was necessary. It also implied, however, a higher benefit than the device actually provides, since, in reality, anyone can use it. If the token is lost, users who believe they have biometric enrollment would not realize that others could use it. One of the participant's mention: *"I guess it is more secure because they make you scan your fingerprint before you can log into your account, but to me it's a bit excessive."*

Confirmation of operation Participants were unable to confirm that the device was working after setup. When users were queried, "How could you test to confirm that your key is working?", a common response was the intuitive "Log out and back in". Unfortunately, since the default during setup is to trust the current computer, users never got to actually experience using their Yubico Security Key outside the set-up process. For single-computer users, this experience could be left until weeks in the future - *"Why didn't you prompt me?! It said it would...maybe I'll just try again."*

5.2 Phase-I Acceptability

The primary drivers of acceptability were the lack of awareness of the risk and the resulting perception that there was no benefit. Here we recommend changes to

increase this acceptability. Several of the participants in our experiments dropped the keys in a shared bin for leftover hardware, often used for mice or cables. This reiterates the importance of theft and loss resistance noted in related work [4, 23]. Participants in the experiment did not have a clear understanding of the possible risk of account subversion. Similar lack of awareness and uncertainty of the risk of their choices has been found in privacy as well as security previously [1].

5.3 Phase-I Recommendations

In response to our results we made specific recommendations in a technical talk presented to Yubico and Google. Some of these recommendations were adopted, either as a result of our work or serendipity. Here we list our recommendations and, in the case of adoption, note the difference.

Finding Instructions The other issue was that people had difficulty finding instructions. The current Yubico web page has vastly improved this, providing icons that link not just to the service but directly to the instructions for Security Key enrollment. We found that the service provider descriptions were easier to follow than the Yubico descriptions for each service provider. Our recommendation for Yubico to provide pointers rather than instructions for each service provider was acknowledged.

Demo versus Reality First time users were not easily able to identify which product they had, or which instructions they were to follow. The "Try out your YubiKey" demo was a source of much confusion. In every experiment condition where a user was directed to the Yubico instructions, they got stuck in a loop with the instructions and required guidance for the next step.

The demo does appear to serve the important goal of providing hypothetical demonstrations to prospective institutional customers. However, when this demo is included as part of the display to those who have already purchased the product, it consistently caused confusion. We recommended that this demo should not be accessible to the end user, as it was a consistent halt point. Though still accessible, the demo has been removed from the installation workflow. As a result this halt point went from confounding nearly every single subject in Phase-I to having negligible impact in Phase-II.

Correctly Identifying the Device In our initial experiment, the instructions asked what Yubico product a user had, but provided little identification guidance. A user's best option was a product comparison table, the top of the website. The table appeared to have been designed to assist in purchase decisions rather than configuration, with prominent price information and technical data.

A new interface offers more prominent pictures and descriptors which allows for easier identification of the device to be used. The title clearly shows the purpose, providing confidence to the subject participants that they had found the correct source for device identification.

Biometric versus Touch A significant change implemented in the Yubico setup instructions is the clear identification that the button is not a fingerprint reader.

Confirmation of Operation During the experiment, participants found it challenging to confirm that a newly registered Security Key was in fact operating

correctly. This confusion was caused by Google's default behavior of marking the browser as a "trusted" device. In this case, users are not required to use a second authentication factor when logging in, even when 2FA is enabled for the account.

The default browser trust defeated subjects' natural inclination to test the newly enrolled device by logging out of their account and logging back in, as there was no prompt to use the key. A subset of the experimental group did arrive at a solution, either using "incognito mode" or clearing cookies from their browser before logging in again. However, not all participants possessed the technical understanding of Google's authentication process necessary to arrive at such a solution.

5.4 Phase-I Acceptability Recommendations

Despite the fact that the chosen research pool was more educated and security aware than the general population, no participant in Phase-I decided to continue to use the token provided. We know that some Yubico security tokens were returned to researchers immediately or placed in the discarded available hardware bin after the experiment. We also piloted the study before deploying it to the participants in phase-I. In contrast, the participants in the pilot phase were all graduate students in security and all of the pilot participants continued to use the tokens.

Communicate the Intrinsic Benefit Rational decision making theories fail to account for observed security and privacy choices, either individual or in the aggregate. Yet people consistently use a set of heuristics in making decisions, such that benefits obtained are greater than the risk. Garg argued that security systems should be designed to take advantage of these theories to encourage more adoption [7]. Applying the observations here, we recommend better risk communication that indicates that the use of the token is a benefit.

Developing appropriate feedback for users has long been recognized as a design challenge [27, 28]. Thus, we recommend the addition of such feedback for users to be aware of the benefits of using the device. This may include confirmation of successful registration on first login, or occasional information about the superior security while content loads.

Communicating the Risks Users did not understand the benefit of the device as compared to a longer, more secure password. Users who chose to return the token expressed confidence in their own security management and length of passwords. Many of the users thought the device was useful in the case of computer theft, but were dismayed to find that the device would remain trusted even when lost.

Create a Cognitive Benefit A major impediment to users' perception of value was the continued need for passwords to authenticate when using trusted devices. During experimental sessions, several people expressed a desire for the authentication device to somehow streamline authentication compared to typical password entries. Other users were surprised that passwords were still needed after setup. Many participants felt that the second factor was an overkill, or too much of a burden in exchange for the no cognitive benefit.

West and Garg had two recommendations that address this major challenge to acceptability: reducing costs associated with security and improving rewards for good decisions [7, 26]. Specifically, we recommend a visible reduction of the cognitive load of passwords in return for use of FIDO to improve acceptability. Streamlining could be achieved by not prompting the user for the full password as long as the proper FIDO token was plugged into a trusted device, or by allowing users to have a shorter, easier to use password when the device is present.

Highlight the Features The FIDO standard is designed so that a single key can be used with multiple accounts without revealing any link between the two accounts even if service providers collude. This feature is crucial to the scalability of U2F for end users; without it they would need to obtain and manage at least one token per account. Unfortunately, only about half of our sample understood that a single U2F token could be used with multiple accounts from different service providers.

5.5 Phase-II Results

The second experimental phase consisted of running an identical protocol with a similar sample of participants. We again focused on analyzing the usability and acceptability of the two-factor authentication tool after the changes described in Section 5.3.

Table 1 and Table 2 show a comparison of halt points and confusion points between the two phases. We observe that Phase-I had a higher percentage of halt points and confusion points when combined.

The demo and going to the incorrect settings were significant halt point in Phase-I, but in Phase-II, most of the participants did not require any intervention from the researchers for resolving this issue.

Several participants expressed confusion over whether the device would operate with Apple devices due to the implementation of the new USB-C port. While this problem can be solved by using YubiKey 4C, it is beyond the scope of the current experiment. It is worth noting that presently YubiKeys are not compatible with browsers other than Chrome or Opera. In this vein, the participants strongly recommended that YubiKeys be made compatible with all other browsers.

Phase-II include two conditions, as with Phase-I. The Google condition in Phase-II directed participants to Google Support's instructions on how to add and register the Yubico security key [24].

5.6 Phase-II Acceptability

In Phase-I of the experiment, no participant chose to continue the use of the token after registering it in our experiment. In Phase-II, most of the participants chose to keep the security keys after the experiment. However, the follow-up survey had low participation and hence, was not coded. Yet we note in Phase-II that five of ten participants reported continued use of the key on the survey. 2FA that requires pairing with a smart device is likely the only exposure to 2FA technology that many students have due to a compulsory 2FA that the

University has implemented on the students to login to use University services. However, lack of participation in the follow-up study could be an indicator of lack of engagement with the token.

Communicate the Intrinsic Benefit Confirmation of operation remains a serious issue underlying acceptability. If any artifact is not seen as working then it will not be seen to have benefits. When asked about continued use, one of the participants said, *"No, my password is secure enough and alerts are active."*

The instructions in the updated Yubico condition included information about the association of the device with other websites such as Facebook, and Salesforce. The links to the other sets of instructions also provided benefit information. Several users pointed out that multiple platforms could be linked and secured by YubiKeys.

Create a Cognitive Benefit Google continues to require use of the full password even with Yubico Security Keys, and does not offer a decreased cognitive burden option. Thus, there was no cognitive benefit for using the device. As noted by a participant in Phase-II, the question arose, *Why is it still asking for a password?*. Remembering a password which adheres to the password rules remains a challenge if one still needs it along with the hardware tokens.

Highlight the Engineering In the first phase, participants believed that they required different tokens for different websites. Phase-II participants indicated some awareness of the potential benefit of Security Key across different websites.

Communicating the Risks One Phase-II participant identified risk mitigation as a reason to use the Security Key, stating that it is, *"more secure against brute force or stolen password."* In Phase-II, due to various alterations in the design, description, and specifically in the demo of the Yubico instructions, participants found it more usable and acceptable in their day-to-day life. We discussed more on how the problems in Phase-I were mitigated in Section 5.3, along with a discussion on further recommendations in Section 7 to make the device more acceptable and to make the online usage experience of users more secure.

5.7 Comparison of Two Phases

The modification of the instructions and other changes mentioned in Section 5.3 made the Security Keys more usable. Table 3 list the statistically significant changes between the two studies. Figure 6 and Figure 6 shows a stark improvement in the halt and confusion points in between the two phases.

The most important changes were the removal of the demo, the presentation of the devices so that they were easily identified, and links to the sites in which the Yubico Security Key can be used. Yubico's removal of their own instructions was a major improvement. Instead, the Yubico website redirected the user to the website the person was seeking to secure. Though the instructions still remains slightly confusing to the participants, which can be improved further.

Yubico Instructions In Phase-I we noted that participants were not happy with the instructions especially those who received the Yubico instructions. We

	Phase-I		Phase-II	
Halt Point(Stop)	Yubico	Google	Yubico	Google
Demo	72.7%	0%	0%	0%
Incorrect Settings	72.7%	20.0%	19.04%	14.29%
Instruction	36.4%	20.0%	4.76%	0%
Form Factor	9%	10%	4.76%	0%
Biometric	9%	0%	0%	0 %
Pressing Button	9%	0%	0%	7.14%

Table 1. Percentage of Participants Encountering Halt Points: Comparison

	Phase-I		Phase-II	
Confusion Point	Yubico	Google	Yubico	Google
Demo	9%	0%	0%	0%
Incorrect Settings	18.2%	0%	4.76%	0%
Instruction	9%	20%	23.80%	71.43%
Form Factor	9%	0%	23.80%	7.14%
Biometric	9%	0%	0%	0%
Pressing Button	9%	10%	23.80%	28.57%

Table 2. Percentage of Participants Encountering Confusion Points: Comparison

found that 72.7% who got the Yubico instructions were restricted by the demo, 36.4% found the instructions unclear and were stopped with the way the instructions were given, and 72.7% of the participants didn't understand which setting to go to setup their device. One of the participants also exclaimed by saying *"This is a horrible web site. I don't know what it wants from me."* In Phase-II, though some of the participants found the instructions verbose, the problems faced by them were reduced.

Biometric versus Touch We noted that participants were more aware of how to press the golden button in Phase-II and no one faced difficulty and asked for assistance from researchers as compared to the 9% in the Phase-I. The instructions helped users in knowing about the keys and the participants expressed awareness that the button was not a biometric as seen in Figure **??**.

Demo In Phase-I, the participants who received the Yubico instructions were confused by the demo setup, resulting in 72.7% of the participants failing to register the keys with their account even with the help of the researchers. Removal of the demo from the instructions in Phase-II removed the halt points found in Phase-I completely. Everyone in the second phase was able to register the keys and associate them to their Gmail account as described in Table 1 and Table 2.

Halt Point	Phase-I Y vs. G	Phase-II Y vs G	Yubico I v. II	Google I v. II
Demo	0.0008	-	0.0033	-
Settings	0.0183	-	0.0033	-
Instructions	-	-	0.0213	0.0988
Form Factor	-	-	-	-
Biometric	-	-	0.1671	-
Pressing Button	-	0.2037	0.1671	-

Table 3. Table of Significance with Kruskal-Wallis Test (p value greater than 0.2 are excluded)

Table 3 shows the results of a Kruskal–Wallis test comparing the two phases. Any p value greater than 0.05 is not significant. Those which are borderline (i.e., between 0.05 and 0.2) are included in the table as these also may be interesting for future experimental evaluations. Those not included were not distinguishable

from random chance and we would not focus on them in future work.

6 Analysis

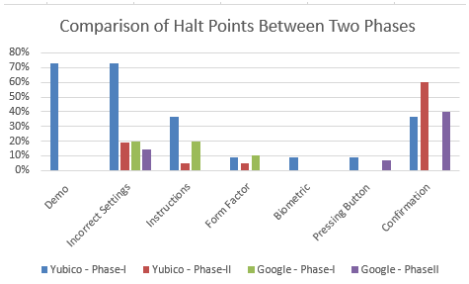
We implemented a set of correlation matrices to examine the potential interaction of the halt points and confusion points. For several people where we observed the correlation between form factor and settings in practice, this took the path of not understanding the nature of the device, in that they treated the device as a memory storage devices rather than an authentication device. Please note, the first phase of the experiment was identification of the device, so that all participants would have seen a description of these as authentication devices. Before having been directed to the settings, the participants searched their computers for this additional memory device and a few also inserted the device upside-down.

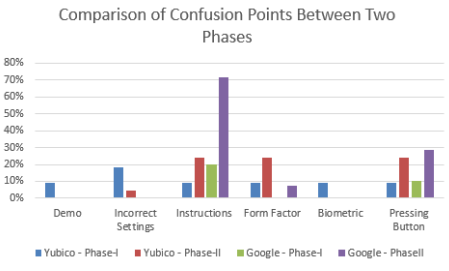
We also examined the correlation of halt points for Phase-II participants who received instructions from Google, finding only one non-zero correlation. In this case, the difficulty of finding settings was correlated with the operation of the button. This means that once the settings were identified, the participants were confused on when to engage with the pressure sensor in the enrollment phase.

Conversely, for those participants who interacted with the Yubico instructions the correlation between the confusion points of finding settings and interacting with the pressure sensor was negative (as shown in the correlation matrices in the Appendix). The only positive correlation was between difficulty in understanding the instruction and operating with the button by touching it at the correct time.

For the confusion points for participants who received the instructions from Google is also provided in the Appendix. The matrix shows an unsurprising correlation between not understanding the form factor and not being able to interact with the button. This is unsurprising as individuals who, for example, placed the device upside down can neither see the illumination nor reach the button itself. A second positive correlation was found between difficulty understanding the instructions and interacting with the button.

We cannot conclude that the confusion and halt points are independent. The correlation of different halt and confusion factors appear inter-related.





7 Discussion and further suggestions

Based on student feedback, we characterize the lack of acceptability of the Security Key as lack of awareness of the risk, lack of knowledge about the benefits, and the fact that the burden of passwords is not mitigated so there is a lack of actual cognitive benefits. This lack of recognition of the intrinsic benefit appears to be a function of defaults and expertise. The members of the security lab implemented a pilot think-aloud protocol to augment experimenter training. None of the members of the lab selected the option to trust the computer. Therefore, the expertise of the employees at Google [13] and Yubico might make this difficulty effectively invisible. Literature on psychology of security illustrates that communicating security as a benefit rather than a cost can be expected to increase acceptability of a technology [7]. Behavioral economics of security indicates that presenting the safety of two factor as an asset that the participants possess, rather than having them experience it as a cost, has the potential to improve perceived value and increase long-term use [8].

As we cannot make all users experts, communication of the benefits and the creation of a cognitive benefit are most promising. Communication of benefit could occur in the form of a validation email and communicating benefit information upon enrollment. There are other possible points of interaction. For example, after a password reset email, a simple message indicating that the participant is safer could be provided. Currently, security information focuses heavily on risk avoidance information and rarely provides benefit information. One possible form of benefit communication could be illustration of the options of different Security Keys.

A suggestion is to remind participants at first login after enrollment that the selected hardware is trusted. Explicit positive benefit messages could include initial congratulations on successful login the first time. After that a periodic reminder that "only this computer can log in without your key" with an image would provide clear indication of benefit. If any login is ever refused from a remote computer, a message to the participant indicating their successful triumph over a potential attacker would clearly identify a benefit.

Even in the second condition, many participants did not understand the benefit of the device as compared to a longer, more secure password. To address this, providers that support 2FA could include a pointer to U2F tokens when there are suspicious login attempts. In both phases, multiple participants expressed disappointment that their full password was still required after configuring the Security Key, even on trusted devices where a second factor was not needed. From a user experience perspective, pressing the single button to activate a U2F token presents a lower physical and cognitive load compared to typing a password [4, 23]. From a security perspective, the authentication provided by the token is stronger than any password a normal participant is likely to choose. As an alternative to this, one can use a shorter password with a few characters along with the Yubico Security Key rather than a password phrase.

Using the security token as a primary authentication factor also offers accessibility benefits. For enterprise customers, it could ease ADA compliance with

respect to authentication requirements for employees. Individuals who can be supported through voice recognition or other alternative means of entering text often still struggle with authentication, particularly when required to submit password phrases. Although still an unusual complaint, an ADA compliance issue could arise in the face of password complexity requirements.

In this experiment we studied the usability and acceptability of the FIDO U2F security key. Further studies could include a range of tokens, not only other Yubico security keys such as YubiKey 4, YubiKey 4 Nano, YubiKey 4C, and YubiKey NEO but also pico and other secure hardware. In addition, a goal of future work is to include vulnerable populations. Such populations are likely to have lower expertise but may have greater awareness of risk. We choose the undergraduates because of their increased skill in relation to computer and security and that they are early adopters of new technology but in the future we hope to expand this study for more diverse age groups including retirees.

8 Conclusion

Through a two-phase experimental study, we investigated the usability issues of the FIDO Security Key. In the first phase, we discovered there were six primary usability concerns or halt points - a confusing demo, going to incorrect account settings, confusing instructions, the hardware's form-factor, validation after setup, and participants' doubts regarding the benefit of the Security Key. In the second phase of the experiment, the usability concerns were mostly mitigated. These included updating installation instructions, removing the demo from the setup work flow, and clarifying descriptions of the registration process. This resulted in a remarkable improvement in the usability of the device - while 33.3% of the users were not able to complete the registration process in the first phase, everyone was successful in the second phase. In the second phase, although the halt points were reduced considerably, many participants were still confused whether the key was working due to the absence of message confirmation at the end of the registration process.

The improvement of usability did not automatically result in improvements in acceptability. Participants continued to express belief in the strength of passwords alone, showing undue faith in their own security acumen. We conclude with compliments to the usability of the 2FA token, and with a warning that even the best designed hardware will not be used if the benefits are not apparent.

9 Acknowledgement

This research was supported in part by the National Science Foundation under CNS 1565375, Cisco Research Support 591000, and the Comcast Innovation Fund. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the the US Government, the National Science Foundation, Cisco, Comcast, or Indiana University.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. "Privacy and human behavior in the age of information". In: *Science* 347.6221 (2015), pp. 509–514. URL: http://science.sciencemag.org/content/347/6221/509.short (visited on 05/04/2017).

[2] Lujo Bauer et al. "A user study of policy creation in a flexible access-control system". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2008, pp. 543–552.

[3] Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. "Graphical passwords: Learning from the first twelve years". In: *ACM Computing Surveys (CSUR)* 44.4 (2012), p. 19.

[4] J. Bonneau et al. "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes". In: *Security and Privacy (SP), 2012 IEEE Symposium on*. May 2012, pp. 553–567. DOI: 10.1109/SP.2012.44. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6234436.

[5] L. J. Camp, J. Abbott, and S. Chen. "CPasswords: Leveraging Episodic Memory and Human-Centered Design for Better Authentication". In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2016 49th Hawaii International Conference on System Sciences (HICSS). Jan. 2016, pp. 3656–3665. DOI: 10.1109/HICSS.2016.457.

[6] Michael Fagan and Mohammad Maifi Hasan Khan. "Why Do They Do What They Do?: A Study of What Motivates Users to (Not) Follow Computer Security Advice". In: *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. 2016.

[7] V. Garg and J. Camp. "Heuristics and Biases: Implications for Security Design". In: *IEEE Technology and Society Magazine* 32.1 (2013), pp. 73–79. ISSN: 0278-0097. DOI: 10.1109/MTS.2013.2241294.

[8] Jens Grossklags and Alessandro Acquisti. "When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information." In: *WEIS*. 2007.

[9] Philip G Inglesant and M Angela Sasse. "The true cost of unusable password policies: password use in the wild". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 383–392.

[10] T. Kelley, P. Rajivan, and L.J. Camp. "An assessment of computer and security expertise". In: *Technical Report*. Mar. 2014.

[11] Saranga Komanduri et al. "Of passwords and people: measuring the effect of password-composition policies". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2011, pp. 2595–2604.

[12] Kat Krol et al. ""They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking". In: *arXiv preprint arXiv:1501.04434* (2015).

[13] Juan Lang et al. “Security Keys: Practical Cryptographic Second Factors for the Modern Web”. In: *Financial Cryptography and Data Security*. Financial Cryptography and Data Security. (Accra Beach Hotel & Spa, Barbados, Feb. 22–26, 2016). International Financial Cryptography Association. Feb. 2016. URL: http://fc16.ifca.ai/preproceedings/25_Lang.pdf.

[14] Juan Lang et al. *Security Keys: Practical Cryptographic Second Factors for the Modern Web*. 2016.

[15] Brett McDowell. *Strong Authentication Canine*. June 2015. URL: https://www.youtube.com/watch?v=sdJ47NFGlgk.

[16] David M'Raihi et al. *Rfc 6238-totp: Time-based one-time password algorithm*. 2011.

[17] D M'Raihi et al. *RFC 4226: HOTP: An HMAC-based one-time password algorithm*. 2005.

[18] *New password guidelines say everything we thought about passwords is wrong*. VentureBeat. Apr. 18, 2017. URL: https://venturebeat.com/2017/04/18/new-password-guidelines-say-everything-we-thought-about-passwords-is-wrong/ (visited on 05/04/2017).

[19] Greg Norcie et al. “Why Johnny Can't Blow the Whistle: Identifying and Reducing Usability Issues in Anonymity Systems”. In: Internet Society, 2014. ISBN: 978-1-891562-37-2. DOI: 10.14722/usec.2014.23022. URL: http://www.internetsociety.org/doc/why-johnny-cant-blow-whistle-identifying-and-reducing-usability-issues-anonymity-systems (visited on 05/11/2017).

[20] Prashanth Rajivan et al. “What Can Johnny Do?–Factors in an End-User Expertise Instrument”. In: *Proceedings of the Tenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2016)*. Lulu. com, p. 199.

[21] Robert W Reeder and Roy A Maxion. “User interface dependability through goal-error prevention”. In: *Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on*. IEEE. 2005, pp. 60–69.

[22] Sampath Srinivas et al. “Universal 2nd factor (U2F) overview”. In: *FIDO Alliance Proposed Standard* (2015), pp. 1–5.

[23] Frank Stajano. “Pico: No more passwords!” In: *International Workshop on Security Protocols*. Springer. 2011, pp. 49–81.

[24] *Use Security Key for 2-Step Verification - Android*. URL: https://support.google.com/accounts/answer/6103523?hl=en&ref_topic=6103521.

[25] Rick Wash et al. “Understanding password choices: How frequently entered passwords are re-used across websites”. In: *Symposium on Usable Privacy and Security (SOUPS)*. 2016.

[26] Ryan West. “The psychology of security”. In: *Communications of the ACM* 51.4 (2008), pp. 34–40. URL: http://dl.acm.org/citation.cfm?id=1330320 (visited on 05/04/2017).

[27] Alma Whitten and J Doug Tygar. “Why Johnny can't encrypt: A usability evaluation of PGP 5.0”. In: *USENIX Security Symposium*. Vol. 99. 1999.

[28] Mary Ellen Zurko and Richard T Simon. “User-centered security”. In: *Proceedings of the 1996 workshop on New security paradigms*. ACM. 1996, pp. 27–33.

10 Appendix

We have provided the correlation matrices of the halt and confusion points for different sets of instructions (Yubico and Google) across the Two Phases. Due to lack of space we have used abbreviations for the halt and confusion points. The abbreviation list are as follows:

1. D: Demo
2. S: Incorrect Settings
3. I: Instructions
4. F: Form Factor
5. B: Bio-metric
6. P: Pressing Button

10.1 Phase-I

$$\begin{array}{c} \begin{array}{cccccc} D & S & I & F & B & P \end{array} \\ \begin{bmatrix} 1 & 1 & 0.04 & 0.19 & 0.19 & 0.19 \\ 1 & 1 & 0.04 & 0.19 & 0.19 & 0.19 \\ 0.04 & 0.04 & 1 & -0.24 & 0.42 & 0.42 \\ 0.19 & 0.19 & -0.24 & 1 & -0.1 & -0.1 \\ 0.19 & 0.19 & 0.42 & -0.1 & 1 & -0.1 \\ 0.19 & 0.19 & 0.42 & -0.1 & -0.1 & 1 \end{bmatrix} \begin{array}{c} D \\ S \\ I \\ F \\ B \\ P \end{array} \end{array}$$

Correlation Matrix of Halt Points for Phase-I participants who received Yubico Instructions.

$$\begin{array}{c} \begin{array}{cccccc} D & S & I & F & B & P \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.38 & -0.17 & 0 & 0 \\ 0 & 0.38 & 1 & 0.67 & 0 & 0 \\ 0 & -0.17 & 0.67 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} Demo \\ S \\ I \\ F \\ B \\ P \end{array} \end{array}$$

Correlation Matrix of Halt Points for Phase-I participants who received Google Instructions.

$$\begin{array}{c} \begin{array}{cccccc} D & S & I & F & B & P \end{array} \\ \begin{bmatrix} 1 & -0.15 & -0.1 & -0.1 & -0.1 & -0.1 \\ -0.15 & 1 & -0.15 & -0.15 & -0.15 & -0.15 \\ -0.1 & -0.15 & 1 & -0.1 & -0.1 & -0.1 \\ -0.1 & -0.15 & -0.1 & 1 & 1 & -0.1 \\ -0.1 & -0.15 & -0.1 & 1 & 1 & -0.1 \\ -0.1 & -0.15 & -0.1 & -0.1 & -0.1 & 1 \end{bmatrix} \begin{array}{c} D \\ S \\ I \\ F \\ B \\ P \end{array} \end{array}$$

Correlation Matrix of Confusion Points for Phase-I participants who received Yubico Instructions.

$$\begin{array}{c} D \quad S \quad I \quad F \quad B \quad P \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.15 & -0.15 & -0.15 & -0.15 \\ 0 & -0.15 & 1 & -0.1 & -0.1 & -0.1 \\ 0 & -0.15 & -0.1 & 1 & 1 & -0.1 \\ 0 & -0.15 & -0.1 & 1 & 1 & -0.1 \\ 0 & -0.15 & -0.1 & -0.1 & -0.1 & 1 \end{bmatrix} \begin{matrix} D \\ S \\ I \\ F \\ B \\ P \end{matrix} \end{array}$$

Correlation Matrix of Confusion Points for Phase-I participants who received Google Instructions.

10.2 Phase-II

$$\begin{array}{c} D \quad S \quad I \quad F \quad B \quad P \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.11 & 0.46 & 0 & 0 \\ 0 & -0.11 & 1 & -0.05 & 0 & 0 \\ 0 & 0.46 & -0.05 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} D \\ S \\ I \\ F \\ B \\ P \end{matrix} \end{array}$$

Correlation Matrix of Halt Points for Phase-II participants who received Yubico Instructions.

$$\begin{array}{c} D \quad S \quad I \quad F \quad B \quad P \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0.68 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0.68 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} D \\ S \\ I \\ F \\ B \\ P \end{matrix} \end{array}$$

Correlation Matrix of Halt Points for Phase-II participants who received Google Instructions.

$$\begin{array}{c} D \quad S \quad I \quad F \quad B \quad P \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.13 & 0.4 & 0 & -0.13 \\ 0 & -0.13 & 1 & -0.31 & 0 & 0.21 \\ 0 & 0.4 & -0.31 & 1 & 0 & -0.31 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -0.13 & 0.21 & -0.31 & 0 & 1 \end{bmatrix} \begin{matrix} D \\ S \\ I \\ F \\ B \\ P \end{matrix} \end{array}$$

Correlation Matrix of Confusion Points for Phase-II participants who received Yubico Instructions.

$$\begin{array}{c} \begin{matrix} D & S & I & F & B & P \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.18 & 0 & 0.4 \\ 0 & 0 & 0.18 & 1 & 0 & 0.44 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.4 & 0.44 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} D \\ S \\ I \\ F \\ B \\ P \end{matrix}$$

Correlation Matrix of Confusion Points for Phase-II participants who received Google Instructions.