On User Choice for Android Unlock Patterns

Marte Loge
NTNU, Norway
marte.loge@gmail.com

Markus Duermuth
Ruhr-University Bochum, Germany
markus.duermuth@rub.de

Lillian Rostad
NTNU, Norway
lilliaro@ntnu.no

***Abstract*—Android Unlock Patterns are one of the most widely used graphical password schemes. However, the scheme's security is limited by users not choosing patterns uniformly but with a specific bias. In this work we take a closer look at this bias, in particular how personal traits influence the chosen patterns. We conducted a user study with 800 participants and demonstrate that certain factors such as age, gender, and experience in IT significantly influence the strength or length of the chosen patterns. This has implications both for how we can help users to select stronger patterns and for forensic applications.**

I. Introduction

Over the last decade, mobile phones have evolved from simple tools for making voice calls to powerful computers which can be used to access emails and social media, make payments, access online banking, and store private as well as work-related sensitive information. User authentication helps protecting the sensitive information stored on the device. Authentication schemes commonly used on smartphones include (i) knowledge-based schemes, mostly PINs and (graphical) passwords; (ii) biometric schemes, mostly fingerprint recognition (recent iPhone models and other high-end models) and face recognition (e.g. offered on Android since version 4.0); (iii) security tokens are rarely used on smartphones (however, a smartphone is commonly used as second factor to authenticate to another device or account). PINs and Android Unlock patterns are the most frequently used schemes, but studies are not conclusive to which one is used more often (cf. [27], [17], [16]).

User-chosen authentication secrets are known to be relatively predictable, regardless if they are PINs [10], passwords [20], or graphical passwords [12], [14], [26], and can therefor be determined by guessing attacks. Even more, certain observable properties of a person have been shown to influence the selected authentication secret. This effect has been demonstrated for PINs [10], where it was demonstrated that knowledge of a person's birthday significantly accelerates guessing the person's PIN, for passwords [11], where it was shown that knowing background information such as birthday, occupation, and friends can improve guessing success by around 5%, and for the graphical password scheme PassFace [12], where it was found that faces were selected with a strong bias based on race and gender.

In this work, we study the effect that personal traits of a user have on his selection of Android Unlock Patterns. This was, to the best of our knowledge, never studied before, the only exception being independent and concurrent work by Aviv et al. [7] which studied the influence of collection methods and personal traits on the collected patterns. However, they only studied the influence of personal traits on specific characteristics of the patterns, such as length, starting point, and occurrence of crosses and knight-moves. In this work, we use a Markov model-based meter to approximate the strength of individual patterns and thus being able to analyze the influence of personal traits on the strength of individual patterns.

We conducted an online survey that asked users to use the Android Unlock Pattern scheme to secure access to (i) a shopping account, (ii) a smartphone, and (iii) a bank account. In addition, we asked participants to answer a questionnaire which contained standard demographic questions, and specifically questions about factors that we believed may influence the strength of the chosen patterns, such as gender, background in IT or IT security, handedness, and others. We show that age and gender have a significant influence on the average strength of the patterns chosen (both male users and younger users choose stronger patterns on average), while somewhat surprisingly having experience in IT or IT security did not have a statistically significant influence (but still had a statistically significant influence on the pattern length).

Our work helps us to understand some of the factors behind weak user-selected authentication secrets, and may hint at directions to help users avoiding weak patterns. Our work also shows directions to speed up the guessing of authentication secrets in forensics, but more work is required before usable results can be obtained.

Outline. We discuss related work in Section II, before describing details about the Android Unlock Pattern scheme in Section III. We present the design of our user study in Section IV and the results in Section V. We discuss these results in Section VI and conclude with some final remarks in Section VII.

II. Related work

Graphical passwords. Graphical passwords have the potential to offer easier-to-use authentication, as there is indication that graphical information is easier to remember by humans [13]. Recently they found wide-spread adoption specifically on

EuroUSEC '16, 18 July 2016, Darmstadt, Germany

http://dx.doi.org/10.14722/eurousec.2016.23001

mobile devices, as they are particularly well-suited for touch-screen use, while text-based passwords are much less suited for devices without a physical keyboard.

The first description of a graphical password scheme goes back to a patent by Blonder [8], which describes a scheme where a user needs to select specific points in an image. This scheme is an example for a *cued-recall based scheme*, other examples include BDAS [15] and PassPoints [29], [30], [31]. Presumably the most widely used cued-recall based scheme is Windows Picture Password, which is quite similar to Blonder's original proposal and to the PassPoints scheme.

The classical example for a *recall-based graphical password scheme* (without a cue) is the draw-a-secret scheme (DAS) [19], where one draws free-handed on a grid. In 2007, Tao and Adams [23] modified this original idea by snapping the drawn lines to the intersections of a grid, thus removing many of the problems of ambiguities of the DAS scheme and making it much easier to use, calling the resulting scheme Pass-Go. This scheme was adopted, with some restrictions, for use in the Android mobile phones in 2008, which we will describe in more detail in Section III.

Finally, *recognition-based schemes* are based on recognizing a previously seen object, instead of recalling information. One of the classical examples is the PassFace scheme, where the user selects several pictures of faces, and has to select these faces among a number of decoy images for authentication. Several related schemes have been deplored, but to the best of our knowledge there is no scheme with significant adoption.

Security of graphical passwords. For the DAS scheme, Oorschot and Thorpe [24] analyzed the security based on mirror symmetric fragments. They constructed dictionaries that improve guessing attacks against graphical passwords and estimated the realistic space of passwords being exponentially smaller than the theoretical space. Jermyn et al. [19] analyzed the security of the DAS scheme for computer-generated passwords. However, computer-generated passwords are in practice only used for very few accounts, problems being user acceptance and low usability.

For the PassPoints scheme, Dirik et al. [14] investigated the distribution of user's choices and found substantial bias based on data collected from human users. Thorpe and Oorschot [25] used a more involved method and used click-points collected in a user-study to seed automated methods for predicting likely click-points, further facilitating and improving this kind of attack. Zhao et al. [32] evaluate the security of the graphical password scheme used in Windows 8 and propose effective guessing algorithms against them.

Android Unlock Patterns. Uellenbeck et al. [26] evaluated the security of Android Unlock patterns and found substantial bias both in the starting point as well as the path chosen by users. They precisely quantified the security of the scheme and found its security to be lower than that of a uniformly chosen 3-digit PIN. They additionally evaluated the influence of a changed layout and found that layout changes indeed have a substantial influence on the security, even with the same number of nodes. The security of variants was also studied by Aviv et al. [4], who compared Android Unlock Patterns both on the standard 3×3 grid and on a 4×4, and found a very limited increase of security on the larger grid. Arianezhad et al. [3] evaluated a gaze-based variant of the scheme using an eye-tracker, and reported statistics about start- and end-points, frequent stroke directions, and similar for several arrangement of contact points.

The influence of strength meters on user-choice for Android Unlock Patterns was tested by Andriotis et al. [1] and by Song et al. [21], who used a very elaborate study setup.

Attacks beyond guessing attacks were considered by Aviv et al. [5], who used "smudges" left on the smartphone screen while entering a pattern to reconstruct the user's secret. Andriotis et al. [2] extended this attack by incorporating statistics about patterns typically chosen by users. The accelerometer built into basically all modern smartphones was shown [6] to leak (partial) information about PINs and patterns entered on a smartphone. Von Zezschwitz et al. [28] measured and compared the usability of (assigned) PINs and Android Patterns under a realistic setting over three weeks.

Individual aspects and security. A particularly interesting aspect is to what extent the authentication secret is influenced by the person choosing it. Specifically, if observable characteristics of a person influences the secret this can potentially be used to speed up guessing attacks. For text-based passwords, Bonneau [9] found different entropy values for different groups of users. However, due to his data collection method he was unable to look at specific password choices and only observed the resulting distribution of choices, so he could not investigate any further details such as the cause for the differences in strength. Castelluccia et al. [11] found that incorporating a few publicly available datapoints about a user can increase the chance of guessing a password correctly by approx. 5%. In the context of graphical passwords, specifically for the PassFace scheme, Davis et al. [12] showed that the bias in choosing faces significantly depends on gender, race, and subjective beauty of the face.

The work closest to our work is recent and independent work by Aviv et al. [7]. They studied the influence of both collection methods and personal traits on the collected patterns. However, they only reported the influence of personal traits on specific characteristics of the patterns, such as length, starting point, and occurrence of crosses and knight-moves, and did not report on the influence of these traits on the actual strength of individual passwords.

III. Android Unlock Patterns

Next we give a brief introduction to the Android Unlock Pattern scheme and describe the pattern strength meter that we used.

A. Description

Android Unlock Patterns (AUP) are a restricted variant of the Pass-Go scheme [23], which in turn goes back to Draw-A-Secret (DAS) [19], one of the early graphical password schemes. They were introduced in 2008, are available on all current Android phones, and are widely used. The most common design, which will be used throughout this paper, uses 9 points arranged in a 3×3 grid. The user selects a path through these points according to the following rules:

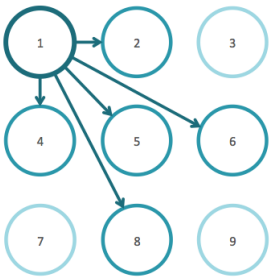


Fig. 1. Points reachable from the top-left node.

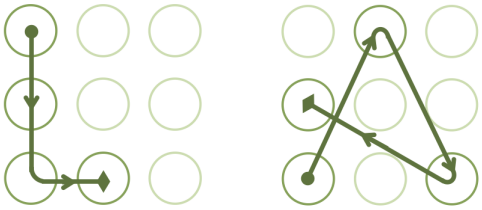


Fig. 2. Strength meter examples with score 0.0361 (left) and $0.114 \cdot 10^{-4}$ (right).

(i) At least four points must be selected,
(ii) No point can be selected more than once,
(iii) Only straight lines are allowed, and
(iv) All points along a path will be connected (unless it was selected before).

The first rule ensure a certain minimal strength of the resulting patterns, albeit little is known about the exact implications on pattern strength. The other rules presumably resolve ambiguities from graphical representations of the patterns, potentially increasing usability. Figure 1 demonstrates the points reachable from the top-left starting position.

B. Measuring pattern strength

One can easily enumerate all possible patterns that adhere to the above rules and determine there are 389 112 valid patterns. However, users do not choose their patterns uniformly from this set, and previous work [26] has established that the resulting strength fall substantially short of the theoretical maximum. Different approaches have been used to determine the strength of patterns. We adapt the approach by Uellenbeck et al. [26], which is based on Markov models. We will describe this approach in the sequel.

Markov Models. Markov models are based on the observation that subsequent tokens, such as letters in normal text or nodes in the Pass-Go scheme, are rarely independently chosen by humans, but can often be quite accurately modeled based on a short history of tokens. For example, in English texts, the letter following a `t` is more likely to be an `h` than a `q`, and for the Pass-Go scheme, nodes which are close to the current node are more frequently chosen than distant ones. In an n-gram Markov model one models the probability of the next token in a string based on a prefix of length $n-1$. Hence, for a given sequence of tokens $c_1, \ldots, c_m$, an n-gram Markov model estimates its probability as

$$P(c_1, \ldots, c_m) \tag{1}$$
$$= P(c_1, \ldots, c_{n-1}) \cdot \prod_{i=n}^{m} P(c_i | c_{i-n+1}, \ldots, c_{i-1}).$$

The required *initial probabilities* $P(c_1, \ldots, c_{n-1})$ and *transition probabilities* $P(c_n | c_1, \ldots, c_{n-1})$ can be determined empirically from the relative frequencies from training data. One commonly applies further post-processing to the raw frequencies: So-called *smoothing* tries to even out statistical effects, in particular it avoids relative frequencies of 0, as these would yield an overall probability of 0 regardless of the remaining probabilities.

Strength-estimation using Markov models. We use Markov models to estimate the probabilities of patterns, and use those probabilities as approximations for their strength. We closely follow the techniques used by Uellenbeck et al. [26]. Their best results were obtained using 3-grams, Laplace smoothing, and using the maximum amount of data available. We train the model on the data collected by Uellenbeck et al. This data was collected in an "adversarial setting", where users chose patterns to protect an account, and were instructed that the account is under attack by other participants. This setup yields one "defensive" pattern, which is used to protect one's account, as well as five "offensive" patterns, used to attack other accounts, per user. We used both the "defensive" and "offensive" dataset, overall more than 600 patterns. In particularly, the model is trained on data which was independently collected from the data that we are considering in this work.

These estimated probabilities $\hat{p}$ can be used directly as a strength measurement. However, a more readable measure is $-\log(\hat{p})$, where logarithms are to basis 2. We use this strength measure throughout this work.

Other strength estimators have been used in previous work. All three are based on readily observable characteristics of the patterns. The meter by Sun et al. [22] uses length, length of the drawn pattern, and the number of intersections. The meter by Andriotis et al. [1] uses the length, number of knight moves, number of overlaps, starting point, and number of changes in direction. The meter by Song et al. [21] uses length, number of intersections, and "non-repeated segments". However, in all three cases there is no theoretical foundation or evaluation for the accuracy of the computation. Thus we refrain from using these metrics.

IV. User study

Next we describe the design and pre-testing of our online study.

A. Study design

We used an online study to collect patterns for the subsequent analysis. Participants were recruited via mailing lists, social networks, and word of mouth. This has the advantage of reaching a relatively large number of participants in a short time, but has the disadvantages that we had little control over participants while filling in the survey (which was mitigated

by rigorous testing of the survey) and little control over the selection of participants (see Section IV-C for statistics about the participants). Data was collected between February and March of 2015 over a time span of 4 weeks.

Different input methods (touchscreen, mouse, pen-on-paper, ...) used by the participant may have an effect on the patterns chosen. For example, using a mouse cursor may allow for a finer control and might facilitate input of more complicated patterns, e.g., those that contain a "knight move". So we wanted to ensure that the users use a smartphone when participating in the study. We used a third-party package[1] to block participants that were not on a mobile device. The package uses a number of heuristics to decide on the device type, including scanning the user agent string transmitted by the participant's browser for specific keywords (such as "mobile", "android", "windows ce", "LG", "wap1.",...), and detecting mobile versions of browsers.

We also wanted to make the study easy to access, specifically without requiring the user to install any additional software. We opted for an HTML/JavaScript web application together with the django/python framework. This means the survey can be accessed using any modern web-browser installed on smartphones, and the look-and-feel can be modeled very similar to that of Android Unlock Patterns, without being restricted to Android Phones.

The survey was structured in four stages: (i) General information, (ii) Short introduction to Android Unlock Patterns, (iii) Pattern collection, and (iv) Questions on demographics and device. We provide more information about these stages in the following sections.

Information. When entering the survey, all participants were presented with a brief introduction to the study, its goals and purposes, the data usage policy, and the researchers behind the project. More detailed information is linked from this screen. No data is collected before a visitor decides to participate by pressing "Start Survey" as illustrated in Figure 3(a). Clicking the green button starts the study.

Introduction to Android Unlock Patterns. Before starting the pattern collection, we need to ensure that participants are familiar with the scheme. Therefore, on the next screen (Figure 3(b)) we provide a brief explanation, and give the participant the possibility to start a more comprehensive training (by pressing "Start training") or continue with the survey (by pressing "Skip training"). In training mode (see Figure 5(c) in the appendix), the participant can test creating patterns as often as she likes, and optical feedback is provided on the validity of the chosen patterns. After selecting "Continue survey", the participant leaves the training mode and continues with the pattern collection as described in the sequel.

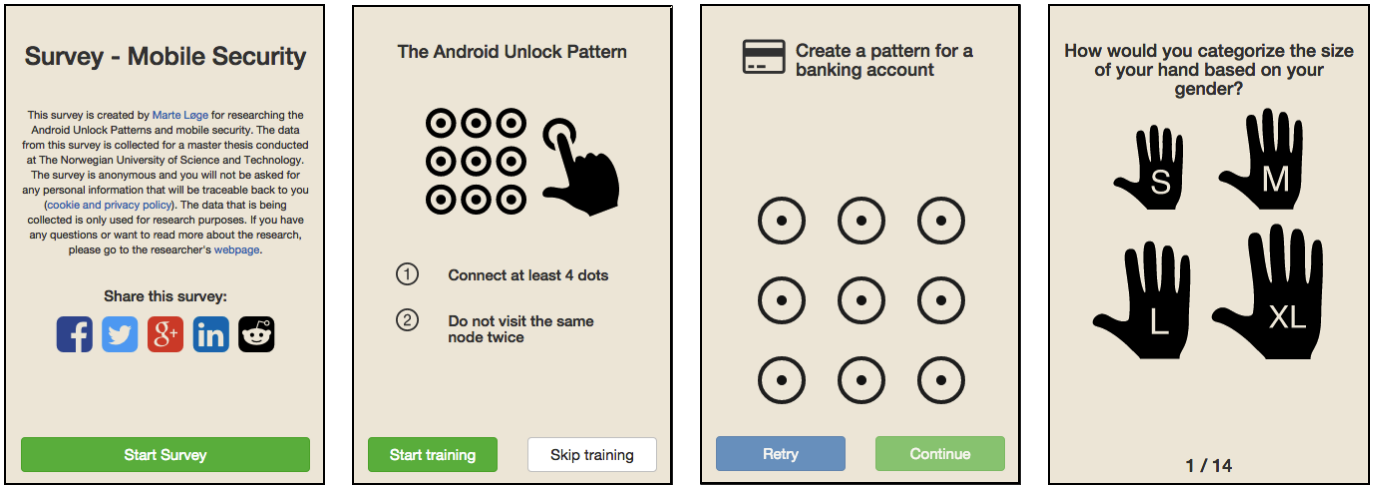
Pattern collection. In the main stage of the study, we ask the participants to create three different patterns for three different scenarios. One pattern for protecting an shopping account, one for unlocking a smartphone, and a third one for protecting a banking account. Those were presented in randomized order. There are two reasons why we ask each participant to create three different patterns: First, this puts pattern creation in a context. The scenarios were selected to cover different situations with different (perceived) security requirements. Thus we avoid problems that one user creates a relatively weak pattern assuming a context with low security requirements (e.g. as she is using the scheme for her smartphone and doesn't value the data on her smartphone very high), whereas another participant assumes a context with high security requirements. Second, we hope this prevents, to some extent, data being submitted by participants that just are curious about the survey and rush to finish the survey, introducing noise into the collected dataset.

The pattern selection step follows the original implementation on Android phones as closely as possible. (Note that, while being functional equivalent, the visual appearance of different Android versions can differ quite a lot.) In a first step, a user selects a pattern that meets the requirements (Figure 3(c)). If a selected pattern fails to meet these requirements, we give visual feedback, as well as a textual description of the condition the pattern failed to meet (see Figure 6(e) in the appendix). Once the user selected a valid pattern, in a second step she is required to confirm this pattern by re-typing it. If the confirmation fails, the system gives visual feedback and allows the participant to try again. If she ultimately fails to re-type the correct pattern, it is possible to go back and create a new pattern. The type-and-re-type approach is the same process used when creating a pattern on a Android device. There are several positive aspects by requiring the respondent to re-type the selected pattern before being able to proceed in the survey. First, it stops users that want to rush through the survey without making an effort to submit an honest answer. Second, it also puts the respondent in a situation where it is needed to create a pattern that is possible for the respondent to actually remember, which is an obvious requirement for real-world patterns.

Demographic questions. Finally, we ask several questions about the participant's demographics as well as the used device. One example screen is shown in Figure 3(d), see Figure 7 in the appendix for a more complete list. In the survey, we ask

- for a subjective assessment of the *hand size of the respondent* based on their gender, ranging from small to extra large, illustrated by icons labeled S, M, L, XL;
- for handedness of the participant using labeled icons for left and right;
- for a subjective assessment of the *screen-size* of the device used, with options S, M, L;
- *which hand is used holding the device* during creation of patterns;
- which *finger was used* when creating the patterns, options were thumb, index finger, other;
- for the usual *reading/writing direction* of the participant, illustrated by an arrow, written text, and an example, options were left-to-right, right-to-felt, and top-to-down;
- for the participant's *gender*, using icons for male and female;
- for the participant's *age*, using a numerical input field;
- if the participant has *experience in IT or IT security*, as a yes-no question;
- for the current type of *screenlock in use*, if any;
- if the participant has any experience with pattern locks as a yes-no question;

[1] https://code.google.com/archive/p/minidetector/



(a) Start screen (b) Android Unlock Patterns introduction (c) Bank pattern (d) Hand size

Fig. 3. Selected screens of the survey. More screens are provided in the appendix.

- for the *mobile OS* on the used smartphone (to simplify this task we tried to automatically detect the OS and asked the question "Is this the mobile operating system on your mobile?"), options were yes, no, I don't know;
- for the *country of origin* of the participant.

We used icons instead of textual lists of alternatives. Our main target device are smartphones, and we believe that icons are easier to interact with on those devices. Icons specifically make it easier for non-natives to quickly complete the survey, and several respondents of a pre-test told us it's more fun to complete the survey with icons. The icons were tested in a pre-test, see Section IV-B. One final screen thanks the participants for their time.

B. Pre-test: Testing the survey

We tested the survey in a controlled environment before releasing it to the public, where we would have little control over the participants. Specifically important for us was testing if the chosen icons where understandable to a broad audience. So even before the pre-test we ran a separate test for the icons only. Test subjects were 12 students, 5 female and 7 male. We showed them the icons used, without the question provided (whereas in the study the questions were stated in English). The only questions that caused some irritation were about the screen lock usage, where some symbols were not readily understandable (we replaced it with a textual list to choose options from), and the question about reading/writing direction, where we added explanation in textual form.

The actual pre-test was conducted with 10 students (5 female, 5 male, a majority with an background in IT or IT Security), in-lab but using their own device. The participants were told (i) to speak aloud during the test about their thoughts and reasons for their choices, (ii) that the test was not about their ability to finish the test, (iii) that they could quit the test at any time if they felt uncomfortable.

Based on the feedback provided by the participants while they interacted with the test we made the following changes to the design (Figures 3, 5, 6, 7 show the final version of the study):

- We simplified the drop-down menu for the country selection, as the original one had graphical flags for each country, which made the component slow and thus hard to use on some devices
- The text for the IT Security question was re-formulated and clarified.
- For the reading/writing direction question we added a textual description and examples to clarify the icons.
- For the screen lock question we replaced the icons with text.
- For the hand-size question, we added that the assessment should be compared to people of the same gender.
- Originally participants did not have to re-type their patterns, and we added that.

C. Participants

A total of 802 respondents completed the whole survey, and 296 more respondents started the survey but did not complete it (81 left before entering any data, 204 started selecting patterns but quit before reaching the questionnaire, and 11 respondents completed creating patterns but did not complete answering questions). Table I provides a summary of the respondents. As some of the respondents did not answer all demographic questions, some questions have more than 802 answers. A majority of the participants was male (66%), between 20 and 29 years old (62%), has some background in IT or IT security (59%), and is from Norway (64%) or the United States (14%). This is a consequence of the recruitment process via social networks and mailing lists, which addressed a proportionally higher number of students and IT security experts. About 88% of the participants is right-handed, which roughly agrees with estimates in the literature [18]. The vast majority reads and writes from left-to-right (98%), which is a consequence of the predominantly western population; we did not use this feature in the following analysis.

		Total	In %
Gender	male	529	66%
	female	278	34%
Handedness	right	690	88%
	left	97	12%
IT or IT security	expert	470	59%
	non-expert	332	41%
Writing-orientation	left-to-right	792	98%
	right-to-left	8	1%
	top-to-bottom	7	1%
Age	16-19	22	3%
	20-24	331	41%
	25-29	169	21%
	30-34	96	12%
	35-39	82	10%
	40-49	73	9%
	50+	30	4%
Hand-size	small	103	13%
	medium	406	50%
	large	255	31%
	extra-large	49	6%
Country	Norway	517	64%
	USA	115	14%
	Germany	33	4%
	Czech Republic	31	4%
	UK	22	3%
	Russia	13	2%
	Rest (<10 each)	75	9%
Total $^{(*)}$		802	100%

TABLE I. STATISTICS OF THE PARTICIPANTS. ($^{(*)}$ NOTE THAT 802 PARTICIPANTS COMPLETED THE ENTIRE STUDY, BUT A FEW PARTICIPANTS ANSWERED SOME QUESTIONS BEFORE LEAVING. THUS SOME QUESTIONS HAVE MORE THAN 802 ANSWERS.)

		Total	In %
Screenlock in use	Android Pattern	202	31%
	4-digit PIN	237	36%
	Fingerprint	116	18%
	Password	44	7%
	slide-to-unlock	28	4%
	Other	28	4%
Screensize	Small	108	13%
	Medium	532	65%
	Large	173	21%
Mobile OS	Android	464	58%
	iOS	321	40%
	Windows	16	2%
	Blackberry	1	0%
Used AUP	Yes	526	65%
	No	278	35%
Total		802	100%

TABLE II. STATISTICS OF THE DEVICES USED BY THE RESPONDENTS

D. Ethical considerations

The ethics committee of NTNU approved the study and the respective contact person was informed. While there is no ethics committee covering this type of user studies at Ruhr-University Bochum (RUB), federal law and privacy regulations must be obeyed. This study complies with these strict regulations. The data we collect about a participant cannot be linked back to a respondent, as the data is in quite broad categories only. We did not collect any identifiers (IP, device ID, name, or similar), and did not use third-party components that still may log such data. Before any data is recorded the respondents are informed about the purpose of the survey and how the contributed data will be managed, and that they can leave the survey at any time.

V. ANALYSIS AND RESULTS

Next we describe the results of analyzing the collected patterns.

Scenario	All	AUP experience	No AUP experience
Shopping	7.06	6.81	7.15
Smartphone	6.45	5.95	7.39
Bank	8.08	8.19	7.69

TABLE III. MEDIAN OF PATTERN CREATION TIMES (IN SEC).

A. Methodology

Most statistical significance test are performed on strength scores. As there is no reason to believe these follow a normal distribution (in fact a Shapiro-Wilk-Test rejects the null hypothesis of normality with $p < 10^{-15}$), we use the Mann-Whitney U-Test for significance testing and Spearman's rank correlation for correlations on strength scores. Similarly, the Shapiro-Wilk-Test rejects the null hypothesis of normality for both the time to choose a pattern and the length of patterns, thus we use the Mann-Whitney U-Test in these cases as well. As we run several tests against the same dataset we use Bonferonni correction. We claim statistical significance for $p < 0.05$, and we indicate possible significant interest for $p < 0.10$. We indicate these in the tables with (**) for $p < 0.05$ and (*) for $p < 0.10$.

Note that, even though we collected three patterns per user (for the three different scenarios), we never use more than one in the comparison, as we test the results for each (fictive) scenario separately.

B. Results for the entire population

First, we report some results for the entire population.

Pattern creation time. The time required to complete a task is one fundamental aspect of the usability of an (authentication) system. We measured the pattern creation time from when the empty grid was displayed on the screen until the user submitted the pattern (separately for each scenario). Table III gives the median creation times for each of the three scenarios that we tested, both for the entire set of users as well as for the subsets of those who reported previous experience with AUP and those that reported no previous experience. (We use the median for its robustness to outliers, as we have encountered some outliers that presumably started the creation process, waited a while, and only returned to their device much later.)

The creation times differ with the (fictive) scenario; it is lowest for the smartphone unlock scenario (6.45 sec), middle for the shopping scenario (7.06 sec), and highest for the bank scenario (8.08 sec). All three differences are statistically significant (as a Mann-Whitney U-Tests show: Shopping vs. Smartphone $p = 0.026736$, Shopping vs. Bank $p < 10^{-5}$, Smartphone vs. Bank $p < 10^{-12}$.) This gives an indication that the (fictive) scenarios used in the study have actually influenced the participants. Also, this gives indication that users invest more effort for accounts with higher (perceived) security requirements, and we will see in the sequel that this increase in effort actually leads to patterns with higher strength.

Interestingly, we find no clear difference in creation times between participants that report experience with the Android Unlock Pattern scheme and those that report no experience. Both for the Shopping and the Bank subsets, we find no significant differences ($p = 1$ in both cases), only in the

	Shopping	Smartphone	Bank	All
#Patterns	841	842	838	2521
Avg. Size	5.541	5.398	5.920	5.619
Avg. Length	5.050	4.920	5.666	5.212
Avg. # Intersections	0.210	0.1769	0.433	0.273
Avg. Overlaps	0.0178	0.014	0.023	0.018
Min.	2.16	2.16	2.16	2.16
1st Qu.	5.84	5.85	6.72	6.18
Median	7.98	8.16	9.35	8.42
Mean	8.86	8.88	10.40	9.37
3rd Qu.	11.12	11.17	13.11	11.72
Max.	32.11	33.19	34.82	34.82

TABLE IV. BASIC STATISTICS FOR THE PATTERN STRENGTH.

	1st Quart. / Median / 3rd Quart.		p
Gender	Female	Male	
Shopping	5.30 7.66 10.18	5.85 8.15 11.83	0.1082
Smartphone	5.66 7.57 10.06	6.02 8.47 11.72	0.0204 (**)
Bank	6.47 8.50 11.38	6.89 9.79 13.42	0.0042 (**)
Handedness	Left	Right	
Shopping	5.28 7.66 10.55	6.07 8.10 11.39	0.8939
Smartphone	5.62 7.84 11.06	6.02 8.29 11.22	1
Bank	6.43 8.90 12.05	6.88 9.53 13.21	0.2570
IT experience	yes	no	
Shopping	5.85 8.15 11.65	5.48 7.66 10.27	0.1577
Smartphone	5.90 8.29 11.80	5.76 7.86 10.06	0.0725 (*)
Bank	6.92 9.43 13.14	6.46 9.09 12.57	0.3205

TABLE V. PATTERN STRENGTH FOR DIFFERENT SUBGROUPS.

	ρ	p
Age		
Shopping	-0.0803	0.1578
Smartphone	-0.0435	1
Bank	-0.1123	0.00986 (**)
Handsize		
Shopping	0.0175	1
Smartphone	0.0408	1
Bank	0.0264	1

TABLE VI. PATTERN STRENGTH FOR DIFFERENT SUBGROUPS.

Smartphone scenario the difference is significant ($p < 10^{-5}$). It is unclear to us why the smartphone scenario behaves differently than both other scenarios.

Pattern strength. Table IV shows the average and median strength of the patterns in the three sets that we collected, as well as several other statistics about the patterns. The (median) strength of the collected patterns differs for the different scenarios, even though they were purely fictional and no consequences followed from it. The strength of the patterns in the Bank scenario were significantly stronger than those in the Shopping scenario ($p < 10^{-8}$) and those in the Smartphone scenario ($p < 10^{-7}$), while there was no significant difference between the Shopping and Smartphone scenario ($p = 1$).

Bias of the patterns. Is has been demonstrated before (e.g. [26], [4], [7] and others) that patterns chosen by humans are biased. To facilitate comparisons with previous work we give some statistics about the structure of the observed patterns in the sequel.

Two aspects that can be used to observe this bias are the distribution of the starting point and the bias of the observed n-grams. The distribution of the starting point is shown in Figure 4(a). This distribution is similar to previously reported numbers: The top-left node is the most frequent one with 44% starting at this particular node (Uellenbeck et al.: 43%), followed by the top-right with 15% (Uellenbeck et al.: 9%) and bottom left with 14% (Uellenbeck et al.: 18%), the remaining nodes ranging from 2% to 9% (Uellenbeck et al.: 2% to 8%).

The most frequent 3-grams are shown in Figure 4(b), where the left figure shows the most frequent 3-grams. The similarities with previous work are striking and show a clear tendency to avoid the middle node, as well as selecting nodes with Euclidean distance one as next node.

A further source of bias is introduced by frequent patterns that resemble common symbols, specifically letters from the Latin alphabet. We inspected the dataset for occurrences of "letters", and found that 385 out of 3393 patterns (11.4%) resembled a letter. Figure 4(c) shows the most frequent cases that we found in the dataset. The most frequent letters where three different versions of the letter "L", as well as "Z", "O", "S", and "U".

C. The influence of personal traits

Next, we present our main results on the influence of specific traits of the user on the resulting pattern strength. An overview can be found in Tables V and VI.

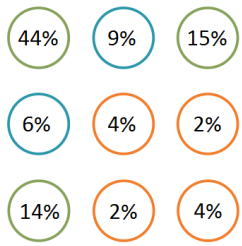
Gender. We found that gender has a significant influence on the pattern strength in the categories Smartphone and Bank ($p = 0.0204$ and $p = 0.0042$, respectively), where female participants chose weaker patterns. The influence in the Shopping scenario is not significant ($p = 0.1082$) (see also Table V). Digging deeper, we see that this is at least in part explained by differences in the patterns length chosen by the participants: female participants choose significantly shorter patterns in the Shopping scenario ($p = 0.0060$) and in the Bank scenario ($p = 0.00072$), but not in the Smartphone scenario ($p = 0.721$). Length is one of the more intuitive factors for pattern strength that should be accessible to a broad audience, but is obviously not the only one.

Handedness. We speculated that the handedness of a participant could have an influence on the chosen patterns, as certain points might be easier to reach than others. This could have an effect on the strength of the chosen patterns. However, we found no significant difference in pattern strength for both groups (see Table V).

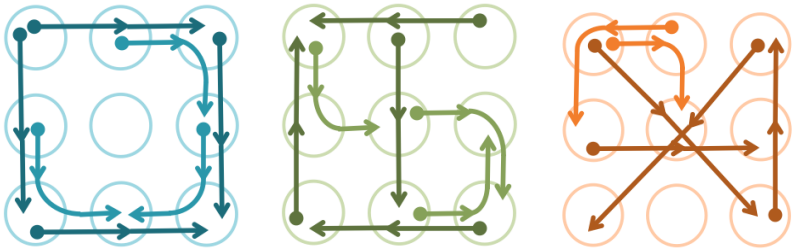
Experience with IT or IT Security. We tested the influence of the (self-reported) experience in IT or IT Security on the pattern strength. We found no statistically significant differences, but we found a significant interest (with $p = 0.0725$) for the Smartphone scenario.

This lack of a clear influence was contrary to our expectations and interesting on its own. To better understand this phenomenon, we also considered pattern length and number of intersections, both which are typically associated with stronger patterns. We found a significant influence of experience on the pattern length in the Banking scenario (Shopping: $p = 0.206$, Smartphone: $p = 0.534$; Bank: $p < 0.0001$), while there was no significant influence on the number of intersections (Shopping: $p = 0.218$, Smartphone: $p = 0.269$, Bank: $p = 1$).

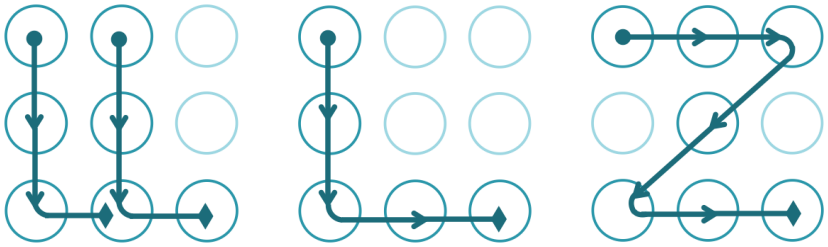
While we have no conclusive explanation for this behavior, it seems plausible that users with experience where trying to



(a) Likelihood of starting points



(b) Most common 3-grams (segments of length 3) found in the data (frequency decreasing from left to right)



(c) Common "letters" found in the dataset: Small "l" (count 94 left + 28 right), large "L" (count 72), "Z" (count 51)

Fig. 4. Basic characteristics of the collected patterns.

choose stronger passwords, but failed in doing so (according to the used strength metric).

Age. The age of the participants has some influence on the strength of the patterns (see Table VI). While the correlation in the Shopping and Smartphone scenario was not significant, we found a significant correlation in the Bank scenario ($p = 0.0264$). The correlation factor for the Bank scenario is moderate ($\rho = -0.113$), i.e., older participants tend to choose weaker patterns. One likely explanation is younger people are generally more technology-affine and thus more used to such schemes.

Handsize. We assumed that a participant's handsize could influence how well she can draw certain complicated patterns (e.g., patterns including a "knight move"), given that mobile devices usually have a very limited screen-size. However, we found no significant correlation of the (self-reported) handsize on the strength of the chosen patterns (see Table VI).

VI. Discussion

Finally, we discuss some limitations and provide an outlook on future work.

A. Limitations

As with all surveys, we rely on the people answering the questions truthfully, and selecting patterns that are realistic. Actually, as our main interest is in comparing strength of different subsets of our dataset, most of our results are invariant to a bias in pattern strength, as long as it affects all collected patterns the same.

As a consequence of our recruitment process via social networks and mailing lists, our participant set is biased towards young (62% are between 20 and 29 years) male (66%) students with a background in IT or IT security (59%) from Norway (64%), thus it does not represent the overall population. As we have seen in Section V-C, specifically age and experience with IT or IT security do influence the pattern strength. However, in the actual comparison the influence of the biased sample should be small, as we are comparing across these subgroups.

B. Future work

We have seen a clear influence of personal traits of a user on the pattern strength. One obvious question regards other measurable properties of users and their influence on pattern strength. Particularly interesting seems the participant's reading- and writing-direction, which we didn't test due to lack of participants with non-western reading-direction.

While in this work we were only concerned with discovering connections between the overall strength and personal traits, there are two directions for future work using these results. Motivated by these findings, one can construct statistical models for individual patterns of a single user, instead of considering the average strength only. Such models can be used first for helping users choose stronger patterns, taking into account their personality, and second for improving the guessing of patterns for the purpose of forensics.

Finally, it would be interesting to extend our findings to other authentication schemes. While some influencing factors are known (see Section II), we are still lacking a more systematic understanding of those factors.

VII. Conclusion

In this work we have shown that personal traits of a user influence the strength of patterns selected for the Android Unlock Patterns. Specifically we have found statistically significant differences in strength based on age and gender, as well several structural properties of patterns. We believe this work is a step towards a more personal treatment of (graphical)

password strength, with the potential to offer more useful password advice for users.

REFERENCES

[1] P. Andriotis, T. Tryfonas, and G. Oikonomou, "Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method," in *Proc. Human Aspects of Information Security, Privacy, and Trust*. Springer, 2014, pp. 115–126.

[2] P. Andriotis, T. Tryfonas, G. Oikonomou, and C. Yildiz, "A pilot study on the security of pattern screen-lock methods and soft side channel attacks," in *Proc. ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*, 2013, pp. 1–6.

[3] M. Arianezhad, D. Stebila, and B. Mozaffari, "Usability and security of gazebased graphical grid passwords," in *Proc. Financial Cryptography and Data Security Workshop on Usable Security (USec)*. Springer, 2013, pp. 17–33.

[4] A. J. Aviv, D. Budzitowski, and R. Kuber, "Is bigger better? comparing user-generated passwords on 3x3 vs. 4x4 grid sizes for android's pattern unlock," in *Proc. Annual Computer Security Applications Conference (ACSAC)*, 2015.

[5] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proc. Workshop on Offensive Technology (WOOT)*, 2010.

[6] A. J. Aviv, B. Sapp, M. Blaze, and J. M. Smith, "Practicality of accelerometer side channels on smartphones," in *Proc. Annual Computer Security Applications Conference (ACSAC)*, 2012.

[7] A. J. Aviv, J. Maguire, and J. L. Prak, "Analyzing the impact of collection methods and demographics for android's pattern unlock," in *Proc. Workshop on Usable Security (USEC)*. Internet Society, 2016.

[8] G. Blonder, "Graphical password," 1996, US Patent 5559961.

[9] J. Bonneau, "The science of guessing: Analyzing an anonymized corpus of 70 million passwords," in *Proc. IEEE Symposium on Security and Privacy (SP)*, 2012, pp. 538–552.

[10] J. Bonneau, S. Preibusch, and R. Anderson, "A birthday present every eleven wallets? The security of customer-chosen banking PINs," in *Proc. Financial Cryptography and Data Security*. Springer, 2012, pp. 25–40.

[11] C. Castelluccia, C. Abdelberi, M. Dürmuth, and D. Perito, "When privacy meets security: Leveraging personal information for password cracking," *CoRR*, vol. abs/1304.6584, 2013. [Online]. Available: http://arxiv.org/abs/1304.6584

[12] D. Davis, F. Monrose, and M. K. Reiter, "On user choice in graphical password schemes," in *Proc. USENIX Security Symposium*. Usenix, 2004.

[13] A. De Angeli, L. Coventry, G. Johnson, and K. Renaud, "Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems," *Int. J. Hum.-Comput. Stud.*, vol. 63, no. 1-2, pp. 128–152, Jul. 2005.

[14] A. E. Dirik, N. Memon, and J.-C. Birget, "Modeling user choice in the PassPoints graphical password scheme," in *Proc. Symposium on Usable Privacy and Security (SOUPS)*. ACM, 2007, pp. 20–28.

[15] P. Dunphy and J. Yan, "Do background images improve draw a secret graphical passwords?" in *Proc. ACM Conference on Computer and Communications Security (CCS)*. ACM, 2007.

[16] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner, "Are you ready to lock?" in *Proc. ACM Conference on Computer and Communications Security (CCS)*. ACM, 2014.

[17] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith, "It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception," in *Proc. Symposium On Usable Privacy and Security (SOUPS 2014)*. USENIX Association, 2014.

[18] C. Hardyck and L. F. Petrinovich, "Left-handedness," *Psychol. Bull.*, vol. 84, no. 3, pp. 385–404, may 1977.

[19] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin, "The Design and Analysis of Graphical Passwords," in *Proc. USENIX Security Symposium*, 1999.

[20] D. V. Klein, "Foiling the cracker: A survey of, and improvements to, password security," *Proc. USENIX Security Workshop*, 1990.

[21] Y. Song, G. Cho, S. Oh, H. Kim, and J. H. Huh, "On the effectiveness of pattern lock strength meters: Measuring the strength of real world pattern locks," in *Proc. Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.

[22] C. Sun, Y. Wang, and J. Zheng, "Dissecting pattern unlock: The effect of pattern strength meter on pattern selection," *Journal of Information Security and Applications*, vol. 19, no. 4–5, pp. 308–320, Nov. 2014.

[23] H. Tao and C. Adams, "Pass-Go: A Proposal to Improve the Usability of Graphical Passwords," *International Journal of Network Security*, vol. 7, no. 2, pp. 273–292, 2008.

[24] J. Thorpe and P. C. Van, "Graphical dictionaries and the memorable space of graphical passwords," in *Proc. USENIX Security Symposium*. Usenix, 2004.

[25] ——, "Human-Seeded attacks and exploiting Hot-Spots in graphical passwords," in *Proc. USENIX Security Symposium*. Usenix, 2007.

[26] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz, "Quantifying the security of graphical passwords: The case of android unlock patterns," in *Proc. ACM Conference on Computer & Communications Security (CCS)*. ACM, 2013, pp. 161–172.

[27] D. Van Bruggen, S. Liu, M. Kajzer, A. Striegel, C. R. Crowell, and J. D'Arcy, "Modifying smartphone user locking behavior," in *Proc. Ninth Symposium on Usable Privacy and Security (SOUPS)*. ACM, 2013.

[28] E. von Zezschwitz, P. Dunphy, and A. D. Luca, "Patterns in the wild: A field study of the usability of pattern and pin-based authentication on mobile devices," in *Proc. International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI)*, 2013.

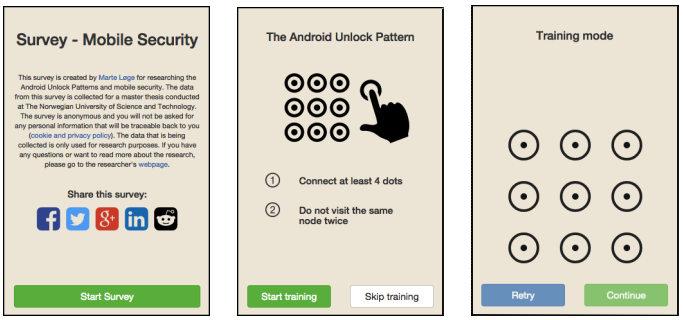
[29] S. Wiedenbeck, J.-C. Birget, A. B. J. Waters, and N. Memon, "Authentication using graphical passwords: Basic results," in *Proc. International Conference on Human-Computer Interaction (HCI International)*, 2005.

[30] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: effects of tolerance and image choice," in *Proc. Symposium on Usable Privacy and Security (SOUPS)*. ACM, 2005, pp. 1–12.

[31] ——, "PassPoints: Design and longitudinal evaluation of a graphical password system," *Int. J. Hum. Comput. Stud.*, vol. 63, no. 1-2, pp. 102–127, Jul. 2005.

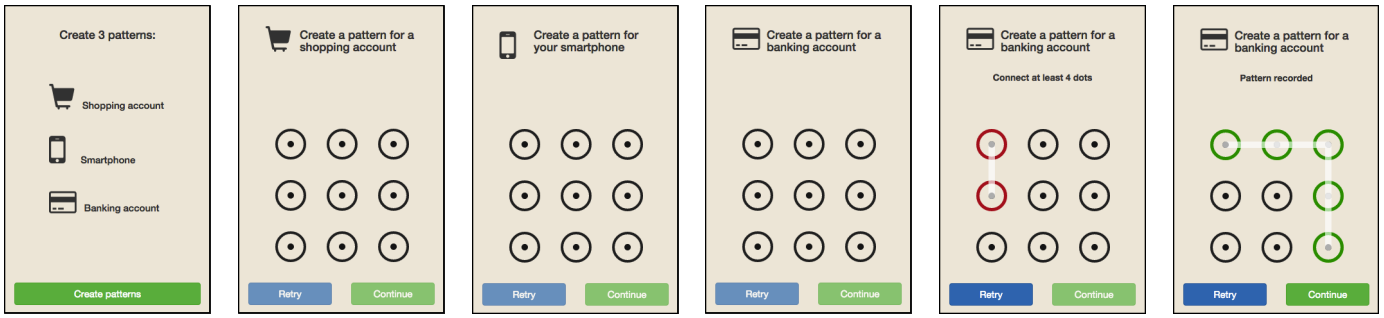
[32] Z. Zhao, G.-J. Ahn, J.-J. Seo, and H. Hu, "On the security of picture gesture authentication," in *Proc. USENIX Security Sympsosium*, 2013.

APPENDIX

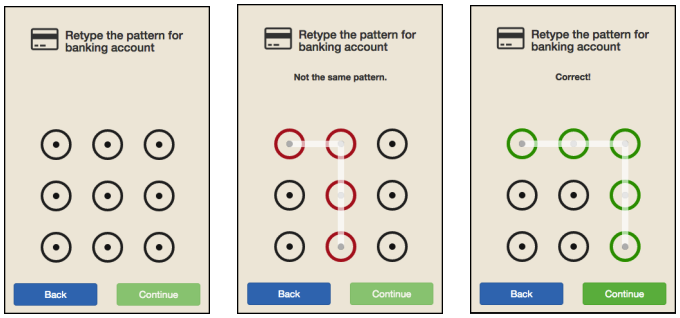


(a) Start screen (b) Android Unlock Pattern introduction (c) Training mode (optional)

Fig. 5. Study design – Introduction

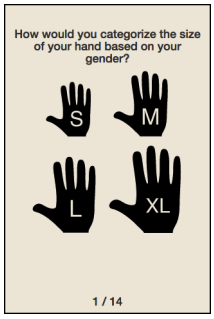


(a) Introduction to patterns (b) Shopping pattern (c) Smartphone pattern (d) Bank pattern (e) Pattern length too short (f) Valid pattern recorded

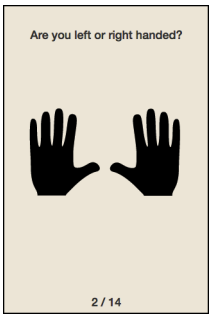


(g) Retype pattern (h) Retype wrong (i) Retype correct

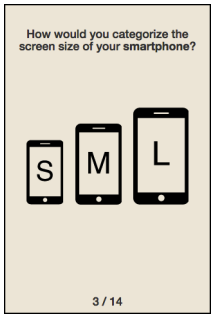
Fig. 6. Study design – Create and retype patterns



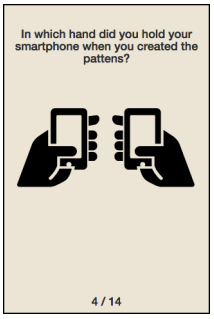
(a) Hand size



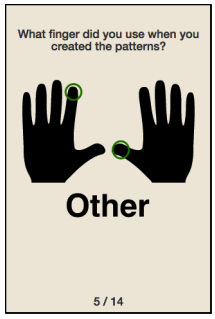
(b) Handedness



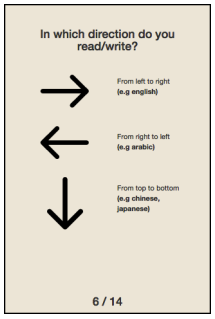
(c) Screen size



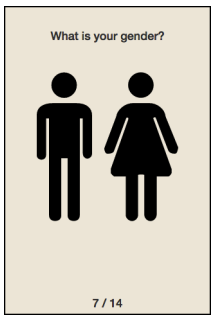
(d) Hand used when creating pattern



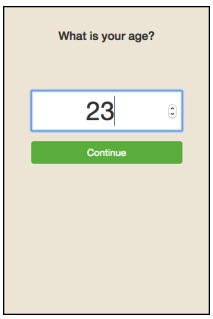
(e) Finger used



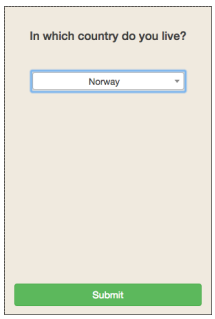
(f) Reading/writing direction



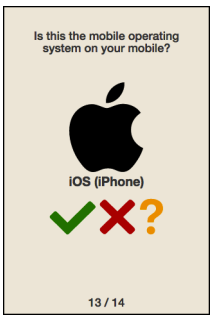
(g) Gender



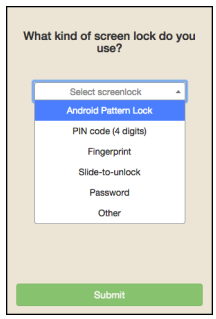
(h) Age



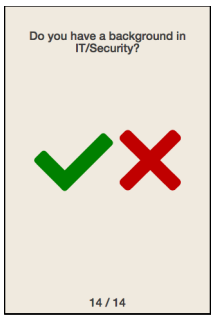
(i) Country



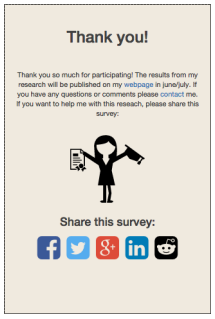
(j) Mobile OS



(k) Screenlock



(l) Experience



(m) Thank you

Fig. 7. Study design – Demographic questions