Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# From hot-spots towards experience-spots: Leveraging on users' sociocultural experiences to enhance security in cued-recall graphical authentication

Argyris Constantinides [a,*], Christos Fidas [b], Marios Belk [c], Anna Maria Pietron [d], Ting Han [d,**], Andreas Pitsillides [e]

[a] University of Cyprus, Cyprus & Cognitive UX LTD, Cyprus
[b] University of Patras, Greece
[c] Cognitive UX GmbH, Germany & University of Cyprus, Cyprus
[d] Shanghai Jiao Tong University, China
[e] University of Cyprus, Cyprus

ARTICLE INFO

*Keywords:*
User authentication
Graphical passwords
Sociocultural experiences
Security
Memorability
User study

ABSTRACT

This paper suggests a novel cued-recall-based graphical authentication method, which leverages on users' sociocultural experiences for improving the security and memorability of selected secrets. We evaluated the suggested approach in the context of three user studies ($n = 139$): *a)* an eye-tracking study ($n = 42$) focusing on security in terms of resistance to brute-force attacks; *b)* a two-week study ($n = 71$) focusing on memorability and login usability; and *c)* a controlled in-lab user study ($n = 26$) focusing on human attack vulnerabilities among people sharing common sociocultural experiences. Analysis of results revealed that the suggested approach influenced visual behavior strategies of end-users, which subsequently resulted in significantly stronger passwords created on images reflecting their prior experiences than on images unfamiliar to them. Simultaneously, both reference and control groups performed similarly in terms of memorability and login efficiency and effectiveness. On the downside, the suggested approach introduces password guessing vulnerabilities in terms of allowing attackers who share common experiences with the end-users to more easily identify regions of their selected secrets. Findings point towards a new direction for delivering personalized cued-recall graphical authentication schemes that depict image semantics bootstrapped to users' real-life experiences.

## 1. Introduction

Picture Gesture Authentication (PGA) is a cued-recall graphical authentication system which requires users to select an image and subsequently draw gestures on that image to create their graphical password. PGA has been introduced in Windows 8™ and utilized by millions of users (Zhao et al., 2013) as a promising alternative authentication experience, given that it leverages on the picture superiority effect (Paivio and Csapo, 1973) and easily adapts to ubiquitous interaction realms.

An important user interface design factor that affects the security strength of user-chosen graphical passwords is the background image used (Thorpe and van Oorschot, 2007; van Oorschot and Thorpe, 2011; Bulling et al., 2012; Alt et al., 2015; Belk et al., 2017b). Research has shown that the selections of images can be predictable since users prefer clear *vs.* incoherent images (Aydın et al., 2013), and choose images that illustrate people (Zhao et al., 2013; Alt et al., 2015) and sceneries (Alt et al., 2015). In addition, users' choices are influenced by human attributes in an image (*e.g.*, race, age, gender (Davis et al., 2004)), image colors and type (Mihajlov et al., 2016).

Prior works (Tullis and Tedesco, 2005; Wiedenbeck et al., 2005; Thorpe and van Oorschot, 2007; van Oorschot et al., 2010; van Oorschot and Thorpe, 2011; Bulling et al., 2012; Alt et al., 2015) investigated the use of *image semantics* and their effects on the security of user-chosen passwords. Images can be broadly categorized as *generic* (*i.e.*, not directly relevant nor familiar to the users, *e.g.*, abstract, nature,

[*] Corresponding author.
*E-mail address:* aconst12@cs.ucy.ac.cy (A. Constantinides).
[**] Co-corresponding author on behalf of the Shanghai Jiao Tong University, China.

https://doi.org/10.1016/j.ijhcs.2021.102602
Received 10 June 2020; Received in revised form 16 January 2021; Accepted 23 January 2021
Available online 2 February 2021
1071-5819/

landscapes, etc.) or *personal* (*i.e.*, directly relevant and highly familiar to the users, *e.g.*, depicting people, objects, or scenes highly personal to users). Studies in (Thorpe and van Oorschot, 2007; van Oorschot and Thorpe, 2011; Bulling et al., 2012) indicate that generic images are susceptible to *hot-spots (points on an image that attract users to select them)*, thus, leading to the creation of easily predictable passwords. Subsequently, several works focused on alleviating the hot-spot issue, mainly by limiting the available choices during password creation to prevent users from making selections on hot-spots (Chiasson et al., 2007, 2008; Bulling et al., 2012).

The use of *personal* images also impacts the security of user-chosen passwords, since it may result to the creation of passwords easily guessable by someone who knows the user (Tullis and Tedesco, 2005; Wiedenbeck et al., 2005; Schaub et al., 2013). The use of images that are familiar to the user (*e.g.*, containing family members) increases the likelihood of certain areas on the image to be selected as part of the password (Bulling et al., 2012). Furthermore, the fact that many users often do not understand security features (Furnell, 2005) may lead to the use of personal photos that violate the privacy of others depicted in the photo, as well as theirs, since private information is revealed during login (Ahern et al., 2007).

Hence, the aforementioned state-of-the-art approaches embrace deficiencies; when image content is delivered randomly, the security of the graphical password is reduced since users, in an attempt to scaffold memorability, tend to choose easy-to-remember and predictable hot-spots (Tullis and Tedesco, 2005; Renaud, 2009); when users are allowed to upload image content, security and privacy considerations also arise since users tend to create easily guessable passwords (Tullis and Tedesco, 2005) and often violate the privacy of people depicted in the uploaded images (Ahern et al., 2007). Therefore, there is a need for a more sophisticated approach within PGA schemes to achieve a better tradeoff between security and memorability (Biddle et al., 2012; Belk et al., 2017a; Katsini et al., 2018; Constantinides et al., 2018a; 2018b; 2020b). A possible direction to achieve this goal, as introduced in this paper, is a *retrospective-based approach* for PGA schemes.

A *retrospective-based approach* for PGA schemes aims at delivering background images to end-users which depict sceneries that reflect users' sociocultural experiences, on different levels of abstractions, thus, expanding the state-of-the-art narrow spectrum (*e.g.*, too generic or too personal) of image content semantics in PGAs.

We suggest a five-tier model (**Fig. 1**) of image content familiarity, namely: *Individual, Group, Organizational, National* and *Global*, bootstrapped to the users' prior sociocultural activities, experiences and declarative memories. At the *individual level*, people have personal experiences (*e.g.*, one's experiences within her neighborhood's cafeteria). At the *group level*, people have shared experiences within the communities they belong to (*e.g.*, one's experiences within the volleyball team she plays for). At the *organizational level*, people have experiences within their working places (*e.g.*, one's experiences within the working space area at the company she works for). At the *national level*, people have nationally shared experiences (*e.g.*, within monuments, landmarks, folklore). At the global level, people can have experiences within places not directly relevant to their culture (*e.g.*, experiences when traveling).

The contributions of this work are summarized as follows:

- We introduce the concept of the retrospective approach during graphical password creation within PGA through a five-tier model of image content delivery.
- We provide empirical data on the coefficient of image semantics and users' declarative memories on their visual behavior during graphical password creation.
- We provide evidence that visual behavior of users with regards to (non-) hot-spot regions of an image is affected by the retrospective approach as highlighted through the security analysis of user-generated passwords.
- We provide evidence that the retrospective approach improves the security of user-generated graphical passwords, while it does not hamper memorability and login usability.
- We provide evidence that the retrospective approach introduces human attack vulnerabilities among people sharing common real-life experiences.

## 2. Related work

### 2.1. Research on influencing security in graphical passwords

A key issue in PGA-like schemes relates to the existence of *hot-spots (points on an image that attract users' attention)* (Thorpe and van Oorschot, 2007; van Oorschot and Thorpe, 2011; Bulling et al., 2012), thus, leading to the creation of easily predictable passwords which are prone to automated attacks (van Oorschot et al., 2010). To prevent users from making poor password selections, prior works focused on limiting the available choices during password creation. For example, Chiasson et al. (2007) proposed a scheme in which users' choices are limited to one click-point per image for a total of five images. In a subsequent work (Chiasson et al., 2008), a viewport was used that highlights a small random area of the image aiming to persuade users selecting passwords that are less likely to include salient regions. Bulling et al. (2012) proposed to hide potential hot-spots using saliency maps, thus, preventing users from selecting them as part of their passwords. Thorpe et al. (2014) used the "presentation effect" that gradually reveals the underlying image to influence users' choices during password creation. Katsini

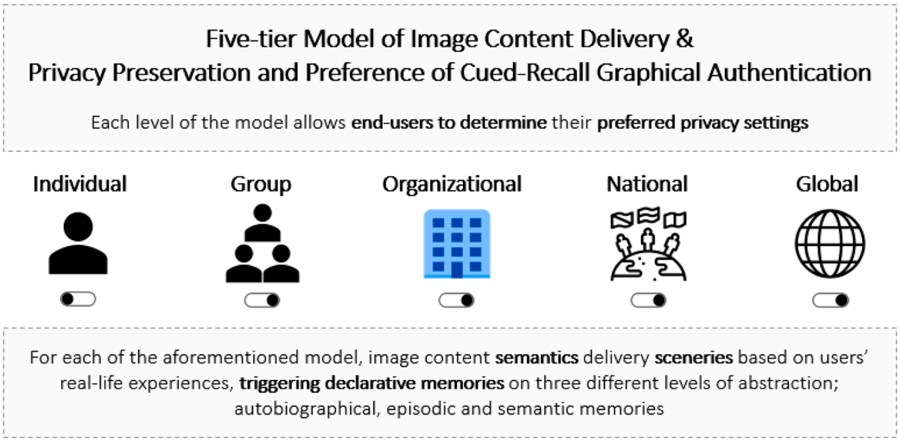

Five-tier Model of Image Content Delivery &
Privacy Preservation and Preference of Cued-Recall Graphical Authentication
Each level of the model allows end-users to determine their preferred privacy settings
Individual
Group
Organizational
National
Global
For each of the aforementioned model, image content semantics delivery sceneries based on users' real-life experiences, triggering declarative memories on three different levels of abstraction; autobiographical, episodic and semantic memories

**Fig. 1.** Five-tier model of retrospective image content delivery inspired by Erez and Gati (2004), indicating that behaviors, attitudes, prior experiences can be represented at various levels in a multi-level model of sociocultural experiences.

et al. (2018) used a similar fade-out effect which starts from the highest saliency mask level and gradually reveals the image based on users' cognitive processing styles.

### 2.2. Research on influencing memorability in graphical passwords

While research on the memorability of PGA-like schemes is rather limited, numerous works in the context of recognition-based graphical authentication schemes suggested that generic image content hinders memorability since users cannot easily connect prior experiences (in their declarative memory) with the depicted content. Renaud (2009) conducted a memorability study to compare the efficiency of three types of images; doodles, generic pictures of random everyday objects, and personal pictures provided by users, showing that the generic pictures were the least memorable because of the lack of strong connection between the users and the pictures. Similar findings were reported by Tullis and Tedesco (2005), in which generic pictures (*i.e.*, random stock pictures) were less memorable compared to personal pictures. The finding was consistent even after a few months had elapsed between the studies, as well as after introducing very similar distractor images. In another study conducted with the same participants over a longer period of time (*i.e.*, six years after password creation), results revealed that twelve out of thirteen participants could still authenticate successfully (Tullis et al., 2011). Furthermore, studies have shown that various background image attributes have an effect on memorability, such as image type (*e.g.*, single objects are more memorable than faces and abstract images) (Mihajlov and Jerman-Blažič, 2011), image properties (*e.g.*, color, shape, and category) (Mihajlov et al., 2016), image distortion (Hayashi et al., 2008, 2011), and interference (Everitt et al., 2009).

## 3. Research motivation

From the aforementioned approaches we can conclude that existing works, in an attempt to enhance security, rather limit the users' visual field by modifying the user interface (*e.g., by hiding hot-spot regions*) or intervening in the users' decision-making process (*e.g., by adopting masking approaches, applying draw-the-curtain effects*, etc.). In addition, in an attempt to increase memorability, users tend to use hot-spot regions on generic images which weakens the strength of selected passwords, or use personal images which can be easily predictable or violate privacy.

We expect that the suggested retrospective approach will potentially trigger users' declarative memories (Tulving, 1972). This might subsequently affect their visual behavior, during PGA password creation, by moving their attention *from hot-spots towards experience-spots* (*i.e.*, regions of the depicted sceneries associated with their real-life experiences), and subsequently improving the security of selected secrets. This could be potentially revealed by analyzing users' visual behavior during password creation by considering eye gaze metrics on hot-spots *vs.* non-hot-spots. We also expect that the retrospective approach will improve memorability aspects since it leverages on the reflection of declarative memories of end-users. On the other hand, we also suspect that the suggested approach might increase guessing vulnerability among people sharing common experiences. Bearing in mind that studies have shown that some people are more concerned about attacks by insiders (*e.g.*, family members, significant others, etc.) (Muslukhov et al., 2013), it is important to shed light on how secure the suggested approach method is against people that are close to the end-user and share common experiences. To the best of our knowledge, such a retrospective-based approach has not been suggested and investigated within PGA schemes so far.

In order to address the aforementioned hypotheses, we designed and implemented three different user studies: *i) User Study A* investigating the effects of the retrospective approach on visual behavior, security and password creation time; *ii) User Study B* investigating the effects of the retrospective approach on memorability and login usability; and *iii) User Study C* investigating guessing vulnerabilities of the retrospective approach in human guessing attacks. Investigating such an interplay could provide important insights on personalizing PGA schemes by considering users' sociocultural experiences as an important personalization factor in the design of adaptive mechanisms.

## 4. User study A – Does the retrospective approach affect visual behavior, security & password creation time?

### 4.1. Research questions

The following main research questions are investigated:

***RQ$_1$***. Is there a significant improvement in the security strength of the created passwords in PGA schemes between the experimental group *(retrospective-based approach)* and control group *(state-of-the-art approach)*? – *Answering this question will provide insights on whether the retrospective approach improves the security strength of created graphical passwords.*

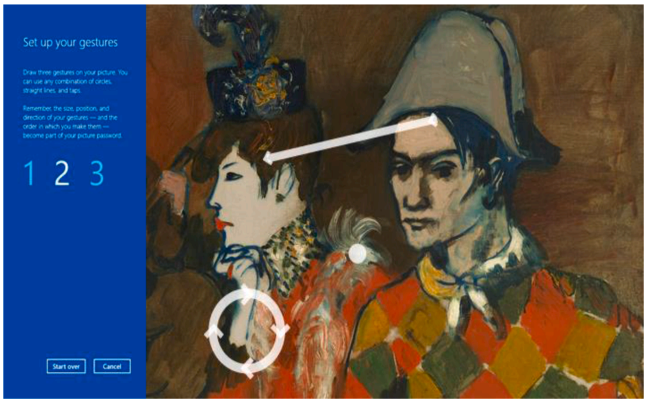
***RQ$_2$***. Is there a significant difference in users' visual behavior between the experimental and the control group, considering also users' fixations on hot-spots? – *Answering this question will provide insights on whether image content reflecting users' memories and experiences will show a decrease of fixations on image hot-spots and simultaneously an increase on experience-spots.*

***RQ$_3$***. Is there a correlation between fixations on hot-spots and user-selected passwords that include hot-spots, between the experimental and the control group? – *Answering this question will provide insights on whether regions that attract the attention of the users are also selected during graphical password creation.*

***RQ$_4$***. Is there an association between the image semantics (retrospective *vs.* generic) and users' password region selection strategy followed (*i.e.*, experience- *vs.* random-driven), between the experimental and the control group? – *Answering this question will provide insights on whether the selected password regions are linked to exclusive personal experiences or are random-driven.*

***RQ$_5$***. Is there a difference in time spent to explore the image between the experimental and the control group? – *Answering this question will provide insights about the efficiency of the password creation phase of the retrospective approach.*

***RQ$_6$***. Is there a difference in image region selections among users that created their graphical password on the same image? – *Answering this question will provide insights to better understand if participants who share common experiences (e.g., go to the same classroom) create similar passwords.*



Draw three gestures on your picture. You can use any combination of circles, straight lines, and taps.

Remember, the size, position, and direction of your gestures — and the order in which you make them — become part of your picture password.

**Fig. 2.** A graphical password illustrating the three gestures allowed in our Web-based graphical authentication scheme.

## 4.2. Study instruments and metrics

### 4.2.1. Cued-recall graphical authentication mechanism

We developed a Web-based PGA-like graphical authentication scheme (Fig. 2), similar to Windows 10™ PGA (Johnson et al., 2014), in which users can create gesture-based passwords on a background image that acts as a cue. Three types of gestures are allowed: taps, lines and circles. Free line gestures are not permitted, hence, they are automatically converted into one of the three permitted gestures.

For the processing of the gestures, the mechanism creates a grid of the image containing 100 squares (segments) on the longest side, and then divides the shortest side by the same scale.[1] Rounding was not applied to any decimal segments, and we allowed 0.25 segments size overflow at the rightmost side of the image. The approach of creating a grid of 100 squares allows for storing the gestures based on their segment position on the grid rather than the coordinates in pixels. For each gesture, the following data are stored: for taps, the coordinates of a point, for lines the coordinates of the starting and ending point, and for circles the coordinates of the center, radius and direction.

Before enrolment in the system, the mechanism provides a demonstration page, in which users can experiment by drawing gestures on a background image. On the enrolment page, the screen is split in two sides (Fig. 2). On the left side, there are instructions about the password creation, and three numbers (1, 2, 3) indicating the current active gesture.

On the right side, there is the background image on which users can create their passwords by drawing three gestures. After each gesture is drawn, the shape of the gesture is temporarily displayed on the screen at the corresponding location, to provide feedback to the user that the gesture has been captured by the mechanism. Users are required to redraw the three gestures to confirm their graphical password.

During the comparison of the two passwords, the mechanism allows for a tolerance distance in terms of the coordinates on the grid (36 segments around each initial selected segment are acceptable[1] (Katsini et al., 2018), thus, building a circle of 3 segments radius). This tolerance allows better accuracy of users' selections during login. However, there is no tolerance regarding ordering, type, and directionality of the gestures.

During the login task, the user is presented with the same page and is required to enter the graphical password by reproducing all three gestures. The mechanism compares the entered password with the stored one and login is considered successful if *(a)* all three gestures (ordering, type, and directionality) match with the stored ones; and *(b)* the tolerance distance between the entered gestures and the stored ones fit in the predefined tolerance threshold.

### 4.2.2. Equipment

An All-in-One HP computer with a 24″ monitor was used (1920 × 1080 pixels, 16:9 aspect ratio). To capture eye movements, we used Gazepoint GP3 eye tracker (GP3 Eye Tracker, 2020), which captures data at 60 Hz. No equipment was attached to the participants. Following the manufacturer's guidelines, the eye tracker was calibrated individually using a 5-point calibration procedure and was positioned at an upwards angle roughly 30 cm below eye level and 65 cm away.

### 4.2.3. Study factors – semantics of image content

To control participants' sociocultural familiarity with the image semantics and thus investigate the research questions, we intentionally chose two specific image sets: *i) Group level retrospective images:* images highly related to the participants' shared, individual and common sociocultural experiences from their daily life context (*i.e.*, depicting sceneries of a University campus such as lecture rooms, lab rooms, cafeteria, etc.), which serve as the experimental factor; and *ii) Global level generic images:* images illustrating generic/artificial content unfamiliar to the users (*i.e.*, depicting generic sceneries and people) in order to avoid users being familiar with a specific scenery, which serve as the control factor. In addition, the scenery of the generic image set reflected state-of-the-art knowledge in PGA user research (Zhao et al., 2015). In order to minimize the bias effect of using one image per group, we provided a set of nine images for each group. Users could select only one image from their corresponding image set. **Fig. 3 (left and right)** illustrates the two image sets used in the study.

The sets of images were based on existing research that has shown that users tend to select images illustrating sceneries (Dunphy and Yan, 2007; Alt et al., 2015; Zhao et al., 2015).

Considering that the number of hot-spots and the image complexity affect the password strength (Wiedenbeck et al., 2005; Katsini et al., 2018), we chose images of similar number of hot-spots and complexity between and within images belonging to the two groups. For doing so, we followed a semi-automated approach to detect the hot-spots regions. To calculate the number of hot-spots, we used a combination of computer vision techniques for object detection[2-4] and a combination of saliency maps[5] and saliency filters[6] (Perazzi et al., 2012) for the salient regions. Furthermore, we assessed the equivalence of the two image sets by calculating the image complexity using entropy estimators[7,8] (Cardaci et al., 2009). Furthermore, we expect that users will make their password selections around objects easily distinguishable from their surroundings on the image (Thorpe and van Oorschot, 2007; Zhao et al., 2013). Therefore, we also considered the objects that “stand out” in the images as potential hot-spots using object detection mechanisms[2-4][2,3,4]. Based on the aforementioned objective measures, we decided on these two image sets (**Fig. 3 left and right**). **Table 1** summarizes the image complexity and number of hot-spots regions.

### 4.2.4. Password strength

We adopted a widely used approach for measuring password strength in PGA schemes (Zhao et al., 2013, 2015) by calculating the number of guesses required to crack the users' passwords. Following existing approaches that consider hot-spots regions (or Points of Interests - PoI) (Sadovnik and Chen, 2013; Zhao et al., 2013, 2015; Katsini et al., 2018), we used a *PoI-assisted Brute-force Attack* model proposed by Zhao et al. (2013) starting from segments covering the hot-spots segments, then checking the neighboring segments, and finally checking the rest of the segments. Aiming to complement the analysis, we further used the *Knowledge-based PoI-assisted Attack* model used in Zhao et al. (2013), which is able to abstract knowledge of user choices in PGA-like schemes, and we generated ranked dictionaries that were used for further assessing the password strength. We intentionally did not adopt a naïve-based brute-force attack since this would not be realistic and relevant to the suggested approach which considers hot-spots segments. Furthermore, we focused on offline attacks since they represent a major and realistic threat to service providers compared to online attacks which are preventable considering the effective key space of PGAs[1] and the ability to lock accounts after 5–10 failed logins. These approaches fit with $RQ_1$ which investigates the strength of users' passwords with respect to the hot-spots segments.

### 4.2.5. Eye gaze metrics

Following common practices for capturing users' visual behavior (Duchowski, 2007; Raptis et al., 2016, 2017; Katsini et al., 2018; Fidas

[1] Microsoft™ Picture Password blog - bit.ly/2SajCDO

[2] Tensorflow - bit.ly/1MWEhkH

[3] Amazon Rekognition - amzn.to/2hm466g

[4] Google Cloud Vision - bit.ly/21xSsUV

[5] Saliency Map - bit.ly/2MuiSZC

[6] Saliency Filters - bit.ly/2QMuQvU

[7] Image Entropy - bit.ly/2wB7Erm

[8] scikit-image Shannon Entropy - bit.ly/2Xx4iBK

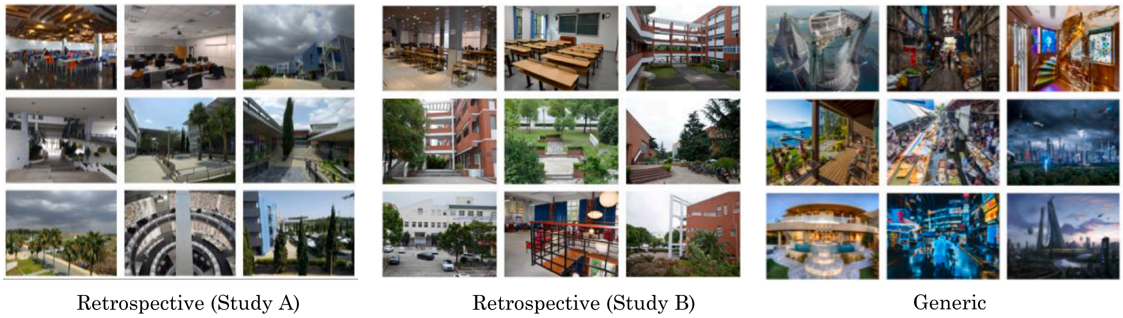


**Fig. 3.** The retrospective set of nine images (left – User Study A; middle – User Study B) illustrating sceneries at the participants' University. The generic set of nine images (right) illustrating generic/artificial sceneries unfamiliar to the participants.

**Table 1**
Similarities of number of hot-spots and image complexity for the image sets used in User Study A.

| Image ID | Complexity in bits (retrospective) | Complexity in bits (generic) | Number of Hot-spots Regions (retrospective) | Number of Hot-spots Regions (generic) |
|---|---|---|---|---|
| 1 | 7.44 | 7.62 | 7 | 7 |
| 2 | 7.48 | 7.39 | 8 | 7 |
| 3 | 7.32 | 7.46 | 7 | 7 |
| 4 | 7.72 | 7.75 | 6 | 6 |
| 5 | 7.51 | 7.58 | 8 | 8 |
| 6 | 7.65 | 7.73 | 6 | 6 |
| 7 | 7.49 | 7.54 | 7 | 7 |
| 8 | 7.19 | 7.24 | 7 | 6 |
| 9 | 7.46 | 7.52 | 8 | 7 |

et al., 2019), we selected *fixation count* and *fixation duration* on the (non-) hot-spots segments of the images. The fixation count metric is the total number of fixations within each area of interest (AOI), by considering visits and revisits to the AOI. The fixation duration is the total duration of fixations within an AOI, considering visits and revisits to the AOI.

### *4.3. Sampling and procedure*

#### *4.3.1. Participants*

A total of 42 individuals (19 females) participated in the study, ranging in age between 20 and 32 years old ($m = 24$, $sd = 3.1$). We recruited undergraduate students from a European University through email invitations and in-class announcements made by colleagues. Participants had no relationship to the researchers to avoid biases. Participants were split into two groups based on the image type (*i.e.*, retrospective user group and generic user group), and the image type was randomly varied across all users. To increase the internal validity of the study, we recruited participants that had no prior experience with PGA-like authentication mechanisms nor knowledge of its security semantics, and participants who had spent the last three years at the University campus, assuming they would have had experiences within the University. Postgraduates and faculty were intentionally not considered aiming to: *i)* control the image semantic familiarity factor by illustrating sceneries where undergraduates engaged with everyday-life activities; and *ii)* retain ecological validity (*i.e.*, give incentives to create secure and memorable credentials). We assured that the technical background of the participants would not bias the experiment by including only those with no knowledge of PGA security semantics.

#### *4.3.2. Experimental design and procedure*

With respect to the ethical aspects of the study, we adopted the University's human research protocol that takes into consideration users' privacy, confidentiality and anonymity. All participants performed the password creation task in a quiet lab room with only the researcher present. To avoid any bias effects, no details regarding the research objective were revealed to the participants. The study involved the following steps: first, participants were informed that the collected data would be stored anonymously and would be used only for research purposes, and they signed a consent form. Next, they completed a questionnaire on demographics and the eye-calibration process followed. Next, participants were introduced to a demonstration page to familiarize themselves with the process of drawing gestures. Participants were then requested to create a user account in order to access an online service. To increase ecological validity and keep security as a secondary task (Egelman et al., 2013), participants were requested to create an account using our PGA scheme, in order to use this account to login for accessing student materials within a University's website.

The retrospective user group received a set of nine retrospective images, and the generic user group received a set of nine generic images. In the first step, they created a username and then they selected one image out of the nine available images, on which they created their graphical password by drawing three gestures using a computer mouse. To confirm their graphical password, they were requested to reproduce the initial three gestures. Finally, semi-structured interviews were conducted with each participant at the end of password creation, covering the following open-ended questions:

- *Did the image scenery impact your password selections?* (To further elicit the effect of declarative memories on the provided sceneries.)
- *What was the rationale behind your password selections?* (To get insights on the experience- *vs.* random-driven selections.)
- *Are you aware of security aspects of PGA?* (To exclude participants having prior knowledge on PGA security. We intentionally did not ask this question during the recruitment to avoid revealing research aspects about security.)

Responses were grouped based on a coding schema relevant to the questions, and the most relevant responses are reported in the form of quotes in the Main Findings' section.

### *4.4. Analysis of results*

In the analyses that follow, data are mean ± standard error. There were no significant outliers in the data.

#### *4.4.1. Differences in security strength of the created graphical passwords ($RQ_1$)*

To investigate $RQ_1$, we ran an independent-samples *t*-test, with the user group (retrospective *vs.* generic) as the independent variable, and the number of guesses needed to crack the password using the *PoI-assisted Brute-force Attack* model as the dependent variable (Fig. 4). The analysis revealed that the passwords created from the users of the retrospective image group required more guesses to crack (14.49 ± 4.69

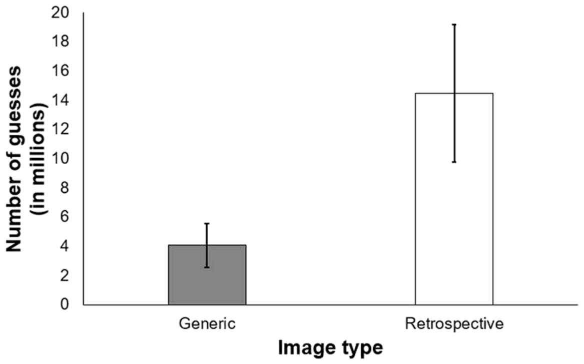


**Fig. 4.** Means of password strength among image types, as assessed by the *PoI-assisted Brute-force Attack* model (Zhao et al., 2013).

million) than users of the generic image group (4.09 ± 1.5 million), a statistically significant difference of 9.8 ± 4.3 million (95% CI, 0.746 to 1.88), $t(21.513)=2.248$, $p=.035$.

Furthermore, the percentage of passwords cracked using the *Knowledge-based PoI-assisted Attack* model was 72.22% for the generic group, while 44.44% for the retrospective group within $2^{18}$ guesses (**Fig. 5**).

To further verify the security strength of the created passwords, we took an extra step to analyze users' individual gestures with respect to hot-spots regions. We ran an independent-samples *t*-test, with the user group (retrospective *vs.* generic) as the independent variable, and the proportion of gestures falling into hot-spots regions as the dependent variable. The analysis (**Fig. 6**) revealed that users of the retrospective image group made a lower proportion of selections falling into hot-spots regions (0.49 ± 0.07) than users of the generic image group (0.83 ± 0.06), a statistically significant difference of −0.332 ± 0.09 (95% CI, −0.533 to −0.131), $t(35)=-3.348$, $p=.002$.

### 4.4.2. Differences in visual behavior during creation of graphical passwords ($RQ_2$)

To investigate $RQ_2$, a one-way multivariate analysis of variance was run to determine the effect of image type on visual behavior. Two measures of visual behavior were assessed: proportion of fixation count on hot-spots and proportion of fixation duration on hot-spots. Users from the generic image group exhibited higher proportion of fixation count on hot-spots (0.417 ± 0.148) than the retrospective image group (0.224 ± 0.171). Also, users from the generic image group exhibited higher proportion of fixation duration on hot-spots (0.515 ± 0.214) than the retrospective image group (0.235 ± 0.209). The differences between the two image type groups on the combined dependent variables was statistically significant, $F(2, 27)=6.217$, $p=.006$; *Wilks' $\Lambda=0.685$; partial* $\eta^2=0.315$. Follow-up univariate ANOVAs showed that both fixation count proportion on hot-spots ($F(1, 28)=10.868$, $p=.003$; *partial* $\eta^2=0.280$) and fixation duration proportion on hot-spots ($F(1, 28)=12.810$, $p=.001$; *partial* $\eta^2=0.314$) were statistically significantly different between the users from the different image type groups.

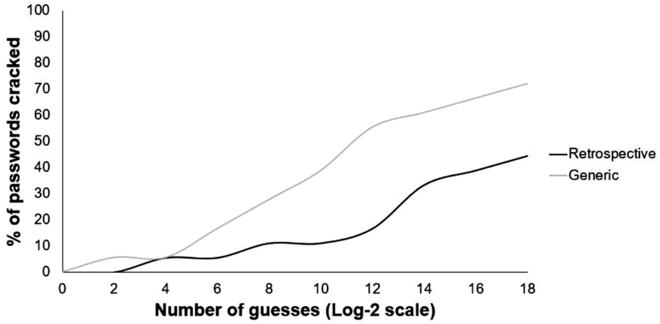


**Fig. 5.** Percentage of passwords cracked, as assessed by the *Knowledge-based PoI-assisted Attack* model (Zhao et al., 2013).

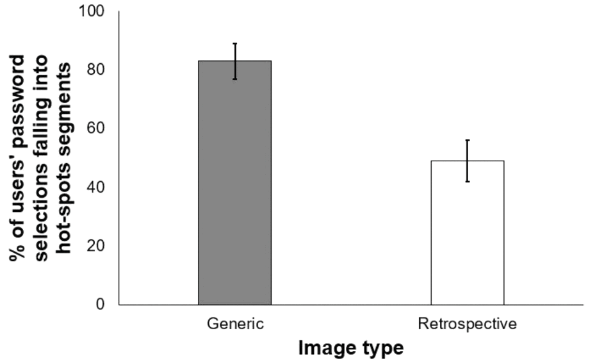


**Fig. 6.** Percentage of users' password selections falling into hot-spots segments among image types.

Furthermore, we conducted a per-second analysis of the total number of fixations and the fixations on hot-spots for each group during the password creation phase (**Fig. 7**). The analysis revealed that while the two groups exhibited a similar number of fixations overall, however, the generic group exhibited higher number of fixations on hot-spots throughout the session. Also, users of the generic group exhibited more and longer fixations on hot-spots *vs.* non-hot-spots compared to the retrospective group, as depicted in **Fig. 8 (left)** and **Fig. 8 (right)** respectively.

### 4.4.3. Correlation of fixations on hot-spots with user-selected passwords that included hot-spots in retrospective vs. generic image groups ($RQ_3$)

To investigate whether users' fixations on hot-spots are correlated with user-selected passwords that included hot-spots in both image groups ($RQ_3$), we performed two Pearson's product-moment correlation tests. The first test was run to assess the relationship between the fixation count proportion on hot-spots and the proportion of hot-spots regions included in user-selected passwords revealing a statistically significant, strong positive correlation between the two, $r(35)=0.531$, $p=.003$. We further assessed the relationship between fixation duration proportion on hot-spots and the proportion of hot-spots regions included in user-selected passwords revealing a statistically significant, strong positive correlation between the two, $r(35)=0.578$, $p=.001$.

### 4.4.4. Association between image semantics and password selection strategy followed across the two image groups ($RQ_4$)

To investigate whether the selected password regions were linked to exclusive personal experiences or were random-driven ($RQ_4$), we conducted an additional analysis on qualitative responses from the semi-structured interviews, by annotating users' quotes as either *experience-* or *random-driven* selections. In the retrospective group, 15 users selected regions based on their experiences related to the image, while 6 users followed a random selection. In the generic group, all users followed a random selection. We conducted a chi-square test for association between image group and password selection approach followed. All expected cell frequencies were greater than five. There was a statistically significant association between image group and password selection approach followed, $\chi^2(1)=23.33$, $p<.001$, with a moderately strong association, $\varphi=0.74$, $p<.001$, suggesting an effect of the depicted image sceneries and semantics, and users' personal experiences towards password selections.

### 4.4.5. Differences in time spent to create a graphical password ($RQ_5$)

The observed differences in users' visual behavior during password creation phase were reflected in the time spent for the completion of the password creation task ($RQ_5$). We ran an independent-samples *t*-test,

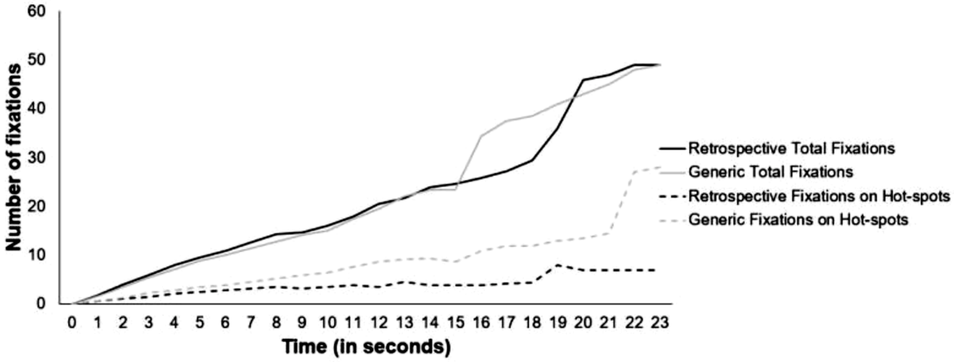


**Fig. 7.** Number of fixations (total and on hot-spots) throughout the password creation phase.

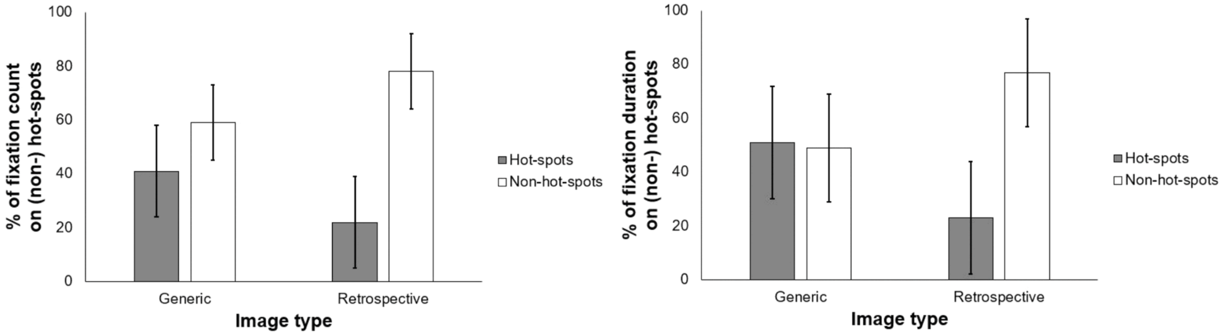


**Fig. 8.** Percentage of fixation count (left) and fixation duration (right) on hot-spots *vs.* non-hot-spots across the two groups.

with the user group (retrospective *vs.* generic) as the independent variable, and the time spent to complete the password creation task as the dependent variable (Fig. 9). The analysis revealed that users of the retrospective image group spent more time to create their password (48.34 ± 7.38 s) than users of the generic image group (28.82 ± 4.80 s), a statistically significant difference of 19.52 ± 8.80 (95% CI, 1.531 to 37.515), $t(29.455)=2.218$, $p=.034$. The analysis also revealed that users of the retrospective image group spent more time to complete the password creation task (68.04 ± 10.69 s), which also included the confirmation phase, than users of the generic image group (43.41 ± 6.71 s), a statistically significant difference of 24.62 ± 12.48 (95% CI, −0.720 to 49.972), $t(35)=1.972$, $p=.05$.

### *4.4.6. Similarities in graphical password regions selections when users created their password on the same image (RQ$_6$)*

To investigate whether individuals who share common sociocultural experiences tend to create similar passwords when they use the same image during password creation, we first split the participants from the retrospective group into subgroups based on the image they used. In the sample of the retrospective group ($n = 18$), 2 out of 9 images from the retrospective image set were not used by any participant. From the remaining 7 images that were selected by participants, 2 images were selected by only one participant, and 5 images were selected by more than one participant, creating five subgroups of participants that had selected the same image.

Given that the implementation of PGA-like mechanisms takes into consideration the order and the type of gestures (*e.g.*, circles are more complex than simple taps but less complex than lines[1]), in order to understand the similarities of users' password selections, we have disregarded the order and the type of the gestures and rather focused on the positions of the password selections. To do so, we simplified the gesture type as follows: For circles, we disregarded the radius and the directionality and kept only the center of the circle as a *x, y* segment, while for lines, we considered only the *x, y* segment of the starting point of the line. Table 2 summarizes the similarities in image regions across users who created their password on the same image. Accordingly, out of 48 gestures made by 16 users who selected the same image, 8 users chose one same region, 4 users chose two same regions, and no user selected all

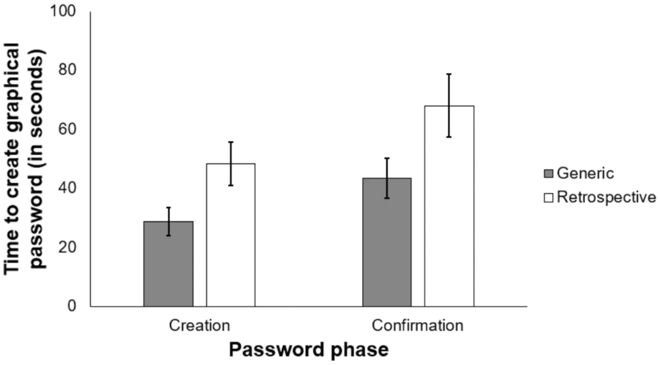


**Fig. 9.** Time spent to complete the password creation and confirmation tasks among image types.

**Table 2**
Summarization of the similarities in image regions across users who created their password on the same image.

| Case # | # of users that selected same image | Common regions in password selections among users | | |
|---|---|---|---|---|
| | | 1 out of 3 | 2 out of 3 | 3 out of 3 |
| 1 | 5 | 4 | – | – |
| 2 | 3 | – | 3 | – |
| 3 | 3 | 2 | 1 | – |
| 4 | 3 | 1 | – | – |
| 5 | 2 | 1 | – | – |
| Total | 16 | 8 | 4 | 0 |

three same regions. We would also like to note that when we consider the exact order of the gestures, there are no observed similarities in image regions across all subgroups and all participants.

#### *4.4.7. Example of users' approaches followed*

Users from the retrospective image group followed a visual behavior on non-hot-spots segments of the image, and subsequently made their selections around these segments, thus leading to a strong password as shown in Fig. 10**(c) top**. On the contrary, users from the generic image group followed a visual behavior around the hot-spots segments of the image, and subsequently made their selections around these segments, thus leading to a weak password Fig. 10**(c) bottom**.

## 5. User study b – does the retrospective approach affect memorability and login usability?

### *5.1. Research questions*

The following main research questions are investigated:

***RQ*₁**. Is there a significant difference in the memorability of graphical passwords between the experimental group *(retrospective-based approach)* and control group *(state-of-the-art approach)*? – *Answering this question will provide insights on whether the retrospective approach improves the memorability of user-created secrets.*

***RQ*₂**. Is there a significant difference in the login time of graphical passwords between the experimental and the control group? – *Answering this question will provide insights about the efficiency of the login phase of the retrospective approach.*

***RQ*₃**. Is there a significant difference in the login failure of graphical passwords between the experimental and the control group? – *Answering this question will provide insights about the effectiveness of the login phase of the retrospective approach.*

### *5.2. Study instruments and metrics*

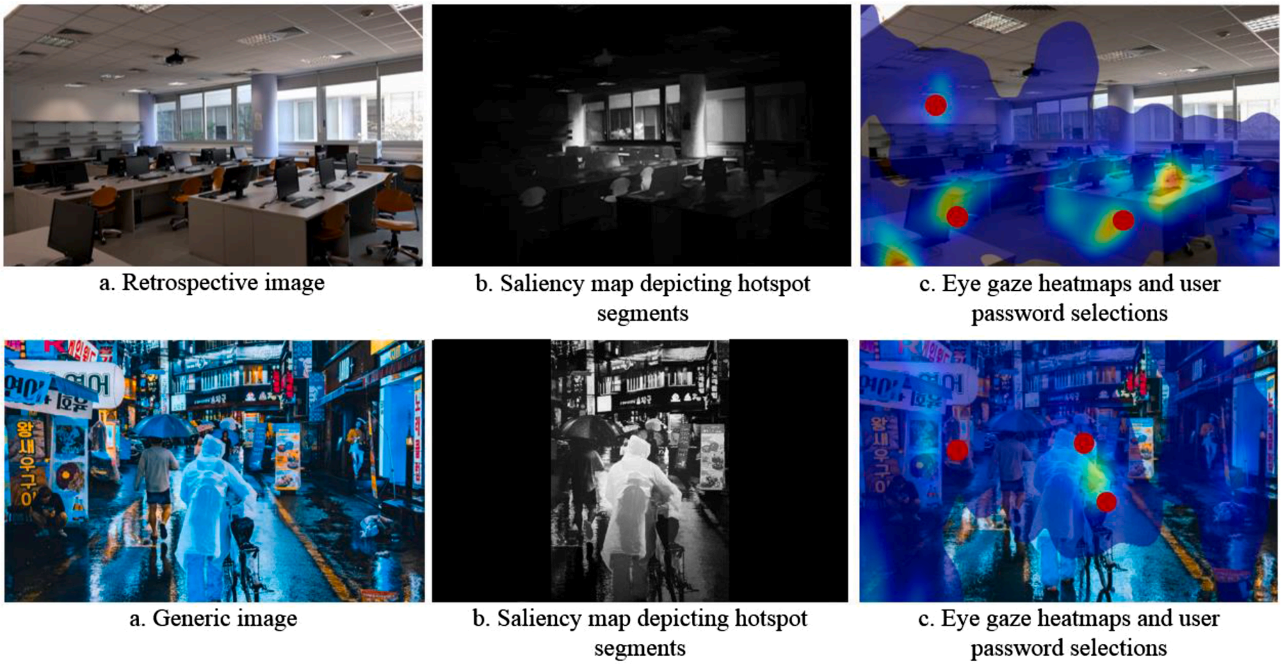
We used the same instruments and equipment as the ones used in *User Study A*. The semantics of image content for the retrospective group were adjusted (Fig. 3 **middle**) to reflect participants' shared, individual and common sociocultural experiences from the daily life context (*i.e.*, depicting sceneries of another University campus). Furthermore, we ensured that the adjusted retrospective image set was similar to the generic set. Table 3 summarizes the image complexity and number of hot-spots regions.

#### *5.2.1. Memorability and login usability metrics*

Following the approach in Stobert and Biddle (2013), we used two metrics as a measure of memorability: *i) memory time*, which is the greatest length of time between a password creation and a successful password login using the same graphical password; and *ii) number of password resets*. With regards to login usability, we measured: *i) login time* which started from the time the image was illustrated to the user, until the user successfully entered the graphical password; and *ii) login failure*, calculated based on the number of sessions that included a failed attempt. A session is considered as failed when more than one attempt is required to login.

**Table 3**
Similarities of number of hot-spots and image complexity for the image sets used in User Study B.

| Image ID | Complexity in bits (retrospective) | Complexity in bits (generic) | Number of Hot-spots Regions (retrospective) | Number of Hot-spots Regions (generic) |
|---|---|---|---|---|
| 1 | 7.51 | 7.62 | 7 | 7 |
| 2 | 7.46 | 7.39 | 7 | 7 |
| 3 | 7.38 | 7.46 | 8 | 7 |
| 4 | 7.69 | 7.75 | 6 | 6 |
| 5 | 7.52 | 7.58 | 7 | 8 |
| 6 | 7.64 | 7.73 | 6 | 6 |
| 7 | 7.51 | 7.54 | 7 | 7 |
| 8 | 7.21 | 7.24 | 6 | 6 |
| 9 | 7.48 | 7.52 | 7 | 7 |



a. Retrospective image

b. Saliency map depicting hotspot segments

c. Eye gaze heatmaps and user password selections

a. Generic image

b. Saliency map depicting hotspot segments

c. Eye gaze heatmaps and user password selections

**Fig. 10.** (a) Original images used by a user from the retrospective group (top) and a user from the generic group (bottom); (b) The saliency maps of the original images depicting the hot-spots segments; (c) Heatmap of fixations during graphical password creation and users' password selections (red circles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### *5.3. Sampling and procedure*

#### *5.3.1. Participants*

A total of 71 individuals (40 females) participated in the study, ranging in age between 18 and 25 years old ($m = 20.76$, $sd$=2.27). We recruited 36 undergraduate students (20 females; $m = 21.58$, $sd$=2.25) from an Asian University, and 35 undergraduate students (20 females; $m = 19.70$, $sd$=1.88) from a European University, through email invitations and in-class announcements made by colleagues. Participants had no relationship to the researchers to avoid biases. Participants were split into two groups based on the image type (*i.e.*, retrospective user group and generic user group), and the image type was randomly varied across all users. To increase the internal validity of the study, we recruited participants that had no prior experience with PGA-like authentication mechanisms nor knowledge of its security semantics, and participants who had spent the last three years at the University campus, assuming they would have had experiences within the University. Postgraduates and faculty were intentionally not considered aiming to: *i)* control the image semantic familiarity factor by illustrating sceneries where undergraduates engaged with everyday-life activities; and *ii)* retain ecological validity (*i.e.*, give incentives to create secure and memorable credentials). We assured that the technical background of the participants would not bias the experiment by including only those with no knowledge of PGA security semantics.

#### *5.3.2. Experimental design and procedure*

The study was split in two phases:

**Phase A.** All participants performed the password creation task in a quiet lab room with only the researcher present. To avoid any bias effects, no details regarding the research objective were revealed to the participants. The study involved the following steps: first, participants were informed that the collected data would be stored anonymously and would be used only for research purposes. Next, they signed a consent form and they completed a questionnaire on demographics. Next, participants were introduced to a demonstration page to familiarize themselves with the process of drawing gestures. Participants were then requested to create a user account in order to access an online service. To increase ecological validity and keep security as a secondary task (Egelman et al., 2013), participants were requested to create an account using our PGA scheme, in order to use this account for accessing a memes website in the following two weeks.

The retrospective user group received a set of nine retrospective images, and the generic user group received a set of nine generic images. In the first step, they created a username and then they selected one image out of the nine available images, on which they created their graphical password by drawing three gestures using a computer mouse. To confirm their graphical password, they were requested to reproduce the initial gestures.

**Phase B.** Following the method in Stobert and Biddle (2013), after Phase A we sent three notification emails on *Day 1, Day 3,* and *Day 6*. Each email directed the participants to the study website and informed them that the website had been updated with new memes. Although the notification emails did not request participants to perform a login, they needed to do so in order to access the updated memes content. The same intervals between the notification emails were also used in the second week of the study (*Day 8, Day 10,* and *Day 13*). The final notification email was sent on *Day 13,* covering the range of a two-week memorability study. With regards to users who failed at dayX and did not reset their password, we considered the range [day1-dayX] for the greatest memory time range calculations. With regards to users who reset their password at dayY, we considered the maximum range MAX([*day*1--*day*Y], [dayY-day13]) for the greatest memory time range calculations. Finally, to check whether participants used any form of external storage during the study to help them remember their graphical passwords (*e.g.*, write down their gestures and look at them during login), we conducted a post-study interview in which participants were interviewed about external password storage to exclude them from the analysis. There were no participants that used any form of external password storage.

### *5.4. Analysis of results*

In the analyses that follow, data are mean ± standard error. There were no significant outliers in the data.

#### *5.4.1. Differences in memorability of graphical passwords across the two image groups ($RQ_1$)*

The maximum memory time that someone could achieve was approximately 336 h (14 days x 24 h). To investigate $\text{RQ}_1$, we conducted an independent-samples *t*-test, with the user group (retrospective *vs.* generic) as the independent variable, and the memory time as the dependent variable (**Fig. 11**). The analysis revealed that memory time between the two user groups was not significantly different ($t(42)=-1.068$, $p$=.292); memory time of the retrospective image group was 216.40 ± 24.22 h, while memory time of the generic group was 249.09 ± 19.31 h.

As an additional measure of memorability, we recorded the number of password resets per participant. The median number of resets for both the retrospective and the generic image groups was 0. A Mann-Whitney U test revealed that the number of resets for the retrospective image group (19 resets; *mean rank*=33.02) was not statistically significantly different than the generic image group (18 resets; *mean rank*=29.89), $U = 431.50$, $z=-0.814$, $p$=.416.

#### *5.4.2. Differences in login time across the two image groups ($RQ_2$)*

Following an existing approach from Belk et al. (2017a), time to login data were analyzed using R (R Core Team, 2015) with the *lme4* package (Bates et al., 2014) using a mixed effects analysis since this enabled us to handle all the variables of the study while accounting for repeated-measures of individuals (6 email notifications across a period of two weeks, yielding a maximum of 426 login observations). Another advantage of such statistical models is that they can handle missing data of users, *e.g.*, a user that has not participated in some sessions across the two weeks of the study can be used in the analysis without requiring removing the user from the sample, as opposed to an analysis of a repeated-measures ANOVA (Pinheiro and Bates, 2006).

For login time differences (*$RQ_2$*), we performed a mixed effects analysis of the relationship between the time to successfully authenticate (by also including any failed attempts that eventually ended in a successful authentication) and the image type. As fixed effects, we entered image type (retrospective and generic) into the model. As random effects, we used subjects in order to account for non-independence of measures. Visual inspection of residual plots revealed that linearity and homoscedasticity were not violated. *P*-values were obtained by likelihood ratio tests of the full model with the effect in

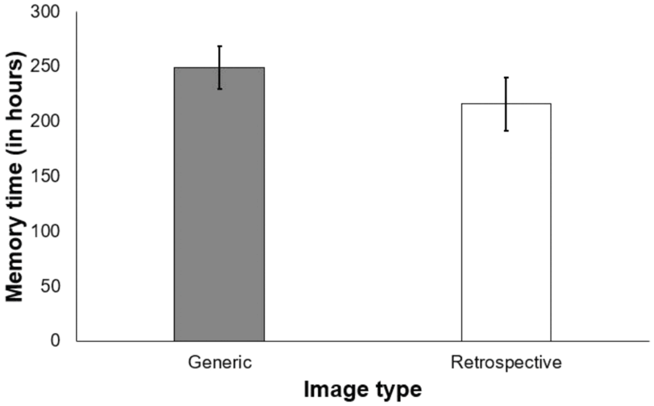


**Fig. 11.** Memory time across the two groups.

question against the model without the effect in question (Winter and Grawunder, 2012). The analysis revealed that the image type had no impact on the time needed to authenticate ($x^2(1)=1.004, p=.31$). During week 1 of the study, the mean login time for the users of the retrospective group was 13.68 ± 2.02 s, while for the users of the generic group was 11.27 ± 1.49 s. During week 2 of the study, the mean login time for the users of the retrospective group was 8.91 ± 1.44 s, while for the users of the generic group was 15.10 ± 4.20 s. Fig. 12 depicts the mean login time across image group for each week.

### 5.4.3. Differences in login failure across the two image groups ($RQ_3$)

Among 224 user authentication sessions, 49 attempts failed (21.875% overall failure rate). For each authentication attempt of each user, we entered a flag indicating whether the particular attempt was successful or unsuccessful. Accordingly, for login failure ($RQ_3$), we performed a mixed logistic regression with the task attempt (successful *vs.* unsuccessful) as the dependent variable. The independent variable was used as fixed effects (image type), and the subjects as random effects. For significance testing we tested the full model against a model without the effects in question by obtaining their likelihood ratio tests. The analysis revealed that the image type had no impact on login failure ($x^2(1)=0.01, p=.90$). During week 1 of the study, there were 20 sessions that failed for the users of the retrospective group, while there were 13 sessions that failed for the users of the generic group. During week 2 of the study, there were 4 sessions that failed for the users of the retrospective group, while there were 12 sessions that failed for the users of the generic group. Fig. 13 depicts the failed attempts across image group for each week of the study.

## 6. User study c – does the retrospective approach introduce guessing vulnerabilities by individuals close to the user?

Bearing in mind that when using the retrospective approach, graphical password selections are based on the users' existing sociocultural experiences, it is probable that the individuals who share common experiences with the end-users might be able to guess their selections. In order to shed light on this aspect, we have conducted a human attack study focusing on guessing vulnerabilities of the approach among people sharing common sociocultural experiences. Each session of the study embraced pairs of participants that were closely related (*e.g.*, friends, couples, relatives, etc.) and who shared common experiences. In each session, we asked both participants to first create a graphical password, and then each participant was asked to guess the password selections of the other participant.

### 6.1. Research questions

The following research question is investigated:

***RQ***. Does the retrospective-based approach within graphical passwords entail guessing vulnerabilities in terms of allowing attackers who share common experiences with the end-users to more easily identify regions of their selected secrets? – *Answering this question will provide insights on whether the retrospective approach introduces guessing vulnerabilities in human guessing attacks.*

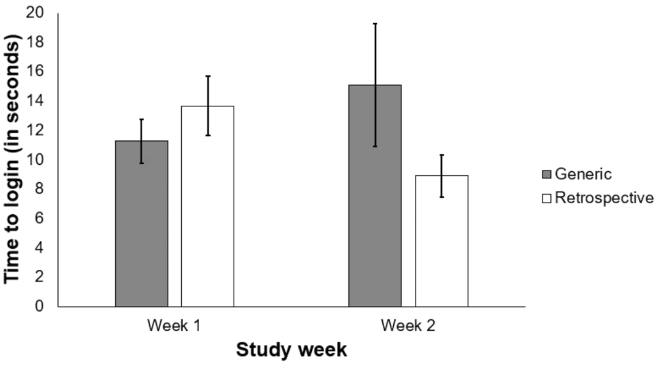


**Fig. 12.** Mean login time across the two groups, during each week of the study.

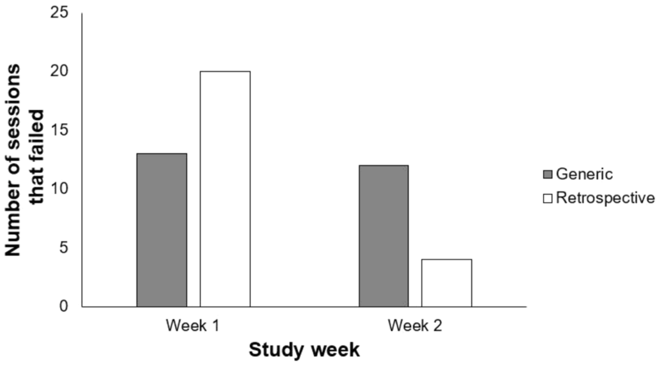


**Fig. 13.** Number of failed attempts across the two groups, during each week of the study.

### 6.2. Study instruments and metrics

We used the same instruments and equipment as the ones used in *User Study A and User Study B*. The semantics of image content for the retrospective group were adjusted to reflect participants' shared, individual and common sociocultural experiences from the daily life context (*i.e.*, working environment in the case of colleagues, café/bars in which couples or close friends usually hang out), as depicted in Fig. 14. For doing so, prior to the study, we asked each pair of participants to provide a set of images from places they share common experiences with. To avoid bias effects, we did not inform the participants about the reason they were providing the images until the end of the study.

Furthermore, we ensured that the adjusted retrospective image set included pairs of images that were similar in terms of image complexity and number of hot-spots regions as shown in Table 4. With regards to calculating the password strength, we adjusted the *PoI-assisted Brute-force Attack* model from *User Study A* to start from segments covering the segments provided by each attacker, then checking the neighboring segments, then checking the hot-spots segments and their neighboring segments, and finally checking the rest of the segments.

### 6.3. Sampling and procedure

#### 6.3.1. Participants

A total of 26 individuals (12 females) participated in the study, ranging in age between 25 and 60 years old ($m$ = 40.03, $sd$=10.23). Since the purpose of this study was to understand how individuals decide on their selections when performing an attack on a password created by another individual with whom they share common experiences within places depicted on retrospective images, we intentionally recruited pairs of participants that are close to each other (5 couples, 3 pairs of close friends, 5 pairs of colleagues). To increase the internal validity of the study, we recruited participants that had no prior experience with PGA-like authentication mechanisms nor knowledge of its security semantics, as assessed by a post-study interview in order to exclude any participants that have prior knowledge on PGA security.

#### 6.3.2. Experimental design and procedure

With respect to the ethical and safety aspects of the study, we adopted state-of-the-art human research protocols that take into consideration users' privacy, confidentiality and anonymity, as well as all the necessary measures against Covid-19 to ensure the participants' safety. Each session of the study entailed a pair of participants that were closely related (*e.g.*, friends, couples, relatives, etc.). Both participants

**Fig. 14.** A subset of images used in the human attack study illustrating sceneries in which participants share common experiences.

**Table 4**
Similarities of number of hot-spots and image complexity for the 5 image sets used in User Study C.

| Image set | Image ID | Complexity in bits | Number of Hot-spots Regions |
|---|---|---|---|
| 1 | 1 | 7.49 | 7 |
| | 2 | 7.45 | 7 |
| 2 | 1 | 7.58 | 7 |
| | 2 | 7.64 | 6 |
| 3 | 1 | 7.37 | 6 |
| | 2 | 7.43 | 7 |
| 4 | 1 | 7.57 | 7 |
| | 2 | 7.52 | 8 |
| 5 | 1 | 7.44 | 7 |
| | 2 | 7.49 | 6 |

were first asked to create a graphical password, and then guess the password of each other. The study was run in a quiet lab room with only the researcher present. The study was split in two phases (password creation phase, and human guessing attack phase) as follows:

**Phase A – Password Creation.** During the first phase, each pair of participants visited the laboratory in a pre-scheduled time within the Covid-19 safety regulations, and were asked independently to create a graphical password in order to access an online service. To avoid bias effects during the attack phase, each participant created a password on a different image provided by them that depicted places in which they share common experiences.

**Phase B – Human Guessing Attack.** In this phase, we switched the image of the pairs and each participant was asked to guess the other participant's secrets as follows: *i)* by first indicating 3 areas (*x, y* segments on the grid) on the image for which they believe that the other participant made their selections around them; and then *ii)* by actually drawing 3 gestures for a total of 3 attempts to guess the actual password (*i.e.*, considering the ordering of gestures and type of gestures). During the attack phase, we adopted the think-aloud protocol aiming to elicit whether the rationale behind the attacker's selections is related to the shared memories and experiences she possesses with the other participant from the same pair. Finally, both participants completed a questionnaire on demographics.

### 6.4. Analysis of results

To investigate the *RQ*, we conducted three analyses: *i)* we calculated the Euclidean distance of the attackers' guessing selections from the end-users' password secret selections; *ii)* based on the first analysis (Euclidean distance), we adjusted the brute-force attack performed in *User Study A* in order to investigate whether users who share common experiences were able to run a more effective attack by starting to guess regions they suspected that the users selected their password; and *iii)* we performed a qualitative analysis based on the participants' responses and researcher's observations to triangulate and better understand the approach followed by attackers on graphical passwords created on retrospective images. In the analyses that follow, data are mean ± standard error. There were no significant outliers in the data.

#### 6.4.1. Euclidean distance of attackers' selections from the end-users' secret selections

To investigate how far the attackers' guessing selections were from the end-users' actual secret selections, we calculated the Euclidean distance between the 3 *x, y* segments provided by the attacker and the 3 *x, y* segments of the end-user. For doing so, we performed three analyses: *i)* disregarding the type of the gesture and the exact order, *i.e.*, the attacker's first segment was compared to the end-user's closest segment, the attacker's second segment compared to the end-user's closest segment, and the attacker's third segment compared to the end-user's closest segment; *ii)* disregarding the type of the gesture, but considering the exact order, *i.e.*, the attacker's first segment was compared to the end-user's first segment, the attacker's second segment compared to the end-user's second segment, and the attacker's third segment compared to the end-user's third segment; and *iii)* considering the type of the gesture and exact order, which in principle, simulates an online guessing attack.

*A. Disregarding the type and the exact order of the gesture.* Fig. 15 depicts the Euclidean distance of each gesture of each participant by disregarding the type and the exact order of the attackers' gestures and the end-users' gestures. For the analysis, we adopted a threshold of 3 segments, by considering the allowed tolerance of the PGA mechanism[1]. Accordingly, among 78 gestures (3 gestures x 26 participants), 16 gestures (20%) were in close proximity with the attacker's guessed selections. Furthermore, we conducted a one-way repeated measures ANOVA to determine whether there was a statistically significant difference in Euclidean distance over the 3 gestures. There were no outliers, and the data were normally distributed at each time point, as assessed by boxplot and Shapiro-Wilk test ($p$>.05), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2)$=0.699, $p$=.705. The segments selected by attackers did not elicit statistically significant changes in Euclidean distance across the 3 gestures, $F$ (2, 50)=1.950, $p$=.153, *partial* $\omega^2$=0.02, with Euclidean distance

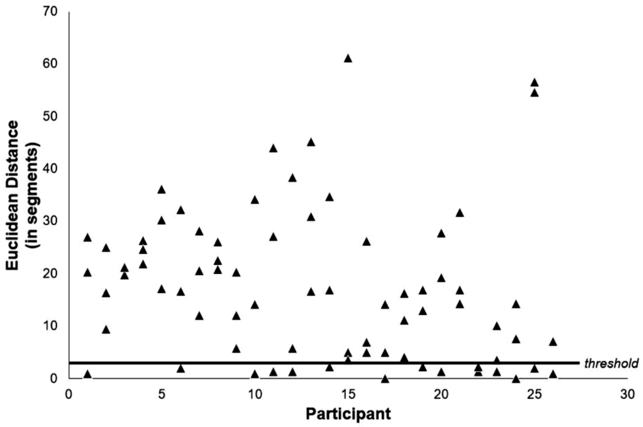


**Fig. 15.** Euclidean distance between attackers' gestures and end-users' gestures, by disregarding the type and the exact order.

decreasing from 21.54 ± 3.10 for the first selection to 14.74 ± 2.79 for the second selection and slightly increasing to 15.32 ± 2.25 for the third selection.

*B. Disregarding the type of the gesture but considering the exact order.* Fig. 16 depicts the Euclidean distance by disregarding the type of selections, but considering the exact order of the attackers' gestures and the end-users' gestures. Applying the same threshold of 3 segments, the analysis revealed that among 78 gestures (3 gestures x 26 participants) made by the participants, 8 gestures (10%) were in close proximity with the attacker's guessed selections. Furthermore, we conducted a one-way repeated measures ANOVA to determine whether there was a statistically significant difference in Euclidean distance over the 3 gestures. There were no outliers, and the data were normally distributed at each time point, as assessed by boxplot and Shapiro-Wilk test ($p$>.05), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2)=0.200$, $p$=.905. The segments selected by attackers did not elicit statistically significant changes in Euclidean distance across the 3 gestures, $F(2, 50)=0.274$, $p$=.761, *partial* $\omega^2=-0.01$, with Euclidean distance decreasing from 32.19 ± 4 for the first gesture to 27.99 ± 4.20 for the second gesture and slightly increasing to 29.53 ± 4.19 for the third gesture.

*C. Considering the type and the exact order of the gesture.* We compared the 3 attempts of each attacker with the end-user's stored password from the same pair of participants. From a total of 78 attacking guesses (3 attempts of each attacker x 26 participants), there was only 1 successful attempt, yielding an online success guessing rate of 0.01%. It is worth noting that the successful online attack contained 3 gestures on 3 hot-spots areas. Although it is a bit surprising for an online attack to succeed with only 3 guessing attempts, considering the complexity and the ordering of PGA-like mechanisms, both the attacker and the creator confirmed that they followed the same approach of selecting 3 areas that attracted their attention initially. In particular, they both drew 3 taps, which are the simplest and least secure types of gestures[1], from left to right on the 3 office chairs depicted on the image.

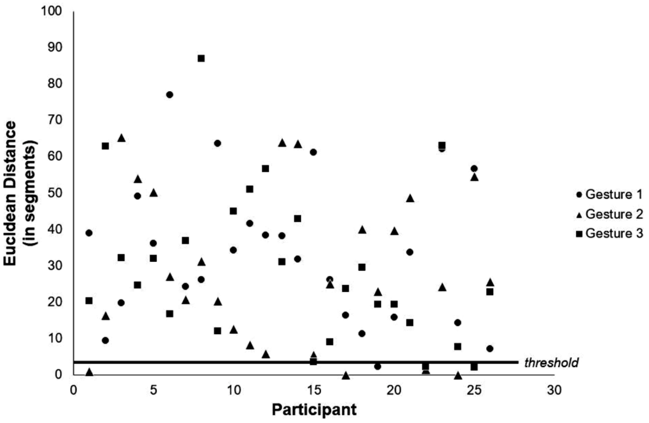


**Fig. 16.** Euclidean distance between attackers' gestures and end-users' gestures, by disregarding the type but considering the exact order.

### *6.4.2. Security strength of the created retrospective graphical passwords based on experience-spot-driven brute-force attack*

To investigate whether the suggested retrospective approach holds against attacks when considering the experience-spots indicated by each participant that acted as an attacker, we conducted an offline attack comparing a POI-assisted brute-force attack (the same attack that considers hot-spots regions as in User Study A) and a personalized POI-assisted brute-force attack that was further enhanced to consider the experience-spots regions as indicated by the human attacker.

*A. Disregarding the type and the exact order of the gesture across all participants.* Given that the implementation of PGA-like mechanisms takes into consideration the type and the exact order of the gestures, which could impact the total guesses required to crack a graphical password (*e.g.*, circles are more complex than simple taps but less complex than lines[1]), it is interesting to first understand how each attack type (*PoI-assisted Brute-force Attack vs. Personalized PoI-assisted Brute-force Attack*) performs when we disregard the type and the exact order of the gestures and rather focus on the positions of the password selections. To do so, we simplify the gesture type as follows: For circles we disregard the radius and the directionality and keep only the center of the circle as a *x, y* segment, while for lines we consider only the *x, y* segment of the start of the line.

We ran an independent-samples $t$-test, with the type of attack (*PoI-assisted Brute-force Attack vs. Personalized PoI-assisted Brute-force Attack* by considering also the experience-spots provided by the attackers) as the independent variable, and the number of guesses needed to crack the password as the dependent variable, without taking into account the order and the type of the gestures across all participants (Fig. 17 **left**). The analysis revealed that the number of guesses required to crack the passwords using the *Personalized PoI-assisted Brute-force Attack* (66.63 ± 21.10 thousand) was not statistically significantly different than the *PoI-assisted Brute-force Attack* (98.04 ± 27.76 thousand), 95% CI, −38.62 to 101.45 thousand, $t(50)=0.901$, $p$=.372. The analysis revealed that, although not statistically significantly different, the *Personalized PoI-assisted Brute-force Attack* required less attempts to crack the passwords than the *PoI-assisted Brute-force Attack*.

*B. Disregarding the type and the exact order of the gesture across participants with at least one gesture containing experience-spot.* We ran an independent-samples $t$-test, with the type of attack (*PoI-assisted Brute-force Attack vs. Personalized PoI-assisted Brute-force Attack* by considering also the experience-spots provided by the attackers) as the independent variable, and the number of guesses needed to crack the password as the dependent variable, without taking into account the type and the exact order of the gestures across participants that recorded at least one gesture containing an experience-spot (Fig. 17 **right**). The analysis revealed that the passwords in which the *Personalized PoI-assisted Brute-force Attack* was performed, required less guesses to be cracked (20.26 ± 5.38 thousand) than the passwords in which the *PoI-assisted Brute-force Attack* was performed (49.76 ± 11.8 thousand), a statistically significant difference of 29.5 ± 12.97 thousand (95% CI, 3.19 to 55.8 thousand), $t(36)=2.274$, $p$=.029.

*C. Considering the type and the exact order of the gesture across all participants.* We ran an independent-samples $t$-test, with the type of attack (*PoI-assisted Brute-force Attack vs. Personalized PoI-assisted Brute-force Attack* by considering also the experience-spots provided by the attackers) as the independent variable, and the number of guesses needed to crack the password as the dependent variable, by taking into account the type and the exact order of gestures across all participants (Fig. 18). The analysis revealed that the number of guesses required to

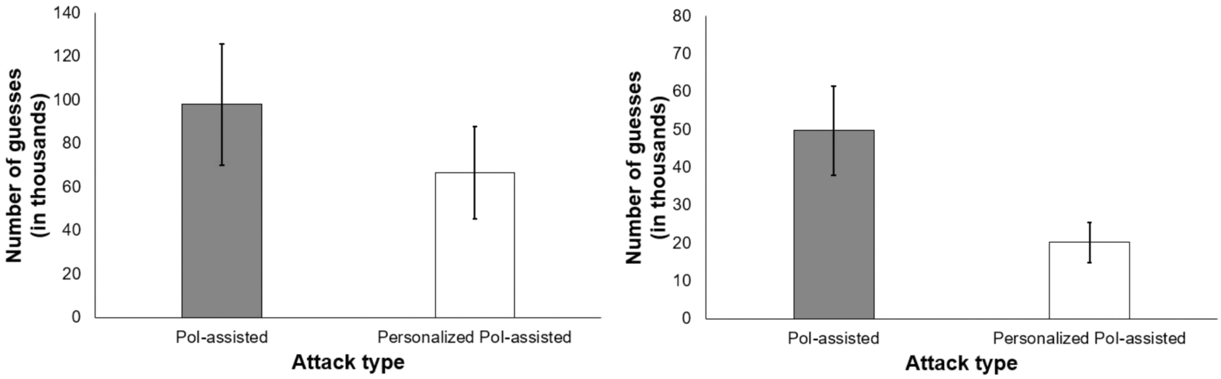


**Fig. 17.** Means of password strength among attack types by disregarding the order and the type of gestures across all participants (left) and across participants with at least one gesture containing experience-spot (right).

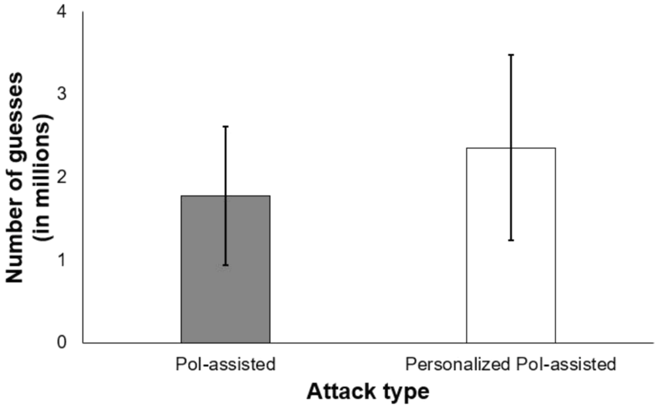


**Fig. 18.** Means of password strength among attack types.

crack the passwords using the *Personalized PoI-assisted Brute-force Attack* (2.36 ± 1.12 million) was not statistically significantly different than the *PoI-assisted Brute-force Attack* (1.78 ± 0.84 million), 95% CI, −3.4 to 2.23 million, $t(50)=-0.417$, $p=.678$. Furthermore, the percentage of passwords cracked using the *PoI-assisted Brute-force Attack* was 55.55%, while it was 38.88% using the *Personalized PoI-assisted Brute-force Attack* within $2^{18}$ guesses (Fig. 19).

### *6.4.3. Qualitative analysis*

To further shed light and understand the approach followed by attackers on graphical passwords created on retrospective images, we used the data gathered from the think-aloud protocol, as well as observations made by the researchers during the attack phase. In many cases, attackers used knowledge about the end-user under attack, related to their habits, preferences and facts about their personality:

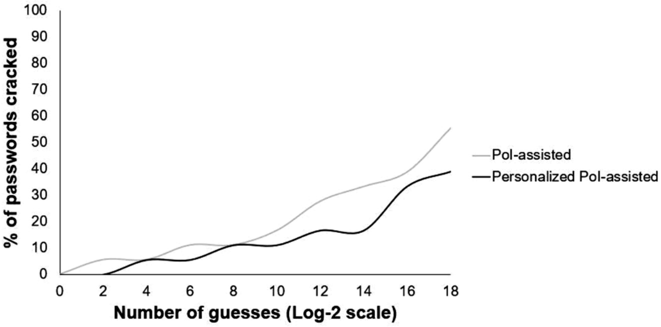


**Fig. 19.** Percentage of passwords cracked among attack types.

*"I go to this Starbucks cafe quite often with my husband. I believe that he must have selected the ashtray since he is a heavy smoker." ~ P02*

*"My son enjoys getting involved in the preparation of the barbecue, that's why I made some of my selections around the grill motor and the meat." ~ P11*

*"I think she must have selected the flower pot at the desk since she likes flowers a lot." ~ P17*

*"She likes bright colors so I believe she must have selected the flowers." ~ P19*

*"I would say that one of her selections must have been the stand for recyclable batteries because she is an environmentally friendly person. Also, she is a religious person, so I believe another of her selections must have been the religious painting on the wall." ~ P20*

*"My girlfriend is a mathematician and I think she should have picked a password that somewhat resembles a pattern of 3 points. Having this in mind, I made my selections in three tables that are close to each other and form a line." ~ P25*

In other cases, it is evident that the scenery depicted on the retrospective images impacted the selections of the attackers. In particular, attackers used a more personalized approach by considering specific information related to their common shared experiences with the end-user under attack within the places depicted on the retrospective images:

*"I think he must have picked the chairs and the table we usually sit at when we visit this café." ~ P02*

*"I picked a place where he usually leaves his car in the parking lot. Also, sometimes we walk together on our way to exit the building towards the parking lot, so I made one of my selections at the end of this route." ~ P08*

*"I believe that he selected the table that we usually book when we go to this bar for a drink. Also, his brother used to work as a bartender at this place a few months ago, so I believe he must have also selected the area that the drinks are being prepared." ~ P09*

*"Also, I selected the place that he uses for storage of his bicycle in the backyard of the house." ~ P11*

*"I believe she selected the hand sanitizer at the office desk because she is obsessed with hands hygiene, especially this period due to the COVID-19 situation." ~ P17*

*"He is always friendly and kind with the office visitors, hence, I made some of my selections at the visitors' area on the image. Also, due to the nature of his responsibilities he is required to make a lot of photocopies every day, so I made my last selection near the photocopy machine." ~ P18*

*"Although the parking lot contains various types of cars, I think he must have selected the pickup trucks we usually use for work purposes out in the fields." ~ P24*

In very few cases, attackers did not employ any sophisticated attack, but rather focused on the obvious hot-spots of the images:

*"I selected the points that attract your attention easily and I believe every person would have selected." ~ P16*

"I think she selected the chair and the side-table because these are the objects that attract someone's attention in the first place." ~ *P21*

The above observations were concentrated in a coding schema relevant to the approach followed by the attackers as follows. Table 5 summarizes the responses about the approach employed by the attackers based on the aforementioned coding schema.

- Habits/Preferences/Characteristics of end-users (*e.g.*, smoking, flowers, religion)
- Common shared experiences (*i.e.*, experiences in the depicted place/scenery)
- Random-guessing approach relying on areas of the image that attract peoples' attention (*i.e.*, hot-spots)

## 7. Main findings

Observations of users' visual behavior allowed us to draw conclusions about their approaches during password creation and the impact of the image content on the strength and memorability of created passwords. Table 6 summarizes the main results.

### *7.1. Finding a – the retrospective approach impacts users' visual behavior related to hot-spots segments during pga password creation*

Users from the retrospective group exhibited more and longer fixations on non-hot-spots regions of the image (Fig. 8 **left and right**). This could be explained by the fact that image content familiar to users' sociocultural experiences triggered users to explore the non-hot-spots regions of the image before making their password selections. It could be further explained by the fact that the depicted content was processed at a deeper and more meaningful way since users could connect prior experiences in their declarative memory (Tulving, 1972), enabling them to fixate on *experience-spots* (*i.e.*, regions of the image that are related to their prior sociocultural experiences). Furthermore, participants from the retrospective image group were able to recall information related to the depicted content.

*"I was excited to create a password on an image from my University's campus. At first, I spent some time recalling my recent experiences within the cafeteria shown on the image and then I made my selections." ~ P02 (User Study A)*

This could be explained by the *"self-reference effect"*, which states that people remember information best when they are personally involved in that information (Rogers et al., 1977). Furthermore, familiarity and relevance of information to one's life have been found to influence password selection (Riddle et al., 1989).

On the contrary, fixations of users from the generic group were mostly concentrated around hot-spots regions of the image. This could be explained by the fact that users could not easily connect prior experiences in their declarative memory and attach a semantic meaning to the depicted generic content, therefore, their visual attention was rather concentrated around specific regions of interest (Bulling et al., 2012; Belk et al., 2017b).

*"Why don't you use more familiar images? Since we spend many hours at the University, I was wondering if you could use images from the University's campus." ~ P14 (User Study A)*

*"The image showed a setting completely strange to me, so I didn't bother searching on the entire image to make my selections." ~ P24 (User Study A)*

**Table 5**
Summarization of the approach followed by the attackers based on the coding schema extracted from data collected during the attack phase.

| Attacking approach followed | Frequency |
|---|---|
| Habits/preferences/characteristics of end-users | 11 out of 26 |
| Common shared experiences | 19 out of 26 |
| Random-guessing approach | 6 out of 26 |

**Table 6**
Summarization of the main results.

| Impact | Metric | Retrospective group (mean ± *SE*) | Generic group (mean ± *SE*) | Differences between the two groups |
|---|---|---|---|---|
| Image content on visual behavior | Proportion of fixation count on hot-spots (combined) | 0.224 ± 0.171 | 0.417 ± 0.148 | *F*(2, 27)= 6.217, *p*=.006; *Wilks' Λ*=0.685; *partial* $\eta^2$=0.315 |
| | Proportion of fixation duration on hot-spots (combined) | 0.235 ± 0.209 | 0.515 ± 0.214 | |
| | Proportion of fixation count on hot-spots (univariate) | n/a | n/a | *F*(1, 28)= 10.868, *p*=.003; *partial* $\eta^2$=0.280 |
| | Proportion of fixation duration on hot-spots (univariate) | n/a | n/a | *F*(1, 28)= 12.810, *p*=.001; *partial* $\eta^2$=0.314 |
| | Time spent to complete password creation task | 68.04 ± 10.69 s | 43.41 ± 6.71 s | 24.62 ± 12.48 s (95% CI, −0.720 to 49.972), *t* (35)=1.972, *p*=.05 |
| Image content on security strength of the created passwords | Number of guesses needed to crack the password | 14.49 ± 4.69 million | 4.09 ± 1.5 million | 9.8 ± 4.3 million (95% CI, 0.746 to 1.88), *t* (21.513)= 2.248, *p*=.035 |
| | Proportion of gestures fallings into hot-spots regions | 0.49 ± 0.07 | 0.83 ± 0.06 | −0.332 ± 0.09 (95% CI, −0.533 to −0.131), *t* (35)=− 3.348, *p*=.002 |
| Correlation of fixations on hot-spots and user-selected passwords that included hot-spots | Proportion of fixation count on hot-spots and proportion of hot-spots regions included in user-selected passwords | n/a | n/a | *r*(35)=0.531, *p*=.003 |
| | Proportion of fixation duration on hot-spots and proportion of hot-spots regions included in user-selected passwords | n/a | n/a | *r*(35)=0.578, *p*=.001 |
| Image content on password selection approach followed | Annotate users' quotes as either experience- or random-driven | n/a | n/a | $\chi^2$(1)=23.33, *p*<.001; $\varphi$=0.74, *p*<.001 |
| Image content on memorability and login usability | Memory time | 216.40 ± 24.22 h | 249.09 ± 19.31 h | *t*(42)=− 1.068, *p*=.292 |
| | Number of resets | n/a | n/a | *U* = 431.50, *z*=−0.814, *p*=.416 |
| | Time to login | n/a | n/a | $x^2$(1)=1.004, *p*=.31 |
| | Failed attempts | n/a | n/a | |

(*continued on next page*)

**Table 6** (*continued*)

| Impact | Metric | Retrospective group (mean ± *SE*) | Generic group (mean ± *SE*) | Differences between the two groups |
|---|---|---|---|---|
| | | | | $x^2(1)$=0.01, *p*=.90 |
| Resistance of the retrospective approach against people that share common experiences with the users | Number of guesses needed to crack the password (considering the type and the exact order of the gesture - all participants) | 1.78 ± 0.84 million (*PoI-assisted*)<br>2.36 ± 1.12 million (*Personalized PoI-assisted*) | n/a<br>n/a | 95% CI, −3.4 to 2.23, *t* (50)=−0.417, *p*=.678 |
| | Number of guesses needed to crack the password (disregarding the type and the exact order of the gesture - all participants) | 98.04 ± 27.76 thousand (*PoI-assisted*)<br>66.63 ± 21.10 thousand (*Personalized PoI-assisted*) | n/a<br>n/a | 95% CI, −38.62 to 101.45, *t* (50)=0.901, *p*=.372 |
| | Number of guesses needed to crack the password (disregarding the type and the exact order of the gesture - participants with at least one experience-spot) | 49.76 ± 11.8 thousand (*PoI-assisted*)<br>20.26 ± 5.38 thousand (*Personalized PoI-assisted*) | n/a<br>n/a | 29.5 ± 12.97 thousand (95% CI, 3.19 to 55.8), *t*(36)= 2.274, *p*=.029 |
| | Euclidean distance between attackers' and creators' gestures (disregarding the type of the gesture but considering the exact order) | n/a | n/a | *F*(2, 50)= 0.274, *p*=.761; *partial* $\omega^2$=−0.01 |
| | Euclidean distance between attackers' and creators' gestures (disregarding the type and the exact order of the gesture) | n/a | n/a | *F*(2, 50)= 1.950, *p*=.153; *partial* $\omega^2$=0.02 |

### 7.2. *Finding b – the retrospective approach impacts users' pga passwords selections related to hot-spots segments*

Users from the retrospective group exhibited a lower proportion of password selections falling into hot-spots segments than users of the generic image group (Fig. 6). Participants from the retrospective image group made their selections based on their memories and experiences, while avoiding selections on the obvious hot-spots.

*"As a Computer Science student, I spend a lot of time studying in this lab room, therefore, I have many experiences in this place. However, my fellow students know which seat I usually take, so I decided not to make any selections around this area because they could easily guess it." ~ P04 (User Study A)*

On the contrary, participants from the generic image group made most of their selections around the easy-to-remember hot-spots segments, which attract users' attention by default.

*"I selected points that I can easily remember, therefore, I made my selections around points that stand out." ~ P22 (User Study A)*

### 7.3. *Finding c – fixations on hot-spots segments correlate with users' selections that included hot-spots segments for both experimental and control groups*

In both user groups, there was a strong correlation between the fixation count on hot-spots regions and user-selected passwords that included hot-spots regions, as well as a strong correlation between the fixation duration on hot-spots regions and user-selected passwords that included hot-spots regions. However, we stress that users who utilized retrospective images had statistically significant lower visual exploration on hot-spots regions (Fig. 8 **left and right**), and hence lower number of proportions on hot-spots regions in their passwords' selections (Fig. 6).

### 7.4. *Finding d – the retrospective approach improves significantly the security of the user-chosen graphical passwords*

Results revealed a main effect of image type towards graphical password strength. Users that created passwords with the retrospective image created significantly stronger passwords than users with the generic image (Figs. 4, 5). This could be explained by the fact that users who utilized retrospective image content not only exhibited limited and shorter attention on the hot-spots segments of the image (Fig. 8 **left and right**), but also avoided making password selections around them (Fig. 10).

*"When I saw this image, I remembered an incident related to a fellow student which took place last year at the University's parking lot. So, what happened was that the security guy placed a wheel clamp on his vehicle for parking violation. For my password selections, I avoided selecting one of the six cars depicted in the parking lot, and I decided to make some of my password selections around the area in which the incident took place because I believe it will be difficult for others to figure out my selections." ~ P07 (User Study A)*

On the contrary, results revealed that participants from the generic image group created less secure passwords (Figs. 4, 5), with most of their selections being around the obvious hot-spots segments.

*"I selected the points which looked attractive at first sight. This way I can remember my password." ~ P25 (User Study A)*

### 7.5. *Finding e – the retrospective approach impacts users' strategies followed during graphical password creation*

The analysis of users' qualitative feedback suggested an effect of the depicted image content on the users' approach followed during password creation. In particular, the majority of users from the retrospective group made their password selections by considering their personal experiences within the depicted image content, whereas all users from the generic group followed a random approach. Similar to *Finding A*, this could be explained by the fact that the depicted content was processed at a deeper and more meaningful way since users could connect prior experiences in their declarative memory (Tulving, 1972). Furthermore, given that the task of creating a graphical password is a perceptual process through the human visual system (Biddle et al., 2012), this finding could be further explained by the fact that people's perceptions are influenced more by prior experiences than by newly arriving sensory information from the eyes (Gonzalez-Garcia et al., 2018).

### 7.6. *Finding f – the retrospective approach did not affect significantly memorability and login usability of graphical passwords*

Analysis of memorability and login usability revealed no main effects on memory time, time to login and login failure (Figs. 11-13).

Nonetheless, a comparison between the two weeks for each group reveals that both time to login and failed attempts for users from the retrospective group were reduced in the second week compared to the first week. On the contrary, in the case of the generic group, no differences between the two weeks were observed with regards to the time to login and failed attempts.

### *7.7. Finding g – the retrospective approach increases guessing vulnerability within human guessing attacks, which however does not compromise the security of the pga system by considering additional measures such as the type and the exact order of the gestures*

The human attack study revealed that vulnerabilities exist in case someone knows the user, since analyses indicate that individuals that share common experiences may spot certain regions that the end-user used to create the graphical password gestures. Nonetheless, based on the personalized brute-force attack on the PGA mechanism as a whole (*i. e.*, when also considering the type and the exact order of the gestures), this didn't affect the security of the retrospective approach since the number of guesses between the POI-assisted attack and the personalized attack (experience-spot-driven) was not statistically significantly different. An implication based on this finding would be to enhance content delivery aspects of the retrospective approach by suggesting to an individual a certain set of images that cannot be leveraged by individuals with whom they share common experiences within the specific context of interaction. For example, in case a user creates a password in the context of her organizational environment, the content delivery mechanism would suggest image semantics from her individual experiences and *vice versa*, etc. to avoid the vulnerability of human guessing attacks.

## 8. Limitations

Despite our efforts to keep the validity of the study, some design aspects of the experiments introduce limitations. We used specific background images in order to control the factors of the study (retrospective *vs.* generic images). Although users' choices may be affected by the content and complexity of the image (Wiedenbeck et al., 2005; Dunphy and Yan, 2007), we provided images of the most widely used image categories (depicting scenery (Dunphy and Yan, 2007) and people (Dunphy and Yan, 2007; Alt et al., 2015)) and of similar complexity. Expansion of our research will consider a greater variety of image categories to triangulate findings with diverse user communities and sociocultural experiences, on different levels of abstractions (**Fig. 1**), and thus increase the validity of the study.

Furthermore, the recruitment of student participants introduces limitations and our future efforts entail extending the sample with varying user profiles and ages. Nonetheless, in order to address internal validity, we used specific background pictures in order to better control the factors of the study with the participants (*i.e.*, we used sceneries from the participants' Universities). Another limitation relates to the nature of the eye-tracking study. Considering that we conducted a controlled in-lab eye-tracking study (*User Study A*), the users' selections might have been influenced, however, no such comment was received from our participants during the discussions that followed the task completion.

Moreover, we compared the proposed retrospective approach against one baseline generic approach. Nevertheless, this was intentional since we wouldn't probably get comparable results had we compared the suggested non-intrusive retrospective approach against other intrusive measures (*e.g.*, "presentation-effect" (Thorpe et al., 2014), hiding salient areas (Bulling et al., 2012) etc.). Additionally, we did not compare our approach against user-uploaded images, since it would be rather mystifying if our analyses relied on images of varying complexity and hot-spots, which would probably be the case had we allowed users to upload their own images.

Finally, memorability was measured by analyzing certain metrics such as memory time and password resets. However, given that memorability is a complex factor, future studies are planned to further triangulate the findings by analyzing gaze behavior of users during password login (*e.g.*, the analysis of users' eye gaze data may assist in detecting high cognitive load during password recall (Katsini et al., 2020)).

## 9. Conclusions and future work

In this paper, we suggested a retrospective approach on the basis of a five-tier model of image content delivery bootstrapped on sociocultural experiences and therefore declarative memories of the end-users. We investigated the effects of the retrospective approach on users' visual behavior during password creation, as well as on the security and memorability of user-chosen graphical passwords.

Results of three user studies ($n = 42$; $n = 71$; $n = 26$) revealed significant differences on the users' inherited password creation strategies which was reflected: *a)* by visual behavior differences during password creation between the experimental and control group (the experimental group moved their attention and perception from hot-spots towards experience-spots); *b)* by subsequent password selections differences on hot-spots *vs.* experience-spots of the background-image; and *c)* by password strength differences between the experimental and control group. Simultaneously, there was no main effect of the suggested retrospective approach on memorability and login usability of the user-chosen passwords, although a weekly comparison revealed a tendency of reduced time to login and failed attempts in the second week compared to the first week for the retrospective group. On the downside, the suggested approach introduces password guessing vulnerabilities in terms of allowing attackers who share common experiences with the end-users to more easily identify regions of their selected secrets.

This work points towards a novel direction of considering users' prior sociocultural experiences as a personalization factor for considering *"best-fit"* image content semantics during graphical password creation activities. This can be achieved on the basis of the five-tier image content semantic delivery model, which entails different levels of sociocultural experiences. Hence, this paper contributes on both theory and application for designing and implementing innovative and personalized approaches for graphical passwords by leveraging on the unique sociocultural experiences of users (Constantinides et al., 2019a, 2019b, 2020a; Belk et al., 2019). Taking into consideration that more than 1 billion devices deploy PGA worldwide (Microsoft, 2021), it is imperative to suggest new methods for finding a better usability-security equilibrium in PGA-like schemes.

We envision that such novel approaches would have many positive implications from the users' point of view. Providing image content related to the users' prior sociocultural experience while preserving users' privacy via the suggested 5-tier privacy preserving model (**Fig. 1**), would help users make memorable image selections (on *experience-spots*), and eventually avoid selecting predictable hot-spots. Further studies are required to evaluate the suggested retrospective approach in the wild to gain further insights.

## Credit author statement

**Argyris Constantinides:** Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Writing – Original Draft, Writing – Review & Editing. **Christos Fidas**: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing. **Marios Belk**: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing. **Anna Maria Pietron**: Investigation, Validation. **Ting Han**: Supervision, Project administration. **Andreas Pitsillides:** Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research has been partially supported by the EU Horizon 2020 Grant 826278 “Securing Medical Data in Smart Patient-Centric Healthcare Systems” (Serums).

## References

Ahern, S., Eckles, D., Good, N.S., King, S., Naaman, M., Nair, R., 2007. Over-exposed? Privacy patterns and considerations in online and mobile photo sharing. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 357–366. https://doi.org/10.1145/1240624.1240683.

Alt, F., Schneegass, S., Shirazi, A.S., Hassib, M., Bulling, A., 2015. Graphical passwords in the wild: understanding how users choose pictures and passwords in image-based authentication schemes. In: Proceedings of MobileHCI 2015, pp. 316–322. https://doi.org/10.1145/2785830.2785882.

Aydın, Ü.A., Acartürk, C., Çağıltay, K., 2013. The role of visual coherence in graphical passwords. In: Proceedings of the Annual Meeting of the Cognitive Science Society, 35.

Bates, D., Mächler, M., Bolker, B. and Walker, S., 2014. **Fitting linear mixed-effects models using lme4**.

Belk, M., Fidas, C., Pitsillides, A., 2019. FlexPass: symbiosis of seamless user authentication schemes in IoT. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA ‘19). ACM Press, New York, NY, pp. 1–6. https://doi.org/10.1145/3290607.3312951. Paper LBW2318.

Belk, M., Fidas, C., Germanakos, P., Samaras, G., 2017a. The interplay between humans, technology and user authentication: a cognitive processing perspective. Comput Human Behav 76, 184–200. https://doi.org/10.1016/j.chb.2017.06.042.

Belk, M., Pamboris, A., Fidas, C., Katsini, C., Avouris, N., Samaras, G., 2017b. Sweet-spotting security and usability for intelligent graphical authentication mechanisms. In: Proceedings of the International Conference on Web Intelligence, pp. 252–259. https://doi.org/10.1145/3106426.3106488.

Biddle, R., Chiasson, S., Van Oorschot, P.C., 2012. Graphical passwords: learning from the first twelve years. ACM Computing Surveys (CSUR) 44 (4), 1–41. https://doi.org/10.1145/2333112.2333114.

Bulling, A., Alt, F., Schmidt, A., 2012. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3011–3020. https://doi.org/10.1145/2207676.2208712.

Cardaci, M., Di Gesù, V., Petrou, M. and Tabacchi, M.E., 2009. A fuzzy approach to the evaluation of image complexity. Fuzzy Sets and Systems, 160(10), pp.1474–1484. 10.1016/j.fss.2008.11.017.

Chiasson, S., Forget, A., Biddle, R., Oorschot, P.V., 2008. Influencing users towards better passwords: persuasive cued click-points. People and Computers XXII 22, 121–130. https://doi.org/10.14236/ewic/HCI2008.12.

Chiasson, S., Van Oorschot, P.C., Biddle, R., 2007. Graphical password authentication using cued click points. In: European Symposium on Research in Computer Security. Springer, Berlin, Heidelberg, pp. 359–374. https://doi.org/10.1007/978-3-540-74835-9_24.

Constantinides, A., Belk, M., Fidas, C., Pitsillides, A., 2019a. On the accuracy of eye gaze-driven classifiers for predicting image content familiarity in graphical passwords. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, pp. 201–205. https://doi.org/10.1145/3320435.3320474.

Constantinides, A., Belk, M., Fidas, C., Pitsillides, A., 2020a. An eye gaze-driven metric for estimating the strength of graphical passwords based on image hotspots. In: Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 33–37. https://doi.org/10.1145/3377325.3377537.

Constantinides, A., Belk, M., Fidas, C., Samaras, G., 2018a. On cultural-centered graphical passwords: leveraging on users’ cultural experiences for improving password memorability. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 245–249. https://doi.org/10.1145/3209219.3209254.

Constantinides, A., Fidas, C., Belk, M., Pitsillides, A., 2019b. “I recall this picture”: understanding picture password selections based on users’ sociocultural experiences. In: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 408–412. https://doi.org/10.1145/3350546.3352557.

Constantinides, A., Fidas, C., Belk, M., Samaras, G., 2018b. On sociocultural-centered graphical passwords: an initial framework. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 277–284. https://doi.org/10.1145/3236112.3236150.

Constantinides, A., Pietron, A.M., Belk, M., Fidas, C., Han, T., Pitsillides, A., 2020b. A cross-cultural perspective for personalizing picture passwords. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, pp. 43–52. https://doi.org/10.1145/3340631.3394859.

Core Team, R., 2015. R: A language and Environment For Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Davis, D., Monrose, F., Reiter, M.K., 2004. On user choice in graphical password schemes. USENIX Security Symposium 13 (2004), 11. -11.

Duchowski, A.T., 2007. **Eye tracking methodology**. Theory and practice. https://doi.org/10.1007/978-1-84628-609-4.

Dunphy, P., Yan, J., 2007. Do background images improve" draw a secret" graphical passwords?. In: Proceedings of the 14th ACM conference on Computer and communications security, pp. 36–47. https://doi.org/10.1145/1315245.1315252.

Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., Herley, C., 2013. Does my password go up to eleven? The impact of password meters on password selection. In: Proceedings of CHI 2013, pp. 2379–2388. https://doi.org/10.1145/2470654.2481329.

Erez, M., Gati, E., 2004. A dynamic, multi-level model of culture: from the micro level of the individual to the macro level of a global culture. Applied Psychology 53 (4), 583–598. https://doi.org/10.1111/j.1464-0597.2004.00190.x.

Everitt, K.M., Bragin, T., Fogarty, J., Kohno, T., 2009. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 889–898. https://doi.org/10.1145/1518701.1518837.

Fidas, C., Belk, M., Hadjidemetriou, G., Pitsillides, A., 2019. Influences of mixed reality and human cognition on picture passwords: an eye tracking study. In: IFIP TC13 Human-Computer Interaction (INTERACT 2019). Springer-Verlag, pp. 304–313. https://doi.org/10.1007/978-3-030-29384-0_19.

Furnell, S., 2005. Why users cannot use security. Computers & Security 24 (4), 274–279. https://doi.org/10.1016/j.cose.2005.04.003.

Gonzalez-Garcia, C., Flounders, M.W., Chang, R., Baria, A.T., He, B.J., 2018. Content-specific activity in frontoparietal and default-mode networks during prior-guided visual perception. Elife 7, 36068. https://doi.org/10.7554/eLife.36068.001.

GP3 Eye Tracker. 2020. [Online] Available at: https://www.gazept.com/product/gazepoint-gp3-eye-tracker/.

Hayashi, E., Dhamija, R., Christin, N., Perrig, A., 2008. Use your illusion: secure authentication usable anywhere. In: Proceedings of the 4th symposium on Usable privacy and security, pp. 35–45. https://doi.org/10.1145/1408664.1408670.

Hayashi, E., Hong, J., Christin, N., 2011. Security through a different kind of obscurity: evaluating distortion in graphical authentication schemes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2055–2064. https://doi.org/10.1145/1978942.1979242.

Johnson, J., Seixeiro, S., Pace, Z., van der Bogert, G., Gilmour, S., Siebens, L. and Tubbs, K., Microsoft Corp, 2014. Picture gesture authentication. U.S. Patent 8,650,636. Retrieved from https://google.com/patents/US8910253.

Katsini, C., Abdrabou, Y., Raptis, G., Khamis, M., Alt, F., 2020. The role of eye gaze in security and privacy applications: survey and future HCI research directions. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20). ACM, pp. 1–21. https://doi.org/10.1145/3313831.3376840.

Katsini, C., Fidas, C., Raptis, G.E., Belk, M., Samaras, G., Avouris, N., 2018. Influences of human cognition and visual behavior on password strength during picture password composition. In: Proceedings of CHI 2018, pp. 1–14. https://doi.org/10.1145/3173574.3173661.

Microsoft, 2021. Microsoft™ by the numbers. [Online] Available at: https://news.microsoft.com/bythenumbers/en/windowsdevices.

Mihajlov, M., Jerman-Blažič, B., 2011. On designing usable and secure recognition-based graphical authentication mechanisms. Interact Comput 23 (6), 582–593. https://doi.org/10.1016/j.intcom.2011.09.001.

Mihajlov, M., Jerman-Blažič, B., Ciunova Shuleska, A., 2016. Why that picture? discovering password properties in recognition-based graphical authentication. International Journal of Human–Computer Interaction, 32 (12), 975–988. https://doi.org/10.1080/10447318.2016.1220103.

Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J., Beznosov, K., 2013. Know your enemy: the risk of unauthorized access in smartphones by insiders. In: Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (MobileHCI ’13). New York,NY, USA. Association for Computing Machinery, pp. 271–280. https://doi.org/10.1145/2493190.2493223.

Paivio, A., Csapo, K., 1973. Picture superiority in free recall: imagery or dual coding? Cogn Psychol 5 (2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7.

Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A., 2012. Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp. 733–740. https://doi.org/10.1109/CVPR.2012.6247743.

Pinheiro, J., Bates, D., 2006. **Mixed-effects Models in S and S-PLUS**. Springer Science & Business Media.

Raptis, G., Fidas, C., Avouris, N., 2016. Using eye tracking to identify cognitive differences: a brief literature review. In: Proceedings of the 20th Pan-Hellenic Conference on Informatics, pp. 1–6. https://doi.org/10.1145/3003733.3003762.

Raptis, G., Katsini, C., Belk, M., Fidas, C., Samaras, G., Avouris, N., 2017. Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies. In: ACM User Modeling, Adaptation and Personalization (UMAP 2017). ACM Press, pp. 164–173. https://doi.org/10.1145/3079628.3079690.

Renaud, K., 2009. On user involvement in production of images used in visual authentication. Journal of Visual Languages & Computing 20 (1), 1–15. https://doi.org/10.1016/j.jvlc.2008.04.001.

Riddle, B.L., Miron, M.S., Semo, J.A., 1989. Passwords in use in a university timesharing environment. Computers & Security 8 (7), 569–579. https://doi.org/10.1016/0167-4048(89)90049-7.

Rogers, T.B., Kuiper, N.A., Kirker, W.S., 1977. Self-reference and the encoding of personal information. J Pers Soc Psychol 35 (9), 677. https://doi.org/10.1037/0022-3514.35.9.677.

Sadovnik, A., Chen, T., 2013. A visual dictionary attack on Picture Passwords. In: 2013 IEEE International Conference on Image Processing. IEEE, pp. 4447–4451. https://doi.org/10.1109/ICIP.2013.6738916.

Schaub, F., Walch, M., Könings, B., Weber, M., 2013. Exploring the design space of graphical passwords on smartphones. In: Proceedings of the Ninth Symposium on Usable Privacy and security, pp. 1–14. https://doi.org/10.1145/2501604.2501615.

Stobert, E., Biddle, R., 2013. Memory retrieval and graphical passwords. In: Proceedings of the ninth symposium on usable privacy and security, pp. 1–14. https://doi.org/10.1145/2501604.2501619.

Thorpe, J., van Oorschot, P.C., 2007. Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords. USENIX Security Symposium 8, 1–8.

Thorpe, J., Al-Badawi, M., MacRae, B., Salehi-Abari, A., 2014. The presentation effect on graphical passwords. In: proceedings of the SIGCHI conference on human factors in computing systems, pp. 2947–2950. https://doi.org/10.1145/2556288.2557212.

Tullis, T.S., Tedesco, D.P., 2005. Using personal photos as pictorial passwords. In: CHI'05 extended abstracts on Human factors in computing systems, pp. 1841–1844. https://doi.org/10.1145/1056808.1057036.

Tullis, T.S., Tedesco, D.P., McCaffrey, K.E., 2011. Can users remember their pictorial passwords six years later. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 1789–1794. https://doi.org/10.1145/1979742.1979945.

Tulving, E., 1972. Episodic and semantic memory. Organization of memory 1, 381–403.

Van Oorschot, P.C., Thorpe, J., 2011. Exploiting predictability in click-based graphical passwords. Journal of Computer Security 19 (4), 669–702. https://doi.org/10.3233/JCS-2010-0411.

Van Oorschot, P.C., Salehi-Abari, A., Thorpe, J., 2010. Purely automated attacks on passpoints-style graphical passwords. IEEE Transactions on Information Forensics and Security 5 (3), 393–405. https://doi.org/10.1109/TIFS.2010.2053706.

Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A., Memon, N., 2005. Authentication using graphical passwords: effects of tolerance and image choice. In: Proceedings of the 2005 symposium on Usable privacy and security, pp. 1–12. https://doi.org/10.1145/1073001.1073002.

Winter, B., Grawunder, S., 2012. The phonetic profile of Korean formal and informal speech registers. J Phon 40 (6), 808–815.

Zhao, Z., Ahn, G.J., Hu, H., 2015. Picture gesture authentication: empirical analysis, automated attacks, and scheme evaluation. ACM Transactions on Information and System Security (TISSEC) 17 (4), 1–37. https://doi.org/10.1145/2701423.

Zhao, Z., Ahn, G.J., Seo, J.J., Hu, H., 2013. On the security of picture gesture authentication. In: Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13), pp. 383–398.