Using Personal Data to Support Authentication: User Attitudes and Suitability

JOLIE BONNER, University of Glasgow, United Kingdom
JOSEPH O'HAGAN, University of Glasgow, United Kingdom
FLORIAN MATHIS, University of Glasgow, United Kingdom and University of Edinburgh, United Kingdom
JAMIE FERGUSON, University of Glasgow, United Kingdom
MOHAMED KHAMIS, University of Glasgow, United Kingdom

Dynamic personal data based on a user's activity, such as recent visited physical locations, browsing history, and call logs, update frequently, making it a promising token for user authentication. However, it is not clear how users perceive this use of personal data and which data types are most suitable for authentication. To investigate this, we conducted an online survey with N=100 participants. For 10 personal data types we asked participants about their comfort with this data for authentication, its perceived security, its impact on behaviour, who has access to it, how frequently it updates, and how memorable they perceive it to be. We found that participants were generally uncomfortable with personal data being used for authentication and, knowing their personal data is used, they may intentionally change their behaviour due to privacy concerns. We discuss the benefits and drawbacks of using personal data as a source of dynamic tokens to complement authentication and conclude with three learned lessons.

CCS Concepts: • **Security and privacy → Authentication**; **Usability in security and privacy**.

Additional Key Words and Phrases: Personal Data Passwords, Usable Security, Dynamic Passwords

ACM Reference Format:
Jolie Bonner, Joseph O'Hagan, Florian Mathis, Jamie Ferguson, and Mohamed Khamis. 2021. Using Personal Data to Support Authentication: User Attitudes and Suitability . In *20th International Conference on Mobile and Ubiquitous Multimedia (MUM 2021), December 5-8, 2021, Leuven, Belgium.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3490632.3490644

1 INTRODUCTION & BACKGROUND

Secrets, such as passwords or responses to fallback questions (e.g., "What is your mother's maiden name?"), are frequently used for user authentication. Digital services such as financial institutions and social media platforms rely on such secrets to allow users access to their personal data and restrict access from non-privileged users. Although users' perception of secure passwords appears to be inline with what password-cracking approaches show [32], users tend to ignore security advice [26, 27], which often leads to insecure choices and puts users' accounts at risk. Users also often reuse their secrets across a variety of accounts [14, 19, 30] and make use of secrets that are either predictable (e.g., "iloathecats123") [33] or easy to discover publicly available information using social networks (e.g., a user's place or date of birth). There are many patterns that are frequently used when coming up with secrets (e.g., keyboard patterns, number sequences, password thematics) [27]. While such alphanumeric, and often static, secrets are still the predominant way to authenticate online, there are alternative authentication systems such as graphical password schemes (e.g., [9, 10, 16, 28, 34]) or autobiographical authentication schemes (e.g., [15]) that achieved promising usability and security. For example, Das et al. [15] showed that a challenge-response authentication system that queries users about their day-to-day experiences can be suitable in risky situations and that their simulation resulted in high user confidence estimates.

Manuscript submitted to ACM

Dynamic secrets, secrets that change based on a user's activities, can improve usability and security by asking "What's the last meal you had?" rather than "What's your favourite meal?". While the latter is easy to guess (e.g., by friends, family members), the former is challenging due to its changing nature. Previous work has explored the use of dynamic secrets. For example, Hang et al. [20] showed that when leveraging a phone's already stored data, data types, such as app installations and communication (e.g., "Who did you call last week?"), are the most promising in achieving a high recall accuracy (up to 100%). In another work by Hang et al. [21], location-based security questions achieved high recall accuracy for fallback authentication while maintaining high security against adversaries. In a similar vein as concluded by Hang et al. [20, 21] but in the domain of graphical authentication research, Chiasson et al. [9] argued that tomorrow's ideal systems leverage preexisting user-specific knowledge, rather than having users memorise entirely new and/or random information. Other work by Constantinides et al. [11] showed that the concept of *retrospective-based authentication* can lead to significantly stronger passwords created on images that reflect a user's prior experiences rather than on images that are unfamiliar to the user. In summary, previous research discussed some promising use cases of event-specific knowledge (ESK) [12] in the usable security research domain, particularly in user authentication [15, 20, 21]. Such dynamic secrets are easy to recall as they rely on users' recent activities [20]. They are also hard to guess, discover, and deduce by others due to the secret's changing nature [15].

To ensure that such systems eventually transition into practice and are adopted by the public (referred to as problem-scoping and problem-solving research [25]), an important step is to investigate a) if users are comfortable to use some of their personal data for security, which will help us understand the privacy-security trade off, b) what personal data types are most suitable (including, but not limited, to data types that are already collected by smartphones like text messages, and app installations [20]), and c) how the use of personal data for dynamic passwords impacts (if at all) users' real-world behaviour. For example, would a user change their behaviour if they know an online authentication service has access to their web browsing history? Previous work (e.g., [20]) argued that the most usable data types (i.e., the categories that lead to a high recall accuracy) are not necessarily the best ones for authentication and that the corresponding privacy implications have been rarely discussed. In this work, we aim to fill this gap and research to what extent users are comfortable utilising some of their personal data as (part of) dynamic authentication tokens. While previous work evaluated the concept of personal data along with a prototype system to shed light on users' correct answers to recall questions [15, 20], we evaluate the idea of using personal data to support user authentication on a more theoretical level and aim to understand users' attitudes and corresponding privacy concerns prior to an actual prototype system evaluation.

We surveyed N=100 users and found that users are more comfortable using their online shopping orders, use of a streaming service, use of a food delivery service, and app downloads to improve the usability and security of their secrets compared to photos and calls/texts, but overall comfort was low for all investigated data types. This means that users do not necessarily feel comfortable in using their activity data for authentication. We also noticed that there are differences in users' update frequency of the specific data types, indicating that some data types are more suitable for frequent authentications (e.g. browsing history, calls/texts), while others (e.g., app downloads, use of a food delivery service) are more likely to be suitable for infrequent authentications. We conclude with an in-depth discussion of our findings and present three lessons learned about the use of personal data for authentication.

Contribution Statement: We contribute with **(1)** empirical data on users' preferred personal data types for personal data secrets, their comfort to let services access personal data in exchange for security (privacy-security trade off), and their opinions about dynamic personal data secrets. **(2)** We provide and discuss three learned lessons about key areas of concern when using personal data for authentication, and how to mitigate said concerns.

2 METHODOLOGY

We designed a survey to explore attitudes towards 10 personal data types which could be used for dynamic authentication. We created an initial list of different data types based on the previous works by Das et al. [15] and Hang et al. [20, 21]), and then added data types from brainstorming session with three researchers. This approach resulted in overall 10 data types which we used for our survey: **apps** downloaded, activity on a **streaming** platform (e.g. Netflix [3]), use of a **delivery** service (e.g. Uber Eats [7]), online shopping **orders**, meta data of **files** (e.g. file names and types of MS Office files [1]), posts interacted with on **social media**, recent physical **locations**, **browsing history**, **calls/texts/DMs/emails**, and taken **photos**.

We then distributed an online survey as described in section 2.1 through word of mouth, internal mailing lists, and social media platforms to investigate how users perceive the use of these personal data types to support authentication.

2.1 Survey Structure

At the beginning of the survey, participants had to confirm that they are aged 16 or above, and they had to provide consent to take part in our study. Participants could withdraw from the study at any point. The survey also provided them with the contact details of the researchers for any questions. We then provided participants with an explanation of personal data authentication and how dynamic personal data (e.g., watched movies on a streaming platform) could be used to support user authentications. We then captured participants' demographics: their age, gender , country of origin general log in frequency. In the first half of our survey, we asked for users' perceived privacy and security of the selected data types when used for authentication. For each data type, we asked on a 5-point Likert scale how comfortable they would be using this data for authentication (1=very uncomfortable, 5=very comfortable) and to optionally justify their answer using an open text box. Participants were also asked how secure they feel this data would be for authentication and if knowing that this data is being used for authentication would they change their behaviour. For example, if their browsing history data is being used as a dynamic token for authentication, would they change their online browsing behaviour? In the second half of our survey, we investigated the efficacy of the different data types. For each data type, we asked participants who would know this information, how often this data is updated, and how easy it is to recall information about each data type.

We closed by asking if they had come across any of our listed data types being used by other applications and if they had any other comments. An overview of our survey's structure can be found in Appendix A.

2.2 Limitations

Through our online survey, we captured users' perceptions towards personal data and how it could be used to enhance user authentication. Although participants' self-reported opinions are valuable and provide first insights into the suitability of personal data for user authentication, we call for future work that evaluates the concept along with a prototype system (e.g., similar to [15]) to shed further light on participants' perception and potential threats to their privacy. We captured the responses of N=100 survey takers, but a comparison between different age groups was out of the scope of our investigation. Our initial investigation of using personal data to support authentication did not aim to compare, e.g., a younger sample with a more senior one, but rather to provide insights into users' attitude towards using personal data for authentication in general. That being said, if authentication systems aim to target specific sub populations, such as graphical password authentication for children [8], we call for future work that considers different age groups and also runs comparisons between those.

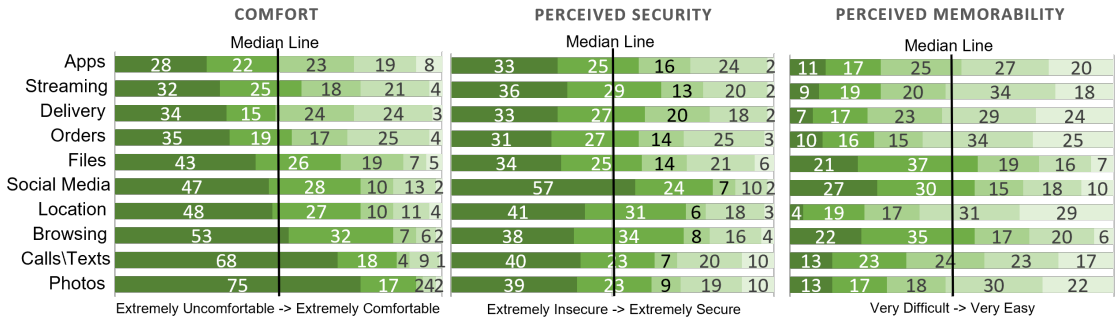

Fig. 1. Participants where asked for each data type to rate on a 5 point Likert scale their comfort(1=extremely uncomfortable, 5=extremely comfortable), perceived security(1=extremely insecure, 5=extremely secure), ease of recall (1=very difficult, 5=very easy).

3 RESULTS

We distributed the survey to N=120 participants. We added an attention check question which 20 participants failed, leaving 100 valid responses (61 female, 36 male, 1 agender, 1 genderqueer, 1 prefer not to say). Participants were aged between 17 and 66 (M=31.54, SD=13.68) and originated from 28 different countries most commonly originating from the UK (49%) followed by the US (10%), Germany (4%), and South Africa (3%). Six participants selected prefer not to say. Participants were asked how often they log into an online account using a PIN or a password, 68% very often, 23% often, 4% sometimes, 3% rarely, 2% never.

Qualitative answers were coded using initial coding [13]. Participants' statements were assigned codes over repeated cycles with the codes grouped using a thematic approach. The lead author performed the coding and reviewed the coding with two additional researchers to resolve unclear codes and discuss the depth and specificity of codes. Overall we completed two coding cycles. We treat responses on Likert scale questions as non-parametric data; therefore, quantitative analysis (sections 3.1 and 3.2) was conducted using Friedman tests with *post-hoc* pairwise comparisons using Nemenyi tests (which control for familywise errors). Cochran's Q tests with Bonferroni corrected follow-up Wilcoxon signed rank tests for *post-hoc* pairwise comparisons were used in sections 3.4 and 3.5 to determine differences on dichotomous dependent variables between three or more related groups. For all tests, a significance level of $p<0.05$ was used.

3.1 Perceived Comfort Using Personal Data for Authentication

Most participants indicated they were either extremely or somewhat uncomfortable with their personal data being used for authentication. Participants were most uncomfortable using their photos (M=1.41, SD=0.87) whilst they were most comfortable using their app downloads (M=2.57, SD=1.29). Participants' comfort using personal data for authentication remains relatively low for all data types with significant differences between the data types $\chi^2(9) = 207.81$, $p<0.001$ (see Figure 1 and Table 1).

We also asked participants optionally to justify their comfort score. We received 57 comments, 10 of which were not meaningful (e.g. P1: *"N/A"*). There were some participants (n=19) who voiced they are uncomfortable using their personal data for authentication due to privacy concerns (e.g., P38: *"too invasive"*). A few others (n=9) mentioned that their data being stored by a third party or used across platforms makes them uncomfortable. For example, P93 voiced that if *"Uber eats asking me about my past order I'm somewhat comfortable, if Instagram asks me where I was at 2pm*

	Apps	Streaming	Delivery	Orders	Files	Social Media	Locations	Browsing	Calls/Texts
Streaming		-	-	-	-	-	-	-	-
Delivery			-	-	-	-	-	-	-
Orders				-	-	-	-	-	-
Files	✓M	✓MB	✓M	✓M	-	-	-	-	-
Social Media	✓CPMB	✓MB	✓CMB	✓PMB	✓P	-	-	-	-
Locations	✓C		✓C			✓MB	✓M -	-	-
Browsing	✓CMB	✓CMB	✓CMB	✓CMB	✓B		✓MB	-	-
Calls/Texts	✓CB	✓CB	✓CB	✓CB	✓CMB	✓PM	✓MB		-
Photos	✓CB	✓CB	✓CB	✓CB	✓CMB	✓CPM	✓CB	✓M	

Table 1. *Post-hoc* significant differences for perceived comfort, security, memorability and likelihood of behaviour change. Differences for: comfort are indicated by ✓C, security by ✓P, memorability by ✓M and likelihood of behaviour change by ✓B.

I'm very uncomfortable". There were only n=6 participants who mentioned that they would be unable to accurately remember most of the proposed data types, and n=5 reported they do not want to be reminded of some of their past activities. A few others (n=6) brought up the privacy concern during authentication. P82, for example, voiced that they are concerned that nearby people could see their private data during authentication: *"depends on who is present in the room when I log in"*. Two participants mentioned that shared accounts (e.g., sharing a movie streaming account such as Netflix) could be problematic as some of the personal data could originate from the other users' history.

3.2 Perceived Security of Personal Data for Authentication

Participants perceived personal data for authentication as insecure, with low scores for all data types. Figure 1 provides an overview of participants' responses on the 5-point Likert scale. Participants perceived social media data as the least secure data type for authentication (M=1.67, SD=1.08), while online orders were perceived as the most secure data type (M=2.42, SD=1.24). Significant differences were found between the data types $\chi^2(9) = 48.81$, $p<0.001$. Post-hoc comparisons are summarised in Table 1.

3.3 Likelihood and Reasoning for Change When Behaviour Is Used for Authentication

Participants reported to alter their behaviour when their personal data is being used for authentication, albeit with some variations (see Figure 2). Using users' taken photos and their browsing history for authentication were the most likely to elicit a behavioural change. In both cases, n=60 participants indicated they would change their behaviour if they knew that the data is being used for authentication. Using participants' activity on streaming platforms (e.g, Netflix) was the least likely (n=20) data type that would induce behavioural change. Significant differences were found between the data types $\chi^2(9) = 170.18$, $p<0.001$. Table 1 provides an overview of all pairwise comparisons.

About half of our participants (n=52) left a comment to justify their change in behaviour, 12 of which were unavailing (e.g. P18: *"don't know"*). Most participants (n=13) said they would change their behaviour due to privacy concerns with the system using their data or because of nearby people potentially seeing sensitive, personal data (e.g. P32: *"Most of these I'd be worried that it'd come on my screen when I was in front of other people"*. Others (n=11) indicated they would change their behaviour due to the feeling of being *"observed"* or *"judged"* by the system while n=5 said that being confronted by their past actions makes them uncomfortable, P15: *"Even though I know this data is already being stored, being even more aware of it and seeing it presented makes one more self conscious"*. A few participants (n=8) mentioned they would want control over which data was tracked and said they would opt-out of some they considered embarrassing, P93: *"even if a human never interacted with the data, it might be embarrassing e.g. condoms, tampons"*. Only n=2 participants voiced they would change their behaviour to make their data more memorable and one participant said they would change their behaviour to increase the security.

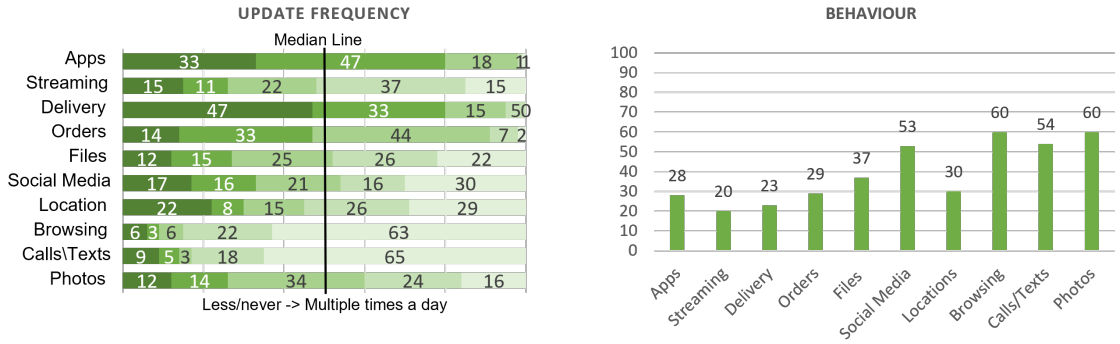

Fig. 2. Participants selected from a scale how often they think the corresponding data type updates (1=multiple times a day, 2=daily, 3=weekly, 4=monthly 5=less/never). We also asked participants if they would change their behaviour when they know this data is being used for user authentication. Plotted is the number of *"yes"* responses.

	Strangers	Acquaintances	Colleagues	Family	Close Friends	Partners	Living with	Nobody
App downloads	9.48%	0.86%	3.45%	7.76%	6.90%	10.34%	2.59%	58.62%
Browsing History	10.19%	0.93%	0.93%	3.70%	0.93%	10.19%	10.19%	62.96%
Calls/texts/DMs/emails	7.86%	3.57%	4.29%	10.71%	7.86%	8.57%	9.29%	47.86%
Meta data of files	6.31%	0.90%	9.91%	5.41%	1.80%	8.11%	6.31%	61.26%
Online Orders	8.57%	0.71%	0.71%	17.14%	7.86%	15.71%	17.86%	31.43%
Photos you have taken	6.01%	4.37%	4.92%	16.39%	15.30%	13.11%	10.38%	29.51%
Social media	12.64%	12.27%	8.92%	13.75%	18.96%	12.27%	9.67%	11.52%
Recent Physical Locations	5.94%	3.20%	3.20%	18.72%	18.26%	15.07%	20.09%	15.53%
Use of a delivery app	9.56%	0.74%	2.21%	12.50%	6.62%	12.50%	16.91%	38.97%
Use of a streaming service	6.29%	0.63%	0.63%	31.45%	10.06%	18.24%	19.50%	13.21%

Table 2. People in the know of users' personal data. The table shows percentages of the total number of selections for each data type.

3.4 Who Else Knows the Users' Personal Data

We also asked participants who else they think has access to each of the personal data types (see Table 3). Close friends, partners, family, and people someone lives with are more likely to have access to another ones personal data. For acquaintances, colleagues, and strangers having access to someones personal data is rather unlikely. Surprisingly, strangers scored as more knowledgeable than acquaintances and colleagues across all data types. The data types that scored most private, i.e., the ones 'nobody' is likely to know, are app downloads, browsing history, meta data of files and calls/texts. The results are summarised in Table 2. Statistical analysis for these results is summarised in Table 3.

3.5 Update Frequency and Perceived Memorability of the Data Types

Participants were asked how often the data they use updates and how easy it would be to recall the information required for authentication (e.g., last watched Netflix movie). Participants reported that (1) the use of a delivery service and app downloads update the least (i.e., on a monthly basis) and that (2) calls/texts and browsing history are data types that update most regularly (i.e., multiple times a day). When asked how easy it would be to recall specific information of a data type (e.g., last visited city), participants reported that location (M=3.62, SD=1.20) would be the easiest to recall for authentication. Online orders (M=2.51, SD=1.19), social media (M=2.54, SD 1.32), and browsing history (M=2.53, SD=1.20) scored the lowest (see Figure 1). Significant differences were found between the data types $\chi^2(9) = 170.17$, $p<0.001$. Post-hoc comparisons are summarised in Table 1.

	Strangers	Acquaintances	Colleagues	Family	Close Friends	Partners
Acquaintances	Delivery	-	-	-	-	-
Colleagues			-	-	-	-
Family	Photos, Location, Streaming	Orders, Photos, Location, Delivery, Streaming	Orders, Photos, Social, Location, Delivery, Streaming	-	-	-
Close Friends	Photos, Location	Orders, Photos, Social, Location, Streaming	Photos, Social, Location, Streaming	Social, Streaming	-	-
Partners	Location	Orders, Photos, Location, Delivery, Streaming	Orders, Photos, Location, Delivery, Streaming		Social	-
Living With	Location, Streaming	Order, Photos, Location, Delivery, Streaming	Orders, Location, Delivery, Streaming		Browsing, Social, Delivery	

Table 3. *Post-hoc* significant differences between types of person for each data type. For example, there is a significant difference between Partners and Strangers in knowing each others' locations.

3.6 Prior Encounters With Personal Data Authentication & General Comments

Only 13 participants reported to have seen the use of personal data for authentication previously. Most of them (n=8) reported to have seen the use of personal data on banking sites, while n=5 reported to have seen it in apps. There were four participants who voiced some concerns regarding using personal data for authentication, particularly due to shared accounts (as already mentioned in Section 3.1) and accessibility reasons. All voiced that shared accounts could make authentication challenging, P90: *"many people share streaming services e.g. Netflix, Prime. If this was used, it would get confusing as I would not be able to verify myself"*. One participant emphasised the corresponding accessibility challenges when personal data is used as a source of dynamic authentication tokens, P87: *"I have a poor memory due to a bunch of chronic illnesses, any type of password that relied on my short term memory is pretty inaccessible to me."*.

4 DISCUSSION

As concluded by Hang et al. [20], the use of personal data secrets requires researchers to consider usability, security, and privacy aspects and, likely presents users with a trade off between those. We encourage future work to look in more detail into the privacy and security trade off when leveraging personal data to support authentication. We envision that the use of personal data can contribute towards more usable and secure user authentication, but finding the sweet spot of usability, security, and privacy is challenging and likely requires evaluating a prototype system instead of collecting self-reported data by participants. Although there are some promising contributions in this field (e.g., the dynamic fallback authentication by Hang et al. [20] or Das et al.'s work about autobiographical authentication [15]), such systems only find usage if, and only if, the benefits (e.g., improved usability and security) are clear to users and outweigh their privacy concerns. While it is important to research how well these approaches perform (e.g., are users able to recall where they have been two days ago?) it seems equally important to research if users are comfortable using different data types for user authentication. We close by discussing the efficacy and privacy of different data types and personal data passwords as a whole, highlighting suggestions future work might explore and presenting three lessons learned.

4.1 Participants Discomfort With Personal Data Authentication

Participants discomfort towards personal data authentication were due to a variety of factors. In particular, participants were uncomfortable using their personal data for authentication due to the sensitivity of some data types. Specifically, that nearby people may see their personal data if displayed on the lock screen (i.e., shoulder surfing [17]), or that

companies would have access to new data, not relevant to its service. Others suggested discomfort due to reminders of past actions they do not want to re-evaluate. In cases where authentication schemes visually represent personal data (e.g., for graphical authentication schemes [22]), we recommend to use only partial elements of the collected data for user authentication or, as discussed by Hayashi et al. [22], make use of distorted images which are still usable while protecting users' privacy. Investigating this in more detail is one direction future work might explore.

4.2 Personal Data for User Authentication and Accessibility

There were notable differences between the update frequencies of the different data types (see Figure 2). The more frequent the update the more suitable that data type is for frequent authentication. A stagnant action looses the potential and advantages of dynamic authentication. Authentication schemes vary in what type of memory they depend on, therefore, it is important to recognise the challenges faced by some individuals. Schemes based on dynamic data, such as our proposed data types, require short term and autobiographical memory [29, 31] which could be inaccessible to some individuals. So while authentications schemes that rely on users' short term or autobiographical memory might be advantageous for some, these systems can be challenging and inaccessible for others.

4.3 Account Sharing: An Opportunity but also a Challenge

Some participants commented that sharing accounts with others (e.g. people within a household sharing a Netflix account) could be problematic for personal data authentication. Although this has only been voiced by a small subset of our participants (n=6), account sharing forms an important challenge in the scope of our investigation. Some services (e.g. Netflix [2]) include functionality to allow multiple users to share a single account by creating a personalised sub-account for each individual. These systems could easily integrate personal data passwords while maintaining sub-accounts by allowing a person logging in to select which sub-user they are. Other services would need to make significant adjustments to their account handling as they do not provide users with personalised sub-accounts. For these cases alternative authentication systems, such as a fallback static PIN or password, may be necessary. While sharing an account or device within the same household is to be expected, users do actively engage in account sharing beyond what is considered appropriate by service providers (e.g., multiple households sharing a single Netflix account [4, 5]). In such situations, personal data passwords can lead to significant authentication issues and significantly impact users' user experience when using said online platforms. Getting such a trade-off between usability, security, and privacy right and conducting a long-term study in the field is likely to be infeasible in an academic setting only.

We call for future work that incorporates both academic research labs and industry stakeholders (e.g., from relevant online services such as Netflix [3] and Spotify [6]) to shed further light on the potential and drawbacks of personal data for user authentication.

4.4 Lessons Learned

We learned that there are three key areas of concern when personal data is used for user authentication.

4.4.1 Lesson #1: Sharing with Third Parties Causes Concern. Participants were concerned about their data being shared across platforms. It is preferable to keep personal data within the authenticating application. For example, online platforms such as Spotify could leverage a user account's past listening history for personal data authentication, but such a platform should avoid using other personal data such as recent physical locations or recent online orders to support user authentication as this likely raises privacy concerns, therefore, users may opt out of existing subscriptions and move to alternative platforms.

4.4.2 Lesson #2: The Data Subject Matters More Than the Data Type. Some personal data is too sensitive regardless of opinions on the data type as a whole. Removing certain pieces of data, before or after creation, from being used for authentication could make users more comfortable. To aid the adoption of such passwords we suggest automatically removing typically sensitive data. For example, supermarkets could remove medical items from the items that are used for personal data authentication. We also recommend to allow for data collection to be paused, and to provide users with options to manually remove or add data points where necessary.

4.4.3 Lesson #3: Avoid Displaying Personal Data When Authenticating. Authentication systems that display personal data are problematic as they could reveal private data to surrounding people (i.e. shoulder surfing [17]). For example, exposing a user´s calls/text history or a user´s recently visited physical locations to those around them when authenticating can be problematic. To preserve a user's privacy, personal data could be obfuscated when used for authentication (e.g., [18, 20, 22]). Hayashi et al. [22] demonstrated that, regardless of their age or gender, users are able to recognise degraded self-selected images. While this is promising in the scope of self-selected images, personal data as studied in our work might not always be presented to users in a highly visual form. At the point where researchers provide users with recognition-based authentication systems (e.g., [23]) that leverage personal data in a degraded version (e.g., snippets of text messages), it seems to be an important first step to investigate to what extent users are able to recognise different data types in their degraded versions and which graphical filters are to be preferred (e.g., similar to [24]).

5 CONCLUSION

We conducted an online survey with N=100 participants to investigate user attitudes on dynamic personal data authentication and the extent to which different data types are suitable for usable and secure authentication. Our results indicated that participants were generally uncomfortable with personal data being used for authentication and that shared accounts can have a notable impact on using personal data to support authentication. We concluded our work with 3 learned lessons to support future work in (personal data) authentication research.

ACKNOWLEDGEMENTS

We thank all participants for taking part in our study. We also thank the reviewers for their valuable feedback. This research was supported by the School of Computing Science at the University of Glasgow. This work was also partially supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships, the EPSRC (EP/V008870/1) and the Royal Society of Edinburgh (award number #65040).

REFERENCES

[1] 2021. *MS Office - If you're looking for Office, you're in the right place.* Retrieved August 12, 2021 from https://www.microsoft.com/en-ww/microsoft-365/microsoft-office

[2] 2021. *Netflix - How to create and edit profiles.* Retrieved August 12 2021 from https://help.netflix.com/en/node/10421

[3] 2021. *Netflix - Watch TV Shows Online, Watch Movies Online.* Retrieved August 12, 2021 from https://www.netflix.com/browse

[4] 2021. *Netflix Tests Cracking Down on Password Sharing.* Retrieved October 04 2021 from https://www.hollywoodreporter.com/tv/tv-news/netflix-password-sharing-4147786/

[5] 2021. *Netflix weighs up crackdown on password sharing.* Retrieved October 04 2021 from https://www.theguardian.com/media/2021/mar/12/netflix-weighs-up-crackdown-on-password-sharing

[6] 2021. *Spotify - Premium Family.* Retrieved October 04 2021 from https://www.spotify.com/uk/family/

[7] 2021. *Uber Eats - What is Uber Eats?* Retrieved August 12 2021 from https://help.uber.com/ubereats/article/what-is-uber-eats-?nodeId=fbf73e2a-c21f-4a48-8333-c874ae195fd1

[8] Hala Assal, Ahsan Imran, and Sonia Chiasson. 2018. An exploration of graphical password authentication for children. *International Journal of Child-Computer Interaction* 18 (2018), 37–46.

[9] Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot. 2012. Graphical Passwords: Learning from the First Twelve Years. *ACM Comput. Surv.* 44, 4, Article 19 (Sept. 2012), 41 pages. https://doi.org/10.1145/2333112.2333114

[10] Sonia Chiasson, P. C. van Oorschot, and Robert Biddle. 2007. Graphical Password Authentication Using Cued Click Points. In *Computer Security – ESORICS 2007*, Joachim Biskup and Javier López (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 359–374.

[11] Argyris Constantinides, Christos Fidas, Marios Belk, Anna Maria Pietron, Ting Han, and Andreas Pitsillides. 2021. From hot-spots towards experience-spots: Leveraging on users' sociocultural experiences to enhance security in cued-recall graphical authentication. *International Journal of Human-Computer Studies* 149 (2021), 102602. https://doi.org/10.1016/j.ijhcs.2021.102602

[12] Martin A Conway and Christopher W Pleydell-Pearce. 2000. The construction of autobiographical memories in the self-memory system. *Psychological review* 107, 2 (2000), 261.

[13] Strauss A. L. Corbin J. M. 1998. *Basics of qualitative research: techniques and procedures for developing grounded theory.* SAGE Publications, Inc.

[14] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and Xiao Feng Wang. 2014. The Tangled Web of Password Reuse.. In *Proceedings of Network and Distributed System Security Symposium (NDSS)*. Internet Society, Reston, VA, USA, 23–26.

[15] Sauvik Das, Eiji Hayashi, and Jason I Hong. 2013. Exploring capturable everyday memory for autobiographical authentication. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 211–220.

[16] Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies* 63, 1 (2005), 128–152. https://doi.org/10.1016/j.ijhcs.2005.04.020

[17] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. *Understanding Shoulder Surfing in the Wild: Stories from Users and Observers.* Association for Computing Machinery, New York, NY, USA, 4254–4265. https://doi.org/10.1145/3025453.3025636

[18] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Andreas Bulling, and Albrecht Schmidt. 2019. Can Privacy-Aware Lifelogs Alter Our Memories?. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313052

[19] Dinei Florencio and Cormac Herley. 2007. A Large-Scale Study of Web Password Habits. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada) *(WWW '07)*. ACM, New York, NY, USA. https://doi.org/10.1145/1242572.1242661

[20] Alina Hang, Alexander De Luca, and Heinrich Hussmann. 2015. *I Know What You Did Last Week! Do You? Dynamic Security Questions for Fallback Authentication on Smartphones.* Association for Computing Machinery, New York, NY, USA, 1383–1392. https://doi.org/10.1145/2702123.2702131

[21] Alina Hang, Alexander De Luca, Matthew Smith, Michael Richter, and Heinrich Hussmann. 2015. Where have you been? using location-based security questions for fallback authentication. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*. 169–183.

[22] Eiji Hayashi, Rachna Dhamija, Nicolas Christin, and Adrian Perrig. 2008. Use Your Illusion: Secure Authentication Usable Anywhere. In *Proceedings of the 4th Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, USA) *(SOUPS '08)*. Association for Computing Machinery, New York, NY, USA, 35–45. https://doi.org/10.1145/1408664.1408670

[23] Korey Johnson and Steffen Werner. 2008. Graphical user authentication: A comparative evaluation of composite scene authentication vs. three competing graphical passcode systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 52. SAGE Publications Sage CA: Los Angeles, CA, 542–546.

[24] Mohamed Khamis, Tobias Seitz, Leonhard Mertl, Alice Nguyen, Mario Schneller, and Zhe Li. 2019. Passquerade: Improving Error Correction of Text Passwords on Mobile Devices by Using Graphic Filters for Password Masking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300916

[25] Florian Mathis, Kami Vaniea, and Mohamed Khamis. 2021. Prototyping Usable Privacy and Security Systems: Insights from Experts. *International Journal of Human–Computer Interaction* (2021), 1–23. https://doi.org/10.1080/10447318.2021.1949134

[26] Shannon Riley. 2006. Password security: What users know and what they actually do. *Usability News* 8, 1 (2006), 2833–2836.

[27] Tobias Seitz. 2018. *Supporting users in password authentication with persuasive design.* Tobias Seitz.

[28] Tobias Seitz, Florian Mathis, and Heinrich Hussmann. 2017. The Bird is the Word: A Usability Evaluation of Emojis inside Text Passwords. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) *(OZCHI '17)*. Association for Computing Machinery, New York, NY, USA, 10–20. https://doi.org/10.1145/3152771.3152773

[29] Larry R Squire and Stuart M Zola. 1998. Episodic memory, semantic memory, and amnesia. *Hippocampus* 8, 3 (1998), 205–211.

[30] Elizabeth Stobert and Robert Biddle. 2014. The password life cycle: user behaviour in managing passwords. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 243–255.

[31] Endel Tulving. 1993. What is episodic memory? *Current directions in psychological science* 2, 3 (1993), 67–70.

[32] Blase Ur, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Do users' perceptions of password security match reality?. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3748–3760.

[33] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added'i'at the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security (SOUPS)*. 123–140.

[34] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. 2005. Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. In *Proceedings of the 2005 Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, USA) *(SOUPS '05)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/1073001.1073002

A ONLINE SURVEY

Our survey started with a participant information sheet, a request of participants´ consent to take part in this research, and an introduction to personal data authentication. The survey continued with the following structure:

(1) **Demographics**
   (a) What is your country of origin?
   (b) How often do you log into any online account?
   (c) What is your age?
   (d) What is your gender?

(2) **Comfort Using Different Personal Data for User Authentication**
   (a) How comfortable would you feel using [data type] for online authentication? *Note that participants responded on 5-point Likert scales from Extremely uncomfortable to Extremely comfortable.*
   (b) Do you have any comments or justifications you would like to make about your answers above?

(3) **Perceived Security of the Different Personal Data Types for User Authentication**
   (a) How secure do you feel [data type] could be for authenticating yourself online? *Note that participants responded on 5-point Likert scales from Extremely insecure to Extremely secure.*

(4) **Likelihood for Behavioural Change When Personal Data is Used for Authentication**
   (a) Would knowing this data was being used as online authentication change the way you behave, e.g. if you were aware that your grocery shopping was part of your online authentication, would you change what you buy? *Note that participants responded with yes/no.*
   (b) If you answered yes to any of the above please briefly describe your reason why?

(5) **Who Else Has Access to the (Personal) Data?**
   (a) Who else has access to [data type] without you directly sharing it, tick all that apply? *Note that participants were asked to tick all entities (i.e., Strangers, Acquaintances, Colleagues, Family, Close Friends, Partners, People you live with, Nobody) that apply to the specific data type.*

(6) **Update Frequency of the Different Personal Data Types**
   (a) How often does [data type] update? For example, how often do you download an app or change your location? *Note that participants responded on a 5-point Likert scale from Less/Never to Multiple times a day.*

(7) **Perceived Memorability of the Different Personal Data Types**

(a) How easy is it to recall information about each data type? *Note that participants responded on 5-point Likert scales from Very difficult to Very easy.*

(8) **Closing Comments**

(a) Prior Encounters With Personal Data Authentication

(b) Any Other Comments

At the end of the survey, participants were also provided with our contact details for further questions.