Replication Study: Cross-Country Evaluation of the Recognition-Based Graphical Authentication Scheme in AR and VR Environments

Naheem Noah
University of Denver
Denver, USA
Email: naheem.noah@du.edu

Peter Mayer
University of Southern Denmark
Denmark
Email: mayer@imada.sdu.dk

Sanchari Das
George Mason University
USA
Email: sdas35@gmu.edu

***Abstract*—Augmented Reality (AR) and Virtual Reality (VR) Head-Mounted Displays (HMDs) handle sensitive data, necessitating user authentication and posing risks of shoulder-surfing in shared spaces. This replication study evaluates the recognition-based graphical authentication scheme – *Things* in AR and VR, extending Düzgün et al. [1], which focused on AR in Germany. We conducted a conceptual replication with 32 U.S. participants using Microsoft HoloLens (AR) and Valve Index (VR). Our findings align with the original study, showing comparable System Usability Scale (SUS) scores and perceived usability across regions. However, our success rate (73%) was lower than the original (90%), while perceived security was higher (3.90 vs. 3.19). Comparing platforms, Valve Index outperformed HoloLens with a higher SUS score (75.16 vs. 70.47), faster authentication ($23.06s$ vs. $49.85s$), and a higher success rate (85% vs. 73%). Participants found Valve Index's controller-based interaction more intuitive than HoloLens' tap gesture. To enhance *Things* in AR, we recommend exploring alternative input methods (e.g., voice commands, gaze-based selection) to reduce physical strain and allowing users to choose password images to improve memorability.**

1. Introduction

Authentication plays a critical role in protecting users' identities and preventing unauthorized access in Augmented and Virtual Reality (AR/VR) environments [2], [3], [4]. While traditional knowledge-based authentication methods, such as passwords and PINs, have been widely used in desktop and mobile computing [5], they may not be well-suited for the unique characteristics of AR/VR systems [6], [7], [8], [9]. As such, graphical authentication schemes have emerged as a promising alternative to traditional methods, offering improved usability and memorability [10], [11]. Of particular note for our work is the *Things* scheme, which is based on the Passfaces [12] scheme and additional ideas like semantic grouping of images as proposed by Weinshall and Kirkpatrick [13]. In this authentication scheme, users learn a set of images during enrolment. During authentication, the user then needs to recognize the learned images among distractors (see Figure 1).

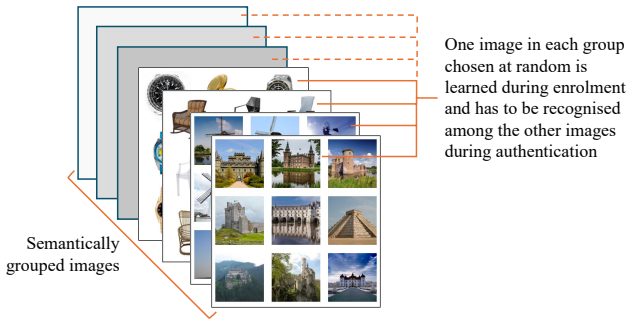

Figure 1. The *Things* authentication scheme: Images are semantically grouped and displayed sequentially in a grid. One image per group is part of the user's password. See Figure 2 for our AR/VR implementation.

Authentication preferences of users can be significantly influenced by cultural differences [14], [15], particularly in emerging technologies like AR/VR. In such, studies have indicated that German users typically demonstrate higher privacy concerns compared to US users [16], which may influence their acceptance of novel authentication methods. These cultural variations extend to biometric authentication acceptance, where research by Zimmermann and Gerber [17] found that German users show stronger preferences for alternatives to biometric authentication due to privacy concerns. The technological landscape and AR/VR familiarity vary across regions. Düzgün et al. [1] first introduced the *Things* scheme in AR, demonstrating strong usability, effectiveness, and resistance to shoulder-surfing in a study with 16 German participants. However, their research was limited to AR HMDs.

Building on this, we conceptually replicate their study while extending the evaluation to VR. Cross-country replication is vital, as "intercultural differences may heavily influence the success of information systems" [18], even among Western nations [19]. Our study broadens the scope in key ways: we implemented the *Things* scheme across AR (Microsoft HoloLens) and VR (Valve Index) to assess cross-platform usability, doubled the participant sample to 32 to accommodate both device groups, and incorporated the NASA Task Load Index (NASA-TLX) [20] along with open-

ended questions to evaluate mental workload and gather qualitative insights on user experience.

2. Related Work

2.1. Authentication in AR/VR Environments

Various knowledge-based authentication schemes for AR HMDs have been proposed, some of which are shoulder-surfing resistant [21], [22] while others are not [23], [24]. Graphical password schemes, including both recall-based and recognition-based approaches, have been explored, in such, Friström et al. proposed a recall-based scheme using gaze gestures to enter a free-form pattern on an AR HMD [25]. Hadjidemetriou et al. also developed a cued-recall scheme for the HoloLens where users enter a pattern on specific positions of an image using hand gestures [26]. In VR HMDS, knowledge-based authentication schemes have also been investigated. Traditional methods such as PIN or pattern lock have been adapted for VR environments [27], [28], [29]. However, researchers have also explored novel authentication schemes tailored to the unique characteristics of VR. George et al. proposed a 3D authentication scheme that leverages the immersive nature of VR, requiring users to interact with virtual objects in a specific sequence [30]. Similarly, Mathis et al. designed a system that combines 3D manipulation and pointing, allowing users to perform authentication gestures using handheld controllers [31]. While first investigations of the shoulder-surfing resistance exist in the VR space [32], the exploration of schemes' resistance to shoulder-surfing attacks in AR is nonexistent.

Recognition-based graphical password schemes are particularly well-suited for AR and VR environments due to their shoulder-surfing resistance and improved usability compared to traditional knowledge-based methods. These schemes, proposed for other platforms [33], [34], [35], [36], use various types of images in different layouts, with the image positions randomized upon each authentication attempt. Hlywa et al. found object images to be more efficient than face images [37]. This finding was corroborated by the findings of Mayer et al. based on a scheme using a semantic grouping of object images in their scheme, which they called *Things* [38]. Building on these findings, we chose to adapt the recognition-based *Things* scheme with semantically-grouped object images to the AR and VR HMD context and evaluate its usability and shoulder-surfing resistance [38].

2.2. Privacy Perceptions in AR/VR Technologies

Studies have highlighted the growing importance of understanding user privacy perceptions in AR/VR environments [39], [40], [41], [42]. De Guzman et al. conducted a comprehensive survey on privacy and security in Mixed Reality (MR), emphasizing the unique challenges posed by these immersive technologies [43]. They found that users are particularly concerned about the collection and potential misuse of biometric data, such as eye-tracking information and hand gestures. Miller et al. observed a shift in user focus from hardware-related privacy issues to concerns about data usage, AI integration, and potential surveillance capabilities of AR devices over a five-year period [44]. This highlights the dynamic nature of privacy perceptions in rapidly evolving technological landscapes. In the context of authentication, Lebeck et al. explored user perceptions of various authentication methods in AR, finding that users often struggle to balance convenience with perceived security [45]. Their work highlights the need for authentication schemes that not only provide technical security but also align with users' mental models of privacy and security in immersive environments.

Our work on the *Things* authentication scheme addresses these privacy concerns by offering a shoulder-surfing resistant method that aligns with users' expectations of security in shared spaces. By evaluating this scheme across different cultural contexts (U.S. and Germany) and in both AR and VR environments, our study contributes to the growing body of knowledge on culturally sensitive, privacy-preserving authentication methods in immersive technologies.

3. Method

We implemented a design similar to that by Düzgün et al. [1] but with several modifications. The original study's authors generously provided access to their anonymized data, enabling us to conduct a comparative analysis.

3.1. *Things* Scheme Configuration

In the original work by Düzgün et al., the authors made several design decisions regarding the configuration of the *Things* graphical password scheme [1]. They chose to use object images with a clear central theme that varies in color, shape, and semantics, as suggested by previous research [13], [37]. The images were selected from royalty-free sources on the web and grouped according to their semantic categories, such as fruits or flowers. To determine the appropriate password space, the authors considered the requirement for the scheme to have a password space at least equal to that of a 6-digit PIN (10^6 = 19.93 bits). They tested various grid sizes and found that a 4x4 grid (16 images) was clearly visible to the user on the HoloLens display. With a grid size of 16 images and a password length of 5, the resulting password space is $16^5 = 1.048.576 \approx 20$ bits, which is close to the security level of a 6-digit PIN. Regarding password choice, Düzgün et al. decided to assign users a randomly generated password to mitigate the predictability issues associated with user-chosen graphical passwords [46] and to leverage the increased memorability of pictures [47], [48].

3.2. Study Design and Procedure

Düzgün et al. conducted an in-lab study to evaluate the usability and perceived security (SUS) of the *Things* scheme on the Microsoft HoloLens 2 AR HMD. The study procedure involved 16 participants and consisted of: Check-in and

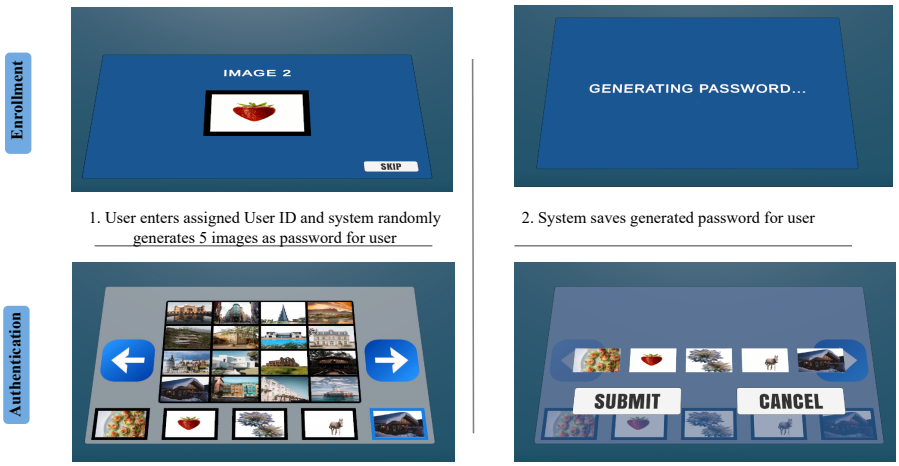

Figure 2. User VR Environment Showing the Enrollment and Authentication Process in the *Things* Scheme

application of hygiene measures due to the COVID-19 pandemic; Informed consent; Questions on previous experience with AR and willingness to use them; Calibration and warm-up with the HMD; Enrollment process with *Things* where participants were assigned a random 5-image password; Authentication process with *Things* repeated three times and the interactions were performed via hand gestures and the information was displayed on the HMD's private display.

We employed a between-subject design where different groups of participants tested the authentication scheme in either AR or VR environments. We conducted a pre-screening survey using Qualtrics to assess the eligibility of potential participants based on the study's requirements. The survey included an informed consent form, demographic questions (age, gender, and educational level), and inquiries about digital usage. Only individuals who met the inclusion criteria, which included being at least 18 years old, resident of the United States, proficient in English, and able to visit the lab for the study, were invited to participate in the in-lab experiment. We recruited a sample of 32 participants to evaluate the usability and perceived security of the *Things* authentication scheme on two distinct platforms: the Microsoft HoloLens for AR and the Valve Index for VR. Upon arrival at the lab, participants completed an initial survey focusing on their previous experience with AR and VR head-mounted displays (HMDs), their willingness to use these technologies in the future, and their prior authentication experiences. Following a brief introduction to the HoloLens and Valve Index, participants viewed a demonstration of the *Things* scheme and proceeded to the experimental phase, which involved memorizing five images that would be used for authenticating the HMDs. After completing the authentication process, participants filled out a post-experiment survey assessing the perceived usability and security of the proposed scheme. The entire procedure was designed to be efficient, with each participant's session lasting no more than 30 minutes. The overview of the participant's recruitment and the research methodology is detailed in. The authentication process are shown in Figure 2.

3.3. Participant Recruitment and Ethics

We recruited participants through a combination of flyer distribution, social media advertising, and mailing list outreach. The study protocol was reviewed and approved by the Institutional Review Board (IRB), ensuring compliance with ethical standards for human subjects research. All participants were provided with an informed consent form, which detailed the study's purpose, procedures, and data handling practices. The researcher was present throughout the study to provide instructions, assist participants, and address any questions or concerns that arose. Participants were informed of their right to withdraw from the study at any time, and in such cases, their data would be promptly deleted. To protect participant privacy, all responses were analyzed in an aggregated and anonymized manner. Data collection and storage were managed using Qualtrics, with all data securely stored in compliance with institutional data protection policies. Any biometric data inadvertently collected during the study was deleted immediately after the experiment and excluded from analysis. Participants received a compensation of 10 USD for their time and effort, an amount determined based on the study duration (max = 30 minutes) and the prevailing minimum wage in the United States at the time of the research.

3.4. Quantitative and Qualitative Evaluation

Düzgün et al. in their work, conducted both quantitative and qualitative evaluations to assess the usability and perceived security of the *Things* authentication scheme on the Microsoft HoloLens 2. The quantitative evaluation focused on measuring the scheme's effectiveness (success rate of correctly entering the password), efficiency (time

taken to provide the secret), and user satisfaction (using the System Usability Scale, or SUS). Similarly, in our study, we conducted quantitative and qualitative evaluations mirroring the original work, while participants interacted with the authentication scheme using the Microsoft HoloLens and Valve Index. To further enhance our understanding of the user experience, we incorporated the NASA Task Load Index for evaluating subjective mental workload, using a 5-point Likert scale.

4. Results

4.1. Participants

The majority of our participants identified as men (81.25%), with ages ranging from 18 to 50 years old. Their educational backgrounds spanned a wide spectrum, from high school graduates (9.4%) to doctoral candidates (3.13%), with a significant contingent pursuing bachelor's (43.75%) and master's (40.6%) degrees. While over half (59.38%) had prior exposure to AR/VR technologies, only 15.63% had entered passwords in such environments before, and a mere 9.38% owned a personal AR/VR device. Table 2 details the demographic information.

Comparison with Düzgün et al. study: They evaluated 16 participants with an average age of 24. Similar to our study, the majority were male (56%), and 44% were female. While 75% had experience with VR HMDs, none had used AR HMDs. Only one participant used an HMD regularly, while the rest did so rarely or not at all. Only one participant owned a VR HMD, and none had ever entered a password on an HMD.

4.2. Effectiveness and Efficiency

The overall SUS score for our study encompassing both AR and VR is 72.81 which falls within the "B-" level, straddling the line between good and excellent usability [49]. The enrollment process was efficient, averaging 25.23 seconds $(SD = 1.54)$. Authentication sessions averaged 36.45 seconds $(SD = 36.80)$, with successful attempts being notably quicker. The overall success rate was 79% $(SD = 0.39)$, with 75% of participants succeeding in all three iterations. For the HoloLens group, the SUS score was 70.47. The average enrollment duration was 25.38 seconds $(SD = 1.63)$, while authentication averaged 49.85 seconds $(SD = 47.53)$. Successful authentications took 29.87 seconds $(SD = 6.42)$ on average. The success rate was 73%, with 68.75% of participants authenticating successfully in all three attempts. The Valve Index group showed superior performance with a SUS score of 75.16. Enrollment averaged 25.01 seconds $(SD = 1.63)$, and authentication took 23.06 seconds $(SD = 9.56)$ on average. Successful authentications were completed in 19.02 seconds $(SD = 4.50)$. The success rate was 85%, with 81.25% of participants succeeding in all iterations.

Comparison with Düzgün et al. study: We found comparable SUS scores (72.81 vs. 74) and the same percentage (75%) of participants succeeding in all three iterations. Our study achieved faster enrollment ($25.23s$ vs. $62.21s$) but slightly longer authentication times ($36.45s$ vs. $32.2s$). Our overall success rate was lower (79% vs. 90%). The consistent outperformance of the Valve Index group over the HoloLens group in terms of SUS score, authentication speed, and success rate suggests that the *Things* scheme may be more effective in VR environments.

4.3. Perceived Security

In our study, we found that participants' overall perceptions of the system's security were considerably high with an average rating of 3.94 $(SD = 1.18)$. The perceived security against shoulder-surfing attacks recorded a "high" security rating of 4.4 $(SD = 0.74)$ indicating that users in our study felt the system provided stronger protection against this type of threat. We also found interesting device-specific differences in user perceptions. The HoloLens group in our study recorded a significantly high general security rating of 3.90 $(SD = 1.22)$ and a perceived shoulder-surfing security rating of 4.25 $(SD = 0.75)$. Remarkably, the Valve Index group exhibited the highest levels of perceived security, both in general (3.94, $SD = 1.14$) and in shoulder-surfing protection, with an exceptional "very high" rating of 4.5 $(SD = 0.71)$.

Comparison with Düzgün et al. study: Participants perceived the general security of the authentication scheme as moderately secure, with a rating of 3.19 out of 5 $(SD = 1.01)$, suggesting room for improvement in the overall security perception. Our study showed considerably higher overall perceptions of the system's security, with an average rating of 3.94 $(SD = 1.18)$. When examining the perceived security against shoulder-surfing attacks, Düzgün et al.'s participants rated it as relatively high, with a score of 3.94 out of 5 $(SD = 1.09)$. However, our study recorded even higher perceived security against shoulder-surfing attacks, with a rating of 4.4 $(SD = 0.74)$, indicating that our participants felt more confident in the scheme's ability to resist such attacks.

4.4. Perceived Usability

The overall ease of use was high, with a rating of 3.81 $(SD = 1.18)$ out of 5. Further, our participants reported the password as easy to remember, with a rating of 3.88 $(SD = 1.39)$. Users in our study also perceived the login process to be faster, rating it 3.94 $(SD = 1.32)$, and they expressed a greater willingness to use the scheme in the future, with a rating of 4.06 $(SD = 1.06)$. For HoloLens users, the perceived ease of use was rated slightly lower at 3.5 $(SD = 1.32)$, while password memorability aligned with the overall findings at 3.81 $(SD = 1.55)$; however, these participants did rate the login speed higher at 3.69 $(SD = 1.26)$ and expressed greater willingness to use the system in the future at 3.88 $(SD = 1.17)$. Remarkably,

	Düzgün et al.'s Study	Our Study Overall	HoloLens Study	Valve Index Study
Participants	16	32	16	16
Effectiveness & Efficiency				
SUS Score	74	72.81	70.47	75.16
Average Enrollment Duration	62.21s (SD = 24.76)	25.23s (SD = 1.54)	25.38s (SD = 1.63)	25.01s (SD = 1.63)
5-second Interval Rating	3.81	4.03	3.81	4.25
Average Authentication Duration	32.2s (SD = 9.39)	36.45s (SD = 36.80)	49.85s (SD = 47.53)*	23.06s (SD = 9.56)*
Average Successful Authentication Duration	-	25.23s (SD = 1.54)	29.87s (SD = 6.42)	19.02s (SD=4.50)
Overall Success Rate	90%	79% (SD = 0.39)	73% (SD = 0.46)	85% (SD = 0.33)
Participants Success in three iterations	75%	75%	68.75%	81.25%
Perceived Security				
Overall System Security	3.19 (SD = 1.01)	3.94 (SD = 1.18)	3.90 (SD = 1.22)	3.94 (SD = 1.14)
System Security against Shoulder-surfing	3.94 (SD = 1.18).	4.4 (SD = 0.74)	4.25 (SD = 0.75)	4.4 (SD = 0.74)
Perceived Usability				
Easy to Use	4.00 (SD = 1.12)	3.81 (SD = 1.18)	3.5 (SD = 1.32)	4.13 (SD = 0.93)
Easy to Remember	4.31 (SD = 0.77)	3.88 (SD = 1.39)	3.81 (SD = 1.55)	3.94 (SD = 1.20)
Fast Login Process	3.19 (SD = 1.38)	3.94 (SD = 1.32)	3.69 (SD = 1.26)	4.19 (SD = 1.33)
Willingness to Use in the Future	3.25 (SD = 1.09)	4.06 (SD = 1.06)	3.88 (SD = 1.17)	4.25 (SD = 0.90)
Perceived Task Load				
Mental Demand	-	2.63 (SD = 1.02)	2.56 (SD = 1.12)	2.69 (SD = 0.92)
Physical Demand	-	2.28 (SD = 1.15)	2.56 (SD = 1.22) -	2 (SD = 1)
Temporal Demand	-	2.47 (SD = 0.83)	2.5 (SD = 0.87) -	2.44 (SD = 0.79)
Successful Accomplishment	-	4.41 (SD = 1.17)	4.19 (SD = 1.29) -	4.6 (SD = 0.99)
Level of Hard work	-	2.5 (SD = 0.97)	2.56(SD = 1.12) -	2.44 (SD = 0.79)
Insecurity or Stress-level	-	1.66 (SD = 0.92)	1.94 (SD = 0.97) -	1.38 (SD = 0.78)

TABLE 1. DETAILS OF SCHEME EVALUATIONS BASED ON EFFECTIVENESS, SECURITY, AND USABILITY. * = STATISTICALLY SIGNIFICANT

the Valve Index group exhibited elevated ratings across all metrics, scoring 4.13 ($SD = 0.93$) for ease of use, 3.94 ($SD = 1.12$) for password memorability, 4.19 ($SD = 1.33$) for login speed, and an impressive 4.25 ($SD = 0.9$) for future usage intentions - substantially higher than both the HoloLens results and the previous study.

Comparison with Düzgün et al. study: Participants rated it as easy to use (4.00 out of 5, $SD = 1.12$). Similarly, our study recorded an overall ease of use rating of 3.81 ($SD = 1.18$), confirming the scheme's user-friendliness. In the previous work, participants found the passwords easy to remember (4.31 out of 5, $SD = 0.77$). While our participants also reported the password as easy to remember, the rating was slightly lower at 3.88 ($SD = 1.39$). The perceived speed of the login process in Düzgün et al.'s work was 3.19 out of 5 ($SD = 1.38$). However, users in our study perceived the login process to be faster, with a higher rating of 3.94 ($SD = 1.32$). Regarding the willingness to use the scheme on a HoloLens device, the previous work reported a moderate level of willingness (3.25 out of 5, $SD = 1.09$) but our participants expressed a greater willingness to use the scheme in the future, with a rating of 4.06 ($SD = 1.06$), indicating a higher level of acceptance and potential for adoption.

4.5. Perceived Task Load

Extending on the pioneering work by Düzgün et al., we leveraged the NASA Task Load Index, a renowned tool for assessing subjective mental workload. We found that the perceived mental demand for HoloLens users was at a medium level with a 2.56 ($SD = 1.12$) rating, similar to the 2.69 ($SD = 0.92$) rating found in the Valve Index group and an overall rating of 2.63 ($SD = 1.02$). Regarding physical demand, HoloLens users also perceived a medium level of exertion ($2.56, SD = 1.22$), while Valve Index users rated it as low at 2.0 ($SD = 1.0$), with an overall score of 2.28 ($SD = 1.15$). The temporal demand of the password scheme was rated as medium for HoloLens at 2.5 ($SD = 0.87$), but the rating for Valve Index was a little lower at 2.44 ($SD = 0.79$), with an overall rating of 2.47 ($SD = 0.83$). When measuring the perceived success

Demographics Info	Düzgün et al's Study	HoloLens	Valve Index	Both
Participants	16	16	16	32
Age Range (years)				
18-24	62.5%	37.5%	25%	31.25%
25-30	31.25%	43.75%	56.25%	50%
31-40	0%	6.25%	12.5%	9.38%
41-50	6.25%	12.5%	6.25%	9.38%
Gender				
Male	56.25%	75%	87.5%	81.25%
Female	43.75%	25%	12.5%	18.75%
Educational Background				
Masters	No info	37.5%	43.75%	40.63%
Bachelors	No info	43.75%	43.75%	43.75%
Doctoral	No info	6.25%	0%	3.13%
High School	No info	6.25%	12.5%	9.38%
Diploma	No info	6.25%	0%	3.13%
Prior Usage of AR/VR Headsets	75%	50%	68.75%	59.38%
Future Use of AR/VR Headset	91.67%	81.25%	93.75%	87.5%
Ownership of AR/VR Headset	6.25 %	12.5%	6.25%	9.38%
Prior password entry in AR/VR Headset	0%	18.75%	12.5%	15.63%

TABLE 2. PARTICIPANT DEMOGRAPHICS AND CHARACTERISTICS

in accomplishing the authentication task, HoloLens users recorded a high-level rating of 4.19 ($SD = 1.12$), while Valve Index participants achieved an even more impressive "very high" rating of 4.60 ($SD = 0.99$), with an overall success rating of 4.41 ($SD = 1.17$). When asked about the difficulty of the task, HoloLens users perceived a medium level of hardship at 2.56 ($SD = 1.12$), but Valve Index participants rated it lower at 2.44 ($SD = 0.79$), with an overall difficulty score of 2.5 ($SD = 0.97$). Most notably, we measured the participants' levels of insecurity, discouragement, irritation, stress, and annoyance, finding a low rating of 1.94 ($SD = 0.97$) for HoloLens and a lower rating of 1.38 ($SD = 0.78$) for Valve Index, resulting in an overall score of 1.66 ($SD = 0.92$).

Comparison with Düzgün et al. study: Perceived task load was not reported.

4.6. User Qualitative Response

To further understand user perception regarding the use of the authentication scheme, we asked users to answer five open-ended questions regarding the perceived usability and security of the scheme.

Overall Experience: The Valve Index users unanimously described the scheme as easy to use. Participant *VI02* noted,

> *"I found the scheme easy to use overall. Once I got going with the system, it was very quick to enter my password, much quicker than it would have been for me to type out a pin or password on a virtual keyboard."*

In contrast, the HoloLens participants, while also acknowledging the ease of use, expressed some notable frustrations. Many found the tap gesture required by the HoloLens to be quite strenuous. One participant (*HL05*) said,

> *"Clicking each item and focusing it was the hardest part. It was having a difficult time sensing clicks and focusing."*

Specific Challenges Encountered: The Valve Index group did not report any significant challenges with using the authentication scheme. One participant (*VI02*) noted,

> *"Nothing specific. The only thing that might make it better would be a slightly longer display time for the images during registration, but that's all I can think of."*

In contrast, the HoloLens users shared several frustrations, primarily centered around the tap gesture and the difficulty of remembering the randomly selected images. One HoloLens participant, *HL03* elaborated,

> *"The main challenge was obviously remembering the images. Since I am used to traditional, [[text-based]] password schemes, I tried to map the images to certain words and assign initials to each image. By assigning those initials I was able to [[remember]] the password for the duration of the experiment. It is highly likely that I will forget it after a while."*

Level of Comfort: We asked participants how comfortable or uncomfortable they felt while using the HoloLens or Valve Index during the experiment. The Valve Index participants largely reported feeling very comfortable with the headset, with one user *VI16* stating,

> *"I was very comfortable with the headset, and did not feel any discomfort."*

However, some Valve Index users did express minor inconveniences, such as the low user perspective, as one participant *VI01* mentioned,

> *"The only inconvenience was that the user perspective was very low, so it felt as though I was looking up at the interface the entire time, which was a bit uncomfortable."*

For the HoloLens group, the feedback was similar in that participants did not experience any motion sickness, but they did report some discomfort associated with the tap gesture interaction. One HoloLens participant, *HL02* mentioned,

"I didn't feel motion sickness just a little fatigued from using my fingers as a button, so having to hold it above 90 degrees for most of the time."

Security & User Friendliness: We asked users their thoughts on how the authentication scheme compares to traditional authentication methods (e.g., passwords, PINs, biometrics) in terms of security and user-friendliness. Based on the participants' feedback, the views on how the authentication scheme compared to traditional methods were quite consistent across both the Valve Index and HoloLens groups. Most users felt that the scheme was easier to use than traditional password or PIN-based authentication, but more difficult than biometric methods. Participant *VI2* noted,

"It seems like it would be just as secure as the other options. I found it easier to use than a password or pin, but more difficult to use than biometrics."

However, some HoloLens users did express reservations about the scheme's security compared to other authentication methods which may be as a result of their misconception about security in traditional authentication schemes as they emphasized that other authentication schemes are traditionally computationally unfeasible to crack, overlooking the vulnerabilities inherent in many conventional password systems. One participant (*HL03*) commented,

"In terms of user-friendliness, and visual appeal, it was definitely very pretty. In comparison it to other authentication methods, I find it to be weaker than them. Maybe because other schemes have certain factors that allow it to be computationally unfeasible or require having a secondary physical device for two-factor auth, and I don't see any such thing being introduced here."

Additional Thoughts or Suggestions: We asked users to share any additional thoughts, suggestions, or concerns about the overall user experience and security of the authentication scheme that we may not have covered in the previous questions. While acknowledging the scheme's usability and user-friendliness, some users felt it should be positioned as an alternative mode of authentication, rather than the sole option. As one Valve Index participant (*VI01*) noted,

"I think that this authentication scheme would make for a good alternative to PINs and passwords, but could be frustrating to users if it was the only option for a system."

Some Valve Index users also expressed a preference for being able to select their own images as passwords, rather than having them randomly generated. One participant (*VI04*) suggesting,

"I think the images at the password creation stage can either be shown twice in succession, or give the user the opportunity to select images they are familiar with."

The HoloLens users echoed similar sentiments, with one participant (*HL11*) suggested,

"Additional thoughts would be that, it will be interesting if the user can custom select their images according to their preference, that way it mitigate the risks of not recalling the images that were automatically presented by the system."

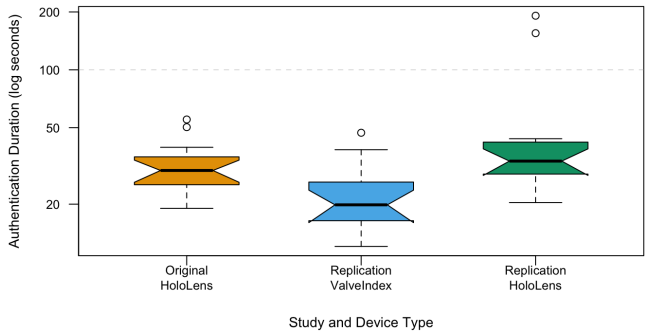

Figure 3. Boxplot showing Authentication Duration based on study and Device Type with outliers.

Comparison with Düzgün et al. study: Users' qualitative responses were not reported.

4.7. Statistical Analysis

We performed a two-way ANOVA on the authentication duration data using R [50] to measure the differences between the studies, as well as between the HoloLens and Valve Index in the replicated work. The main effect of the study (original vs. replication) on authentication duration is not statistically significant ($p = 0.6352$). This suggests that there is no significant difference in authentication durations between the original and replication studies, irrespective of the device type used. However, the main effect of device type (HoloLens vs. Valve Index) on authentication duration is statistically significant ($p = 0.0134$) indicating that there is a significant difference in authentication durations between the HoloLens and Valve Index devices. The boxplot visualization of the authentication duration data grouped by the interaction of Study and Device Type, shown in Figure 3, provides further insights into the distribution of the data.

5. Discussion and Implications

5.1. Comparing Replicated Study

In terms of effectiveness and efficiency, our study recorded an overall SUS score of 72.81, comparable to the original study's score of 74, indicating similar levels of usability. The enrollment process in our study proved to be more efficient, with an average duration of 25.23 seconds compared to 62.21 seconds in the original work. The 5-second interval for displaying password images was rated higher in our study than in Düzgün et al.'s with fewer participants expressing a preference for a longer duration. While our study recorded a slightly longer average authentication session duration (36.45 seconds) compared to the original work (32.2 seconds), successful authentication

sessions were notably quicker in our study. The overall success rate in our study (79%) was slightly lower than the original study (90%), but the percentage of participants succeeding in all three iterations was consistent (75% in both studies). Regarding perceived security, our study found higher ratings for both overall system security (3.94) and shoulder-surfing resistance (4.4) compared to the original work (3.19 and 3.94 respectively). This suggests that users in our study felt more confident in the scheme's ability to protect against unauthorized access and shoulder-surfing attacks. The perceived usability metrics in our study were generally comparable to those reported by Düzgün et al., with slight variations. Our participants found the scheme easy to use (3.81 vs. 4.00 in the original study), the password easy to remember (3.88 vs. 4.31), and expressed a greater willingness to use the system in the future (4.06 vs. 3.25).

The demographic differences between our replication study and Düzgün et al.'s original work warrant careful consideration when comparing results. Our study had a higher proportion of male participants (81.25% vs. 56%) and a broader age range (18-50 years vs. average age of 24), which could influence technology adoption rates and comfort levels with AR/VR devices. The higher percentage of participants with prior AR/VR experience in our study (59.38% vs. 75% for VR only in the original) might contribute to easier adaptation to the authentication scheme, potentially affecting usability ratings and task performance. The broader educational background of our participants could also influence cognitive approaches to the authentication task. Additionally, cultural differences between the U.S. and German populations could impact user behavior and preferences in authentication methods. For instance, the higher perceived security ratings in our study might reflect differing cultural attitudes toward privacy and technology.

5.2. Comparing HoloLens and Valve Index Study

We found that in terms of effectiveness and efficiency, the Valve Index group demonstrated better performance compared to the HoloLens group. The Valve Index users achieved a higher SUS score (75.16) than the HoloLens users (70.47), indicating a more positive overall usability experience. The enrollment process was slightly more efficient on the Valve Index, with an average duration of 25.01 seconds compared to 25.38 seconds on the HoloLens. Valve Index users also rated the 5-second interval for displaying password images higher than HoloLens users. The authentication process on the Valve Index showed an average authentication duration of 23.06 seconds and an average successful authentication duration of 19.02 seconds. These figures significantly outperformed the HoloLens trial, which had an average authentication duration of 49.85 seconds and an average successful authentication duration of 29.87 seconds. The authentication success rate was also higher on the Valve Index (85%) compared to the HoloLens (73%).

Regarding perceived security, both device groups exhibited high levels of confidence in the scheme's ability to protect against unauthorized access and shoulder-surfing attacks. The Valve Index group rated the overall system security (3.94) and shoulder-surfing resistance (4.5) slightly higher than the HoloLens group (3.90 and 4.25 respectively). The perceived usability metrics were consistently higher for the Valve Index group compared to the HoloLens group. Valve Index users found the scheme easier to use (4.13 vs. 3.5), rated the login process as faster (4.19 vs. 3.69), and expressed a greater willingness to use the system in the future (4.25 vs. 3.88). Both groups found the password comparably easy to remember (3.94 for Valve Index and 3.81 for HoloLens).

The difference between the HoloLens and Valve Index groups could be attributed to the differences in the interaction methods and user experiences offered by the two devices. The Valve Index's controller-based interaction was found to be more intuitive and comfortable compared to the HoloLens' tap gesture, which some users reported as tiring and unresponsive. The Valve Index's high-resolution display and immersive environment may have also contributed to better user engagement and faster recognition of password images. These factors likely influenced the Valve Index group's superior effectiveness, efficiency, and user satisfaction ratings compared to the HoloLens group.

5.3. Cognitive and Physical Demands

Our study extended Düzgün et al.'s work by incorporating the NASA Task Load Index, revealing low to medium levels of mental (2.63), physical (2.28), and temporal (2.47) demand across both devices. These scores suggest that the *Things* scheme doesn't impose significant cognitive or physical burdens, a crucial factor for long-term adoption in immersive environments. Key differences emerged between devices: HoloLens users reported slightly higher mental (2.56 vs 2.69) and physical (2.56 vs 2.0) demands compared to Valve Index users, likely due to the more strenuous tap gesture interaction. Both groups reported high levels of perceived success (HoloLens: 4.19, Valve Index: 4.60) and low levels of frustration (HoloLens: 1.94, Valve Index: 1.38), indicating good overall user experience despite the differences. These findings have significant implications for AR/VR authentication design, suggesting the need for ergonomic input methods, consideration of cognitive accessibility, and platform-specific optimizations. The consistently low cognitive and physical demands indicate the *Things* scheme's potential for long-term adoption and user retention.

5.4. Cultural and Regional Differences

The security perception ratings showed differences, with our US AR participants rating security at 3.90 ($SD = 1.22$) compared to German participants' 3.19 ($SD = 1.01$). This variance aligns with research by Cyr and Trevor-Smith showing that cultural differences significantly impact user interaction with technology interfaces, even between Western countries [19]. German users' historically stronger emphasis on privacy and security validation may explain

their more conservative security ratings [16]. The willingness to use the system showed an interesting contrast, with US AR participants expressing higher future use intention $(3.88, SD = 1.17)$ compared to German participants $(3.25, SD = 1.09)$. This aligns with findings from Zimmermann and Gerber that US users generally show greater openness to adopting new authentication methods [17], while German users typically demonstrate stronger preferences for established authentication mechanisms, influenced by their cultural emphasis on privacy protection.

Our study confirms that the *Things* authentication scheme compares favorably to existing methods in immersive environments. Hlywa et al.'s study on graphical passwords reported average authentication times of 22.55 seconds $(SD = 10.02)$ for object images and 35.96 seconds $(SD = 18.1)$ for face images [37]. Our Valve Index implementation achieved similar average times (23.06 seconds, $SD = 9.56$) but faster successful authentication times (19.02 seconds, $SD = 4.50$). Kim et al.'s grid-based scheme [51] recorded average enrollment and authentication times of 22.1 and 35.6 seconds, respectively. Our Valve Index implementation showed slightly longer enrollment times (25.01 seconds) but faster authentication times (23.06 seconds, 19.02 seconds for successful attempts).

6. Limitations and Future Work

While the original study utilized the HoloLens 2 for AR, our replicated work explored the scheme using the HoloLens 1 in AR and the Valve Index in VR. This variation may have contributed to some differences in the results, although the studies had similar outcomes for AR. The primary difference between the HoloLens 1 and 2 is the inclusion of the second-generation Holographic Processing Unit (HPU), which enhances computing power from both graphics and processing standpoints which had no direct impact on our study. However, in the future, we plan to investigate the performance of the *Things* scheme across other AR and VR platforms such as the Apple Vision Pro which supports both AR and VR modes of operation.

7. Conclusion

The extensive adoption of AR/VR in shared spaces has underscored the necessity for enhanced security measures in these immersive environments. In this study, we conceptually replicated and validated the effectiveness, efficiency, and user acceptance of the *Things* authentication scheme in both AR and VR settings, corroborating the findings of Düzgün et al.'s original work on AR [1]. The scheme demonstrated comparable performance across different geographic regions (US and Germany) and device types, with the Valve Index group exhibiting superior results in terms of authentication success rate, duration, and user satisfaction compared to the HoloLens group. Specifically, Valve Index participants provided higher ratings for ease of use (4.13), login speed (4.19), and willingness to use the scheme in the future (4.25) compared to the HoloLens group's ratings for those metrics. Qualitative feedback provided valuable insights into user challenges and preferences, emphasizing the need for intuitive interaction methods and customizable password image selection to improve user experience and acceptance. Our results suggest that the *Things* scheme holds considerable promise, particularly in VR settings, where it has shown remarkable potential for adoption.

References

[1] R. Düzgün, P. Mayer, and M. Volkamer, "Shoulder-surfing resistant authentication for augmented reality," in *Nordic Human-Computer Interaction Conference*, 2022, pp. 1–13.

[2] N. Noah, S. Shearer, and S. Das, "Security and privacy evaluation of popular augmented and virtual reality technologies," in *Proceedings of the 2022 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence, and Neural Engineering (IEEE MetroXRAINE 2022)*, 2022.

[3] Y. Zhang, C. Slocum, J. Chen, and N. Abu-Ghazaleh, "It's all in your head (set): Side-channel attacks on {AR/VR} systems," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 3979–3996.

[4] R. Zhang, N. Zhang, C. Du, W. Lou, Y. T. Hou, and Y. Kawamoto, "Augauth: Shoulder-surfing resistant authentication for augmented reality," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[5] S. Das, *A risk-reduction-based incentivization model for human-centered multi-factor authentication*. Indiana University, 2020.

[6] A. Kupin, B. Moeller, Y. Jiang, N. K. Banerjee, and S. Banerjee, "Task-driven biometric authentication of users in virtual reality (vr) environments," in *MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I 25*. Springer, 2019, pp. 55–67.

[7] H. Riyadh, D. Bhardwaj, A. Dabrowski, and K. Krombholz, "Usable authentication in virtual reality: Exploring the usability of pins and gestures," in *International Conference on Applied Cryptography and Network Security*. Springer, 2024, pp. 412–431.

[8] N. Noah and S. Das, "From pins to gestures: Analyzing knowledge-based authentication schemes for augmented and virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, 2025.

[9] J. M. Jones, R. Duezguen, P. Mayer, M. Volkamer, and S. Das, "A literature review on virtual reality authentication," in *Human Aspects of Information Security and Assurance: 15th IFIP WG 11.12 International Symposium, HAISA 2021, Virtual Event, July 7–9, 2021, Proceedings 15*. Springer, 2021, pp. 189–198.

[10] R. Biddle, S. Chiasson, and P. C. Van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys (CSUR)*, vol. 44, no. 4, pp. 1–41, 2012.

[11] R. Düzgün, N. Noah, P. Mayer, S. Das, and M. Volkamer, "Sok: A systematic literature review of knowledge-based authentication on augmented reality head-mounted displays," in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 2022, pp. 1–12.

[12] P. Corporation, "The Science Behind Passfaces," Passfaces Corporation, Tech. Rep., 2006.

[13] D. Weinshall and S. Kirkpatrick, "Passwords you'll never forget, but can't recall," in *CHI'04 extended abstracts on Human factors in computing systems*, 2004, pp. 1399–1402.

[14] I. Y. Alhasan, "Human factors in cybersecurity: A cross-cultural study on trust," Ph.D. dissertation, Purdue University Graduate School, 2023.

[15] M. Merhi, K. Hone, and A. Tarhini, "A cross-cultural study of the intention to use mobile banking between lebanese and british consumers: Extending utaut2 with security, privacy and trust," *Technology in Society*, vol. 59, p. 101151, 2019.

[16] H. Krasnova and N. F. Veltri, "Privacy calculus on social networking sites: Explorative evidence from germany and usa," in *2010 43rd Hawaii international conference on system sciences*. IEEE, 2010, pp. 1–10.

[17] V. Zimmermann and N. Gerber, "The password is dead, long live the password–a laboratory study on user perceptions of authentication schemes," *International Journal of Human-Computer Studies*, vol. 133, pp. 26–44, 2020.

[18] P. Roessger, "An international comparison of the usability of driver-information-systems: tools, results and implications," *SAE transactions*, pp. 776–779, 2003.

[19] D. Cyr and H. Trevor-Smith, "Localization of web design: An empirical comparison of german, japanese, and united states web site characteristics," *Journal of the American society for information science and technology*, vol. 55, no. 13, pp. 1199–1208, 2004.

[20] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[21] D. V. Bailey, M. Dürmuth, and C. Paar, "Typing passwords with voice recognition: How to authenticate to google glass," in *Proc. of the Symposium on Usable Privacy and Security*, 2014, pp. 1–2.

[22] M. Funk, K. Marky, I. Mizutani, M. Kritzler, S. Mayer, and F. Michahelles, "Lookunlock: Using spatial-targets for user-authentication on hmds," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–6.

[23] A. Khare, V. Kulkarni, and A. Upadhayay, "A collaborative augmented reality system based on real time hand gesture recognition," *Global Journal of Computer Science and Technology. Global Journals, US*, pp. 47–51, 2012.

[24] M. R. Islam, D. Lee, L. S. Jahan, and I. Oakley, "Glasspass: Tapping gestures to unlock smart glasses," in *Proceedings of the 9th Augmented Human International Conference*, 2018, pp. 1–8.

[25] E. Fristrom, E. Lius, N. Ulmanen, P. Hietala, P. Kärkkäinen, T. Mäkinen, S. Sigg, and R. D. Findling, "Free-form gaze passwords from cameras embedded in smart glasses," in *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*, 2019, pp. 136–144.

[26] G. Hadjidemetriou, M. Belk, C. Fidas, and A. Pitsillides, "Picture passwords in mixed reality: Implementation and evaluation," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–6.

[27] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann, "Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality," in *Proceedings of the 2017 Symposium on Usable Security and Privacy (USEC '17)*. NDSS, 2017.

[28] I. Olade, H.-N. Liang, C. Fleming, and C. Champion, "Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr)," in *Proceedings of the 2020 4th international conference on virtual and augmented reality simulations*, 2020, pp. 45–52.

[29] F. Mathis, K. Vaniea, and M. Khamis, "Can i borrow your atm? using virtual reality for (simulated) in situ authentication research," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 301–310.

[30] C. George, M. Khamis, D. Buschek, and H. Hussmann, "Investigating the third dimension for authentication in immersive virtual reality and in the real world," in *2019 ieee conference on virtual reality and 3d user interfaces (vr)*. IEEE, 2019, pp. 277–285.

[31] F. Mathis, J. H. Williamson, K. Vaniea, and M. Khamis, "Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing," *ACM Transactions on Computer-Human Interaction (ToCHI)*, vol. 28, no. 1, pp. 1–44, 2021.

[32] T. Länge, P. Matheis, R. Düzgün, M. Volkamer, and P. Mayer, "Vision: Towards fully shoulder-surfing resistant and usable authentication for virtual reality," 2024.

[33] S. Brostoff and M. A. Sasse, "Are passfaces more usable than passwords? a field trial investigation," in *People and computers XIV—usability or else! Proceedings of HCI 2000*. Springer, 2000, pp. 405–424.

[34] P. Dunphy, A. P. Heiner, and N. Asokan, "A closer look at recognition-based graphical passwords on mobile devices," in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, 2010, pp. 1–12.

[35] B. O. Okundaye, "A tree grammar-based visual password scheme," Ph.D. dissertation, University of the Witwatersrand, Faculty of Science, School of Computer Science, 2016.

[36] S. Subramaniyam, "Sketch recognition based classification for eye movement biometrics in virtual reality," Ph.D. dissertation, Texas A&M University, 2019.

[37] M. Hlywa, R. Biddle, and A. S. Patrick, "Facing the facts about image type in recognition-based graphical passwords," in *Proceedings of the 27th annual computer security applications conference*, 2011, pp. 149–158.

[38] P. Mayer, M. Volkamer, and M. Kauer, "Authentication schemes-comparison and effective password spaces," in *Information Systems Security: 10th International Conference, ICISS 2014, Hyderabad, India, December 16-20, 2014, Proceedings 10*. Springer, 2014, pp. 204–225.

[39] A. Alghamdi, "Exploring the security landscape of ar/vr applications: A multi-dimensional perspective," Ph.D. dissertation, University of Central Florida, 2024.

[40] D. Adams, A. Bah, C. Barwulor, N. Musaby, K. Pitkin, and E. M. Redmiles, "Ethics emerging: the story of privacy and security perceptions in virtual reality," in *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, 2018, pp. 427–442.

[41] S. Kaur, S. Rajvanshi, and G. Kaur, "Privacy and security concerns with augmented reality/virtual reality: a systematic review," *Augmented Reality and Virtual Reality in Special Education*, pp. 209–231, 2024.

[42] J. O'Hagan, P. Saeghe, J. Gugenheimer, D. Medeiros, K. Marky, M. Khamis, and M. McGill, "Privacy-enhancing technology and everyday augmented reality: Understanding bystanders' varying needs for awareness and consent," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–35, 2023.

[43] J. A. De Guzman, K. Thilakarathna, and A. Seneviratne, "Security and privacy approaches in mixed reality: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1–37, 2019.

[44] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson, "Personal identifiability of user tracking data during observation of 360-degree vr video," *Scientific Reports*, vol. 10, no. 1, p. 17404, 2020.

[45] K. Lebeck, K. Ruth, T. Kohno, and F. Roesner, "Towards security and privacy for multi-user augmented reality: Foundations with end users," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 392–408.

[46] D. Davis, F. Monrose, M. K. Reiter *et al.*, "On user choice in graphical password schemes." in *USENIX security symposium*, vol. 13, 2004, pp. 11–11.

[47] A. Paivio and K. Csapo, "Picture superiority in free recall: Imagery or dual coding?" *Cognitive psychology*, vol. 5, no. 2, pp. 176–206, 1973.

[48] A. Paivio, T. B. Rogers, and P. C. Smythe, "Why are pictures easier to recall than words?" *Psychonomic Science*, vol. 11, no. 4, pp. 137–138, 1968.

[49] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[50] R Core Team, "R Project for Statistical Computing," https://www.r-project.org/, accessed on May 4, 2024.

[51] S.-H. Kim, J.-W. Kim, S.-Y. Kim, and H.-G. Cho, "A new shoulder-surfing resistant password for mobile environments," in *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, 2011, pp. 1–8.