User Perceptions of the Usability and Security of Smartphones as FIDO2 Roaming Authenticators

Kentrell Owens[⋆,◇], Olabode Anise[⋆], Amanda Krauss[⋆], Blase Ur[†]
⋆ *Duo Security,* ◇ *University of Washington,* † *University of Chicago*

Abstract

The FIDO2 standard aims to replace passwords with public-key cryptography for user authentication on the web. Doing so has benefits for both usability (e.g., not needing to remember passwords) and security (e.g., eliminating phishing). Users can authenticate with FIDO2 in one of two ways. With platform authenticators, users authenticate to trusted hardware on the same device on which they are accessing a website. However, they must re-register for each website separately on each device. With roaming authenticators, such as USB security keys, they only need to register once, transferring the security key across devices. However, users might not be willing to pay for a USB security key, carry it around, or figure out how to plug it into different devices. These drawbacks have driven recent efforts to enable smartphones to serve as roaming authenticators. We conducted the first user study of FIDO2 passwordless authentication using smartphones as roaming authenticators. In a between-subjects design, 97 participants used either their smartphone as a FIDO2 roaming authenticator (via a prototype called Neo) or a password to log into a fictitious bank for two weeks. We found that participants accurately recognized Neo's strong security benefits over passwords. However, despite Neo's conceptual usability benefits, participants found Neo substantially less usable than passwords both in objective measures (e.g., timing to accomplish tasks) and in perception. Their critiques of Neo included concerns about phone availability, account recovery/backup, and setup difficulties. Our results highlight key challenges and opportunities for spurring adoption of smartphones as FIDO2 roaming authenticators.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Virtual Conference.

1 Introduction

For decades, the standard method of online authentication has involved a username and password [7, 21]. Unfortunately, password-based authentication on the web has key weaknesses. For instance, most online data breaches are caused by weak or reused passwords [48]. As a result, for decades researchers and practitioners have attempted to develop alternative authentication schemes that are more secure than passwords, yet no harder to use or deploy [7]. Consider, for example, federated identity systems like single sign-on (*SSO*). In concept, SSO eases the burden of remembering numerous passwords while also being more secure by reducing password reuse. Nonetheless, it has seen limited adoption outside organizational contexts due in part to users' privacy concerns [5, 40, 42]. Similarly, password managers have gained in popularity due to recommendations from security experts [36]. They help users generate, store, and enter unique passwords for each online account, reducing instances of password reuse or weak passwords [31]. However, adoption rates of password managers have also remained low [41].

In this line of attempts to replace passwords for web authentication, the recent FIDO2 standard [3] is a particularly promising approach that leverages public-key cryptography in place of passwords. When a user registers on a website, instead of entering a username and password, they use an *authenticator* (dedicated hardware or software following the FIDO2 specification) to generate a public-private keypair. Their client then shares the public key with the web application. FIDO2 has key benefits. In terms of usability, users no longer have to remember a password. In terms of security, users are protected from remote attacks like credential stuffing and phishing. In terms of privacy, FIDO2 does not have the centralized privacy risks of federated identity systems [13].

Major browsers Chrome, Firefox, Edge, and Safari all already support FIDO2 [33], as do a growing number of websites [47]. To use FIDO2, a user must have an authenticator, of which there are two key types. *Platform authenticators* are integrated with a broader-purpose client device and enable

authentication only on that device. For example, one can use Apple's Touch ID as a FIDO2 platform authenticator to log into websites from an iPhone or Mac laptop, or Windows Hello as a FIDO2 platform authenticator from a Windows laptop. Unfortunately, the user must re-register for a given website separately on each of their devices. In contrast, *roaming authenticators*, like USB security keys, are portable. A single roaming authenticator can be used across all of a user's devices [44]. While roaming authenticators offer usability benefits, such as enabling users to authenticate on different devices, prior work has shown users are reluctant to carry around USB security keys for authentication [10, 26]. Additionally, users may not be willing to pay for a security key.

This scenario has driven recent technical efforts to enable smartphones to be used as roaming authenticators. Over 81% of Americans own a smartphone [32], so using smartphones as roaming authenticators is likely to overcome key barriers faced by USB security keys. To this end, several proposed modifications to the FIDO2 specification are in progress, including caBLE (cloud-assisted Bluetooth Low Energy) and the closely related Network Transport [28]. Similarly, Duo Security is experimenting with a software-based mobile authenticator that we refer to as Neo. Because these implementations are recent, there has yet to be a usable security evaluation of the use of smartphones as FIDO2 roaming authenticators.

To understand user perceptions of the security and usability of Neo relative to passwords, we conducted a longitudinal user study. In a between-subjects design, participants were assigned to use either a password (termed *Password* participants) or their own smartphone as a FIDO2 roaming authenticator via the Neo prototype (termed *Neo* participants) to log into a fictitious bank from their own computer daily for two weeks. A total of 97 participants completed the full protocol and all daily tasks. We asked a series of research questions:

- **RQ 1**: Neo involves non-trivial setup relative to passwords. How difficult do users find Neo's initial setup?

By both objective and subjective measures, participants found Neo's setup process difficult. More than half of *Neo* participants dropped out of the study before completing the setup process, whereas under 10% of *Password* participants did so. Even among those who did complete setup, it took the median *Neo* participant over fifteen minutes to configure the software. While some of this difficulty was due to Neo being a research prototype, other aspects were inherent in using smartphones as roaming authenticators. Even beyond one-time setup costs, the recurring steps in account creation took longer for *Neo* participants than for *Password* participants.

Neo participants also perceived the setup process as less usable than *Password* participants. In particular, *Neo* participants rated the setup process 20 points lower on the 100-point system usability scale (*SUS*) than *Password* participants.

- **RQ 2**: In daily authentication, how does the usability of Neo compare to passwords (after Neo has been set up)?

Passwords can be forgotten or mistyped, whereas Neo simply requires access to a smartphone. Thus, we expected daily authentication to be easier for Neo than for passwords. We found the opposite, however. *Neo* participants were more likely than *Password* participants to be *unsuccessful* at logging in, typically because they could not authenticate to their phone (e.g., with their fingerprint sensor) or their phone seemed not to receive the push notifications that are part of the protocol. Unsurprisingly, then, *Neo* participants were less likely to rate daily sign-ins as easy than *Password* participants.

- **RQ 3**: Overall, how do users perceive the security and usability of Neo relative to passwords? Are they correct?

Overall, participants perceived Neo as both secure and usable. Notably, participants correctly perceived Neo as *more secure* than passwords. However, they also perceived Neo as *less usable* than passwords even beyond their direct experiences with setup and authentication. Nonetheless, over half of the participants who used Neo reported being "likely" or "very likely" to use Neo over passwords for five of the six account types (all except banking) that we asked about.

- **RQ 4**: Collectively, what are the barriers to user adoption of smartphones as FIDO2 roaming authenticators?

Neo participants frequently expressed concerns about not having their phone available or accessible when they hoped to log in. Notably, one-third of participants reported misplacing their phone at least once a day. They also worried about account recovery and losing access to their account, whereas they could simply write their password down somewhere safe. As a result, many participants expressed reluctance to adopt Neo for their own accounts even after using it.

- **RQ 5**: Does a user's prior experience with two-factor authentication (*2FA*) influence their perceptions of Neo?

We found that *Neo* participants who had prior 2FA experience rated its usability more highly (in terms of SUS score) than those who had never used 2FA.

Collectively, our work contributes the first user-centered understanding of smartphones as FIDO2 roaming authenticators. We uncovered a number of usability drawbacks, both in actuality and in perception, that will likely hamper the adoption of systems like Neo even as they move from research prototypes to being directly integrated with browsers. We thus highlight key challenges and opportunities for spurring adoption of smartphones as FIDO2 roaming authenticators.

Our paper proceeds as follows. We first detail how FIDO2 works (Section 2) and then present our user study's methodology (Section 3). Next, we present our study's results (Section 4). We then discuss these results (Section 5), compare them with related work (Section 6), discuss both limitations and future work (Section 7), and finally conclude (Section 8).

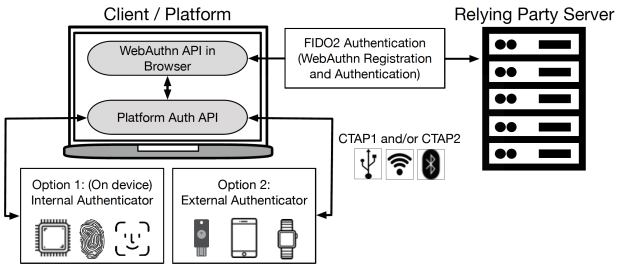

Figure 1: FIDO2 authentication with WebAuthn and CTAP2. This diagram is taken from Lyastani et al. [26].

2 Background

In this section, we further detail FIDO2 and its constituent protocols on a technical level. We particularly focus on efforts to support smartphones as roaming authenticators. Figure 1 summarizes the FIDO2 authentication process.

2.1 FIDO2: WebAuthn and CTAP2

The FIDO2 standard includes two key protocols. The Web Authentication API (*WebAuthn*) is a standard jointly developed by the FIDO Alliance and the W3C [33]. The WebAuthn API enables web applications (termed *relying parties*) to leverage public-key cryptography to authenticate users. Instead of a password, a unique public/private key pair is generated for each website registration using an authenticator. The private key is stored on the user's authenticator. The public key, along with a randomly generated credential ID, is stored on the web application's server. Credentials are scoped to the web application through the use of a relying party identifier that identifies the server. The user can then authenticate to that web application by interacting with their authenticator.

The other half of FIDO2 is *CTAP2*, a protocol being developed by the FIDO alliance. It is used when a relying party is interacting with a roaming authenticator [3], such as mobile devices like smartphones. The two salient parts of the protocol are the Authenticator API and the transport-specific bindings, referred to as *transports*, that can be used. The Authenticator API details how an authenticator should interact with a relying party when making a credential (i.e., public/private key pair) and creating assertions that provide proof of an authentication and a user's consent. The protocol defines how each of these operations should take place given the capabilities of the authenticator. The transports are how messages are conveyed from the host to a roaming authenticator. Currently, the modes that are supported are USB, NFC, and Bluetooth. The next section details implementations using these transports.

2.2 Mobile Roaming Authenticator Efforts

Next, we summarize three recent efforts that enable mobile devices to be used as FIDO2 roaming authenticators.

simFIDO is an implementation of FIDO2 by Chakraborty et al. [8] that uses a SIM-card-based Trusted Platform Module (TPM) called simTPM [9] to allow Android devices to serve as hardware authenticators. The authors introduced a new Android system service called *External FIDO Request Receiver Service* (XFRR) that forwards CTAP commands to the simTPM. Unlike typical implementations where credentials are bound to a particular device and cannot be removed, a SIM card (the authenticator) can be moved across devices.

caBLE (Cloud-Assisted BLE) is a proposal by Google that would extend CTAP2. It attempts to overcome some of the disadvantages of system BLE pairings, such as client-implemented preference syncing. The caBLE proposal allows mobile devices to serve as a roaming mobile authenticator by establishing a secure channel to pass CTAP2 messages between the authenticator and the client (e.g., the Chrome browser) [28]. The latest version of this proposal, caBLEv2, permits both temporary and permanent pairings between devices. The latter is appropriate for a personal device.

Neo is a prototype developed by Duo Security that allows mobile devices to serve as roaming authenticators. To use Neo, the user first pairs their mobile device with a Chrome browser with the aid of a mobile application and Chrome extension. The pairing process between the mobile device and the client takes place through a QR code generated by the extension. The QR code contains a shared secret. After the successful pairing, the client communicates with the mobile device through proxying of the WebAuthn API actions via the Chrome Extension to an intermediary server. Whenever the user attempts to authenticate on a website, they will receive a push notification to their smartphone that they can accept or reject after unlocking their device (if their phone is locked). Ongoing work aims to add an HTTPS-based transport (Network Transport) to the list of CTAP2 transports [2, 28]. With the addition of Network Transport to the CTAP2 specification, the Chrome extension would no longer be necessary during assertion or pairing for Neo or similar efforts; CTAP2 authenticators could communicate with the client directly.

There is debate about when user *presence*, versus user *verification*, should be required for authentication. Yubico recommends *presence* for 2FA and *verification* for passwordless authentication (like Neo) [46]. Since simple possession of a device is insufficient for authentication, we predict that similar schemes will (like Neo) require users to unlock their device before responding to an authentication push request; not doing so facilitates many attacks. Account sharing is easy with FIDO2 — simply register multiple phones with one account. Shared phones would be a security risk and potentially not possible if biometric user verification is required.

3 Methodology of Our User Study

To understand users' initial perceptions of the security and usability of using a smartphone as a FIDO2 roaming authen-

ticator, as well as how those perceptions might change after extended use, we conducted a longitudinal, between-subjects study. We compared the relative usability of passwords and Neo using both qualitative and quantitative methods. This study was conducted between May 2020 and July 2020.

3.1 Recruitment

We recruited participants on Amazon Mechanical Turk (*MTurk*), an online crowdsourcing platform that has been frequently used in usability studies. Redmiles et al. found that MTurk users are often more diverse in terms of age, income, education level, and geography than traditional social science pools [34]. Because of the requirements for using Neo (Chrome extension and Neo app), we required participants to have an Android mobile phone, access to a computer, and Google Chrome installed on that computer. Participants had to complete a screening survey verifying that they met the requirements for participating in the study, including 1) having an Android mobile phone running Android version 9+, 2) being located in the US, 3) using Google Chrome, and 4) having a fingerprint scanner on their phone. We used a free web service to validate the location of participants, following techniques outlined by Kennedy et al. [22]. Participants also had to have a 95% approval rating on MTurk.

3.2 Study Design

Eligible participants (n=247) were randomly split into two groups, with each group assigned one authentication method (*Password* or *Neo*). We informed participants that they were participating in a study about online authentication and that they would perform a series of ten tasks on a fictitious banking application over the course of two weeks. The banking application we used was a fork of the one used by Reese et al. [37] in a prior study [38], modified to support FIDO2.

We then instructed each group on how to register for an account with our web application using their assigned authentication method. Participants assigned to the *Password* condition were instructed to choose a username and password. We required that the password chosen contain at least 8 characters, without any further restrictions. Similar to Lyastani et al. [26], we chose this password-composition policy, which is the simplest NIST-recommended password policy, to avoid skewing usability perceptions with a potentially frustrating password-composition policy [43]. Choosing a more complex password-composition policy may have led to different perceptions of both the security and usability of passwords. Furthermore, to replicate participants' current approach to passwords, we neither encouraged nor prohibited the use of a password manager. Participants assigned to the *Neo* condition were given instructions on how to install the mobile application and Chrome extension needed for the Neo prototype to work, how to complete the pairing process between the mobile phone and the Chrome Extension, and how to register for an account on the banking website.

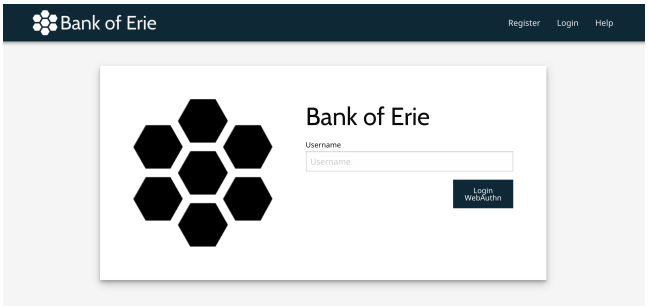

Figure 2: A screen shot of the simulated banking application to which participants authenticated throughout the study.

After successfully registering for their account and logging into the web application, participants were then instructed to complete one of the ten required tasks by using their assigned method to log into the banking application. Participants then completed the System Usability Scale (*SUS*) [16], evaluating the usability of setting up their assigned authentication method. In addition, they answered questions concerning their experience with setup and provided demographic information.

Over the two-week period, participants were sent a daily reminder to complete one of the ten required tasks by authenticating to the banking application. Full participation in the study required completing these ten tasks within 14 days. At the conclusion of the study, participants in both groups completed an exit survey where they completed another SUS questionnaire and answered open-ended questions about their authentication experience during the two weeks.

3.3 Attrition

During the development of the study protocol, we expected participant attrition in both the *Password* and *Neo* groups because of the longitudinal nature of the study. For participants in *Neo*, specifically, we hypothesized that there were steps that could prove to be challenging, causing additional attrition. Two such steps were app installation and fingerprint enrollment. For participants to install the Neo prototype on their Android, they would have to sideload the application. Depending on their version of Android, doing so necessitated enabling a setting to install unknown applications. While we detailed this process in our onboarding instructions, it is possible that participants did not feel comfortable enabling the setting. To use Neo, participants would also have to register a fingerprint if they had not already done so. This fingerprint was used to authenticate to their phone, thus instructing their phone to use their private key to authenticate to the banking application. Given common misconceptions about biometrics [6, 25], we anticipated that some participants assigned to *Neo* would not feel comfortable enrolling a fingerprint. Not

enrolling a fingerprint would prevent them from completing onboarding and participating in the remainder of the study. Sections 4.2–4.3 detail the attrition rates observed in practice.

3.4 Data Collected

We hypothesized that one of the barriers to the adoption of FIDO2 in general, and specifically a roaming mobile authenticator, would be a poor setup experience. For *Neo* participants, there are several steps where a participant could get stuck or have difficulty, such as installing the Chrome extension or pairing the Chrome extension and the mobile application. To understand which steps proved to be the most problematic, we collected detailed timing data on the following parts of the setup process for *Neo*: 1) downloading the Chrome extension, 2) enrolling a fingerprint, 3) downloading the mobile application, 4) pairing the Chrome extension and mobile application, and 5) registering a credential on the experiment platform. Since password-based authentication is something that *Password* participants would be familiar with, we only collected timing data for the account creation step. We attempted to capture each participant's initial impressions of the usability of their assigned authentication method through both the SUS questionnaire and a series of open-ended questions.

During the longitudinal portion of the study, we had three goals. First, we wanted to understand how long and error-prone the login experience was over time. To do so, we collected data on how long it took participants to authenticate during each of the ten sessions, recording failed authentication attempts (whether due to timeouts/cancellations in the *Neo* condition or incorrect password entries in the *Password* condition). Second, we wanted to understand how participants felt in the moment after each authentication. We accomplished this by implementing a diary-style Likert item where we asked participants to rate their agreement ("strongly disagree" through "strongly agree') that "logging into this application is easy." Lastly, we wanted to understand how participants' opinions of their assigned authentication method changed over time. Thus, we again asked participants to complete an SUS questionnaire and answer relevant open-ended questions.

3.5 Data Analysis Methods

We conducted both qualitative and quantitative data analyses. For free-response data, we conducted qualitative content analysis. In particular, two researchers independently read through the full survey data, each making a broad list of topics participants raised. They discussed the list and jointly created a code book combining topics under closely related themes. They iterated on this code book and reached consensus on the codes they would use. Using these codes, they both independently coded one-third of the responses. Cohen's κ, a measure of inter-coder agreement [11], was 0.85 between the two researchers. Fleiss et al. [19] consider values of κ over 0.75 as excellent agreement. If the two researchers disagreed on a code, they subsequently discussed the disagreement and reached consensus. After observing this acceptable value of κ, one researcher independently coded the remaining responses.

Many of our quantitative analyses were comparisons between the *Neo* and *Password* groups, such as in timing or SUS scores. Because most of this data was not normally distributed, we typically used the Mann-Whitney U test (*MWU*, also known as the unpaired two-samples Wilcoxon test). We also sought to understand how participants' success at authenticating, time required to authenticate, and Likert-scale responses both changed over time (across the ten authentication sessions) and varied between the *Neo* and *Password* groups. Because this data was not independent, we built mixed-effects regression models with the participant as a random effect. Based on the outcome data types of these three longitudinal models, we respectively built mixed-effects logistic, linear, and ordinal regression models. For all statistics, $\alpha = .05$.

3.6 Ethics

While Duo Security is not an academic institution and does not have a formal IRB, the study protocol and mechanisms used to conduct the study underwent an internal privacy review. As part of the screening process, participants had to consent to their data potentially being published externally. To conduct the study, we needed to store usernames and hashed versions of participants' passwords, but that data was discarded at the conclusion of the study. Participants were eligible for up to $30 in compensation based on whether they completed the survey that followed the setup process, the ten daily authentication tasks, and the exit survey. We chose that number to compensate participants at roughly $15/hr.

3.7 Pilot Study

Prior to conducting the study on MTurk, we conducted an internal, eight-person pilot to uncover potential problems in our study setup (e.g., survey questions, mobile app user interface, time allocated, bugs in the experiment platform), as well as to identify additional questions to ask. At the conclusion of the pilot, we corrected bugs in the code for the experimental platform. We also made our setup instructions more clear.

4 Results

In this section, we describe our participants and then report our results, grouped chronologically and by research question.

4.1 Participants and Their Demographics

Overall, 97 participants completed all parts of the protocol: onboarding, initial survey, ten daily tasks, and the exit survey. While the initial assignment to groups was randomized

Table 1: The demographics of the 97 participants who completed all surveys and parts of the full longitudinal protocol.

	Password	*Neo*
Gender		
Female	31	10
Male	28	17
No Answer	7	4
Age		
18–24 years old	2	5
25–34 years old	26	12
35–44 years old	26	7
45–54 years old	8	4
55–64 years old	1	2
No Answer	3	1
Race		
American Indian or Alaska Native	0	1
Asian	5	5
Asian, White	1	0
Black or African American	2	2
Black or African American, Hispanic	1	0
Hispanic	6	1
Hispanic, White	3	1
White	45	20
No Answer	3	1
Education		
High School Diploma/GED	3	3
Some College But No Degree	13	6
Associate's Degree	10	5
Bachelor's Degree	24	12
Professional Degree	13	4
No Answer	3	1
CS Background		
Yes	4	8
No	59	22
No Answer	3	1
TOTAL	66	31

and approximately equal, 66 participants in *Password* and 31 participants in *Neo* completed all parts of the protocol. We discuss this unequal attrition between groups further in Sections 4.2–4.3. All analyses other than those of participant attrition report on these 97 participants.

Table 1 summarizes participants' demographics. Participants in both conditions were more educated than the broader United States population, with 59% of *Password* participants and 53% of *Neo* participants having attained a bachelor's degree or higher. We also asked which of five 2FA methods (SMS, TOTP, pre-generated codes, push-notification based, and security keys) participants had used before; we included example images of these different methods to make them easily identifiable. Among participants, 94% had used SMS, 54% had used TOTP, 45% had used push notifications, 26% had used pre-generated codes, and 3% had used security keys. The proportions of each were similar between conditions.

Table 2: Summary of the time (in seconds) to set up Neo.

Step	10th %-ile	Median	Mean	90th %-ile
Install Chrome Extension	30.4	56.7	283.5	155.4
Enable Fingerprint	5.5	12.1	73.8	162.5
Install Phone Application	48.1	96.6	220.9	414.1
Pair Device	52.7	148.2	218.0	505.0
Create Account	17.9	105.3	139.4	233.9
Total Time	432.9	1000.1	1488.9	2316.3

4.2 Initial Setup (RQ 1)

As mentioned in Section 3.4, we measured the time it took for *Neo* participants to complete each step in the setup process. The timing data was collected through Qualtrics as participants progressed through the setup guide. The median time to complete setup for *Neo* was 1,000.1 seconds (16 minutes and 40.1 seconds). The step with the highest median time was pairing the participant's mobile device with the browser, which took 148.2 seconds. The step that took the least time was enabling fingerprint authentication for participants who did not already have it enabled. The majority (23/31) of participants in *Neo* reported that they already had their fingerprints enrolled on their smartphone prior to beginning the study. Additional timing results for *Neo* can be found in Table 2.

The only step in the setup process that *Neo* and *Password* both shared was account creation. The median time for account creation was 74.4 seconds for *Password* and 105.3 seconds for *Neo*, though this difference was not statistically significant (MWU, $U = 1269$, $p = .142$). In addition to the timing data, we analyzed the SUS scores that participants submitted after they completed setup. The median SUS score for *Password* was 88.6, while for *Neo* it was 66.6. This difference was significant (Mood's median test, $p < .001$).

Of the 31 *Neo* participants who completed the setup process, 11 nonetheless described challenges they encountered. They described it as too complex, particularly the process of downloading the mobile app and pairing the phone with the browser. P5 managed to get it working, but expressed frustration with the process: *"I never really understood exactly what I was doing or what was required when logging in. I figured out the steps to make it work but don't understand the meaning or process."* Five participants said the installation and setup process should be simpler. One participant suggested adding video instructions to the text ones provided, while another suggested that the additional app download and the extension should be eliminated entirely, if possible. Those steps could indeed be eliminated if Neo were supported natively in future web browsers, though usability challenges would remain.

An important usability finding is that we observed a high rate of participant attrition and drop-out, particularly among *Neo* participants during the setup phase. To control access to the study, we utilized MTurk qualifications. After assign-

ing groups randomly among eligible participants from the screening survey, there were 123 participants in the *Password* condition and 115 participants in the *Neo* condition. At the conclusion of the setup phase, only 45% (52) of the participants assigned to *Neo* remained. Comparatively, 91% (112) of the participants assigned to *Password* remained. This difference in attrition during onboarding across conditions was significant ($\chi^2(1) = 64.596, p < .001$).

At each stage of the setup process for *Neo*, we saw participants leave the study. The two steps that resulted in the worst attrition were installing the application (16 participants dropped out) and creating an account (15 participants dropped out). As mentioned in Section 3.3, we posited that installing the application would cause problems. However, we did not predict that account creation would trail so closely. It is possible that participants did not have a compatible device to complete the credential creation process, or they could have reached a threshold of frustration with the entire onboarding process. In Section 4.1, we detailed that our final sample for *Neo* consists of 31 participants who completed all phases of the study. The remaining attrition occurred longitudinally.

4.3 Daily Authentications (RQ 2, RQ 5)

In the longitudinal phase, participants authenticated ten times over 14 days. We measured the additional participant attrition, errors logging in, the time authentication took, and participants' perceptions of the usability of logging in.

Attrition: During the longitudinal portion of the study, participants left the study at similar rates across the *Password* and *Neo* groups ($\chi^2(1) < .001, p = 1.000$). Among participants who successfully completed the setup process, 60% of *Neo* participants and 59% of *Password* participants completed all remaining parts of the full protocol.

Authentication Errors: Participants attempted to log into the banking application ten times in 14 days, and we recorded each time whether they successfully authenticated using their assigned mechanism. Although we provided no training for the *Neo* condition, Figure 3's jump in authentication success rate from Day 1 to Day 2 demonstrates quick learning. Across authentication attempts for all days, 98% of attempted *Password* authentications were successful, while 87% of attempted *Neo* authentications were successful.

To quantify how authentication failures varied across groups and changed over time, we created a mixed-effects logistic regression model. Authentication success was the dependent variable, while the assigned group, the day in the study, and the interaction of those terms were the independent variables. As this data is not independent, the participant was modeled as a random effect. Table 3 presents our model. As suggested above, we found that *Neo* participants were less likely than *Password* participants to authenticate successfully

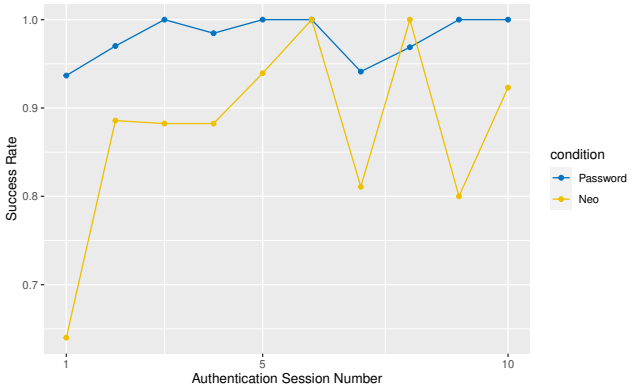

Figure 3: The success rate of authentication attempts over the ten authentication sessions, split by condition.

($OR = 0.222$, $p = .028$). We observed a marginally significant effect in that a participant was more likely to authenticate successfully the further the participant was in the fourteen-day longitudinal protocol ($OR = 1.179$, $p = .056$). Figure 3 shows that the lowest rate of successful *Password* authentications (94%) occurred during the first authentication session of the study. Comparatively, the authentication success rate was 64% for *Neo*. We defined an authentication session as all authentication attempts that occurred in a 10-minute span.

Fingerprint scans were one cause of errors for *Neo* participants. Participants noted that the reliability of a fingerprint scanner varies, and they may not work in certain scenarios (e.g., when one's finger is wet). A few participants mentioned that they did not use biometrics on their phone before this study, and others mentioned issues with their fingerprint scanners during the study. P11 described how they *"had to add extra finger scans into [their] phone in order to get it to work better."* Some participants brought up reliability issues with the Neo platform prototype itself, saying that they sometimes had to try multiple times to receive push notifications.

Timing Data: For each authentication attempt, we logged when the user landed on the login page and when they completed authentication (pressing submit or approving the push). The average times to authenticate for *Neo* and *Password* were 20.9 seconds and 8.1 seconds, respectively. In our mixed-effects linear regression model (Table 4), we found that *Neo* participants took significantly longer to authenticate than *Password* participants ($\beta = 13.708$, $p < .001$). We also observed a marginally significant result that authentication took slightly less time as the study progressed ($\beta = -0.217$, $p = .083$).

Ease of Authentication: After each authentication, participants responded on a Likert scale ("strongly disagree" to "strongly agree') to the statement "logging in to this application is easy." We again built a model, this time a mixed-effects ordinal regression model (Table 5). We found that *Neo* par-

Factor	Baseline / (Type)	Odds Ratio	95% CI	σ	z	p
Group: *Neo*	Password	0.222	[0.058, 0.850]	0.685	-2.198	**.028**
Day in Study	(Continuous variable)	1.179	[0.996, 1.397]	0.086	1.914	.056
Group: *Neo* * Days in Study	(Interaction effect)	0.909	[0.745, 1.108]	0.101	-0.947	.344

Table 3: A mixed-effects logistic regression model of participants' success (1) or failure (0) logging in on each day of the longitudinal study. The independent variables (*IVs*) were the participant's assigned group and how many days into the longitudinal study they were, as well as the interaction between the two. We report the odds ratio and 95% confidence interval of the odds ratio (*95% CI*). We also note the baseline (for categorical predictors) or the data type of the IV, as applicable.

Factor	Baseline / (Type)	β	95% CI	SE	DF	t	p
Group: *Neo*	Password	13.708	[8.574, 18.841]	2.627	168.773	5.219	**<.001**
Day in Study	(Continuous variable)	-0.217	[-0.462, 0.028]	0.125	910.706	-1.734	.083
Group: *Neo* * Days in Study	(Interaction effect)	-0.219	[-0.673, 0.235]	0.232	907.666	-0.943	.346

Table 4: A mixed-effects linear regression model of the time it took participants to authenticate on each day of the longitudinal study. The IVs and terminology are the same as in Table 3.

ticipants consistently had lower agreement that logging in was easy compared to *Password* participants ($OR = 0.012$, $p < .001$). As the study progressed, participants had higher agreement that logging in was easy ($OR = 1.134$, $p = .002$), even more so in the *Neo* group ($OR = 1.140$, $p = .022$). When examining the data over time for *Neo*, we found that the lowest percentage of responses agreeing or strongly agreeing that logging in was easy (73%) occurred on the first day of the study. Across the entire study, 90% of the *Neo* responses agreed or strongly agreed that logging in was easy. This number was 99% for *Password*.

Usability: At the study's conclusion, we asked participants to complete an additional SUS questionnaire to understand if their perceptions of the usability of Neo had changed over the course of the study. Figure 4 summarizes the results from the exit SUS. The average *Neo* SUS score was 81.3 in the exit survey, compared to 66.6 in the initial survey. When transforming those scores using the adjective scale from Bangor et al. [4], Neo received an "OK" rating in the initial survey and a "Good" rating in the exit survey. Comparatively, passwords received an "Excellent" rating in both surveys, with average SUS scores of 88.6 and 90.4 for the initial and exit surveys, respectively. Exit survey SUS scores for *Password* were higher than for *Neo* (MWU, $U = 1393.5$, $p = .002$).

SUS scores for *Neo* were also higher in the exit survey than in the initial survey (Paired Wilcoxon, $W = 636.5$, $p = .003$). As P12 said, *"While [Neo] may seem unfamiliar, it is quick to set up and easy to learn, and it makes getting into your account quick and straightforward."* Although *Neo* received lower SUS scores than passwords, 23 of 31 *Neo* participants nonetheless described Neo as "simple," "easy-to-use," or "straightforward" in free-response data. We also found that participants with prior experience with push notifications for 2FA found Neo more usable (MWU, $U = -1.965$, $p = .050$).

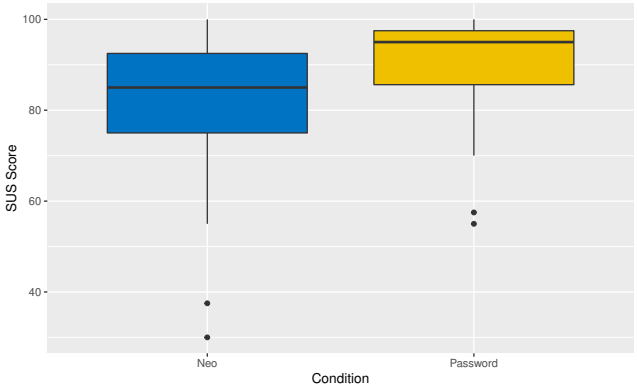

Figure 4: SUS scores from exit survey by condition

4.4 General impressions of *Neo* (RQ 3, RQ 4)

Security: All *Neo* participants expressed a belief that Neo was secure or trustworthy because of either its requirement for physically possessing one's phone or the use of biometric fingerprint scans.

> P31: *"I definitely think this authentication method makes online accounts safer. The fact that it sends a notification to your phone and requires your fingerprint makes me feel that my account is safe and secure because only I can authenticate the logins."*

P27 even mentioned that they believed Neo protected them from *"the problem of SIM card hacking."*

Availability: Participants mentioned a number of concerns about using Neo for authentication. The most common concern was phone availability (18 participants). Some participants described needing to have your phone physically nearby as annoying: *"It's a minor annoyance to try to log in when*

Factor	Baseline / (Type)	Odds Ratio	95% CI	σ	z	p
Group: *Neo*	Password	0.012	[0.002, 0.064]	0.832	-5.270	**<.001**
Day in Study	(Continuous variable)	1.134	[1.045, 1.230]	0.042	3.024	**.002**
Group: *Neo* * Days in Study	(Interaction effect)	1.140	[1.019, 1.275]	0.057	2.298	**.022**

Table 5: A mixed-effects ordinal regression model of participants' agreement (5 = "strongly agree"; 1 = "strongly disagree") that "logging in to this application is easy" on each day of the study. The IVs and terminology are the same as in Table 3.

your phone is across the room or in another room charging" (P4). Participants also described other availability challenges, such as a phone running out of battery, not having internet access, or being broken. For example, P24 wrote, *"If the phone breaks or is forgotten somewhere (I know this is probably uncommon), I didn't really see an alternative way to log in or secure your account."* When asked at the end of the study how frequently they were generally unable to access their mobile device when they needed it, 10 participants said "once a day," while the rest said at most "once a week." Notably, 10 other participants said that this "almost never" happened.

Related to availability challenges, participants raised concerns about account recovery or a backup authentication method. They pointed out how Neo lacked an obvious recovery/backup method, and this could cause them to be locked out of their accounts. For example, P32 wrote, *"If I can't authenticate with my phone and there is not a backup login procedure, then I can't login to my account."*

Privacy: Three participants were concerned about the privacy implications of using Neo. Regarding fingerprints, P13 wrote, *"[Neo] requires a thumbprint on the phone currently to use it, so people concerned about the privacy of that cannot use it."* One participant mentioned that they feel like Neo gives the banking institution too much information, while another said they would not be comfortable setting up Neo on someone else's client while away from their computer.

Deployment: Fifteen participants offered concrete suggestions for improving Neo. Participants mentioned wanting alternatives to using a fingerprint for locally authenticating to their smartphone (as part of the process of the phone serving as a roaming authenticator). They suggested facial recognition, a PIN, or behavioral authentication. P15 wrote, *"This is probably far-fetched but maybe in the future . . . it just knows you are the one holding the phone. Instead of giving me a popup to select an action, it simply registers your fingerprint when holding the phone and logs you in."* Participants voiced the need for account recovery/backup methods to be available. Finally, some participants commented that the UI was plain and should be improved.

Adoption: Eight participants mentioned (unprompted) that they would use Neo if it were widely available for authentication. One additional participant said they would use it for 2FA, but not as their primary form of authentication. P24 wrote, *"For relatively unimportant account (like dating or streaming services), this is already enough for me to use it as long as I feel like Neo is a trustworthy and secure company."* Four participants explicitly stated that they believe the benefits of Neo outweigh the additional effort it requires. For example, P27 wrote, *"It is a little more 'difficult' than just entering a password, since it involves another step (grabbing your phone and opening up the authentication app), but the added security makes it worth it."*

When asked how likely ("very likely" to "not likely") they would be to use Neo over passwords for six different account types (dating services, streaming services, social media, health care services, banking, and email), over half of *Neo* participants said they were "likely" or "very likely" to use Neo over passwords for all account types except for banking. Streaming services ranked the highest with 61%. Banking ranked the lowest, with less than half (39%) of *Neo* participants being likely to use Neo over a password.

4.5 Comparisons with Alternatives (RQ 4)

Some *Neo* participants made comparisons between Neo and other authentication schemes they had used. Thirteen participants described benefits they perceived Neo as having relative to passwords. These benefits (in order of decreasing prevalence) included not having to remember/store passwords, the security benefits of using biometrics (instead of a password that might be cracked), and ease of use. For example, P13 wrote, *"It's also very easy to use because you just have to use your thumbprint to verify that it's you rather than taking the time to type out a password and guessing which password you used for which account."* Individual participants also mentioned that they found Neo easier to use than password managers or email/SMS PINs. Conversely, one participant described Neo as frustrating relative to alternative schemes:

> P11: *"I didn't really like it. I thought it was a bunch of extra unnecessary steps just to log in and do some simple tasks. It got easier to use, but was still clunky and I really didn't like it . . . There are easier and better ways to do authentication that aren't as frustrating or unnecessary."*

Some participants described Neo as being similar to 2FA. For example, P18 wrote, *"It provides authentication like 2FA. I feel it makes things somewhat safer."*

Password participants confirmed findings from prior work on passwords [7], including that they found passwords simple, familiar, and easy to use. Some participants specifically called out how learnable passwords are, even for people with little technical expertise: *"The authentication was easy to use and not too complex . . . It is easy to use for those with very little computer knowledge or skills"* (P49). A few *Password* participants mentioned that they liked that they did not need a second device to authenticate. Others described the historical resiliency of passwords as a sign of its strength as an authentication scheme: *"It's been proven to be quite secure (when done properly) over decades of use"* (P95) and *"No one has yet come up with something worth the trade-offs"* (P91).

When asked about disadvantages of passwords, *Password* participants overwhelmingly mentioned security. Their concerns regarding password security included weak passwords, others learning one's passwords, password reuse, and the risks of browsers' auto-fill login if someone gains access to their devices. Several participants said that our fictitious banking application's password policy should have been stronger to ensure they created secure passwords. One participant also discussed the lack of a CAPTCHA on the login page, mentioning how bots could hack accounts. Several participants raised the lack of multi-factor authentication (*MFA*) as a weakness of our implementation of password authentication:

> P40: *"Nowadays, it feels a little vulnerable for something like banking not to require a two-step validation process using a texted or emailed validation code. . . . If there was no second step to verify the user generally, I might be a little concerned."*

Participants suggested different forms of MFA that could improve passwords, including an email/SMS code, security questions, physical presence, and biometrics (facial recognition, fingerprints). As P76 wrote, *"I still think two-factor authentication can protect the safety of online accounts better on top of the traditional password authentication method. [Add] face recognition, SMS/email/app authentication, physical authentication assure the users that they're more protected."*

5 Discussion

In this section, we discuss our results' implications for efforts to spur adoption of smartphones as roaming authenticators.

5.1 Separating Setup and Day-to-Day Use

Reynolds et al. recommended that researchers study setup and day-to-day authentication separately when evaluating authentication schemes so that problems during the setup phase would not impact participants' perceptions of usability for day-to-day use [39]. In our initial study plan, we intended to help participants set up Neo in-person. However, the COVID-19 pandemic shifted our study online and changed plans such that participants had to set it up themselves. Given the improvement in SUS scores for both conditions between our initial and final surveys, we believe that the overall experience for participants in both condition was impacted by their setup experience. This is specifically evident in the significant portion of participants who dropped out of the study before completing the setup process. In our analysis of the longitudinal data and initial SUS scores, we found that *Neo* participants had comparatively worse experiences authenticating at the beginning of the study. However, as the study progressed, participants found authenticating somewhat easier, authenticated somewhat faster, and made somewhat fewer authentication errors. Moreover, *Neo* participants found the scheme more usable at the conclusion of the study than at the beginning. Separating setup and daily use also enabled us to disentangle perceptions of Neo from general perceptions of using smartphones as roaming authenticators. As different technical approaches develop for using smartphones as roaming authenticators, their implementations may have significantly different setup processes. To better understand perceptions of smartphones as roaming authenticators after continued use, researchers will need to evaluate the day-to-day use of FIDO2 implementations like Neo separately from setup.

5.2 Security vs. Usability

Timing data showed that *Password* participants authenticated more quickly than *Neo* participants at every point in the study. We attribute the authentication speed for *Password* participants to both familiarity and the use of auto-fill capabilities by password managers and browsers. 25% of *Password* participants reported that they used a password manager, browser, or other tool to generate new passwords. These timing results are similar to the findings of Farke et al. [18]. Neo's consistent underperformance relative to passwords raises the question of whether highlighting an authentication method's security benefits is enough to encourage adoption.

Unlike in prior work on security keys, in which participants did not fully understand the potential benefits of security keys [14, 26], participants in our study reported that Neo was substantially more secure than passwords, yet found passwords more usable. Nonetheless, the majority of participants who used Neo during the study reported being likely to use Neo over passwords for all account types we asked about other than banking accounts. It is possible that users of password managers already receive FIDO2's best non-security related attributes (e.g., memorylessness, decreased cognitive load during registration). To counter similar arguments and spur adoption, implementers will need to underscore the flaws of passwords, such as the threat of phishing, credential stuffing, and data breaches, highlighting how FIDO2 avoids them.

5.3 Availability/Account Recovery

The most common concern for *Neo* participants was phone availability. For people to feel comfortable adopting smartphones as roaming authenticators, system designers must solve availability issues that arise from using a smartphone to authenticate. That is, if the user's smartphone is their only authenticator and their smartphone is inaccessible for any of the reasons detailed below, the user will not be able to log into any websites. Lyastani et al. and others have identified analogous problems for USB security keys, particularly the difficulty of account recovery and revocation if the key is lost or stolen [1, 10, 26]. Smartphones, just like security keys, can be stolen or lost, temporarily or permanently. Identifying appropriate methods for recovering from authenticator loss is still an open problem, although FIDO2 recommends registering multiple authenticators to avoid being completely locked out [20]. As participants mentioned, though, smartphones raise additional availability issues. Unlike security keys, phones can run out of battery, making it impossible for the owner to authenticate without charging the phone. Phone availability can also be impacted by limited wireless reception. Finally, phones are also higher-value targets for theft.

Shortcomings in accessibility can also present availability challenges. For example, schemes that require biometrics to verify user identity (e.g., as might be required after confirming a push notification as we did for Neo) could cause problems for people who cannot touch a security key's capacitor, who cannot use a fingerprint scanner, or for whom facial recognition is not reliable. It is important for system designers to consider a variety of ways for users to verify their identity, potentially including some that could cause their systems to lose some of their security benefits (e.g., PINs).

When asked at the end of the study how frequently they were generally unable to access their mobile device when they needed it, a third of participants said "once a day," a third said "almost never," and the other third was "once a week" or less frequently. The variety of these results make it difficult to provide general recommendations regarding these challenges. Currently, the best approach to availability challenges may be nudging or requiring users to register multiple authenticators.

Another potential way to allow users to enjoy the security benefits of authentication methods like Neo while also considering account recovery is to enable email-based account recovery, as is typical for passwords. If a user breaks or no longer has their registered authenticator, they could receive a link in their email account to register another type of authenticator. Of course this means users have to remember at least one password, similar to using a password manager. Of course they would need to *not* be using Neo on their email account, and they would need to have a strong password for their email account. However, if they have backup unregistered authenticators at hand (e.g., old phones or platform authenticators on desktop devices), this approach could provide the usability benefits of password managers while providing the security benefits of using smartphones as roaming authenticators within FIDO2 passwordless authentication.

6 Related Work

Oogami et al. [29] conducted the first study evaluating the usability of smartphones as WebAuthn-enabled *platform authenticators*. In 2018, their website (yahoo.co.jp) was the first commercial portal to let users choose to log in from their smartphones using WebAuthn with a fingerprint. Conversely, we evaluated the use of smartphones as *roaming authenticators*. Out of their 10 participants, only three were able to complete the registration process without assistance. Although their registration process was significantly different than ours, participants in our study similarly struggled with setup.

Lyastani et al.'s [26] between-subjects lab study (N = 94) evaluated the usability of security keys with FIDO2 passwordless authentication. The authors sought to understand users' perceptions, acceptance, and concerns when using security keys for FIDO2 passwordless authentication. They found that passwordless authentication with security keys was seen as both more usable and more acceptable than passwords. Our study builds on this work, but focuses on using smartphones (instead of security keys) as roaming authenticators. While participants in their study preferred passwordless authentication with a security key to passwords, they were also concerned about account recovery and account revocation. Our participants raised these same concerns. Farke et al. [18] conducted a similar experiment in the context of a small company. Like us, they found that participants were concerned about the availability of their authenticators (security keys) in terms of physical location (e.g., losing the authenticator) and functionality (e.g., a malfunctioning authenticator).

Owens et al. [30] presented a framework for evaluating authentication schemes that use smartphones as FIDO2 roaming authenticators. They specifically highlighted user perceptions of phone availability and account recovery challenges as potential focus areas for researchers. The data we collected included the types suggested by their framework. Bonneau et al. [7] proposed a framework for evaluating web authentication schemes. This framework used 25 properties to rate 35 password-replacement schemes on usability, deployability, and security. Their expert evaluation found that no scheme analyzed offered the same benefits as passwords. Prior work has evaluated WebAuthn using this framework [13, 18, 26, 27].

To address the challenge of account recovery, Conners and Zappala [13] proposed the Let's Authenticate alternative to FIDO2 based on certificates instead of keys. Certificates are issued after users prove ownership of an account with a username and password, facilitating re-issuance if an authenticator is lost. Credential recovery and revocation problems are critical for roaming authenticators like security keys and smartphones. While Let's Authenticate eliminates the burden

of registering an authenticator with every web service, it also introduces a new trusted third party.

Klieme et al. [23] proposed an extension to FIDO2/WebAuthn that would allow continuous authentication over BLE. They created a proof-of-concept Android application to serve as a roaming authenticator and implemented a custom relying party that supported their extension. Due to browsers' lack of support for custom FIDO2/WebAuthn extensions, they *simulated* (rather than tested) extension processing by adding functionality to their custom relying party. We work around this current browser support challenge by using a Chrome browser extension to simulate Network Transport functionality.

A number of researchers have studied the usability of mobile phones as a *second* factor for authentication, including via SMS codes, TOTP codes, and push notifications [12, 15, 24, 38, 45]. Weidman and Grossklags [45] studied a transition from token-based 2FA to a push-notification 2FA system, finding that employees preferred the token-based system to the Duo app. Colnago et al. [12] studied the deployment of 2FA via the Duo app at their university. They found that 2FA adopters found it annoying, yet easy to use. Neo uses a similar push notification mechanism for authentication.

7 Limitations and Future Work

As in many user studies, our findings are somewhat limited in their generalizability by the small sample size. Additionally, in both experimental conditions, participants logged into a fictitious banking website. Consequently, they did not experience any real risk or incentive during authentication. This could cause *Password* participants to create weaker passwords than they otherwise might, and generally cause participants to behave differently than they might in real-life scenarios. We attempted to simulate risk by having participants perform simulated transactions within the banking application. However, there was no reward associated with protecting the assets in the accounts. To better simulate risk, future work could adapt the approach from Redmiles et al. [35] and assign a probability of a participant's account being "compromised" based on the characteristics of the password they created. Moreover, it is possible that users may have exhibited different behaviors or perceived things differently if the study website had a different focus (e.g., social media). Future work could explore those differences by conducting a between-subjects experiment with additional conditions that mimic other well-known web services.

Our participant pool also likely impacted our results. Because users with prior 2FA experience with push notifications are over-represented (45% vs. 19% in the general US population, according to Engler [17]), and we found that this prior experience made participants view Neo as more usable, the results from this study could be seen as overly optimistic. Although MTurk users are often more diverse in terms of age, income, education level, and geography than traditional social science pools, they are also younger, Whiter, and more tech-savvy than the general US population [34]. Because we required that participants live in the US (to reduce confounding factors), our results are not reflective of global populations. Future work should study more diverse populations.

As previously discussed, *Neo* overall had a far greater attrition rate than *Password* despite random assignment. We listed several potential causes in Section 3.3. Some amount of the dropout was likely a result of the difficulty associated with setting up Neo. Thus, our final set of participants may be biased and present overly optimistic results. However, if this effect were strongly present, one might expect Neo to have been found to be more usable than passwords. We found the opposite. Future work should study the setup process and daily use separately, even more closely tracking attrition. We also observed a jump in the *Neo* authentication success rate from Session 1 to 2. We speculate that this jump reflects the learning effect for a new system, although we cannot speak definitively about what aspects were barriers in Session 1. Studying this increase is an avenue for future work.

We tested only a simple password-composition policy. Future work should test a variety of different policies and separately test password managers to further understand how WebAuthn compares to various password use cases. While our participants were Chrome users, our study platform did not enforce the use of Google Chrome when registering or during authentication. This means that participants in *Password* could have used other browsers, introducing a confound. Finally, future work should study using platforms like Neo across multiple websites. A user has to register separately each time they want to add Neo as an authenticator on a new website; we only studied its usability on a single website.

8 Conclusion

We conducted a between-subjects (N = 97), longitudinal study of FIDO2 passwordless authentication with smartphones as roaming authenticators. Participants recognized the security benefits of the Neo smartphone-based passwordless authentication scheme, yet still found passwords to be more usable. Nonetheless, many participants were willing to use Neo over passwords for five of the six account types we asked about. Participant were acutely aware of challenges associated with losing an authenticator and stressed the need for account recovery methods. Participants suggested that the setup process for Neo be simplified, that different ways of verifying user presence (e.g., PIN, facial recognition) be made available for authentication, and that account recovery/backup methods be added. While some of the concerns participants had (e.g., setup issues) were unique to the design of Neo, we believe our findings highlight issues and opportunities designers of smartphone-based FIDO2 passwordless authentication must consider when implementing new schemes.

Acknowledgments

We would like to thank our anonymous shepherd and reviewers for their helpful feedback. We would also like to thank multiple people at Duo Security for their support of this project, including Bronwyn Woods, Jeremy Erickson, Nick Mooney, Nick Steele, Rich Smith, Brian Lindauer, and the entire Data Science Team.

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. EuroUSEC*, 2017.

[2] Eldridge Lee Alexander, James Leslie Barclay, Nicholas James Mooney, and Mujtaba Hussain. Identity Services for Passwordless Authentication. US Patent US20200403993A1, December 2020.

[3] FIDO Alliance. FIDO2: WebAuthn & CTAP, May 2020. https://web.archive.org/web/20200512070231/https://fidoalliance.org/fido2/.

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

[5] Lujo Bauer, Cristian Bravo-Lillo, Elli Fragkaki, and William Melicher. A Comparison of Users' Perceptions of and Willingness to Use Google, Facebook, and Google+ Single-sign-on Functionality. In *Proc. DIM*, 2013.

[6] Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. Biometric Authentication on iPhone and Android: Usability, Perceptions, and Influences on Adoption. In *Proc. USEC*, 2015.

[7] 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 *Proc. IEEE S&P*, 2012.

[8] Dhiman Chakraborty and Sven Bugiel. simFIDO: FIDO2 User Authentication with simTPM. In *Proc. CCS*, 2019.

[9] Dhiman Chakraborty, Lucjan Hanzlik, and Sven Bugiel. simTPM: User-centric TPM for Mobile Devices. In *Proc. USENIX Security*, 2019.

[10] Stéphane Ciolino, Simon Parkin, and Paul Dunphy. Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling. In *Proc. SOUPS*, 2019.

[11] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[12] 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 *Proc. CHI*, 2018.

[13] James S. Conners and Daniel Zappala. Let's Authenticate: Automated Cryptographic Authentication for the Web with Simple Account Recovery. In *Proc. WAY*, 2019.

[14] Sanchari Das, Andrew Dingman, and L. Jean Camp. Why Johnny Doesn't Use Two Factor: A Two-Phase Usability Study of the FIDO U2F Security Key. In *Proc. FC*, 2018.

[15] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *Proc. USEC*, 2014.

[16] Department of Health and Human Services. System Usability Scale (SUS), Sep 2013. https://web.archive.org/web/20200201012855/https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

[17] Maggie Engler. 2019 State of the Auth: Experiences and Perceptions of Multi-Factor Authentication. Duo Security E-book, December 2019. https://duo.com/assets/ebooks/state-of-the-auth-2019.pdf.

[18] Florian M. Farke, Lennart Lorenz, Theodor Schnitzler, Philipp Markert, and Markus Dürmuth. "You still use the password after all"–Exploring FIDO2 Security Keys in a Small Company. In *Proc. SOUPS*, 2020.

[19] Joseph L. Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 2013.

[20] Hidehito Gomi, Bill Leddy, and Dean H. Saxe. Recommended Account Recovery Practices for FIDO Relying Parties. *FIDO Alliance*, 2019. https://web.archive.org/web/20210520070746/https://fidoalliance.org/recommended-account-recovery-practices/.

[21] Troy Hunt. Passwords Evolved: Authentication Guidance for the Modern Era, 2020. https://web.archive.org/web/20200501185526/https://www.troyhunt.com/passwords-evolved-authentication-guidance-for-the-modern-era/.

[22] Ryan Kennedy, Scott Clifford, Tyler Burleigh, Ryan Jewell, and Philip Waggoner. The Shape of and Solutions to the MTurk Quality Crisis. *SSRN*, 2018. https://www.ssrn.com/abstract=3272468.

[23] Eric Klieme, Jonathan Wilke, Niklas van Dornick, and Christoph Meinel. FIDOnuous: A FIDO2/WebAuthn Extension to Support Continuous Web Authentication. In *Proc. TrustCom*, 2020.

[24] 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 *Proc. USEC*, 2015.

[25] Leona Lassak, Annika Hildebrandt, Maximilian Golla, and Blase Ur. "It's Stored, Hopefully, on an Encrypted Server": Mitigating Users' Misconceptions About FIDO2 Biometric WebAuthn. In *Proc. USENIX Security*, 2021.

[26] Sanam Ghorbani Lyastani, Michael Schilling, Michaela Neumayr, Michael Backes, and Sven Bugiel. Is FIDO2 the Kingslayer of User Authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication. In *Proc. IEEE S&P*, 2020.

[27] Robbie MacGregor. Evaluating the Android Security Key Scheme: An Early Usability, Deployability, Security Evaluation with Comparative Analysis. In *Proc. WAY*, 2019.

[28] Nick Mooney. Addition of a Network Transport, 2020. https://github.com/w3c/webauthn/issues/1381.

[29] Wataru Oogami, Hidehito Gomi, Shuji Yamaguchi, Shota Yamanaka, and Tatsuru Higurashi. Poster: Observation Study on Usability Challenges for Fingerprint Authentication Using WebAuthn-enabled Android Smartphones. In *Proc. SOUPS Posters*, 2020.

[30] Kentrell Owens, Blase Ur, and Olabode Anise. A Framework for Evaluating the Usability and Security of Smartphones as FIDO2 Roaming Authenticators. In *Proc. WAY*, 2020.

[31] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why People (Don't) Use Password Managers Effectively. In *Proc. SOUPS*, 2019.

[32] Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States, 2019. https://web.archive.org/web/20210606233708/https://www.pewresearch.org/internet/fact-sheet/mobile/.

[33] Suby Raman. Guide to Web Authentication, 2021. https://webauthn.guide.

[34] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *Proc. IEEE S&P*, 2019.

[35] Elissa M. Redmiles, Michelle L. Mazurek, and John P. Dickerson. Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions. In *Proc. EC*, 2018.

[36] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In *Proc. USENIX Security*, 2020.

[37] Ken Reese. 2FA Banking Website. https://bitbucket.org/isrlauth/ken-bank-thesis/src/master/, 2020.

[38] Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons. A Usability Study of Five Two-factor Authentication Methods. In *Proc. SOUPS*, 2019.

[39] 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 *Proc. IEEE S&P*, 2018.

[40] Scott Ruoti, Brent Roberts, and Kent Seamons. Authentication Melee: A Usability Analysis of Seven Web Authentication Systems. In *Proc. WWW*, 2015.

[41] Aaron Smith. Americans and Cybersecurity. Pew Research Center, January 2017. https://web.archive.org/web/20210514203628/https://www.pewresearch.org/internet/2017/01/26/2-password-management-and-mobile-security/.

[42] San-Tsai Sun, Eric Pospisil, Ildar Muslukhov, Nuray Dindar, Kirstie Hawkey, and Konstantin Beznosov. What Makes Users Refuse Web Single Sign-on? An Empirical Investigation of OpenID. In *Proc. SOUPS*, 2011.

[43] Joshua Tan, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Practical Recommendations for Stronger, More Usable Passwords Combining Minimum-strength, Minimum-length, and Blocklist Requirements. In *Proc. CCS*, 2020.

[44] W3C. Web Authentication, 2019. https://www.w3.org/TR/webauthn/.

[45] Jake Weidman and Jens Grossklags. I Like It, But I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. In *Proc. ACSAC*, 2017.

[46] Yubico. User Presence vs. User Verification, 2021. https://web.archive.org/web/20210605113506/https://developers.yubico.com/WebAuthn/WebAuthn_Developer_Guide/User_Presence_vs_User_Verification.html.

[47] Yubico. Works with YubiKey Catalog (FIDO2), 2021. https://www.yubico.com/works-with-yubikey/catalog/#protocol=fido2&usecase=all&key=all.

[48] Tin Zaw and Richard Yew. 2017 Verizon Data Breach Investigations Report (DBIR) from the Perspective of Exterior Security Perimeter. Verizon Media Platform, 2017. https://web.archive.org/web/20200409012027/https://www.verizondigitalmedia.com/blog/2017-verizon-data-breach-investigations-report/.

A Screening Survey

[This survey was sent to MTurkers who accepted our HIT.]

Please complete the following 1 minute screening survey to see if you qualify for a two-week longitudinal study on online authentication. This study may require you to install a mobile application and/or a Chrome extension. You will be asked to login into a web application ten times over the course of two weeks, completing a simple task each time. Completing all of the tasks over two weeks should not take more than 75 minute total. After completing the first task you will take an initial survey. After two weeks pass, we will send you a link to a final survey. Upon adequate completion of the final survey, you will receive your $30 in compensation as a bonus.

Do not take this survey unless you meet the following criteria:

- Have an Android phone with a fingerprint sensor
- Have Android version 9.0+ (to check what version of Android you have, go to your phone's "Settings," and search "Android version," or you can visit the following website from your mobile phone: https://whatismyandroidversion.com/)
- Have Google Chrome installed on your computer
- Are an adult currently living in the USA

If you take this screening survey and do not meet the above criteria, you will not receive compensation.

If you qualify for the survey, we will message you with more details about the study and send the $30 as a bonus upon completion of the longitudinal study.

By taking the survey, you are agreeing to the non-disclosure agreement found at the following link: https://bankoferie.com/nda ◯ I agree ◯ I do not accept and will not participate in this study

Please enter your MTurk ID below. Please ensure that it is correct to ensure that we are able to compensate you for your participation. _____

Do you have an Android mobile device? ◯ Yes ◯ No

If you answered yes to the above question, what Android software version do you have do have? To check this, go to your phone's "Settings," and search "Android version" or you can visit https://whatismyandroidversion.com/ from your mobile phone. _____

Are you an adult (18+ years old) currently living in the United States? ◯ Yes ◯ No

Do you have Google Chrome installed on your computer? ◯ Yes ◯ No

Does your Android phone have a fingerprint sensor? ◯ Yes ◯ No

B Survey Instrument

[Below we highlight the difference between the initial/exit surveys received by participants in the Neo and passwords conditions after the completion of their tasks.]
Please enter your MTurk ID below. Please ensure that it is correct to ensure that we are able to compensate you for your participation. _____

Please enter the username that you used for the study. Please ensure that it is correct to ensure that we are able to compensate you for your participation. _____

B.1 System Usability Scale

In the following survey, the word "system" refers to the [mobile phone-based or passwords-based] authentication method you used to log into your account. Please state your level of agreement or disagreement for the following statements based on your experience with this system. There are no right or wrong answers.

[Response choices: ◯ Strongly agree ◯ Agree ◯ Neither agree nor disagree ◯ Disagree ◯ Strongly disagree]
Questions:

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very awkward to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

B.2 Additional questions

These questions are about your experience with [setup or day-to-day use] of the [mobile phone-based or passwords-based] method you used to log into your account in the web application. Please answer them thoroughly and honestly. There are no right or wrong answers.

How would you describe your general experience with the authentication method you used? _____

What advantages do you see with using this authentication method? _____

What disadvantages do you see with using this authentication method? _____

[Final, Passwords only] Do you think using this authentication method is the best available for protecting the safety of your online accounts? _____

[Final, Neo only] If you were to recommend Neo to a friend, how would you describe its benefits? _____

[Neo only] How likely are you to choose Neo over passwords for the following types of accounts, if Neo were widely available?

[Final, Neo only] What changes would need to be made to Neo to make you more likely to use it? _____

[Final, Neo only] Did you previously visit the website (https://webauthn.guide) mentioned in the tutorial to learn more about WebAuthn? ◯ Yes ◯ No ◯ I don't remember

	Very likely		Neutral		Not likely	N/A.
Dating services (e.g. Bumble, OkCupid)	◯	◯	◯	◯	◯	◯
Streaming services (e.g. Netflix, Hulu)	◯	◯	◯	◯	◯	◯
Social media	◯	◯	◯	◯	◯	◯
Healthcare services	◯	◯	◯	◯	◯	◯
Bank	◯	◯	◯	◯	◯	◯
Email	◯	◯	◯	◯	◯	◯

[Final, Neo only] What other sources, if any, did you use to learn about WebAuthn (if you didn't use any input N/A)? _____

[Final, Neo only] Do you think using this authentication method makes an online account safer? _____

Generally, how frequently have you not been able to access your mobile phone when you needed it? ◯ Once per day ◯ Once per week ◯ Once per month ◯ Once per year ◯ Almost never ◯ Other _____

[Initial only] How do you typically choose your password for a new email account? ◯ Reuse an existing password ◯ Modifying an existing password ◯ Create an entirely new password on my own ◯ Randomly generate an entirely new password with browser/password manager/other tool ◯ I prefer not to answer ◯ Other _____

[Final, Neo only] Did you have fingerprint enabled on your Android phone PRIOR to beginning this study? ◯ Yes◯ No ◯ Other _____

[Final only] Anything else you'd like to add? _____

[Initial only] Have you ever been a victim of account compromise/hacking? ◯ Yes (please briefly describe the incident)_____ ◯ No

B.3 Demographic Info

[Initial only] Please choose the range that includes your age. ◯ 18-24 years old ◯ 25-34 years old ◯ 35-44 years old ◯ 45-54 years old ◯ 55-64 years old ◯ 65-74 years old ◯ 75+ years old

[Initial only] Please choose your race/ethnicity (select all that apply). ◯ American Indian or Alaska Native ◯ Asian ◯ Black or African American ◯ Hispanic ◯ Native Hawaiian or Other Pacific Islander ◯ White ◯ Other _____

[Initial only] What is your gender? _____

[Initial only] Please indicate if you have a computer science background. ◯ Yes ◯ No

[Initial only] Please indicate your highest educational degree. ◯ High School Diploma/GED ◯ Some college but no degree ◯ Associate's degree ◯ Bachelor's degree ◯ Professional degree (e.g. Master's, PhD, MD, JD) ◯ Other _____

[Initial only] What forms of two-factor authentication have you used in the past, if any? ◯ SMS/Text Message ◯ TOTP code generator app (e.g. Google Authenticator, Authy, DUO Mobile) ◯ Pre-generated codes (that you printed or wrote down to use later) ◯ Push notification based mobile app (e.g. Google Prompt, Authy OneTouch, DuoMobile) ◯ Physical security keys (e.g. YubiKey, Titan) ◯ Other _____

C Selected screenshots from Neo setup guide

We made a setup guide to help participants successfully register and authenticate using Neo. Participants were required to pair the mobile application with their browser to share a secret via a QR code.

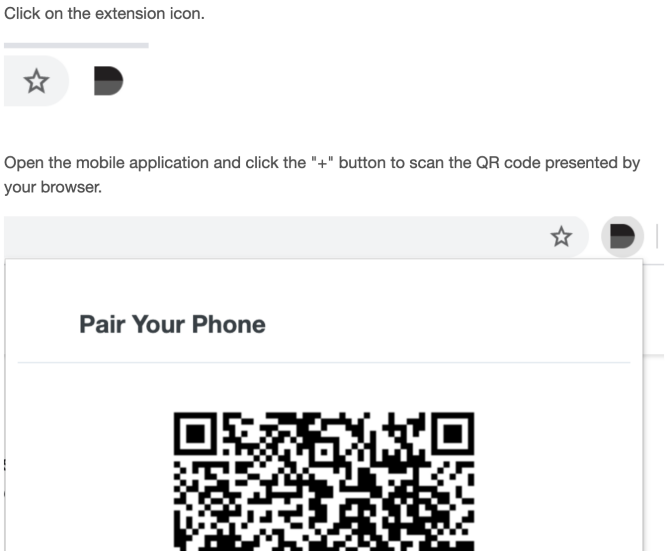

Figure 5: A screenshot of the Chrome extension during account registration from the Neo setup guide.

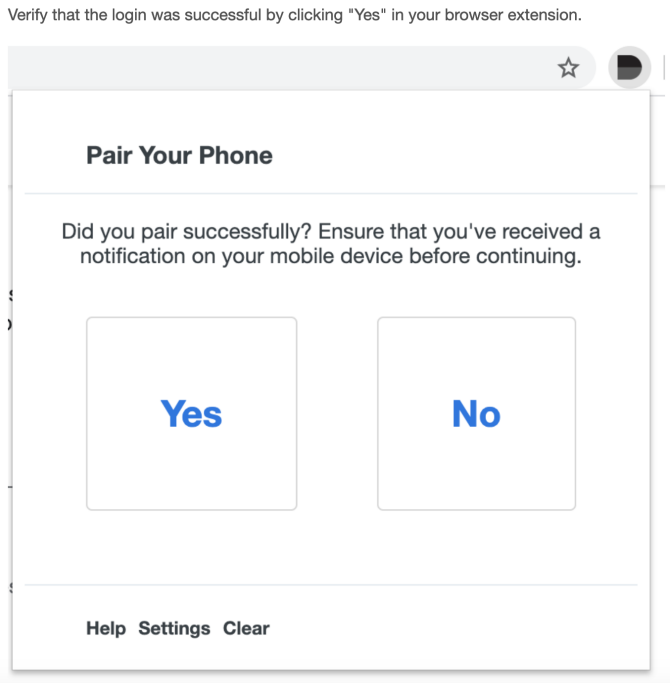

Figure 6: A screenshot of the Chrome extension prompting a user to confirm that the pairing was successful.

D Supplemental Material

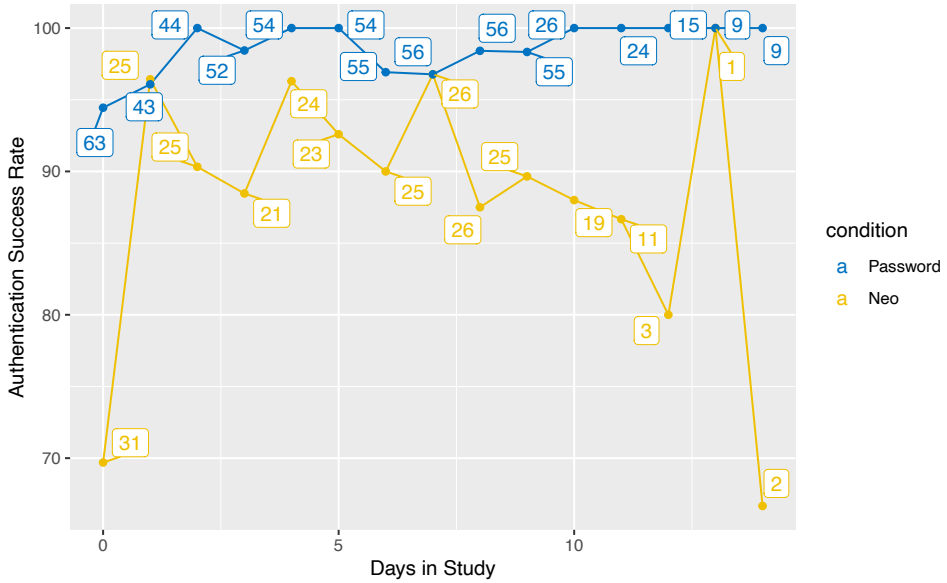

Figure 7: Aggregate authentication success rate over time by condition. The labels indicate the number of unique participants who authenticated on that day in the specified condition. We required that participants log in on ten days within a fourteen day window, and many participants simply logged in for the first ten days of the study. Note that the number of authentication *attempts* (reflected in the authentication success rate) can be greater than the number of participants in the case of failed authentication attempts. For instance, on Day 14 there were two Neo participants, one of whom logged in successfully on the first attempt and one of whom had a failed authentication attempt followed by a successful attempt.