A Qualitative Study on Usability and Acceptability of Yubico Security Key

Sanchari Das
Indiana University Bloomington
Bloomington, Indiana 47408
sancdas@indiana.edu

Gianpaolo Russo
Indiana University Bloomington
Bloomington, Indiana 47408
russog@indiana.edu

Andrew C. Dingman
Indiana University Bloomington
Bloomington, Indiana 47408
adingman@indiana.edu

Jayati Dev
Indiana University Bloomington
Bloomington, Indiana 47408
jdev@iu.edu

Olivia Kenny
Indiana University Bloomington
Bloomington, Indiana 47408
okenny@iu.edu

Dr. L. Jean Camp
Indiana University Bloomington
Bloomington, Indiana 47408
ljcamp@indiana.edu

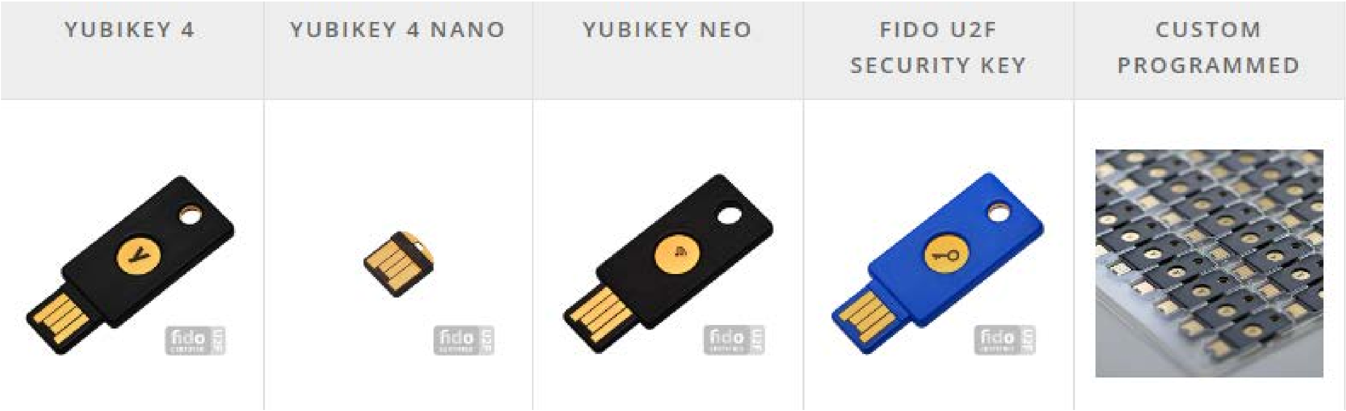


Figure 1. Two factor authentication security keys

Abstract

Individual concerns about account takeover and subversion are well-documented. Surveys indicate that concerns for the privacy and security of online accounts are widely shared. Adopting Two-Factor Authentication (2FA) is an action that individuals can take to secure their own accounts, including many popular consumer-facing services. Given that, why is two-factor hardware not more widely adopted? What usability and acceptability factors drive the adoption, or lack of adoption of 2FA in the form of trusted hardware? Passwords are inherently misaligned with human cognition, and hardware keys designed for ease of use are readily available in the marketplace. Yet passwords remain the dominant online authentication method. In order to better understand relevant issues driving or impinging adoption of Two-Factor Authentication, we implemented a two-phase study of the Yubico FIDO U2F security key.

The Yubico security key is a 2FA device designed to be user friendly. We examined the usability of the device by implementing a think-aloud protocol, and documented the halt and confusion points. We provided this analysis to Yubico, who implemented many of the recommended changes. We then repeated the study in the same context; noting significant improvements in usability. However, increase in usability did not affect the acceptability of the device, affecting the prolonged usage of the device. In both phases we interviewed the study participants about the acceptability of the device, finding similar concerns about lack of benefits and the invisibility of risk. A source of opposition to adoption is the concern for loss of access, with participants prioritizing availability over confidentiality. Another concern is that these do not lessen or simplify interaction with services as passwords are still required. We close with open questions for additional research, and further recommendations to encourage online safety through the adoption of 2FA.

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

ACM Reference format:
Sanchari Das, Gianpaolo Russo, Andrew C. Dingman, Jayati Dev, Olivia Kenny, and Dr. L. Jean Camp. 2017. A Qualitative Study on Usability and Acceptability of Yubico Security Key. In *Proceedings of , Florida, USA, December (STAST 2017),* 12 pages.
DOI: 10.475/123_4

1 Introduction

Authentication for an online account is generally understood as *something you know*, *something you are*, or *something you have*.

STAST 2017, Florida, USA
© 2017 ACM. 978-1-4503-6357-0/17/12…$15.00
DOI: 10.1145/3167996.3167997

These correspond to passwords, PINs, or passphrases as knowledge; biometric identity as intrinsic to self; and some form of physical token as possession respectively. There are other forms of authentication such as *somewhere you are* which includes providing location-based access [1], or *someone you know* such as social authentication schemes [2]. When any two such factors are required together for authentication, it is commonly referred to as Two-Factor Authentication (2FA).

Despite the wide range of authentication options, passwords continue to dominate online authentication, illustrated by Ruoti and Seamons [3]. Passwords suffer from widespread security flaws and the sheer amount of passwords generated increases the risk. In May 2017, one billion password and username credential sets were added to *Have I Been Pwned* [4]. Consumer-facing accounts and even workplaces continue to use this single-factor authentication technique despite well-documented misalignment with human cognition, difficulty in developing appropriate policies, and vulnerability to social engineering. 2FA is being increasingly adopted, but a simple examination of the risks and benefits would argue for wider popular adoption [5].

To examine the possible reasons for the limited diffusion of hardware tokens for 2FA for personal use, we implemented a two-phase usability and acceptability evaluation. The particular USB token we tested was the *Yubico security key*. The device we tested is the fourth from the left in Figure 1, labeled the *FIDO U2F Yubico security key*, which we will refer to henceforth as the 'Security Key'. The key is labeled 'FIDO U2F' because it is an implementation of the Fast Identity Online (FIDO) Alliance's Universal Second Factor (U2F) standard [6]. We chose the U2F Yubico security key due to the design focus on usability and privacy. According to the Executive Director of the FIDO Alliance, Brett McDowell, "We fail if FIDO is not more usable than all the other options you have used before" [7], which reaffirms Yubico's design priority towards usability. Privacy could also be an issue in adoption of 2FA tokens. The security key removes the potentially confounding factor of privacy risk, although the issue of privacy risk perception was explored and is addressed in our study. The security key is designed as a consumer-facing device for use with Google, Dropbox, and GitHub.

Specifically, we implemented a think-aloud protocol to identify stop points, perceived benefits, and perceived costs. We reported the findings along with recommendations to Yubico and documented the consequent changes for a second iteration of the study implementing these modifications. We focused on participants with above average technical literacy by recruiting students from STEM degree programs. Our goal was to identify difficulties that might be barriers to adoption for technically literate participants, particularly those who were likely to use GitHub, DropBox, or other sharing platforms.

We conducted the entire experiment in two-phases. In both the phases we asked the participants to configure a FIDO U2F security key for their Google account. Significant improvements in usability were noted in Phase-II over Phase-I. However, the overall acceptability did not change. Subsequently, we provided additional recommendations, such as confirmation of successful completion of the login, and the need to communicate the benefits of the device.

Our contributions are the specific suggestions for Yubico, the instrument we developed for evaluating perceived costs and benefits, the coding for these results, and the final analysis indicating the primary reasons for individuals not adopting 2FA. The specific suggestions are immediately applicable. The coding of the results allows for construction of multiple choice or other easier to scale instruments for evaluation the costs and benefits of 2FA which might be of interest to other researchers as well. The overall results can inform interactions or communications that are targeted at increasing 2FA acceptability, usability, and adoption.

Our study design laboratory analysis focused on usability and acceptability of two-factor hardware. There has been significant research on usability of passwords, passphrases, and other two-factor approaches which ensured that our research could be well-grounded in usable security practices. A qualitative approach not only offers insights into individual perceptions but may also inform future quantitative research.

In the following sections, beginning with some related work in usable security, we describe the experiment design in Section 3, and then provide the findings in section 4 and section 6. We close with a set of further recommendations in Section 9 both for Yubico security keys and Two-Factor Authentication in general.

2 Related Work

As mentioned earlier, our current work is grounded in prior research on usable authentication through passwords and passphrases.

Our initial evaluation of the security key was based on frameworks for evaluating authentication options. For example, Bonneau et al. argued that in order for an authentication technology to be broadly acceptable as a password replacement, it must outperform passwords on multiple fronts such as cognitive burden, physical burden, scalability, and privacy preservation [8]. Before Bonneau et al.'s listing was published, Stajano provided a set of recommendations for any authentication system through research grounded in the Pico hardware token authentication project [9]. Stajano proposed five core attributes for the token: secure, memoryless, scalable, loss resistant, and theft resistant. While the security key does introduce a physical burden, it is lightweight, and is physically effortless as its operation is only a button press. It is secure, scalable, as well as unlikely to be lost or stolen. It is also compatible with passwords.

The security key is cognitively effortless as well once enrolled for a given service. Lang et al. refer to the use of a security key as "brainless", which seems to indicate a belief that there are no stop points in security key adoption [10]. However, the only documentation of adoption benefits we could locate was the one from Google. It included neither qualitative components nor human subject experiments. The Google report examined the benefits of adoption in quantitative terms of password support costs. By definition, technologies which are mandatory for continued employment will be used by all employees, but this does not make them inherently acceptable in the larger market. Thus, we complement this study with qualitative insights.

In contrast to the quantitative record of performance of security keys when adoption is mandated, previous academic research and market penetration numbers have shown low user acceptance of two-factor authentication in absence of employment requirements. A previous human-centered evaluation of 2FA found that users perceived twice the utility from *avoiding* 2FA compared to adopting it [11].

Our research design is informed by experiments on usability of access control [12], firewalls [13] and PGP [14]. These previous threads of research found that traditional usability principles can not always usefully inform security design. Specifically security is often a secondary task, so there is rarely continuous user focus. Security is also concerned with risk, which is stochastic, so a direct connection from action to consequence cannot be made. As with previous research, we recruited students from STEM degree programs and tested setup instruction sets using a think-aloud protocol.

The evaluations of password usability also motivated our research. In a particularly salient work, Sasse and Inglesant have shown that there are very real costs to passwords, particularly with burdensome policies [15]. Later work illustrated how improved security may reduce compliance and increase workarounds [16].

The canonical *Why Johnny Can't Encrypt*, examined the use of PGP [14] using think aloud protocol in the laboratory and proposed a set of guidelines that could inform design of secure systems. PGP secures emails; 2FA enables access control for services including emails so these guidelines may be applicable. Whitten et al. proposed that people are aware of needed security tasks; can perform these tasks without making dangerous errors; and find the interaction protocols such that they will continue to use the security technology. Here we consider the first two of these usability, and the third acceptability.

We followed the same process as the two phase examination of Tor by Norcie et al. [17]: a think-aloud protocol resulting in design recommendations and changes, followed by an additional study. Norcie refined design heuristics for security systems focusing on anonymity systems, specifically discussing the following heuristics. The first heuristic, *installation precedes operation*, refers to the fact that the system must be easy to install because without complete installation it will never be used. The second heuristic identified was to *ensure users are aware of trade-offs.* The third heuristic proposed for anonymity systems was *Say why, not how.* Our research addressed acceptability as well as usability. That is, not only did we evaluate if it were possible to use security keys, we asked if the participants perceived the 2FA hardware as desirable. Significant work has been done to evaluate the sources of, and the ability to alter, user behavior in terms of security practices [11]. Yet in contrast with the workplace study by Albrechtsen we did not find high levels of motivation by our participants to be secure [18].

In addition to these heuristics, there is literature that explicitly addresses the need to present benefits of security measures like costs as immediately apparent. Increased perceptions of the benefits could increase likelihood of long-term use [19]. Clear framing of security as providing positive benefits in addition to reducing possible harm, or framing adoption as a gain instead of a loss, can increase acceptability of a new technology [20, 21].

3 Methodology and Experiment Design

We asked experiment participants to configure a FIDO U2F security key for their Gmail accounts and then observed as they succeeded or failed. From these observations we developed a set of possible recommendations to improve user experience and provided these to Yubico. Yubico made a subset of the recommended changes, and then we repeated the study. Here we detail how we investigated the end user experience of configuring and using the security key by combining a think-aloud protocol and surveys.

We began the experiment design by evaluating the framework provided by Bonneau et. al. for testing the authentication technologies as possible substitutes for passwords [8]. This framework identified important attributes such as, deployability, security, and usability considerations [8]. Although deployability and security aspects of FIDO U2F are beyond the range of this study, we do observe that under any security analysis U2F would be an improvement over passwords alone. For an authentication technology to be broadly accepted by the users as a password replacement, it must outperform passwords on at least several aspects of the framework, such as cognitive burden, physical burden, scalability, and privacy preservation. The security key design goal is low cognitive burden, high ease of use, and privacy preservation compared with single sign-on, and high acceptability.

Security keys appear to meet all of the requirements set forward by Bonneau and Stajano. Previous work has shown that FIDO security keys are easy to deploy. A study by Google of internal adoption found that security keys have significantly decreased user support requirements when compared to one-time passcodes delivered via mobile app or SMS [10]. There is no obvious intrinsic reason for low levels of adoption of the security key based on these popular frameworks. Thus, we moved forward with the usability analysis in the lab.

The experiment consisted of a preliminary survey, a think aloud protocol, then an interview, and finally a follow-up survey. In a think-aloud protocol the subject narrates their actions, providing a real time description of their decisions, choices, or motivations. The surveys were online, while the think-aloud protocol and the interviews took place in a University computer laboratory.

The study conducted is a two-phase experiment and to avoid inconsistency across the two phases, we recruited participants from the same undergraduate course. The course selected was an introductory level security course. To keep similarity between the participant pool they answered the same pre-survey which asked about questions related to their background, knowledge, and skill set. To develop the questionnaire we followed the questionnaire used by Rajivan et al. who validated the result by implementing it in a large scale multiple-population study [22].

To participate in our study in both the phases, the participants were screened by the four following criteria. They were required to be eighteen years old, have a personal Gmail account, have their own laptop with them for the experiment, and finally, they were required to own a smartphone.

The participants were purposefully selected to have more security and computer expertise than the general population. This matches the current and targeted population. Consider that in 2017, Yubico indicates that two of the top eight uses of the security key are to interact with a master password system [23]. The top three uses are Facebook, Google, and GitHub. The use of password managers and the fact that GitHub is in the top three indicates that the early adopters are more technical than the general population, supporting our demographic choice.

After the preliminary survey, a coin flip was used to randomly divide the participants into two groups. In one group, participants were given the short URL from the Yubico packaging, which took them to the official Yubico security key instructions. The other group was directed to the security key instructions provided by Google. The instructions provided by Google have a longer URL, so we used a university-run URL shortening service to facilitate manual entry.

The think-aloud protocol began by giving each participant a Yubico security key, as shown in Fig. 2. 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. The researcher took notes, but did not offer additional guidance unless requested, or when the participant was unable to proceed without some guidance. Each researcher had previously participated in a pilot, where all the researchers concurrently went through the experiment and then discussed the observations. The goal of the pilot was to provide agreement on note-taking and language used in interacting with the participants. The pilot data included two rounds, but is not included here as it was in the experiment development phase.

Each researcher observed the configuration of the two-factor authentication and documented various stop points. Each researcher intervened when the participant came to a complete stop, or when the participants believed they had completed a task but had not. The purpose of the think aloud protocol was to identify stop points, in order to create recommendations for improvement.

Figure 2. Original device identification

After the task was complete, participants were asked to describe the use of the security key. We used a closed interview format, with seven questions asked in the same order by each researcher.

1. How could you test to confirm that your key is working?
2. If your key were 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 multiple goals in this closing interview. The first 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 these research subjects were capable of removing 2FA without researcher assistance. We also ensured that they had the contact information of the team and a specific researcher before they left. In addition, they were told to ask the teaching assistant or professor if difficulties arose. While the researchers were not instructors in the course, the instructors had previously agreed to direct any question to the researchers. The participants were allowed to keep the security keys as a token of appreciation for their participation. This gesture also helped us in continuing the next phase of the follow-up study.

One month after the end of the think-aloud protocol, the subjects received a follow-up survey about their continued use of 2FA. The follow-up survey was sent over email. The participation in the follow up survey proved that this was not a useful step, due to very low responses. We therefore do not include it in the analysis below.

We implemented the protocol in two phases. Similar methodology was followed in both the phases to avoid inconsistencies. There were 21 participants in the first, and 35 participants in the second. There was no overlap of participants in the two phases to avoid biased behavior. In Phase-I, we discovered that the most significant stop point was the confusion resulting from a Yubico demonstration tool. The participants were given instructions from both Google's site as well as that from the Yubico. Yubico had built a tool clearly illustrating how to configure 2FA for Gmail. Participants went through the demo and believed that they had completed the installation process. No participant in the experimental group who was directed to the Yubico demo was able to realize they needed to continue and complete the installation. The majority of the participants either believed that they had completed the installation or could not find where else to go.

Phase-I concluded with a set of recommendations about the instructions, visualizations, device identification, and guidance provided to users. The details of the recommendations are described in Section 5. We repeated the experiment to test the efficacy of the adopted recommendations after Yubico implemented a subset of these. We also revisited the recommendations that were not implemented from Phase-I, to determine if those changes were still needed after a subset of the changes were made.

In the following section we provide details on the analysis of the results, particularly the coding and analysis procedure. We then show the results themselves, and the recommendations derived from it.

3.1 Coding and Analysis

As discussed in the Section 3, we implemented the preliminary survey to screen participants based on our selection criteria. Thereafter the participants registered the security keys provided to them as a part of the experimental task given to them. While registering the key the participants followed a think-aloud protocol to discuss about their problems faced while registering the keys. An interview followed after the task to discuss about their problems and to find out reasons whether they find the security keys useful or not in their daily life. Later on a follow-up survey after a month concluded our survey.

The thorough method resulted in the survey data, transcribed data from the interview, measurements from the experiment, and analyzed qualitative data. Audio recording of the think-aloud protocol was carried on which was only stored in University's secure storage. We deleted the audio recording of the participants as soon as the transcription was complete. Two research assistants who were involved with the IRB protocol transcribed and rechecked the transcribed data. The transcribed data was then coded by research assistants. The same protocol was followed in both phases.

The transcription of the think-aloud protocol was of the recording that started with as soon as the participants were handed the security keys. We wanted to capture what the participants thought by looking at the security keys as well. The transcription ended as soon as the participants enrolled with the 2FA and they answered the questions asked by the interviewers. The questions asked by the interviewers were open ended questions and are discussed in Section 3.

The researchers were trained in qualitative coding methods and each of them individually coded a subset of the transcribed data in both the phases. After which discussion on any discrepancies was made. After the discussion the three researchers went through the transcribed data again and coded the comments of the participants in three categories or themes: the halt points, the confusion points, and the value points.

The halt points were coded where the participants were driven to a complete halt and could not proceed without the help of the researcher allotted to them. The confusion points were coded were the participants stopped for a while but did not need the help of any researcher to proceed further. The value was coded on the comment where the participants expressed their own view and opinion about the product and which could enhance the usability and acceptability of the device (our recommendations have acknowledged such value points). As standard in qualitative research, the themes were compiled into a code book.

In Phase-I the initial inter-coder agreement was 89.5% and in Phase-II it was 87.9%. The recorded halt points in Phase-I were clustered around four major issues: form factor, a setup demo, validation of configuration, and security benefits of the device. In Phase-II the stop points were significantly reduced with only 2.9% of the participants engaging with the demo in a confusing manner, in comparison with 63.2% of the Yubico instruction group being confused at the demo in Phase-I.

In both experiments, many participants recognized the potential value of the security key in theory, but not for themselves in practice. The details of the two phases are described in the following two sections, and then a discussion that addresses both follows the two sections.

4 Phase-I

As reported in Section 3, our participants for both the phases were recruited from the same course, albeit different semesters, to ensure that the sample was moderately security savvy. Because this was a sample of students, they were young. Six were between 18 and 20, 16 were between 21 and 23, 4 were 24-26, and one was over 30. There were 20 male students, and 7 female students, a 75% to 26% split. Every student was enrolled in at least one information security or computer science class, by definition.

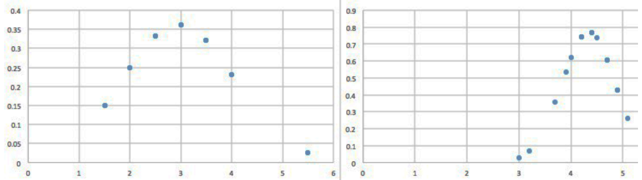


Figure 3. Phase-I Participant Expertise

Figure 3 shows the computer expertise and security expertise on the right and left, respectively. The mean security expertise was 2.96 of 5 and the mean computing expertise was 4.34. Compare this with a general population survey of 593 where the results was a mean security expertise (using the same calculation) of 1.7 and a mean computing expertise of 1.77 [24]. This illustrates that the population was more highly technical than consumers in general. As a result, it is reasonable to assume that any stop points encountered by this population could also occur in a less technical and less educated population. Recall that our expertise was measured using the survey and factor analysis as developed in [22].

We asked the participants about the password practices they use to secure their email and/or other websites. We provide and compare these with the second set of participants in Table 2. The survey measured skills and expertise, the question on password behavior was intended to measure behaviors, and was drawn from Egelman's study [25].

4.1 Phase I Findings and Usability

From the observations made by the use of the think-aloud protocol, we observed a set of common failures. The problems are discussed below.

The form factor of the device is of a small USB. As a result it can physically be inserted upside down. This was not expected by the participants. When inserted upside down, the device does not connect and does not work. If it were able to connect and interact with the machine, it would not have been possible to see or interact with the touch sensor. Thus, many participants inserted the device upside down and were confused when the device did not work.

The Yubico security keys were inserted through the USB ports. Thus, after insertion of the device the participants usually waited for the device manager to pop up a message without proceeding it to the browser due to their lack of experience with the security keys. Once the participants had moved to their browsers, there was confusion regarding the stated instructions as well. For example, the instructions mentioned to go to the settings, and instead of exploring the account settings, the participants explored the browser settings.

The participants were provided with two sets of instructions, the Yubico official instruction set and the Google instruction set on how to register their account with the security key. The participants in the first phase who got the Yubico instructions found the Yubico landing page to be difficult to understand. They also were unable to identify where to go to understand which device they were referring to in the instructions. Despite having the original packaging for the device, participants generally were not confident about which model of 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. The decision was often accompanied by a comment, such as *"Well, my key is blue and this one is blue, so I think it's the right one"*. No subjects mentioned using the images on the button to differentiate security key models. The different types of security key models are shown in the Figure 1.

Finding the correct key instructions wasn't the only set of problems faced by the participants. In fact once subjects had determined which model of key they were using, the next challenge was finding the correct setup instructions. This step presented the greatest challenge to our subjects in the first phase. 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. This was reasonable, as it was labeled, "Try out this key". In reality, the link took participants to an application designed to demonstrate the enrollment and use of the device, but not connected to any account. The removal of this ambiguous link was our strongest recommendation. The demo gave the participants a perception that the link registered their keys and associated their account with two-factor authentication.

Another major problem as given example above is finding the correct accounts settings instead of the browser settings which they were using. For successful setup, participants were required to follow a non-linear path through the control panel, and at each page offered a large array of options. This presented many opportunities for confusion and abandonment of setup altogether for several participants. Many participants found it difficult to remember all the steps and going back to the instruction manual was considered to be tedious and tiresome.

For participants having found the instructions and having interacted with the key, there were two primary results. To activate the security key, either for enrollment or authentication, participants had to 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. A steady light might indicate that the device is ready to authenticate. A short series of flashes might confirm that the button had been activated, substituting for the tactile feedback a capacitive button lacks or a text during enrollment.

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. This was clear to all participants, particularly since the device lights up when touch is needed. It also implies, however, a higher benefit than the device actually provides, since, in reality, anyone can use it. If the token is lost, participants who believed they have bio-metric enrollment were unaware of the risk. To support this view, one of the participants mentioned, *"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"*. In fact, several of the participants in our experiments dropped the keys in a shared bin for leftover hardware, often used for mice or cables. We did not directly observe this but did quickly recover the devices to avoid risk to the participants. We therefore cannot say if this was correlated with the belief in the touch sensor as a bio-metric. This redefines the importance of theft and loss resistance noted in related work [8, 9].

The demo was a particularly problematic stop point. 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 hault point. As one participant noted, *"The web site is kinda confusing because I do not know what it wants me to do."*

Participants were unable to confirm that the device was working after setup. When participants 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, participants never got to actually experience using their security key outside from the set-up process. As a result, the first time they would use their security key as part of their normal authentication pattern was divorced from the setup process, and left to a future point in time when they would attempt to log in from a new computer. For single-computer participants, 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."*

As noted above, where we discuss participants simply dropping the keys in the loose hardware box, few participants kept the device. The primary drivers of acceptability were lack of awareness of the risk, and the resulting perception that there was no benefit. Here we recommend changes to increase acceptability. 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 [26].

4.2 Summary

In Phase-I we have identified and classified stop points as halt, confusion, and value points. We found significant usability challenges and low acceptability in a participant population that was non-expert but more expert than the general population. In the next section we enumerate the recommendations corresponding to these stop points, also inherently describing the difference between the Yubico condition in Phase-I and Phase-II. After the Phase-II discussion we compare the results between the phases. For summary statistics please see Table 1.

5 Recommendations

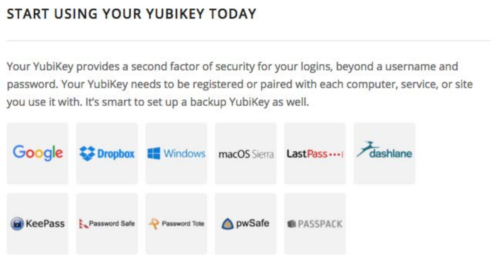
In response to our results we made specific recommendations in a technical talk presented to Yubico and Google. Some of these recommendations were then quite quickly adopted, either as a result of our work or serendipity. Here we list our recommendations and, in the case of adoption, note the difference.

One recommendation was not adopted. Specifically, we proposed that first time a Yubico key is inserted, the browser could open a dialogue so that the participant could easily match the device and find the desired supported service. However, the other recommendations were actionable for Yubico.

Participants had difficulty finding instructions. The improved Yubico web page had vastly improved. The page provides icons that link not just to the service but directly to the instructions for security key enrollment. The table of instructions provided during Phase-II are shown in Figure 4.

The updated service provider descriptions proved easier to follow than the Yubico descriptions, as discussed below. This is not unexpected, given the relative expertise each service provider has over its own service. Our recommendation for Yubico to provide pointers rather than instructions for each service provider had an effect.

Our first recommendation was in regard to the landing page to which participants are directed by the product packaging. First time participants were not able to easily identify which product they had, or which instructions to follow. The "Try out your security key" demo was a source of much confusion. Subsequently, the demonstration link led participants to erroneously believe that they are enrolling their key in a desired service, rather than simply using a demonstration application. In every experiment condition where a participant was directed to the Yubico instructions, they got stuck in a loop with the instructions and required guidance to reach a further step. Many participants looped through the demo for up to 10 minutes before receiving intervention.



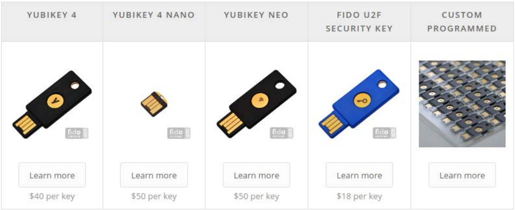
START USING YOUR YUBIKEY TODAY

Your YubiKey provides a second factor of security for your logins, beyond a username and password. Your YubiKey needs to be registered or paired with each computer, service, or site you use it with. It's smart to set up a backup YubiKey as well.

Figure 4. Clear Links to Instructions

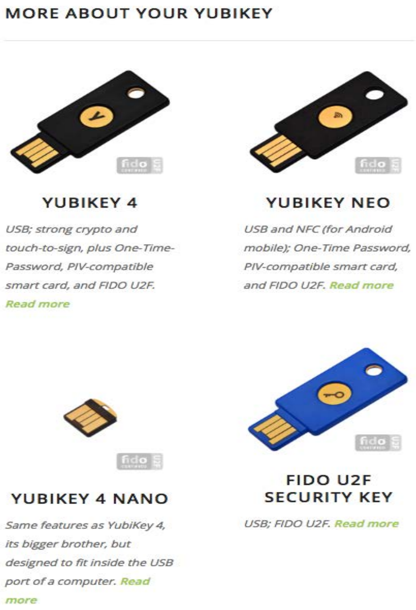
The demo does appear to serve the important goal of providing hypothetical demonstrations to prospective institutional customers. However, when this demo is included as part of the display to those who have already purchased the product, it consistently caused confusion. We recommended that this demo should not be accessible to the end participant, as it was a consistent stop-point. The demonstration cycle has since been removed from being directly in the participant's workflow, though it is still accessible. As a result this hault point went from confounding every single subject in the Yubico condition in Phase-I to having very low impact in Phase-II.

At the time of our initial experiment, individuals had difficulty determining which device they were using. The instructions asked which Yubico product a participant has, but provided little identification guidance in answering. A participant's best option was a product comparison table, the top of which is shown in Figure 5. The table appeared to have been designed to assist in purchase decisions rather than configuration, with prominent price information and technical data. A new interaction, pictured in Figure 6, 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. A significant change is the clear identification in the Yubico setup instructions that the button is not a fingerprint reader. The new description is shown in Figure 7. The resulting interaction is extremely clear on this point. At the time of 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, participants are not required to use a second authentication factor when logging in, even when 2FA is enabled for the account. The enrollment process did include a confirmation screen with a check-box, which made it possible to refrain from making the current browser a trusted device. Few participants noticed the box or understood its implications.



YUBIKEY 4	YUBIKEY 4 NANO	YUBIKEY NEO	FIDO U2F SECURITY KEY	CUSTOM PROGRAMMED
Learn more	Learn more	Learn more	Learn more	Learn more
$40 per key	$50 per key	$50 per key	$18 per key	

Figure 5. Original device identification



MORE ABOUT YOUR YUBIKEY

YUBIKEY 4

USB; strong crypto and touch-to-sign, plus One-Time-Password, PIV-compatible smart card, and FIDO U2F. Read more

YUBIKEY NEO

USB and NFC (for Android mobile); One-Time Password, PIV-compatible smart card, and FIDO U2F. Read more

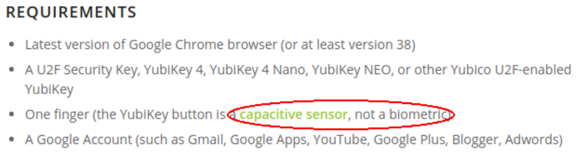
YUBIKEY 4 NANO

Same features as YubiKey 4, its bigger brother, but designed to fit inside the USB port of a computer. Read more

FIDO U2F SECURITY KEY

USB; FIDO U2F. Read more

Figure 6. The larger labels and more clarity in labeling offer the promise of improved device identification.



REQUIREMENTS

- Latest version of Google Chrome browser (or at least version 38)
- A U2F Security Key, YubiKey 4, YubiKey 4 Nano, YubiKey NEO, or other Yubico U2F-enabled YubiKey
- One finger (the YubiKey button is a capacitive sensor, not a biometric)
- A Google Account (such as Gmail, Google Apps, YouTube, Google Plus, Blogger, Adwords)

Figure 7. Setup requirements now explain what the button is

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.

This difficulty has since been partially addressed by a change in Google's defaults. As of January 2017, adding a security key to a Google account does not appear to make the browser "trusted", and a participant can therefore test the new sign-in procedure by signing out and back in. However, this only works if the participant had not previously marked the browser trusted – the check box for which is consistently prechecked at login. Notice that this was not a problem in the focus group of security researchers. This indicates that participants with a wider range of expertise are needed to evaluate products.

5.1 Acceptability Recommendations

We specifically picked up a research pool who were concerned about security and were enrolled in a non-technical security class as well.

The security and expertise score in Figure 3 indicates that, however none of the participants decided to continue using the security keys. We provided the security keys as a token of appreciation for their participation and still they discarded some of them which were collected in a bin by the researchers. We specifically collected the registered security keys to avoid exposing the participants to any potential risks. We also sent a follow-up survey to the Phase-I participants and none responded positively.

This depressing result can stem from the fact that the participants thought they do not need two-factor authentication since they do not have any confidential information in their email as indicated by them in the think-aloud protocol. However, we did the pilot study on 15 graduate students who were in the field of security and all continue the use the tokens. They did encounter a subset of the registration problems described in Section 4.

However, we cannot deny the fact that graduate students with more expertise in security accepted the device better than those without much security expertise. Thus, indicating that expert users understand and acknowledge the need of better security benefits. Additionally, as a company which manufactures such security tools should communicate the risks and the benefits of such device such that non-experts can evaluate the need for such tools as well. Two-factor authentication is a tool that should not just be subjected to a more technical friendly population but rather should be directed towards everyone to improve the security and our study aimed at improving the usability of the device even from the start and to provide better and effective means of communication to increase adaptability of the device. The registration of the keys do not make it compulsory for the users to use the device and unless risks are communicated in a detailed way the users will not be interested in the device even if that is effective.

Our above argument is supported from Gard et al.'s work where they discussed about the heuristics that user adopt to avoid risks and how a design changes can be utilized to increase the security of an individual or an organization [20]. Here, we do not indicate in influencing the users, whereas we talk about how users can take a more informed decision if the risks are communicated in a better way. For example, participants in our study register the keys as a part of their activity. However, they did not continue their usage and even refused to take a security key given to them free of cost since they were unable to visualize the benefits associated with the hardware token.

The overview *Psychology of Security* [21] and the precursor *Heuristics and Biases* [20] note that while people frequently make poor security decisions, and often undermine mechanisms in place for their own protection, they do so in a systematic manner. These works enumerated the psychological factors contributing to this phenomenon, explaining how people discount and fail to understand risks, require positive feedback for good decisions, and prefer the risk of large losses over small but certain costs. All of these impinge decisions about abstract, probabilistic nature of losses contrasted with concrete security costs. And our recommendations are grounded in the phenomena addressed in these works. Adding occasional positive feedback could improve acceptability.

In the initial study, participants did not understand the benefit of using the the device over a longer, more secure password. Participants who chose to return the token expressed confidence in their own security management and length of passwords. Many of the participants also thought that the device would be useful in case of computer theft. But once they understood the the FIDO key would not protect a trusted device any more than a stored password (i.e., not at all), they were dismayed. Participants did not understand the nature of mutual authentication. No participant in either phase indicated awareness of the online protection schemes like mutual authentication to combat phishing provided by FIDO capability.

There were some participants who are not concerned about the gmail accounts, so it did not matter if security key was safer as there was nothing to protect. This was specific to gmail for the participant said that they *"Probably not [on] gmail is not important. Would have used for work"*. Another proposed that his email was not actually useful or valuable, stating that, *"For my use, No, it is inconvenient to use. The reason is that I don't have any sensitive information."* Thus, for a number of users it was apparent that using the security key beyond work might not have been very useful. However, given the value of an email account to attackers, this seems somewhat unlikely. For example, popular security journalist Krebs illustrates the value of a stolen email as including the ability to spam, implement trusted traveler frauds, and calendar spam in high. So in this case, the issue may be incentive alignment for the participant, which has already been described as a chronic problem by various researchers [27, 28].

Neither the risks of weak passwords and phishing nor the value of email to an attacker were invisible to our participants, so it is critical to communicate the existence of risks that need to be mitigated to potential adopters.

One of the participants had recently lost a phone. He suggested that enrollment with 2FA through security key could allow recovery as well as protection of the phone too. His ideal was that a lost phone could be located and forced to return, once the owner called by potentially 'bricking' the phone with a combination of a security key, password, and security questions. This remote control without centralized registration is an interesting possible feature, and was suggested but was too complex to consider a recommendation.

Participants could find lesser value with 2FA since the participants still need to use passwords to login to their devices. Their perception of the use of tokens was the removal of the need of passwords. When the participants were informed that the passwords are still needed, they often made comments such as *"well... I don't really understand the point of the key if I still need to enter my username and password."* Many participants felt that the second factor was overkill, or too much of a burden in exchange for the no cognitive benefit. One participant suggested that there was no benefit other than amusement, saying he might *"use it out of curiosity, [as it] might not be practical."*

West et al. and Garg et al. in their respective works had two recommendations that address this major challenge to acceptability: reducing costs associated with security and improving rewards for good decisions [20, 21]. We addressed both of the comments by providing the keys free of cost thus indicating only one feasible challenge which is rewards for a good decision. Specifically, we recommend a clear immediate benefit by reducing the cognitive load of passwords in return for use of FIDO can improve acceptability. This could be achieved either by not prompting the participant for a password or not requiring the entire password when FIDO is being used. The current design is the reverse of this: participants must enter the password, but the device is not required. The change (no password, partial password, or choosing a PIN instead of a password) would also confirm verification of correct configuration.

Discussing about rewards for using the keys one of the major lack of communicating the benefits of using the security keys were only half of our participants in Phase-I were aware that the single key can be linked between multiple accounts and multiple keys can be linked with a single account. Due to lack of such knowledge the participants were scared of losing their tokens and in return losing access to their accounts. In fact, 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. Among those participants who understood the working principle of the key with multiple accounts none expressed awareness that this could be done without revealing that both accounts belonged to a single person. This benefit needs to be clearly mentioned and can be one of the selling points of such important security tools.

Despite the challenges discussed above, it is noteworthy that participants in our study did not object to the need for the token. In previous work [11], the most significant participant complaint about 2FA was the need for a second device. The fact that such a complaint did not arise at all in our observations suggests that Yubico's design and engineering efforts have resulted in a substantially more acceptable physical form factor.

6 Phase-II

We continued the next phase of the study with 35 participants. In both the phases we followed the same study protocol and there was no overlap of students between the two phases. Students of the same undergraduate class was studied but there were no students repeating the course.

6.1 Phase-II Demographics

As with Phase-I, in Phase-II the participants were students recruited from a non-technical computer security course. A major change, uncontrolled by the researchers, was that the University made 2FA mandatory for student university accounts. In Phase-I, students may have had limited previous interaction with 2FA. Thus, in Phase-I, rejection of the security key would mean that there was no 2FA interaction. In Phase-II a few students who were employed by the university had been required to enroll in 2FA for their jobs, but none had a physical token, only SMS-based authentication. We cannot isolate the effect of the changes of individual experience and that of the changes in the conditions.

Figure 8 shows the computer expertise and security expertise on the right and left, respectively. In Phase-II the expertise was 2.95 of five and the mean computing was 4.22. Recall that in Phase-I the respective means were 2.96 and 4.34, with the same distributions as shown in Figure 3. Again, it is reasonable to assert that our participants have more security and computing expertise than the general population. The differences in the means were not significant.

6.2 Phase-II Experiment & Results

In the second experiment we ran an identical protocol in the same class one year later. The survey and the instructions were the same. The coding processes were used on the think aloud protocol. 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 [29].

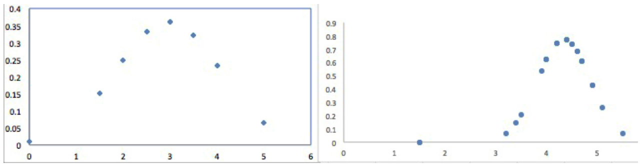


Figure 8. Phase-II Participant Expertise

We found significant differences in this second examination of the the usability and acceptability of the two-factor authentication token. In Phase-I the Google instructions were far more effective at guiding individuals to installation. In Phase-II participants in the Yubico condition were more successful. Table 1 shows a comparison of halt points and confusion points between the two phases. Phase-I had different halt and confusion points, as well as more overall such points.

The demo and going to the incorrect settings were significant stop-point in Phase-I, but in Phase-II, most of the participants were able to locate the settings. The demonstration video remained, but it was not longer the most visible component. The new setup description is shown in Figure 9, appears to be an improvement in terms of usability.

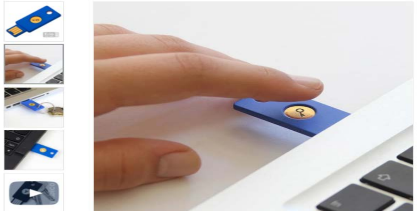


Figure 9. Setup Description

A major constraint on acceptability was the concern about loss of the key. The risk of being denied access to the account was more salient than the risk of losing access to an attacker. During the interview 23.5% of the participants mentioned were confused about how to recover their account were the key lost. The majority of the users were able to remove the key in the exit interview.

Some participants thought the devices were branded. Specifically, the form factor of the security key we were using was not compliant with the newer Apple devices because of Apple's adoption of USB-C only. Participants also strongly recommended that security keys be made compatible with other browsers.

7 Comparisons Between Phases I & II

The modification of the instructions and other changes mentioned in Section 5 made the security keys more usable. However, we cannot distinguish between the changes in enrollment and the 2FA mandate for any changes in acceptability. Table 1 lists the statistically significant changes between the two phases of the study.

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 security key can be used. Yubico's removal of their own instructions was a major improvement. Instead, the Yubico website redirected the participant to the website the person was seeking to secure.

In Phase-I we note that participants have difficulty both locating and following the instructions, especially those who received the Yubico instructions. We found that 72.7% who got the Yubico instructions were stopped by the demo, 36.4% found the instructions too unclear to follow, and 72.7% of the participants could not follow the instructions enough to find the settings, much less interact with them. One of the participants expressed this sentiment quite clearly, exclaiming, *"This is a horrible web site. I don't know what it wants from me."* In Phase-II, while some participants found the instructions redundant, overall participants could locate the instructions and these new instructions were much easier to follow.

Every participant in Phase-II was able to locate and press the single button that is a core component of the security key interaction. None of the Phase-II participants required guidance, while 9% of participants in Phase-I needed help in interaction with "the golden button". In the Phase-I, participants believed the security key to have a fingerprint reader. All Phase-II participants expressed awareness that the button was not a biometric, and we believe this was the result of Figures 7 and 9.

In Phase-I the participants who received the Yubico instructions were confused by the demo setup, resulting in over 72% of the participants being unable to register their keys. Removal of the demo from the instructions in Phase-II removed the corresponding halt points. All participants in Phase II were able to register the device and associated it with their Gmail accounts as shown in Table 1.

	Phase-I	Phase-II	Yubico	Google
Halt Point	Y vs. G	Y vs G	I v. II	I v. II
Demo	0.0008	-	0.0033	-
Settings	0.0183	-	0.0033	0.7771
Instructions	-	-	0.0213	0.0988
Form Factor	-	-	-	-
Biometric	-	-	0.1671	-
Pressing Button	-	0.2037	0.1671	-

Table 1. Table of Significance with Kruskal-Wallis Test

Table 1 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.

8 Acceptability

In Phase-I of the experiment, no participants chose to continue to use the token after it had been provided. In contrast, in a pilot study consisting of members of the research security group, every participant chose to keep the token. This difference suggests the expertise level of the students as one of the causes indicating differences in their rational choices but we did not find statistical correlation of this within the experimental groups.

In the second phase usability was increased. Acceptability was also increased; however, some of the students were required to use 2FA in their University employment. This was a change in environmental conditions that the researchers could not control. As a result, in the first phase participants could reject 2FA entirely; in the second phase some students could but choose which type of 2FA. The university supports tokens, but requires students to purchase their own. The primary 2FA used by the university is the phone app Duo for one-time SMS codes. The security keys removed the dependency on the phone but the participants still perceived they cannot use their account if the key is lost.

We conducted a follow-up survey that had low participation, so it was not coded. Yet we can note some qualitative responses, and that in Phase-II, five of ten participants reported continued use of the key on the survey. One participant stated, *"It's a convenient way of utilizing two-factor authentication without needing to look at my mobile device"*. However, lack of participation in the follow-up study can be an indicator of lack of engagement with the token. Another participant noted, *"I find it much easier to use on Duo instead of using my phone. I just keep it in my laptop case."* indicating acceptance and continued use.

Table 2 shows that the password usage behaviour was not significantly different. Authentication patterns in the populations were similar. However, there was a change - an increase in the use of password managers as opposed to writing passwords on paper. The participants who indicated that they used password managers primarily used the in-browser password manager. In terms of convenience, writing on paper seems to have lost out to the password management functionality in browsers.

Password Behaviour	Phase-I	Phase-II
Same password for every website	5.3%	0%
A few passwords I use interchangeably	84.2%	61.8%
One password that I use for important sites and another password I use for less important sites	21.1%	17.6%
Different passwords for each site	26.3%	26.5%
Web browser's password manager to store my passwords	52.6%	41.8%
Write passwords down on paper	15.8%	0%
A program to store my password	5.5%	29.4%

Table 2. Comparison of Phase-I and Phase-II Password Behaviour

As with Phase-I, the halt points were not significantly correlated with the reported security behaviors so no statistics are reported.

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 a benefit. When asked about continued use, one participant said, *"No, my password is secure enough and alerts are active."* Acceptability can be improved if the interaction communicates the intrinsic benefit of risk mitigation, or alternatively communicates the intrinsic risk of account takeover.

Google continues to require use of the full password even with security keys. As a result there was no cognitive benefit for using the device. As an annoyed participant queried in Phase-II, *"Why is it still asking for a password?"*. One participant stated that there would be a benefit if there were no password requirement. At this University the password requirements are 15 characters, one uppercase letter, one lowercase letter, and one non-letter (which can be a number or control character). The core challenge of *hard to remember, multiple, long passwords* remains. Adopting organizations can improve acceptability by creating a cognitive benefit by design.

Figure 10. Security key compatibility with different websites

In Phase-I, many participants thought that they required different tokens for different websites. In Phase-II, the participants had better knowledge of how the security key worked and found them to be more acceptable, after looking at the potential benefits of using the same key for 2FA across different websites. The instructions in the Phase-II 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 participants pointed out that multiple platforms could be linked and secured by security key as shown in Figure 10. One part of communicating benefits is highlighting the privacy-providing and applicability of the token. A major concern of many participants is denial of access to the account. Mitigating that concern is another possible enhancement to acceptability.

9 Discussion and General Suggestions

Our evaluation examined usability and acceptability of the security key as informed by previous evaluations. Specifically, we were informed by previous work [14, 30] which proposed the following six heuristics for usable, acceptable secure systems: i) installation precedes operation; ii) ensure users are aware of trade-offs; iii) say why, not how; iv) awareness of needed security tasks is required; v) can perform these tasks without making dangerous errors; and vi) interactions result in continued use of the security technology. The importance of *installation precedes operation* was reiterated in both phases of the experiment. Installation must be easy and usable for adoption.

Based on feedback from the participants, we have found that risk awareness, knowledge of potential benefits, and cognition about the importance of passwords were the critical factors determining the acceptability of the security key. Thus the need to *ensure users are aware of trade-offs* was supported by this work, despite the fact that we were not testing an anonymity technology per se. The interaction did not result in continued use of the technology, in part because participants did not see a clear benefit. This also impinges the *say why, not how.*

Although we do not report on the pilot in formal terms, it is worth noting that there was notable divergence between the members of the security lab and research participants. To begin with, all security lab members who were provided a security key continued to use it. None were turned in. The acceptability differences were stark. None of the members of the security lab chose the option to 'trust the computer'. This meant that confirmation of operation was simple, the security lab member logged out, then logged in, and this required use of the key. Many lab members also browsed in incognito mode by default, and thus verification was trivial. In contrast, the research participants chose to trust the computer. These differences in behavior may have been present in the developer population for the security key, given the predictions of perfect usability. [31]

Our findings do not perfectly align the requirement that *awareness of needed security tasks is required.* Participants were already well aware of the need to authenticate to their accounts. Thus we cannot assert that participants did not know the task is required, rather they seemed unaware of the existence of the risks. This, and the need for knowledge of trade-offs, both indicate that there would be benefits to integrating risk communication during installation and use. Unfortunately, our possible ideas about risk communication at this point could not be tested in the scope of this experiment. Security messages during installation were recommended. One was to send a text message of positive congratulations, "Thank you for using security key! You now do not need to enter a code to be secure." This would be useful for the workplace to indicate a benefit. Security benefits can be communicated by trusted hardware prompts, congratulatory messages of secure initial login, and periodic reminders like *'Only this Computer Can Login Without Your Key'.* Users can also confirm such benefits when unsuccessful login attempts made from unregistered or remote computers are communication, demonstrating their triumph over potential attacks. Benefit communication through suggestions such as *'Make 2-Step Easy'* with reference to Figure 5 for different security keys improve acceptability.

It clear that in Phase-II the participants were capable and *can perform the 2FA authenticating task without making dangerous errors.* It was also clear that none would choose to do so without these being mandated.

Not having to remember a password was a specific benefit sought by multiple participants. In fact, for 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. [8, 9] 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.

To the extent that two-factor authentication remains optional, it would seem to make sense to offer the U2F token as a single factor even on untrusted devices. Such *interactions would result in an increase in continued use of the security technology.* Users would have the usability benefit of a cognitively-effortless single factor, while still retaining significant security benefits. While a lost key would provide full account access, tokens are substantially harder to steal than passwords and the requirement that the attacker know the associated account provides the potential to identify the misuse as attack on the server end. Stealing a token requires physical access rather than cheap scalable methods such as malware or phishing. U2F tokens also provide a measure of mutual authentication, since an illegitimate site cannot generate a correct U2F challenge.

Using the security token as a primary authentication factor also offers accessibility benefits. For enterprise customers, this 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 passphrases. Although still an unusual complaint, an ADA compliance issue could arise in the face of password complexity requirements. As noted in our future work section, we seek to include older adults to determine if this single-factor authentication is desirable by those participants.

10 Conclusion

From the survey results of Phase-I, it was found that most users were uncomfortable with the usability of the FIDO security key. They indicated confusion about the demo, the hardware's form-factor, setup validation and the end-user security benefits as major stop points in using the FIDO Key. Thus, about 33% of the users were unable to complete the registration process in the first Phase. There was a significant increase in usability, but we cannot assert any corresponding increase in acceptability. Our results tell a different story from the glowing quantitative records of performance observed in enterprise contexts where adoption is mandatory.

However, notable improvements were observed in the second Phase when the usability issues were resolved. Updated and clear installation instructions during setup, including removal of the demo, made Yubico Security keys more usable. The registration process was also made to be more descriptive. As a result everyone was able to complete the registration process in Phase-II. To few participants remained confused at the end of Phase-II because the confirmation message showing successful registration was absent.

A fair number of participants continued to believe in the superiority of passwords over the FIDO Security key after the survey was completed. They also demonstrated greater faith in their own security acumen, which led us to conclude that acceptability of the Yubico key did not result from improvement in usability. Thus, it is evident that even if the best design practices are kept in mind to improve the usability of the 2FA token, the benefits have to be made apparent to the users for it to be widely used.

Future studies could include a range of tokens, not only other Yubico security keys such as, Yubikey 4, Yubikey 4 Nano, Yubikey 4C, Yubikey 4C Nano, Yubikey NEO, 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.

11 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] Dorothy E Denning and Peter F MacDoran. Location-based authentication: Grounding cyberspace for better security. *Computer Fraud & Security*, 1996(2): 12–16, 1996.

[2] John Brainard, Ari Juels, Ronald L Rivest, Michael Szydlo, and Moti Yung. Fourth-factor authentication: somebody you know. In *13th ACM Conference on Computer and Communications Security*, pages 168–178. ACM, 2006.

[3] Scott Ruoti, Jeff Andersen, and Kent E Seamons. Strengthening password-based authentication. In *WAY@ SOUPS*, 2016.

[4] Troy Hunt. Password reuse, credential stuffing and another billion records in have i been pwned.

[5] Joseph Bonneau, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. Passwords and the evolution of imperfect authentication. *Communications of the ACM*, 58(7):78–87, 2015.

[6] Sampath Srinivas, Dirk Balfanz, Eric Tiffany, and FIDO Alliance. Universal 2nd factor (u2f) overview. *FIDO Alliance Proposed Standard*, pages 1–5, 2015.

[7] Brett McDowell. Strong authentication canine. Cloud Identity Summit, June 2015. URL https://www.youtube.com/watch?v=sdJ47NFGlgk.

[8] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *IEEE Symposium on Security and Privacy*, pages 553–567, May 2012.

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

[10] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security keys: Practical cryptographic second factors for the modern web, 2016.

[11] 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.

[12] Lujo Bauer, Lorrie Faith Cranor, Robert W Reeder, Michael K Reiter, and Kami Vaniea. A user study of policy creation in a flexible access-control system. In *Conference on Human Factors in Computing Systems*, pages 543–552. ACM, 2008.

[13] Robert W Reeder and Roy A Maxion. User interface dependability through goal-error prevention. In *Dependable Systems and Networks, 2005*, pages 60–69. IEEE, 2005.

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

[15] Philip G Inglesant and M Angela Sasse. The true cost of unusable password policies: password use in the wild. In *Conference on Human Factors in Computing Systems*.

[16] Sacha Brostoff, Philip Inglesant, and M Angela Sasse. Evaluating the usability and security of a graphical one-time pin system. In *24th BCS Interaction Specialist Group Conference*, pages 88–97. British Computer Society, 2010.

[17] Greg Norcie, Jim Blythe, Kelly Caine, and L Jean Camp. Why johnny can't blow the whistle: Identifying and reducing usability issues in anonymity systems. In *Proceedings 2014 Workshop on Usable Security. https://doi. org/10.14722/usec*, 2014.

[18] Eirik Albrechtsen. A qualitative study of users' view on information security. *Computers & security*, 26(4):276–289, 2007.

[19] 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.

[20] V. Garg and J. Camp. Heuristics and biases: Implications for security design. 32 (1):73–79. ISSN 0278-0097. doi: 10.1109/MTS.2013.2241294.

[21] Ryan West. The psychology of security. 51(4):34–40. URL http://dl.acm.org/citation.cfm?id=1330320.

[22] Prashanth Rajivan, Pablo Moriano, Timothy Kelley, and Jean Camp. What can johnny do?–factors in an end-user expertise instrument. In *Tenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2016)*, page 199. Lulu. com.

[23] Compare yubikeys — strong two-factor authentication for secure logins. Top Applications of Yubico, June 2017. URL https://www.yubico.com/products/yubikey-hardware/compare-yubikeys/. [Online; accessed 3-Oct-2017].

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

[25] Serge Egelman, Marian Harbach, and Eyal Peer. Behavior ever follows intention?: A validation of the security behavior intentions scale (sebis). In *2016 Conference on Human Factors in Computing Systems*.

[26] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. 347(6221):509–514. URL http://science.sciencemag.org/content/347/6221/509.short.

[27] R. Anderson. Why Information Security is Hard - an Economic Perspective. In *Computer Security Applications Conference*. IEEE, 2001.

[28] Tyler Moore, Richard Clayton, and Ross Anderson. The economics of online crime. *The Journal of Economic Perspectives*, 23(3):3–20, 2009.

[29] Use security key for 2-step verification - android. URL https://support.google.com/accounts/answer/6103523?hl=en&ref_topic=6103521.

[30] Greg Norcie, Jim Blythe, Kelly Caine, and L. Jean Camp. Why johnny can't blow the whistle: Identifying and reducing usability issues in anonymity systems. Internet Society. ISBN 978-1-891562-37-2. doi: 10.14722/usec.2014.23022.

[31] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security keys: Practical cryptographic second factors for the modern web. In *Financial Cryptography and Data Security*. International Financial Cryptography Association, February 2016. URL http://fc16.ifca.ai/preproceedings/25_Lang.pdf.