Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling

Stéphane Ciolino, ***OneSpan Innovation Centre & University College London;***
Simon Parkin, ***University College London;*** **Paul Dunphy,** ***OneSpan Innovation Centre***

https://www.usenix.org/conference/soups2019/presentation/ciolino

This paper is included in the Proceedings of the Fifteenth Symposium on Usable Privacy and Security.

August 12–13, 2019 • Santa Clara, CA, USA

ISBN 978-1-939133-05-2

Open access to the Proceedings of the Fifteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling

Stéphane Ciolino
OneSpan Innovation Centre & University College London
stephane.ciolino.17@ucl.ac.uk

Simon Parkin
University College London
s.parkin@ucl.ac.uk

Paul Dunphy
OneSpan Innovation Centre
paul.dunphy@onespan.com

Abstract

Security keys are phishing-resistant two-factor authentication (2FA) tokens based upon the FIDO Universal 2nd Factor (U2F) standard. Prior research on security keys has revealed intuitive usability concerns, but there are open challenges to better understand user experiences with heterogeneous devices and to determine an optimal user experience for everyday Web browsing. In this paper we contribute to the growing usable security literature on security keys through two user studies: (i) a lab-based study evaluating the first-time user experience of a cross-vendor set of security keys and SMS-based one-time passcodes; (ii) a diary study, where we collected 643 entries detailing how participants accessed accounts and experienced one particular security key over the period of one week. In the former we discovered that user sentiment towards SMS codes was typically higher than for security keys generally. In the latter we discovered that only 28% of accesses to security key-enabled online accounts actually involved a button press on a security key. Our findings confirm prior work that reports user uncertainty about the benefits of security keys and their security purpose. We conclude that this can be partly explained by experience with online services that support security keys, but may nudge users away from regular use of those security keys.

1 Introduction

User authentication mechanisms are based on one of three factors: *knowledge* (e.g., password), *ownership* (e.g., token), or *inherence* (e.g., fingerprint). Combining factors (e.g., passwords and tokens) is widely recognized as an effective technique to protect both corporate and personal online accounts against account hijacking threats. Indeed, there are already examples of citizens being advised to use Two-Factor Authentication (2FA) by government agencies (as in the UK [35]). The most common second factor is a One-Time Passcode (OTP) received via a text message to a mobile device [5]. While this technique conveniently leverages pre-existing telecommunications infrastructure and mobile devices that users already own, it is vulnerable to Person-In-The-Middle (PITM) attacks [8, 23, 32], such as phishing attacks. Hardware-based authentication tokens have historically been deployed for 2FA in closed user communities such as corporations, or for banking customers, particularly in territories such as Europe.

Recently, efforts have gathered pace to position tokens as a general purpose second factor for end-users to secure a range of online accounts. The Fast IDentity Online (FIDO) Alliance was founded in 2012 to reduce reliance on passwords for Web authentication by moving to new authentication standards underpinned by public key cryptography that are resistant to phishing [2]. *Security keys* are commercially available authentication tokens based on the Universal 2nd Factor (U2F) standard created by FIDO [3]. They are currently supported by more than 30 Web service providers [47] including Dropbox, Facebook, GitHub, Google, and Twitter.

However, there are barriers to the widespread uptake of security keys for 2FA. These include the need for an improved setup process [16], and inherent concerns about losing the devices [1, 38]. In the broader debate, there are mixed views about the decisive factors that might influence the uptake of security keys for everyday use. Research has shown that some users struggle to see the utility in security keys [16], yet other work reports user satisfaction with the devices [38].

In this paper, we aim to further understand user perceptions of utility and security of the security keys. We firstly compared the setup and login experience for three cross-vendor security keys: Feitian ePass FIDO® NFC (ePass), OneSpan DIGIPASS SecureClick (SecureClick), Yubico YubiKey 4 Nano (YubiKey); and SMS-based OTPs. Participants used

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

each mechanism on two representative Web services: Gmail and Dropbox. We discovered that the security keys generated diverse usability issues and that the Web service user interface could also impact the efficiency of the setup process. We built on the lab-based study with a week-long diary study of one specific security key (the SecureClick) involving fifteen participants, each using a SecureClick with free choice of their online accounts. We collected 643 diary entries and found that participants only used security keys in 28% of logins that *could have* used a security key. Also, we found that button presses on the security key decreased by 50% from the first day of the study to the last.

The rest of the paper is arranged as follows. Related Work is discussed in Section 2. We describe the protocol and results of our laboratory study in Section 3. The protocol and results for the diary study are described in Section 4. We close the paper with Discussion and Conclusion (Sections 5 and 6 respectively).

2 Related Work

2.1 2FA Mechanisms

Research is increasingly exploring the usability and security of tokens that provide 2FA based on a pre-shared secret between the token and the service provider. De Cristofaro et al. [18] surveyed the usability of various forms of 2FA: codes generated by dedicated tokens or smartphone apps (e.g., Google Authenticator), or received via email or SMS. Respondents' perception of 2FA usability correlated with distinguishing characteristics of each mechanism; the actual 2FA technology deployed or the context of use had less of an influence. Three metrics were argued to be necessary for rating the usability of 2FA technologies: ease-of-use; required cognitive effort; and trustworthiness of the device.

Other studies have examined 2FA in the context of online banking (e.g., [7, 30, 44, 45]). Weir et al. [44] studied three different tokens used by banking customers and found that participants preferred those with higher perceived convenience and usability, at the expense of perceived security. A follow-up study [45] contrasted password-based authentication against token- and SMS-based 2FA, finding that convenience, personal ownership, and prior experience were key factors in selecting an authentication mechanism. Krol et al. [30] report that the mental and physical workload required to use tokens influenced user strategies for accessing online banking (e.g., how often they would be willing to log in). Althobaiti and Mayhew [7] conducted an online survey across students studying abroad, identifying higher perceived usability for tokens over SMS-based authentication.

Weidman and Grossklags [43] examined the transition from an authentication token to a 2FA system using DuoMobile on employees' personal mobile devices within an academic institution. Users rated the DuoMobile solution more negatively compared to the token-based solution, as users resented using their personal mobile devices in a work context.

Gunson et al. [25] recruited banking customers to contrast knowledge-based one-factor authentication (1FA) and token-based 2FA for automated telephone banking. No single 1FA or 2FA approach stood out as a preferred authentication method. However, a trade-off between usability and security was identified, with 1FA judged more usable but less secure than 2FA. Sasse et al. [40, 42] examined authentication events involving passwords and RSA tokens in a US governmental organization – authentication disruptions reduced staff productivity and morale, to the extent that work tasks were arranged to minimize the need to authenticate.

2.2 FIDO U2F and 2FA Security Keys

Lang et al. [32] applied the usability framework established by Bonneau et al. [11] to assess a range of security keys, alongside authentication activity data from Google. The authors identified that security keys evidenced quicker authentication and fewer support incidents in a work environment, as compared to alternative tokens. Das et al. [16, 17] conducted a two-phase laboratory study with students, to improve the usability of setting up a Yubico security key with Gmail. Participants did not perceive benefits to using the Yubico security key in their everyday lives and were most concerned with the potential of losing access to their account. Colnago et al. [15] examined the adoption of Duo 2FA at a university. Security keys were one of four 2FA options offered to users, with less than 1% choosing this option. Reynolds et al. [38] explored usability perceptions of Yubico security keys during enrolment, and in everyday use (by way of a diary study). They found that participants experienced problems to set up the security keys with services but perceived them as usable for regular activities. As also found in the work by Das et al. [16, 17], losing the security key was also highlighted as a concern.

2.3 Open Challenges

Authentication tokens have been prevalent for many years in closed and centralized deployments, e.g., in workplaces or for individual banking services in some countries. These represent service-centric technologies which are centralized and orchestrated by the service provider. Security keys are intended to support *user-centricity* [10] which is a term that has specific connotations in digital identity of: decentralization, user control, selective disclosure, and interoperability. With security keys, this user-centricity is achieved through public key cryptography: the security key can generate private keys that are stored confidentially on the device and can create digital signatures that may be shared with a service provider to attest to ownership of a given public key.

Adoption of security keys has eradicated account takeover at Google [28]. However, user adoption of security keys more generally is low, with evidence that 1% of observed logins across one entire user population leveraged security keys [21], and 1% of users in a sample at a university were using these devices [15]. Furthermore, there are strong technical arguments against the use of other more popular 2FA mechanisms today due to the threat of person-in-the-middle-attack, particularly SMS-based OTPs [6, 27, 48]). The threat against SMS-based OTPs has taken on a new dimension in recent times due to the emerging prevalence of mobile device SIM swap attacks [46].

Research up to now has been valuable to provide an early understanding how the form and function of security keys themselves impact adoption, but it has limitations: the lab work of Das et al. [16, 17] wholly focused on the setup of one YubiKey with Gmail; Reynolds et al. [38] conducted a between-subject lab study for device setup with one YubiKey and a diary study limited to Gmail, Facebook and Windows 10 with a YubiKey that did not capture specific login events; the main insight about security keys in the work of Colnago et al. [15] is that they were rarely adopted. While there is no single answer to the low adoption of security keys, we were interested in obtaining a new perspective on the user experience of security keys and compatible online services in everyday Web browsing, driven by the following research question: *Are there 2FA usability issues that security keys perpetuate, or new issues that they introduce in an everyday context, at setup, login, and service use?*

Our work, presented in the following sections, contributes to the state of the art in the following ways: (i) it compares several 2FA mechanisms with each other; (ii) the diary study accommodates free choice of Web services and captures daily interaction data; and (iii) it begins to shed light on findings from prior work why users might fail to see a benefit in the use of security keys.

3 Lab-based Usability Study

We conducted a lab-based study in August 2018, to capture the main factors that affect the usability and security perceptions of security keys at the setup and login phases of use.

3.1 Method

We conducted a within-subjects research lab-based study to compare several 2FA methods directly against each other in a way that maximizes the number of data points collected per participant. We recruited 15 participants via flyers posted on the university campus. Convenience sampling was employed, with no pre-screening applied. Each lab session lasted approximately 45 minutes and involved a series of tasks and a debrief discussion to be completed. Participants received a complimentary £10 Amazon voucher upon completing the study.

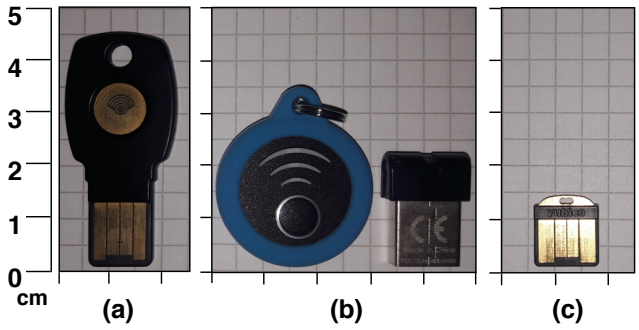

Figure 1: Visual comparison of each security used in the laboratory study: (a) ePass, (b) SecureClick, and (c) YubiKey.

The study required participants to use each of the following four 2FA mechanisms:

- ePass security key
- SecureClick security key
- YubiKey security key
- SMS-based OTPs

The security keys, shown in Figure 1, were chosen as the manufacturers (Feitian, OneSpan, and Yubico) are part of the FIDO Alliance board, and the devices themselves are diverse in form factor. The ePass and SecureClick are dual-mode devices that work with both laptops/desktops, via USB, and mobiles, via Near Field Communication (NFC) or Bluetooth Low Energy (BLE). Our study wholly focused on laptop/desktop usage. With the ePass and YubiKey, users need to plug the security key into a USB port and press the button/handle on the device to execute its functionality. The SecureClick comes in two parts, with a USB Bluetooth Bridge (as in Figure 1), requiring installation of a browser extension to link the Bluetooth Bridge and SecureClick before first use on laptops/desktops; users then only need to plug the Bluetooth Bridge into a USB port and press the button on the SecureClick itself. The study also included SMS-based OTPs since the mechanism is typically present in a 2FA choice architecture [37] competing with other methods of 2FA, and may affect users' perception or preference towards security keys.

Participants used each authentication mechanism with one of two mainstream Web services that support the above 2FA techniques: Dropbox or Gmail. We randomly assigned eight participants to use Dropbox while we assigned Gmail to the other seven participants. An earlier pilot test informed the decision to focus each participant on one of the Web services rather than both, to reduce the risk of fatigue. Participants used email accounts created solely for the study, with one per Web service (Dropbox, Gmail) and 2FA method (SMS-based OTPs, ePass, SecureClick, and YubiKey). We chose the

usernames to be easy to recall and type for the participants, and the password was the same for all of the accounts.

An abstraction of the 2FA setup and login processes for both Web services are illustrated in the Appendix, in Figure 5. This diagram relates the technical mechanisms and activities under observation [41] to user experiences and perceptions emerging from the studies (revisited in the Discussion in Section 5).

3.1.1 Procedure

We followed the procedure below during the lab study:

1. **Preparation:** The participant sat at a desk in the laboratory room (the same desk for each session). The experiment moderator followed a script to explain the study to the participant. We provided an information sheet and consent form to the participant, who was then allowed time to read the forms before providing their consent. The participant would be encouraged to *think aloud* during the subsequent tasks [14].

2. **Testing Different 2FA Methods:** The main part of the laboratory session consists of four 2FA tasks. Each task has a 'set-up' phase (for enrolling a 2FA mechanism with a Web service), and a 'login' phase (using the 2FA mechanism for login with the Web service):

   - *Task A:* 2FA using SMS-based OTPs.
   - *Task B:* 2FA using ePass.
   - *Task C:* 2FA using SecureClick.
   - *Task D:* 2FA using YubiKey.

   The instructions we gave to participants are detailed in Appendix A. We varied the order of tasks A, B, C, and D across participants to minimize ordering effects on participant preference and behavior. The participants also completed a System Usability Scale (SUS) assessment of the technology immediately after each task.

3. **Debrief:** After the structured tasks, the researcher debriefed the participant in a semi-structured interview, shared the study goals, and encouraged a focused discussion. Debrief questions explored issues around 2FA, participant satisfaction/dissatisfaction with security keys, and perceptions of where security keys could be useful (or not).

3.1.2 Test Equipment

Participants performed all tasks on a Dell Latitude E5540 laptop using the Windows 7 operating system, the Google Chrome browser, and a Motorola XT1100 Nexus 6 mobile phone (for SMS-based OTPs). We used a voice recorder to capture 'think-aloud' responses and the debrief dialogue, to facilitate transcription at a later stage. Interactions with the Web services were also video-recorded for timing purposes, recording only the laptop screen and page/screen transitions (not the participant).

3.1.3 Research Ethics

The study was approved through the sponsor university's IRB-equivalent research ethics committee, Project ID 5336/010, and raised no specific cited concerns nor requested corrections. After we completed the study, we thoroughly debriefed participants and compensated them immediately for their time.

3.1.4 Demographics

We recruited fifteen participants for our study (6 female, 9 male). The ages of the participants were between 21- and 37-years old (median 25.5). Eight were postgraduate, three were graduate, and four were up to undergraduate level. Nine already had experience of using 2FA (either SMS-based OTP or mobile-based authentication app). None had any familiarity with security keys.

3.1.5 Limitations

Participants' behavior and views of the authentication mechanisms may have been shaped by the laboratory conditions (controlled to uphold internal validity [31]). Furthermore, the lab study did not present a real risk to the participants' personal data (where this is addressed in the diary study, Section 4), and required participants to use machines and accounts provided by the experimenters. Participants were comparing a relatively new authentication method (security keys) to a well-known incumbent (SMS-based OTPs), where evaluating a security technology against others in the same session has the potential to encourage more critical feedback [29]. Although our sample of 15 participants is above the recommended minimum of 12 participants to achieve data saturation [24] (achieved after 11 interviews in our case), the sample could be considered as modestly sized. We aimed to mitigate this concern in this study by capturing a detailed range of data points with our within-subjects study design and debrief interviews.

3.2 Results

The following sections present quantitative results (Sections 3.2.1 to 3.2.3) and qualitative results (Sections 3.3.1 to 3.3.3) pertaining to our research question.

3.2.1 Phase 1: Setup

We captured critical events that prevented participants from progressing with a task (without further assistance), and present a timing analysis of setup interactions with each

Issue	Source	Severity	Frequency
Bluetooth pairing errors on SecureClick	Device	Major	12
Generally unsure how to achieve their goal based upon available instructions	Web Service & Device	Major	12
Confusion due to SMS setup brought to the fore before mention of security key	Web Service	Major	6
Unsure whether to allow Chrome browser to see make and model of security key	Browser	Minor	6
Animated circle misinterpreted as loading by users	Web Service	Major	5
SMS-based OTP not received to set up secondary authentication	Web Service	Major	2
Unable to locate the button on a specific device	Device	Major	2
Inserting the YubiKey the wrong way up	Device	Major	2

Table 1: Issues encountered when setting up 2FA technologies with a Web service, alongside their severity and frequency of occurrence. All but one issue is rated as having 'Major' severity.

security key on each Web service.

Usability Roadblocks: Table 1 lists the most common roadblocks that users encountered during the setup of the security keys. We use the Nielsen rating system to categorize the severity of those usability roadblocks [36]. 'Major' usability problems may cause a lot of confusion or result in the incorrect use of the system, whereas 'Minor' usability problems indicate a delay or inconvenience in the completion of a task.

Participants generally needed guidance to pair the SecureClick with the Bluetooth Bridge, citing a specific need for more clarity in the instructions and error messages displayed.

Twelve participants needed guidance to navigate the Dropbox and Gmail Web service interfaces to activate 2FA. One crucial issue was that both of the Dropbox and Gmail interfaces prioritized the process with the activation of SMS-based OTPs. The option to use a security key was not salient to participants who were unsure if they were on the correct path to activate a security key. Once participants had finally discovered the correct option (by ignoring their initial intuition), several minor user interface design issues disrupted the user journey. First, both services displayed an animated spinning circle while waiting for the user to touch the button on their security key after users inserted it into the USB port. At this point, we saw participants conclude that the website was loading rather than prompting for the button to be pressed on the security key. Five participants specifically asked for help as to whether they were required to do anything at that stage.

Another issue was that once users pressed the button on their security key, a pop-up window appeared in the browser asking the user to confirm that the Web service had permission to 'See the make and model of your security key'; six users were unsure whether to allow this since they were not forewarned that it would occur, and weren't sure if this option would breach their privacy beyond the basic use of the security key itself.

Learnability and Efficiency: We used the video recordings to retrospectively measure setup timings for the 2FA techniques on each online service. The measurement started when participants accessed the login page of the Web service and ended when participants viewed confirmation from the Web service that the 2FA setup was complete. The timings to set up 2FA with different mechanisms and Web services are shown in Table 2.

The median time to set up the ePass was 2min 29s and 2min 49s on Dropbox and Gmail respectively.

The median time to set up the SecureClick was 2min 23s and 2min 25s on Dropbox and Gmail respectively. Also, there was a one time process required to download the DIGIPASS SecureClick Manager and pair the SecureClick with the Bluetooth Bridge (median time was 3min 06s).

The median time to set up the YubiKey was 5min 20s and 2min 06s on Dropbox and Gmail respectively. The timings on Dropbox were impacted by device form factor and user interface issues: participants were confused about the location of the button on the YubiKey (7 participants); had to be guided through the Dropbox user interface (4); inserted the YubiKey the wrong way around (1); didn't receive the SMS-based OTP and had to restart the process (1).

The median time to set up SMS-based OTPs was 2min 33s and 1min 41s on Dropbox and Gmail respectively.

We had no a priori hypotheses about significant interactions that could emerge between the devices and services. However, we noted patterns in the data that led us to conduct post hoc analysis. A Kruskal-Wallis test uncovered significant differences in the setup times, considering the specific Web service and device as factors: $\chi^2 = 18.0366$ $p < 0.05$. Pairwise comparisons yielded significant differences between (i) YubiKey on Dropbox and YubiKey on Gmail ($p < 0.05$); (ii) YubiKey on Dropbox and SMS-based OTPs on Gmail ($p < 0.01$). The p-values included Bonferroni Correction for multiple comparisons.

3.2.2 Phase 2: Login

As with the setup phase, we used the video recordings to measure 2FA login timing. We started the measurement when

	Dropbox		Gmail	
2FA Method	**Setup**	**Login**	**Setup**	**Login**
ePass	149 (76)	22 (11)	169 (99)	24 (11)
SecureClick	143 (80)	28 (20)	145 (92)	33 (8)
YubiKey	320 (139)	29 (15)	126 (42)	25 (7)
SMS OTPs	153 (116)	50 (20)	101 (43)	38 (21)

Table 2: Median timings (and interquartile range) in seconds for each 2FA method and each Web service tested. Timings rounded to the nearest integer.

participants accessed the login page of the Web service and stopped once the participant had successfully logged in.

The timings for participants to perform 2FA login with different mechanisms and Web services are shown in Table 2. Excluding outliers, it seems that logging in using security keys is faster than using SMS-based OTPs. This is presumably due to the time the user must wait to receive the SMS-based OTP and then type it on the user interface.

The median time to perform 2FA login using the ePass was 22s and 24s on Dropbox and Gmail respectively.

The median time to perform 2FA login using the SecureClick was 28s and 33s on Dropbox and Gmail respectively. The timings were impacted by one participant holding the button down for too long on the SecureClick, and another waiting idle before pressing the button.

The median time to perform 2FA login using the YubiKey was 29s and 25s on Dropbox and Gmail respectively. The timings were impacted by three participants waiting some time before touching the handle on the YubiKey.

The median time to perform 2FA login using the SMS-based OTPs was 50s and 38s on Dropbox and Gmail respectively.

3.2.3 2FA SUS Scoring

Participants completed an SUS rating scale after each of the four tasks. 2FA using SMS-based OTPs, the ePass and the YubiKey were all deemed 'acceptable' [9] with a mean score of 85.17 ($SD = 8.37$, 95% $CI = \pm 4.24$), 80.5 ($SD = 19.58$, 95% $CI = \pm 9.91$) and 73 ($SD = 28.16$, 95% $CI = \pm 14.25$) respectively. The SecureClick had a mean score of 61.5 ($SD = 22.93$, 95% $CI = \pm 11.60$) which is deemed 'marginal'. The distribution of the SUS scores for each mechanism are illustrated in Figure 2.

3.3 Qualitative Results

The data was analyzed using thematic analysis [12]. All participant responses were coded in several stages by one researcher, initially as low-level labels, moving to higher-level analytical categories. We identified seven high-level categories (with the three most prominent presented in the following sub-sections),

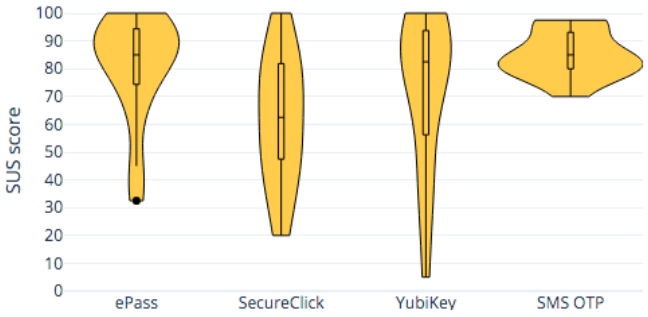

Figure 2: System Usability Scale (SUS) scores for each method of 2FA.

and 31 sublevel codes. Analysis is presented here, including notable quotes (we refer to individual participants using PL##).

3.3.1 Effort, Convenience, and Fearing the Worst

Nine participants were concerned about being locked out of their accounts if using 2FA, should they not be able to present the security key or SMS-based OTP. One solution proposed to mitigate this problem was keeping a backup security key (PL06): *"so like when you get a car you get a spare key, this is something that you would need for me I think, as a backup if you lose one."*

Indeed there was an intuitive awareness that account lockout and recovery issues are easier to resolve with SMS-based OTPs. Participants also expressed being comfortable with 2FA using SMS-based OTPs, with one participant in particular (PL07) believing this to be an appropriate approach for people who are not *"tech-savvy."* Delays in receiving an SMS-based OTP, or having to swap SIM cards or enable roaming while traveling abroad, were issues voiced about SMS authentication.

It was widely recognized among participants that security keys removed the wait that is inherent to the delivery of SMS-based OTPs; PL04: *"much more efficient than codes verification and stuff like that," "it kind of removes all that leg work," "you can kind of just tap in and it's done."* However, security keys are not as versatile as mobile devices and comprise an extra object that users would have to procure, protect and carry around, as noted by PL13: *"Nowadays everyone has a phone that you carry around with you, for me to carry an extra piece which is this, it doesn't have any function other than just stick in into a computer. I mean a phone is something you need, you need it to call, it has multi-functions, it is a multi-functional thing."* The use of SMS-based OTPs (PL14) *"works well since it is using your phone, it's not something that requires an extra piece of equipment or hardware."*

One participant (PL08), who reported only using password-based 1FA, generally sees 2FA as an extra step slowing down

every single login process: *"I'm bound to using the two-step every time I want to log in, again it adds on a few extra seconds to the login process."* Others had encountered issues setting up 2FA on other online services which have created a negative perception of all 2FA procedures generally, e.g., PL09: *"I try to avoid two-step verification because I once did it as my Apple and then it got really messed up, so it's a bit hard because my phone didn't get the text so I disabled it, so just to avoid that I don't do it."*

Seven participants mentioned that having to pair the SecureClick with its Bluetooth Bridge was a drawback, e.g., PL03: *"I need to install things before I get to use it, for the moment I seem to be able in most devices."*

3.3.2 Size Matters: Loss, Breakage and Design Choices

The form factor of specific devices influenced perceptions of usability. Thirteen participants commented on the unusually small size of the YubiKey, e.g., PL12: *"this one is more discrete, you can't really see it,"* with five expressing concern as a result, e.g., PL06: *"I would lose it, or I'd forget it because I would forget that I plugged it in a desktop computer because I can't see it."* Some saw this as a potential security threat, fearing that if the security key were always plugged in then an attacker could also use it.

Conversely, some participants equated a more substantial form factor with an increase in usability. Six participants commented on this aspect regarding the ePass, e.g., PL06: *"I would feel better about using it because it's like a USB."* Greater size, however, fuelled concerns about breaking the device, as it protrudes from the USB port, e.g., PL11: *"I did feel kind of like it could snap."*

The security keys all rely on a single touch-button interaction. However, this simple format created challenges for participants; all but one participant (PL10) failed to realize that the gold area on the YubiKey was the 'button' or 'gold disk' referred to by the Dropbox and Gmail user interface. In comparison, only a few participants failed to notice the 'button' on the ePass, e.g., PL02: *"because I just didn't see it as a button, it's flat."*

The form factor of the security keys also informed perceptions of when they could be used. Participants generally saw the ePass device as well suited to be carried around, whereas the SecureClick and the YubiKey devices were judged to be better suited to be attached to one computer.

3.3.3 Rationalizing the Security Benefits of 2FA

Seven participants perceived that security keys provide additional security, but experienced challenges to articulate exactly how they provided added protection, and the threat they protected against, e.g., PL07: *"[it's] like an added protection, basically it is trying to identify it is you that is opening that account."* Only two participants recognized that security keys primarily defend against phishing attacks. One participant (PL03) perceived that having no visible association between devices and Web services adds further security, as opposed to for example bank tokens that are branded and thus more vulnerable to attacks: *"The issue [with bank tokens] is that it's all branded and everything so if it gets stolen, someone who's really desperate to have it work for him can actually do that, and for what I've seen with this, yes it's kind of branded but I could easily fit this in to my key holder and only I know what it's for."*

A few participants argued against the security provisions of security keys, for instance conveying that SMS-based OTPs were just as secure, but furthermore that Web service providers (such as Gmail and Yahoo) send real-time email notifications of any suspicious activity on the user account, e.g., PL07: *"[Gmail and Yahoo] send me a code and I have to log in and then they also send me 'You logged in from another device', so I guess because that happens automatically, I don't really have to bother myself, and then I feel a lot more secure when it happens."* One participant (PL04) added that locking security keys with biometric authentication would make it comparable to modern smartphones, for instance *"in case it gets lost [...] you kind of have that biometric control and power."*

4 Diary Study

To examine the fit of security keys with users' everyday practices, we conducted a diary study in January-February 2019. We focused specifically on the SecureClick security key since we were more knowledgeable about this device than the others, and could better support participants during day-to-day use (also discussed in Section 4.1.2). By asking participants to link a security key with personal online accounts, we hoped to capture more realistic usage data than a lab study could provide [31]. We chose a study time period of one week in order to minimize the burden on participants and hence an adverse effect on participation [33, 39].

4.1 Method

4.1.1 Procedure

We recruited fifteen participants for the diary study, via a flyer/advert and electronic newsletters distributed across the sponsor university campus, and flyers shared with a nearby partner university. We also advertised the study on Twitter. There was no overlap in recruited participants with those recruited for the lab-based study.

Potential participants were directed to a pre-screen questionnaire, to provide basic details about the online services that they normally use. It was imperative to recruit participants that actively use U2F-compatible services (e.g., Dropbox, Facebook, GitHub, Google, Twitter, etc.). Participants should also actively use a desktop Web browser that supports

the use of security keys. No experience with security keys was necessary. Participants were compensated with their choice of a complimentary Amazon or Love2shop voucher worth £30 upon completing the study.

Each participant took part in a briefing session, lasting approximately 25 minutes. The experimenter followed a script to explain the study and would provide an information sheet and consent form; the participant was given time to read the documents before providing consent. To begin the study, we briefly discussed a participant's current authentication practices (revisiting the pre-screen responses). A researcher then guided the participant to set up a unique SecureClick security key (pre-paired with its Bluetooth Bridge) with up to two of their existing (U2F-compliant) accounts. Participants were free to set it up with their other accounts during their participation in the study if they wished.

Instructions were given on how to complete the diary journal (shown in the Appendix, Figures 6-7), and submit daily entries to the research team towards building rapport and to ensure participants remained motivated to complete the diary exercise [26, 39]. Participants who had not submitted their diary entries at the end of a day were reminded to do so the following day, in a single short message from the researcher. We managed communications with participants via email or WhatsApp, at the preference of each participant. After the diary exercise, each participant took part in a debrief interview, lasting approximately 15 minutes, to discuss their experiences and clarify uncertain entries provided during the diary exercise. Finally, online accounts linked to the SecureClick security key were restored to their initial state if requested by the participant. In addition to the financial incentive offered to take part in the study, participants were also offered to keep the SecureClick only at the time of leaving; 12 opted to do so.

Participants' answers during the brief and debrief interviews were audio-recorded to facilitate transcription at a later stage.

We finalized our study design by running a pilot study with one extra participant before the main study. We concluded that participants should use a personal computer in the briefing session as opposed to an unfamiliar device since we discovered this might serve as a deterrent to participation.

4.1.2 Research Ethics

The diary study received ethical approval as part of the same project that included the lab-based study (Section 3).

Participants registered interest in taking part in the study using a pre-screen online form, managed from a survey platform operated from within the host university. We required a contact email address for the duration of the study and collected demographic information: age, gender, education level.

During the briefing session, up to two of a participant's online accounts were set up with a security key. Only researchers involved in the study had access to the dedicated email address and phone number, and we stored transcript data using a pseudonymous participant number.

To mitigate harm to study participants [19], we provided instructions on how to contact the research team with any issues during the study. The contact phone number for submitting diaries by WhatsApp, and for contacting researchers with issues, was terminated after the study.

4.1.3 Demographics

We recruited fifteen participants for the study (5 female, 10 male). Participants ages were between 21- and 44-year old (median 24). Five were postgraduate, five were graduate, and five were up to the undergraduate level. Thirteen participants were from the host university and the remaining two from a partner university. Nine already had experience of using 2FA (either SMS-based OTPs or mobile-based authentication app). None of the participants had any familiarity with security keys.

4.1.4 Limitations

The diary entries are self-reported data, which can be prone to under-reporting [34]. Participants may have adjusted their login behavior or diary completion to align with the perceived interests of the researchers. However, we had no leading hypothesis that could be leaked to the participant implicitly or otherwise that could prime participant behavior in one direction or another. Efforts were made to minimize interruption to participants, by asking for short entries in a structured table for each relevant event during the day, complemented with open-ended reflection only at the end of each day and at the end of the week (see Appendix). Our sample size is modest; however, our study design was structured to generate a useful volume of data irrespective of that. Crucially, none of our participants were familiar with security keys.

4.2 Results

We present the analysis of the diaries themselves and debrief interviews in the sections that follow. We refer to notable quotes from specific participants using numeric identifiers: e.g., PD##.

4.2.1 Diary Entry Analysis

Overall Activity: We recorded 643 diary entries across all participants. The median number of diary entries per participant was 38.

The median number of Web services that a participant registered their security key with was 3 ($IQR = 2$). As illustrated in Figure 3, the services for which participants reported the most frequent logins were: Gmail (321 events, 50%), Facebook (114, 18%), Dropbox (95, 15%) and

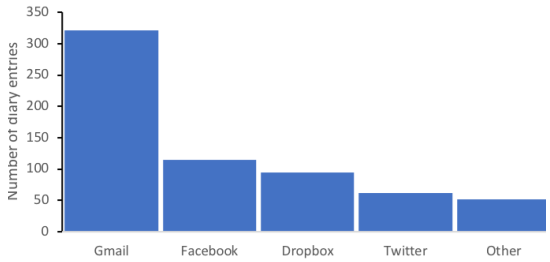

Figure 3: Chart of the services for which participants chose to register their security key and for which we recorded login events.

Twitter (61, 10%).

Locations of Events: The median number of different locations that participants used the security key was 4 ($IQR = 4$). Participants reported that the vast majority of login events took place in a home environment (396 events, 62%); followed by the workplace (80, 12%); whilst in transit (e.g., on trains, buses) (52, 8%) and at a university (57, 8%); in public spaces, e.g., cafes (48, 7%); and finally, at a friends' house (10, 2%).

Computer Use: The median number of computers that participants accessed accounts from where the security key was enabled was 2 ($IQR = 2$). The most common device used with the security key was a personal laptop (345 events, 54%), followed by an own mobile device (191, 30%), a personal tablet (41, 6%), a work desktop (32, 5%), and public computers (13, 2%). Other devices comprising less than 1% of accesses included the devices of friends/family.

Device Management and Portability: Concerning how participants managed the two parts of the SecureClick, eight participants always kept both the button and USB parts of the SecureClick together, whereas the remaining seven kept them separately. There was a mix of attitudes regarding where participants kept the parts: keyring (button part only), laptop (USB part only), wallet, original box, clear plastic case, bag, pocket, safe place at home or work.

In terms of portability, ten participants generally carried both parts of the SecureClick and thus always had access to it if needed. On the other hand, three participants carried only the button part and left the USB part plugged in their laptop at home at all times, and the remaining two participants always left both parts of the SecureClick in a safe place at home.

Login Methods: Table 3 illustrates the most common means by which participants accessed security key-enabled accounts during the study. The most common type of login event recorded was where users utilized the 'automatic login' functionality of a Web service (63%). This is where a Web service remembers a successful login on a specific device and does not prompt the user to authenticate again for a time, such as 30 days. Thus the user is logged in transparently and without any authentication friction. The combination of a security key and password appeared in only 28% of the captured logins. A password and alternative 2FA, such as SMS, was used 5% of the time, and circumstances were possible where users reported being able to access a service with only a password – a possibility that appears specific to Gmail. Participants abandoned 2% of the sessions due to issues with accessing a service.

Login Type	**Percent**
Automatic login	63%
Password & Security Key	28%
Password & Other 2FA	5%
Password only	2%
Abandoned sessions	2%

Table 3: Frequency of the different types of captured logins.

Figure 4 shows the 2FA login methods used on each day of the study across all participants as a proportion of all login events, for online accounts with an enabled security key. Usage of automatic login increased over time at the expense of security keys – automatic login and the combination of password and security key were used 38% and 44% of the time respectively on the first day of the study, whereas the figures were 75% and 16% respectively at the end of the week.

Satisfaction: At the end of each day participants were asked to respond to the question *"On a scale of 1 (very bad) to 9 (very good), how would you rate your experience of using the security key today?"* The median response was 7 ($IQR = 3$), and there was no discernible relationship with these scores and the progression of the study. An example of a free-text response to a day with a score of 3 would be (PD07) *"annoyances started when attempting to log in using a mobile phone, Git pushing from Ubuntu terminal, as well as logging in with different browsers on public computers like Firefox."* The example of GitHub is of specific interest since a security key can be enabled on GitHub for login as long as SMS-based OTPs are also enabled. Then, when using the command line interface, if 2FA is enabled, the user must register an SSH key to authenticate repository updates (since the browser cannot capture 2FA from the command line). Another low score (4) included the comment: *"[Facebook] still sent an SMS message. This doesn't feel particularly secure if every time I use the security key, I get an SMS. Why not just use the phone if it's going to communicate with the phone, anyway?"* (PD01). The comment refers to Facebook's practice of sending an SMS-based OTP even if a user is being prompted to use a security key. An example of comments associated with a high score (9 out of 9) was (PD03) *"I only realized now this solution is excellent when using public computers."*

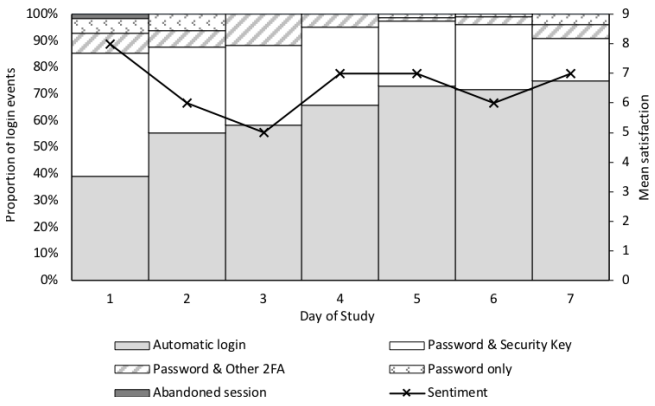

Figure 4: Proportion of login methods used on each day of the study across all participants, with Sentiment, Very Bad (1) to Very Good (9), on right-hand y-axis. Usage of 'automatic login' increases over time at the expense of security keys. Sentiment towards the security key stayed relatively constant over the course of the week.

The sentiment towards the security key stayed relatively constant over the week, as highlighted in Figure 4. We noted this effect despite the usage of the security key decreasing over time.

4.2.2 Qualitative Results

We analyzed the qualitative data using thematic analysis [12]. All participant responses were coded in several stages by one researcher, initially as low-level labels, moving to higher-level analytical categories. We identified five high-level categories – device form factor was omitted as a repeat of lab study findings, leaving the four themes presented in the following sub-sections – and 16 sublevel codes.

Threats, Context, and the Purpose of FIDO U2F

In terms of perceived threats, participants were generally not concerned about losing their passwords via phishing emails, with only one participant (PD05) in contrast conveying that they were *"massively worried"* about it. Five participants were worried about using public machines (PD02, PD03, PD04, PD05, PD13), attributing this to potential loss of credentials via malware or shoulder surfing. Three participants (PD05, PD09, PD15) mentioned concerns about losing their credentials when using public WiFi. Two participants (PD05, PD10) were worried about losing their laptop or having it stolen, lest it permits an attacker to access their online accounts.

In terms of general authentication practices away from the study, participants predominantly used 1FA to access their online accounts, with 2FA via SMS-based OTPs seldom used when forced to do so by specific Web services, e.g., banks or work Virtual Private Network (VPN). To facilitate access to Web services, eleven participants reported using a password manager (dedicated, or credentials saved in the browser), with a further three using automatic login. The remaining two participants reported re-entering their credentials each time they accessed their online accounts.

Three participants mentioned proactively taking extra steps to secure their online accounts. PD01 implemented a bespoke password manager, as they did not trust commercially available tools. PD02 reported using a widely available password manager to increase the *"entropy"* of their passwords, also only using their own 'trusted' personal machine, having *"hardened it and [I] don't let people put random USB in."*

Thirteen participants perceived the security key as useful only in specific contexts, generally to protect accounts holding sensitive information (e.g., emails, work, banking). The remaining two participants did not see a use for the security key. Five participants thought it could be useful when accessing accounts using public machines or another person's machine, although there was a concern that the machine owners would conversely be uneasy about allowing an 'unknown' technology to interact with their machine. Two participants (PD06, PD14) speculated that the security key could be useful to secure work-related accounts on a specific device. PD03 also mentioned that security keys could be useful to secure physical objects: *"This device, I don't know if you could use it in a way to secure storage."*

Security Key as a Perceived Barrier to Login

Three participants (PD01, PD08, PD12) explained that the security key affected or reduced their inclination to access online services because *"it does make a conscious barrier between you [when you] log in onto a site, because it's a different action."* (PD01). This effect was particularly noted during access of social media sites: *"It has reduced my usage by quite a lot [...] I think I had at the back of my mind that I would need to go back in my bag, get the key out, put it in, go onto the thing."* (PD01); for some this barrier was not unwelcome, e.g., PD08: *"I wasted less time on Facebook whenever I was in the library."* Four other participants (PD05, PD09, PD10, PD11) thought that the login delay introduced by using the security key could be frustrating. Issues with 'authentication fatigue' have been reported elsewhere as factors in employees reducing their use of computers [40]. Only one participant (PD05) reported a perceived increase in their usage of Google services as a result of using the security key: *"I've already started to save some stuff to Google, which I wasn't doing before, because it felt safer now."*

Challenges in Configuring and Using FIDO U2F

An inability to make the security key work with a mobile phone was a recurring issue affecting six of the fifteen participants.

Other set up issues concerned poor or complete lack of

phone reception when setting up mandatory backup authentication mechanisms (PD02, PD03); a participant's browser not supporting U2F by default (PD05, PD11); an inability to find an option to set up the security key with Google (PD01, PD13) or specific applications (Thunderbird (PD04), Apple Mail (PD12)); and the participant's computer OS not supporting U2F by default (PD02).

Usage issues specific to the SecureClick were sporadic. Two participants (PD06, PD12) experienced inability to, or had issues with, getting the SecureClick to work on a new computer due to problems 'installing' the USB part when plugging it into the new computer. Two other participants (PD09, PD15) complained about being unable to leave the USB part plugged into their computer, as they have to pull the dongle out and put it back in for it to work again after restarting their computer. Two participants (PD11, PD12), where the SecureClick ran very low on battery and effectively stopped working, failed to recognize this was the case, as the light feedback still operated as usual in these instances (where insufficient battery life for a security key has been seen as an issue elsewhere [40]).

Other general challenges to using security keys lay in the limited support of U2F amongst browsers (PD05, PD07, PD10, PD15) and operating systems (PD07). Two participants mentioned the barrier to the use of security keys with Git CLI for GitHub (PD02, PD07).

Some of these issues caused two participants to remove the SecureClick from some of their accounts, e.g., PD07: *"I disabled it on GitHub, because the pushing and pulling part was just getting a bit annoying."*

User Choices Lack Support

User selections of authentication mechanisms were often not respected or persisted. Participants were initially frustrated if provided 2FA options did not meet their expectation of needing to use the security key. Eight participants were unaware that some Web services (e.g., Gmail) enabled a 'remember this security key' option by default on the first login, sometimes leaving participants puzzled as to why they were not prompted for the security key again during subsequent login events, e.g., PD10: *"I think it just surprised me when it did it, and I'm not 100% sure if it was done by my computer or if it was the token that initially did that."* Two participants (PD01, PD04) had issues with some Web services sending them an SMS-based OTP at the same time that they were using the security key, e.g., PD01: *"Facebook spammed my iPhone with texts suggesting I needed a code - even when the key was working."*.

Six participants thought that services offering a 'remember this security key' option, or sending an SMS-based OTP at the same time as a security key was being used, rendering the keys redundant, e.g., PD04: *"Facebook always sends an SMS code at the same time, so I didn't need the key to do it, so the key feels useless at that point."*

Three participants (PD02, PD06, PD09) mentioned that using a security key should come with weaker or less stringent password policies. Regarding the choice of login methods, four participants (PD02, PD07, PD09, PD13) preferred persistent login sessions, two (PD05, PD08) did not, and three (PD06, PD10, PD12) would make different decisions depending on the context. Two participants (PD07, PD12) also mentioned that they would like a choice of different login policies for different devices. These results allude to the decisions about whether to use a particular combination of 2FA technologies being more complicated than which two to use, but how to combine them to reach a level of security that is satisfactory to the user.

5 Discussion

5.1 Comparisons between 2FA techniques

Revisiting our overarching research question (Section 2.3), our lab-based study evaluated a diverse cross-vendor set of security keys alongside SMS-based OTPs. The security key setup process is generally not efficient for novice users to complete [38], but we found that the devices were deceptively heterogeneous, and created their own specific challenges. Setup times were considerably larger than login times for the security keys. The median overall setup time was 146 seconds ($IQR = 101$), and the median login time was 30 seconds ($IQR = 17$). No particular device was significantly more successful than another in enabling greater efficiency, despite device-specific features being known to impact efficiency for users, e.g., the Bluetooth pairing required by the SecureClick, and the (small) size of the YubiKey. The ePass was free of both issues, but required drivers to be installed (on Windows platforms) before use (creating and contributing to setup delays). These findings challenge the often remarked claims that these are simple, *'one tap'* devices.

SMS-based OTPs were never rated below acceptable by participants through SUS ratings, whereas security keys often were. The high ratings that participants provided for SMS-based OTPs were not in line with measurements of time efficiency during our laboratory study. This result could be symptomatic of participants trusting the familiar SMS technology, or could indicate that users anticipate security keys impacting on convenience and account recovery. Also, configuring backup 2FA (typically SMS-based OTPs) was often a pre-requisite for setting up security keys, which could have led participants to attach more significance to the role of SMS, rather than security keys.

5.2 Everyday Experiences of Security Keys

Prior work [38] has reported that users were generally satisfied when using a security key; we obtained similar results through SUS ratings: mean=75.83 ($SD = 14.81$, 95% $CI = \pm 7.49$), or

'acceptable' [9]. However, participants only used the security keys in 28% of the recorded login events in which security keys were active. On the final day of the diary study, the daily usage of the security key had declined as a proportion of all login events by 50% compared to the first day. This decrease could partly explain why prior work highlights acceptable user satisfaction with security keys in field studies [38], yet greater challenges in lab-based studies focusing on the keys themselves [16].

Participants were generally using or willing to use 2FA with Web services, at least for accounts that they deemed to be critical. However, there was a perceived risk of being locked out of one's account, should the second factor be unavailable when needed (the activity of locating the key, as in Figure 5 – see Appendix). This risk was a major concern for participants, and the form factor of security keys may exacerbate such fears. Participants perceived some security keys as more suited for use in one place only (e.g., at home, or in a work environment), whereas others were judged acceptable to be carried around and used for login from different computers on the move. Some participants felt that it was inevitable to require ownership of several security keys for this reason. However, the diversity of such suggestions hints that users struggled to spot an identifiable 'universal' usage proposition for the security keys. It may be that a use case for security keys is as devices for infrequent use in bootstrapping a set of trusted devices, for subsequent transparent logins. If distinct use cases were to emerge, this would require device manufacturers to set different expectations about how users should optimally use the device.

5.3 Service Providers and Managing Friction

Security keys are user-centric [10] technologies that are decentralized. As such, there is no natural recovery mechanism that can be provided by the service provider should a device be lost, except to provide a toolbox of ready 2FA alternatives. Service providers encounter a conflict between reducing authentication friction for their customers to access services easily, and to enable users to protect their accounts. The same conflict has been noted with alphanumeric password strength for online services, where those with the largest customer bases had the least stringent password requirements [22].

But there may be risks to completely removing friction from security key usage. Specifically, user trust in the devices may decline due to the way that the user interface prioritizes other 2FA options. As an example, if 2FA is enabled, Facebook sends an SMS to the user at the point of login, even if the user previously selected to use a security key. Similarly, Gmail occasionally requests only a password to login to a device where the user used a security key in the past and, indeed, security keys are at the bottom of the list of alternative 2FA methods in the choice architecture for Gmail 2FA. Finally, '*Remember me'* was a feature that constituted the majority of service accesses in our diary study. Each of these examples illustrate how the user perception of the importance of pressing the button on a security key is undermined since that preference is under constant challenge by the presentation of 2FA alternatives at crucial moments. These events may act like a 'nudge' of the user towards a preferred 2FA [37], rather than the display of a choice architecture that promotes informed decisions for a particular user [13].

The transition towards FIDO2 [4] may alleviate some of these challenges, through closer integration of U2F with mobile devices. In the long-term, it may be that these standards are necessary not only to move toward seamless mobile device support, but also to support service providers to optimize the design of their infrastructure around future iterations of U2F or even decentralized identifiers from emerging decentralized identity schemes [20].

6 Conclusion

Security keys are 2FA technologies that are resistant to phishing, whereas ubiquitous SMS codes are not. However, uptake of security keys for general Web browsing is generally low. In this paper, we conducted two empirical studies to better understand the user experience of security keys for purposes of everyday Web authentication.

Firstly, in a laboratory study, we evaluated a diverse cross-vendor set of security keys alongside SMS-based OTPs, to capture factors affecting the usability and security perceptions of security keys during setup and login. We found that the setup time for security keys was considerably greater than login time. Also, SMS-based OTPs were never rated below 'acceptable' by participants using an SUS scale, whereas security keys often were.

Secondly, we conducted a diary study over one week, to capture user experience challenges encountered in everyday use of a security key. We found that only 28% of accesses to security key-enabled online accounts involved pressing a button on a security key, and use of a security key decreased as a proportion of all account accesses as the study progressed. Inadvertently nudging users away from explicit use of security keys likely erodes the perception of utility of security keys which is seen in prior work [16].

Our research demonstrates the importance of considering security key usage in the broader context of other competing 2FA technologies and the nature of 2FA choice architectures provided by Web services.

Acknowledgments

We would like to thank the SOUPS reviewers for their comments and support in preparing the paper for the conference. Stéphane Ciolino was supported in part by OneSpan. Study incentive and security key costs were supported by OneSpan.

References

[1] Seb Aebischer, Claudio Dettoni, Graeme Jenkinson, Kat Krol, David Llewellyn-Jones, Toshiyuki Masui, and Frank Stajano. Pico in the Wild: Replacing Passwords, One Site at a Time. In *Proc. European Workshop on Usable Security (EuroUSEC 2017)*. Internet Society, 2017. URL: https://www.internetsociety.org/sites/default/files/eurousec2017_17_Aebischer_paper.pdf, doi:10.14722/eurousec.2017.23017.

[2] FIDO Alliance. About The FIDO Alliance. URL: https://fidoalliance.org/about/overview/.

[3] FIDO Alliance. Approach & Vision. URL: https://fidoalliance.org/approach-vision/.

[4] FIDO Alliance. FIDO2: Moving the World Beyond Passwords using WebAuthn & CTAP. URL: https://fidoalliance.org/fido2/.

[5] FIDO Alliance. 2017 State of Authentication Report, October 2017. URL: https://fidoalliance.org/2017-state-authentication-report/.

[6] FIDO Alliance. MakeUseOf: It's Time to Stop Using SMS and 2FA Apps for Two-Factor Authentication, January 2018. URL: https://fidoalliance.org/time-stop-using-sms-2fa-apps-two-factor-authentication/.

[7] M. M. Althobaiti and P. Mayhew. Security and usability of authenticating process of online banking: User experience study. In *2014 International Carnahan Conference on Security Technology (ICCST)*, pages 1–6, October 2014. doi:10.1109/CCST.2014.6986978.

[8] All Things Auth. SMS: The most popular and least secure 2FA method, February 2018. URL: https://www.allthingsauth.com/2018/02/27/sms-the-most-popular-and-least-secure-2fa-method/.

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

[10] Abhilasha Bhargav-Spantzely, Jan Camenisch, Thomas Gross, and Dieter Sommer. User centricity: a taxonomy and open issues. In *Proceedings of the second ACM workshop on Digital identity management - DIM '06*, page 1, New York, New York, USA, 2006. ACM Press. URL: http://portal.acm.org/citation.cfm?doid=1179529.1179531, doi:10.1145/1179529.1179531.

[11] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symp. on Security and Privacy*, pages 553–567, May 2012. URL: http://ieeexplore.ieee.org/document/6234436/, doi:10.1109/SP.2012.44.

[12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. URL: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa, doi:10.1191/1478088706qp063oa.

[13] Pamela Briggs, Debbie Jeske, and Lynne Coventry. Behavior change interventions for cybersecurity. In *Behavior Change Research and Theory*, pages 115–136. Elsevier, 2017.

[14] Elizabeth Charters. The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal*, 12(2), July 2003. URL: https://brock.scholarsportal.info/journals/brocked/home/article/view/38, doi:10.26522/brocked.v12i2.38.

[15] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. "It's not actually that horrible": Exploring Adoption of Two-Factor Authentication at a University. In *CHI 2018*, pages 1–11, Montreal, QC, Canada, April 2018. ACM Press. URL: http://dl.acm.org/citation.cfm?doid=3173574.3174030, doi:10.1145/3173574.3174030.

[16] Sanchari Das, Andrew Dingman, and L Jean Camp. Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key. *Preproceedings Financial Cryptography and Data Security 2018*, 2018.

[17] Sanchari Das, Gianpaolo Russo, Andrew Dingman, Jayati Dev, Olivia Kenny, and L Jean Camp. A Qualitative Study on Usability and Acceptability of Yubico Security Key. *Proceedings of , Florida, USA, December (STAST 2017)*, December 2017.

[18] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *Proceedings 2014 Workshop on Usable Security*, San Diego, CA, 2014. Internet Society. URL: https://www.ndss-symposium.org/ndss2014/workshop-usable-security-usec-2014-programme/comparative-usability-study-two-factor-authentication, doi:10.14722/usec.2014.23025.

[19] David Dittrich, Erin Kenneally, et al. The Menlo Report: Ethical principles guiding information and communication technology research. Technical report, US Department of Homeland Security, 2012.

[20] P. Dunphy and F. A. P. Petitcolas. A First Look at Identity Management Schemes on the Blockchain. *IEEE Security Privacy*, 16(4):20–29, July 2018. doi:10.1109/MSP.2018.3111247.

[21] Duo. Bringing U2F to the Masses. URL: https://duo.com/blog/bringing-u2f-to-the-masses.

[22] Dinei Florêncio and Cormac Herley. Where Do Security Policies Come From? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 10:1–10:14, New York, NY, USA, 2010. ACM. URL: http://doi.acm.org/10.1145/1837110.1837124, doi:10.1145/1837110.1837124.

[23] Bennett Garner. Why 2FA Matters & the Best Types of 2FA, April 2018. URL: https://coincentral.com/why-2fa-matters-the-best-types-of-2fa/.

[24] Greg Guest, Arwen Bunce, and Laura Johnson. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, 18(1):59–82, February 2006. URL: http://journals.sagepub.com/doi/10.1177/1525822X05279903, doi:10.1177/1525822X05279903.

[25] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, 30(4):208–220, June 2011. URL: http://www.sciencedirect.com/science/article/pii/S0167404810001148, doi:10.1016/j.cose.2010.12.001.

[26] P. G. Inglesant and M. A. Sasse. Studying Password Use in the Wild: Practical Problems and Possible Solutions. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, 2010. URL: https://cups.cs.cmu.edu/soups/2010/user_papers/Inglesant_passwords_in_wild_USER2010.pdf.

[27] Kaspersky. SMS-based two-factor authentication is not safe - consider these alternative 2FA methods instead, October 2018. URL: https://www.kaspersky.co.uk/blog/2fa-practical-guide/14589/.

[28] Brian Krebs. Google: Security keys neutralized employee phishing, 2018. URL: https://krebsonsecurity.com/2018/07/google-security-keys-neutralized-employee-phishing/.

[29] Kat Krol, Simon Parkin, and M. Angela Sasse. Better the Devil You Know: A User Study of Two CAPTCHAs and a Possible Replacement Technology. In *Proceedings of NDSS Workshop on Usable Security (USEC 2016)*, San Diego, CA, USA, 2016. Internet Society. URL: https://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2017/09/better-the-devil-you-know-user-study-of-two-captchas-a-possible-replacement-technology.pdf, doi:10.14722/usec.2016.23013.

[30] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M. Angela Sasse. "They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking. In *USEC '15*, San Diego, CA, USA, February 2015. Internet Society. URL: https://www.ndss-symposium.org/ndss2015/ndss-2015-usec-programme/they-brought-horrible-key-ring-thing-analysing-usability-two-factor-authentication-uk-online, doi:10.14722/usec.2015.23001.

[31] Kat Krol, Jonathan M Spring, Simon Parkin, and M Angela Sasse. Towards robust experimental design for user studies in security and privacy. In *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)*, pages 21–31, San Jose, CA, May 2016. USENIX Association.

[32] 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*, Lecture Notes in Computer Science, pages 422–440. Springer, Berlin, Heidelberg, February 2016. URL: https://link.springer.com/chapter/10.1007/978-3-662-54970-4_25, doi:10.1007/978-3-662-54970-4_25.

[33] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Elsevier, Cambridge, MA, 2nd edition edition, 2017. OCLC: 1030364616.

[34] Shrirang Mare, Mary Baker, and Jeremy Gummeson. A Study of Authentication in Daily Life. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, SOUPS'16, pages 189–206, Berkeley, CA, USA, 2016. USENIX Association. URL: http://dl.acm.org/citation.cfm?id=3235895.3235912.

[35] National Cyber Security Centre (NCSC). Setting up two-factor authentication (2FA), 2018. URL: https://www.ncsc.gov.uk/guidance/setting-two-factor-authentication-2fa.

[36] Jakob Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380. ACM, 1992.

[37] Karen Renaud and Verena Zimmermann. Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies*, 120:22–35, 2018.

[38] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent Seamons. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 872–888, San Francisco, CA, May 2018. IEEE. URL: https://ieeexplore.ieee.org/document/8418643/, doi:10.1109/SP.2018.00067.

[39] John Rieman. The Diary Study: A Workplace-oriented Research Tool to Guide Laboratory Efforts. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 321–326, New York, NY, USA, 1993. ACM. URL: http://doi.acm.org/10.1145/169059.169255, doi:10.1145/169059.169255.

[40] M. Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. The Great Authentication Fatigue - And How to Overcome It. In P. L. Patrick Rau, editor, *Cross-Cultural Design*, Lecture Notes in Computer Science, pages 228–239. Springer International Publishing, 2014.

[41] Jonathan M. Spring and Phyllis Illari. Building General Knowledge of Mechanisms in Information Security. *Philosophy & Technology*, Sep 2018. URL: https://doi.org/10.1007/s13347-018-0329-z, doi:10.1007/s13347-018-0329-z.

[42] Michelle Steves, Dana Chisnell, Angela Sasse, Kat Krol, Mary Theofanos, and Hannah Wald. Report: Authentication Diary Study. Technical Report NIST IR 7983, National Institute of Standards and Technology, February 2014. URL: https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7983.pdf, doi:10.6028/NIST.IR.7983.

[43] Jake Weidman and Jens Grossklags. I Like It, but I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference on - ACSAC 2017*, pages 212–224, Orlando, FL, USA, 2017. ACM Press. URL: http://dl.acm.org/citation.cfm?doid=3134600.3134629, doi:10.1145/3134600.3134629.

[44] Catherine S. Weir, Gary Douglas, Martin Carruthers, and Mervyn Jack. User perceptions of security, convenience and usability for ebanking authentication tokens. *Computers & Security*, 28(1-2):47–62, February 2009. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167404808000941, doi:10.1016/j.cose.2008.09.008.

[45] Catherine S. Weir, Gary Douglas, Tim Richardson, and Mervyn Jack. Usable security: User preferences for authentication methods in eBanking and the effects of experience. *Interacting with Computers*, 22(3):153–164, May 2010. URL: https://academic.oup.com/iwc/article-lookup/doi/10.1016/j.intcom.2009.10.001, doi:10.1016/j.intcom.2009.10.001.

[46] Wired. How to Protect Yourself Against a SIM Swap Attack. URL: https://www.wired.com/story/sim-swap-attack-defend-phone/.

[47] Yubico. Works with YubiKey. URL: https://www.yubico.com/solutions/.

[48] Yubico. OTP vs. U2F: Strong To Stronger, February 2016. URL: https://www.yubico.com/2016/02/otp-vs-u2f-strong-to-stronger/.

Appendices

A Lab-Study Task: 2FA Using *[Tested Mechanism]* on Laptop

- ['Set-up phase'] On laptop, ask participants to:
  - [SecureClick only] Install OneSpan DIGIPASS SecureClick Manager and pair SecureClick with its Bluetooth Bridge.
  - open Chrome.
  - login onto Web service.
  - set up 2FA using *[Tested Mechanism]* on Web service.
  - logout of Web service.
  - close Chrome.
- ['Login' phase] On laptop, ask participants to:
  - open Chrome.
  - login onto Web service.
  - logout of Web service.
  - close Chrome.
- [SUS] Ask participants to fill in the SUS questionnaire about their experience of 2FA using *[Tested Mechanism]* on laptop.

B FIDO U2F Mechanisms

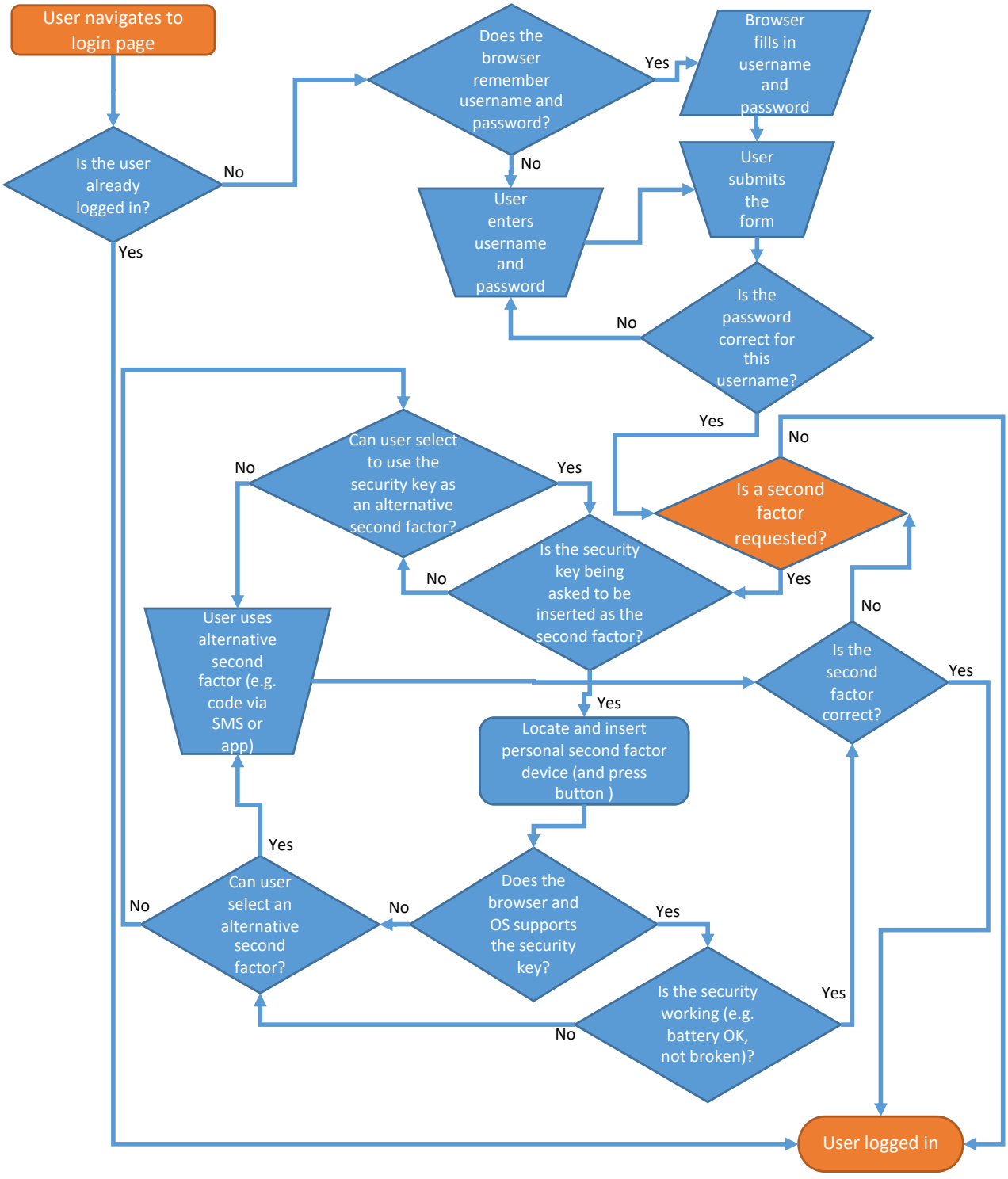

Figure 5: Mechanisms and activities in the FIDO U2F study.

C Diary Forms

Per Access

For each attempted access (successful or otherwise) to a web service where the DIGIPASS SecureClick (security key) is enabled, please fill in the following information in a new row.

#	Time	Web service you were accessing *(Gmail, Facebook, Twitter, Dropbox, GitHub, etc)*	Location you were accessing it from *(home, work, internet café, etc)*	Device you were accessing it from *(personal laptop, work desktop, mobile, etc)*	Successful access? *(Yes/No)*	How many unsuccessful attempts did you have before successful access or abandoning?	If the access was successful, which authentication method(s) did you use to access your account? *(Tick all that apply)*			
							Security key	Code (via SMS, email, other)	Password	Nothing. Automatic login
0	13:15	Twitter	Library	Public desktop	Yes	0	☒	☐	☒	☐
1							☐	☐	☐	☐
2							☐	☐	☐	☐
3							☐	☐	☐	☐
4							☐	☐	☐	☐
5							☐	☐	☐	☐
6							☐	☐	☐	☐
7							☐	☐	☐	☐
8							☐	☐	☐	☐
9							☐	☐	☐	☐
10							☐	☐	☐	☐

(Continues next page if needed)

At the End of the Day

1. Which of the following, if any, did you experience today? *(Please tick any that apply and specify where required)*

a. I could readily get hold of the security key whenever I wanted to use it: ☐ Yes ☐ No

If you misplaced or could not readily get hold of some part(s) of the security key when you wanted to use it, please specify...

☐ Button part ☐ USB part For how many hours:

b. The security key always worked as I expected throughout the day: ☐ Yes ☐ No, specify:

c. I used an alternative authentication method to the security key during at least one access attempt: ☐ Yes ☐ No

If Yes, please specify which method(s) you used instead of using the security key:

☐ Code via SMS ☐ Code via mobile authenticator app ☐ Other, please specify:

d. I felt it was quick to get the security key ready to use, and to use it: ☐ Yes ☐ No How long did it typically took:

e. I found the feedback (e.g. light) from the security key clear: ☐ Yes ☐ No, please specify:

f. Other, please specify:

2. What best describes where your security key has been today? *(Please tick any that apply and specify where required)*

a. Part(s) of the security key have been on my person: ☐ Yes ☐ No

If Yes, please specify: ☐ Button part ☐ USB part For how many hours:

b. Part(s) of the security key have been somewhere safe, but not on my person: ☐ Yes ☐ No

If Yes, please specify: ☐ Button part ☐ USB part For how many hours:

c. Other, please specify:

3. Please leave any further comments you may have about your experience of using the security key today.

4. On a scale of 1 to 9, how would you rate your experience of using the security key today? *(Please tick which best applies)*

Very bad				Neither bad nor good				Very good
1	2	3	4	5	6	7	8	9

Figure 6: Daily diary form for the FIDO U2F diary study.

End of Week Questionnaire

We are now focusing on your overall personal experience of using the security key during the last 7 days. Please keep this in mind when answering the following questions.

1. Did you set up (or wanted to set up) the security key with any other web service(s)? ☐ Yes ☐ No

If Yes, please specify with which web service(s):

2. Did you remove (or wanted to remove) the security key from any web service(s)? ☐ Yes ☐ No

If Yes, please specify with which web service(s):

3. Did using the security key affect the way you access web services (e.g. accessing web services less/more often than usual, or from different locations and/or devices)? ☐ Yes ☐ No

If Yes, please specify how it affected your behavior:

4. Thinking specifically about web services where the security key is enabled, some offer an option to 'remember' the security key on devices you trust. Did you use this option? ☐ Yes ☐ No ☐ Don't know

If Yes, please specify with which web service(s):

5. Thinking specifically about web services where the security key is enabled, did you need to look at instructions or get help from someone to access your account(s)? ☐ Yes ☐ No ☐ Don't know

6. You have now been using the security key for a week. Overall, on a scale of 1 to 9, how would you rate your experience of using the security key? *(Please tick which best applies)*

Very bad				Neither bad nor good				Very good
1	2	3	4	5	6	7	8	9

7. Following this study, to what extent would you see yourself using a security key if you had one? *(Please tick which best applies)*

☐ I would never use it ☐ I would use it only in some specific contexts ☐ I would use it in any context

8. Please leave any further comments you may have about your overall experience of using the security key.

Figure 7: End-of-week diary form for the FIDO U2F diary study.