Evaluating Voxel-Based Graphical Passwords for Virtual Reality

Prashant Rawat[*] Dalhousie University
Rumeysa Turkmen[†] Kadir Has University
Chukwuemeka Nwagu [‡] Dalhousie University
Kissinger Sunday [§] Dalhousie University
Mayra Donaji Barrera Machuca [¶] Dalhousie University

Abstract

Previous work has proposed using voxel-based graphical passwords (VGPs) for Virtual Reality (VR) as a secure, easy-to-remember way to authenticate users. Moreover, eye-tracking technology adds another level of security, as it avoids observational threats when entering the password. However, previous work has yet to evaluate the user performance, usability, and memorability of different combinations of VGPs. In two user studies, we first identified the best combination of shape and volume for VGPs. Then, we compare 3D versus 2D VGPs. Our results show that a cube is the best shape regarding usability and user preference. We also identified that 2D VGPs are easier to remember than 3D VGPs, as shown by a higher password accuracy and lower error rate. Our results inform the implementation of VGPs and other graphical passwords in VR.

Index Terms: D.4.6 [Security and Protection]: Authentication—; H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—; H.5.2 [User Interfaces]: Evaluation/methodology—;

1 Introduction

Graphical passwords are authentication methods utilizing images [9], which makes them secure [8] and memorable [5]. These characteristics have made them popular for Virtual Reality (VR) authentication [14]. In this paper, we focus on one type of graphical passwords called voxel-based graphical passwords (VGPs), which are either 2D or 3D shapes made of cubes. See Fig. 1. These types of passwords allow users to select a combination of cubes in the shape of their password and use their color and position as memorability aids. Moreover, past work has proposed using eye-gaze as an input method for VGPs [18], which solves two problems with authentication in VR: 1) observation threats when entering the password by hiding the user eye-gaze behind the VR head-mounted display (HMD) [11], and 2) improving the interaction performance, as eye-gaze is an accurate and fast input method [16]. However, the effect of a VGP shape, type, and volume on user performance, memorability, and usability is unknown [8].

In this paper, we first evaluated the usability of different shapes and volumes of 3D VGPs and how they affect user preference. Our results show an interaction between shape and volume, where a cube with a small volume was the best shape regarding task completion time, selection accuracy, and rotations. We also found that users prefer a cube and star over a sphere and a pyramid. The second study focuses on understanding the user performance, preference, and memorability of VGPs by comparing 2D and 3D VGPs. Our results show that 2D VGPs provide the best user experience and are easier to remember than 3D VGPs, as shown by a higher password accuracy and lower error rate. These results extend previous work investigating authentication within VR using graphical passwords [1, 14]. Finally, we made recommendations for future designers of authentication methods for VR.

2 Related Work

2.1 Authentication Methods for VR

There is a wide range of VR authentication methods, including knowledge-based [14], token-based [21], multifactor approaches [22], and graphical passwords [14, 18]. For graphical passwords, past work has suggested requiring users to select an area of a 3D shape [14, 18] or specific environmental elements [2].

Regarding evaluating the usability of graphical passwords for VR, Mathis et al. [14] found that interacting with a 3D cube had high authentication speed, resistance against observational threats, and seamless integration into established VR applications. Other work has studied the usability of cue-based authentication using graphical passwords [1]. Finally, other work have compared different input methods for authentication in VR [11]. Yet, we did not find an evaluation of different graphical passwords regarding user performance, usability and memorability.

2.2 Gaze User Interfaces

Previous work has found an advantage of using eye-gaze for authentication [11, 19]. For example, De Luca et al. [6] assessed three eye gaze interaction techniques for PIN entry and showed that incorporating eye gaze increased accuracy and enhanced the overall usability of their system. Yet, they recommended further research in this domain. VoxAuth [18] is a VGP that uses eye gaze as an input method. Yet, the authors did not evaluate their system. Here, we extend these past works by assessing the user performance and memorability of VGPs.

3 Motivation & Research Questions

We aim to understand better the use of VGPs for authentication in VR, as their impact on memorability remains unexplored [1]. For VGPs, it is also unclear the differences in usability between shapes and volume regarding user preference and performance [12]. Thus, our research questions are: **RQ1** *what is the best volumetric shape for 3D VGPs?* **RQ2** *does 3D VGPs improve user performance and memorability over 2D VGPs?* And **RQ3** *how do 3D VGPs affect the user experience?* By answering RQ1, we identify how different shapes and volumes affect their usability and user preference. RQ2 and RQ3 help us identify the differences in user performance, memorability and preference amongst 2D VGPs and 3D VGPs.

4 User Study 1

4.1 Voxel-based Graphical Password System

We adapted a 3D VGP called VoxAuth [18] to include four shapes and three volume levels. Past work found that familiarity helped with the memorability of a shape [15], which is why we chose a pyramid, a sphere, a star, and a cube. Each shape had three different volume levels, i.e., the number of voxels making up the shape: easy,

[*] e-mail: prashant.rawat@dal.ca
[†] e-mail:rumeysa.turkmen@stu.khas.edu.tr
[‡] e-mail:cnwagu@dal.ca
[§] e-mail:kissinger.sunday@dal.ca
[¶] e-mail:mbarrera@dadl.ca

medium, and hard. The number of voxels in a shape influences the password length, and by evaluating three volume levels, we can understand how the number of voxels in a shape affects its usability. The voxel count increases consistently from easy to hard, but each shape has a different number of voxels to preserve its distinctive features. Keeping the distinctive features of each shape helped with memorability, as the unique features are always noticeable. Finally, each voxel face had a different color-blind accessible color to help identify the different faces. Fig. 1 show all the 3D VGPs combinations used in the study with the number of voxels

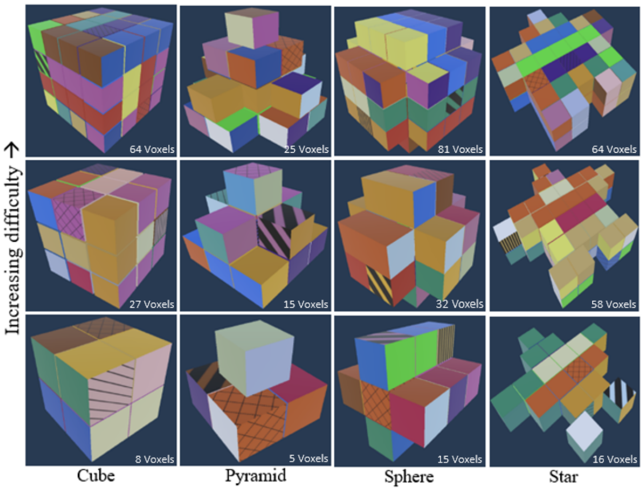


Figure 1: The voxel-based shapes. The participants select the voxels with non-solid textures in a specified order.

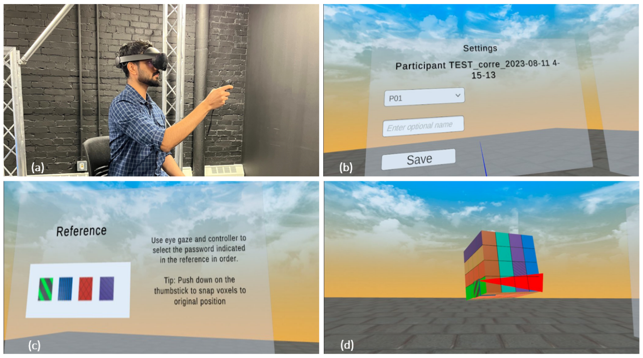
We developed the system using Unity 2021.3.16f1. We used the Meta Quest Pro with its controllers as the VR headset.

4.2 Methodology

Participants We recruited 12 participants (7 female, 5 male) from the local university. Their ages ranged from 18 to 35 (M=23.416, STD=4.733). No participant suffers from color blindness. Regarding their experience with graphical passwords, five were reported as beginners, two as novices, four as intermediates, and only one as an expert. Yet, no participant had used VGPs before. One participant had no VR experience, two experienced VR 1-3 times, and the rest experienced VR more than five times.

Experimental Design We designed a two-factor within-subject study with four **Shapes** (4_S = cube, pyramid, sphere, and star) and three **Volume levels** (3_{VL} = easy, medium, hard). For password length, all passwords were four voxels long. Each participant performed four shapes with three volume levels ($4_S \times 3_{VL}$ = 12 conditions (12_{Co}) each) with three repetitions (3_{rep}), which resulted in ($12_{Co} \times 12_{part} \times 3_{rep}$) 432 tasks in total. The order of conditions across within-subject dimensions was counter-balanced using a Latin Square.

Procedure The experiment happened in a noise-free room. Upon arrival, the participants signed the consent form and underwent the Ishihara Color-blindness test. During the study, participants remained seated and used the Quest Pro Controller with their dominant hand. See Fig. 2 (a). Participants had to point to a specific voxel using their eye-gaze and select them by pressing a button on the controller, see Fig. 2 (a) and (d). Participants were also allowed to rotate the shape with the thumbstick as needed. All 12 conditions (Fig. 1) had a fixed password length of four voxels, which participants had to select in a specific order. Participants had to select the voxels with the correct textures in the shape in the same order as the reference image (Fig. 2 (c)). The participants could not unselect a voxel once selected to understand each condition's selection difficulty better. After choosing all 12 passwords, the participants rested for two minutes between each repetition. Finally, after completing all the tasks, participants completed the post-experiment questionnaires.



Settings
Participant TEST_corre_2023-08-11 4-15-13
P01
Enter optional name

Reference

Use eye gaze and controller to select the password indicated in the reference in order.

Tip: Push down on the thumbstick to snap voxels to original position

Figure 2: (a) Participant performing experiment, (b) Participant settings, (c) Reference screen showing voxel images in order for password selection, and (d) Eye-gaze pointing on 3D VGP

4.3 Evaluation Metrics

We averaged the three repetitions by condition to get the final data. Next, we describe the data collected and its use:

- **Task Completion Time:** We logged the time when the user starts and finishes a new condition, which measures the user performance, e.g., how fast they found all the voxels.

- **Selection Accuracy:** We considered whether participants selected the correct voxels in the proper order. Accuracy was calculated as 100% when the participants selected all voxels correctly. For every mistake, we deducted 25%, as participants could only select four voxels.

- **Error Rate:** We considered whether the system granted access, e.g. if participants selected the correct voxels in the correct order. If the system gave access, the error rate was calculated as 0% and 100% if not. We use this info to understand the difficulty of correct authentication using 3D VGPs.

- **User Click Accuracy:** We recorded the number of times the user pressed the trigger button. If participants made four selections, click accuracy is 100%. For each extra voxel selection, click accuracy is reduced by 25%. We use the user click to identify eye-gaze accuracy.

- **Rotation:** We recorded the number of times the user rotated the shape in a task. We use rotations to identify how easy it was to find the correct voxel, e.g., more rotations mean the participant had more trouble finding the voxel.

- **User Experience:** Participants answered a post-study survey with open questions, where we asked their preferences and reasons for them.

4.4 Results

The data were analyzed using Repeated Measures (RM) ANOVA in SPSS 27. To check the normality of data, Skewness (S) and Kurtosis (K) were used where S and K values are within the range of ±1 [10]. None of the data was normally distributed, so we used aligned rank transformation (ART) [20] to normalize data.

Table 1: Two-Way RM ANOVA results for study 1. Grey background shows a statistically significant result.

	Shape (S)	Vol. Level (VL)	SxVL
Task Completion Time	F(3,33) = 5.167, p = 0.005, η^2 = 0.320	F(2,22) = 0.119, p = 0.888, η^2 = 0.011	F(6,66) = 6.049, p < 0.001, η^2 = 0.355
Selection Accuracy	F(3,33) = 2.522, p = 0.075, η^2 = 0.186	F(2,22) = 1.676, p = 0.21, η^2 = 0.132	F(6,66) = 7.692, p < 0.001, η^2 = 0.412
Error Rate	F(3,33) = 4.462, p = 0.01, η^2 = 0.289	F(2,22) = 0.310, p = 0.737, η^2 = 0.027	F(6,66) = 1.742, p = 0.125, η^2 = 0.137
User Click Accuracy	F(3,33) = 0.668, p = 0.578, η^2 = 0.57	F(2,22) = 3.138, p = 0.063, η^2 = 0.222	F(6,66) = 2.537, p = 0.029, η^2 = 0.187
Rotation	F(3,33) = 1.490, p = 0.235, η^2 = 0.119	F(2,22) = 8.753, p = 0.002, η^2 = 0.443	F(6,66) = 12.828, p < 0.001, η^2 = 0.538

Task Completion Time The results show significant differences in task completion time for shapes but not volume level. A post hoc analysis shows that participants were faster with the cube than the star (p = 0.005) and pyramid than the star (p = 0.046). See Fig. 3(a). We also found a statistically different interaction between shape and volume level. Our results show that for the cube, participants were faster with the easy ($p < 0.05$) and medium ($p < 0.05$) volume than the hard volume. For the sphere, participants were faster with the medium than with the hard ($p < 0.05$) volume. Lastly, for the star, participants were faster with the hard than the medium volume ($p < 0.05$). See Fig. 3(d) for the results.

Selection Accuracy The results show a significant interaction between shape and volume level. A post hoc analysis shows that for the cube, password accuracy is higher with the easy volume than the hard volume ($p < 0.05$). Also, for the pyramid, accuracy is higher with the easy than the medium volume ($p < 0.001$) and than the hard volume ($p < 0.05$). See Fig. 3(e) for the results.

Error Rate There is a significant difference between shapes in terms of error rate. A post hoc analysis shows that the error rate is less with the sphere than the star (p=0.035) (Fig. 3(b)).

User Click Accuracy There is a significant interaction between shape and volume level. For the cube, click accuracy is higher with the easy volume than the medium ($p < 0.05$) and than the hard ($p <$ 0.05). For the pyramid, click accuracy is higher with the easy volume ($p < 0.05$) and the hard volume ($p < 0.05$) than the medium volume. For the sphere, click accuracy is higher with the easy volume than the hard volume ($p < 0.001$). Lastly, for the star, click accuracy is higher with the easy volume than the hard volume ($p < 0.05$). See Fig. 3(f) for the results.

Rotation Count There is a statistical difference between volume levels, but not for shapes. The post hoc results show that participants rotated more on easy than on medium volume level (p = 0.003). The interaction between shape and volume level also shows a statistical difference. For the cube, participants rotated more with the easy volume than the medium volume ($p < 0.05$) and the hard volume than the medium volume ($p < 0.05$). For the sphere, participants rotated more with the easy volume than the medium volume ($p < 0.05$) and hard volume than the medium volume (p < 0.001). Finally, for the star, participants rotated more with the medium volume than the easy volume ($p < 0.05$) and than the hard volume ($p < 0.001$). See Fig. 3(g) for the results.

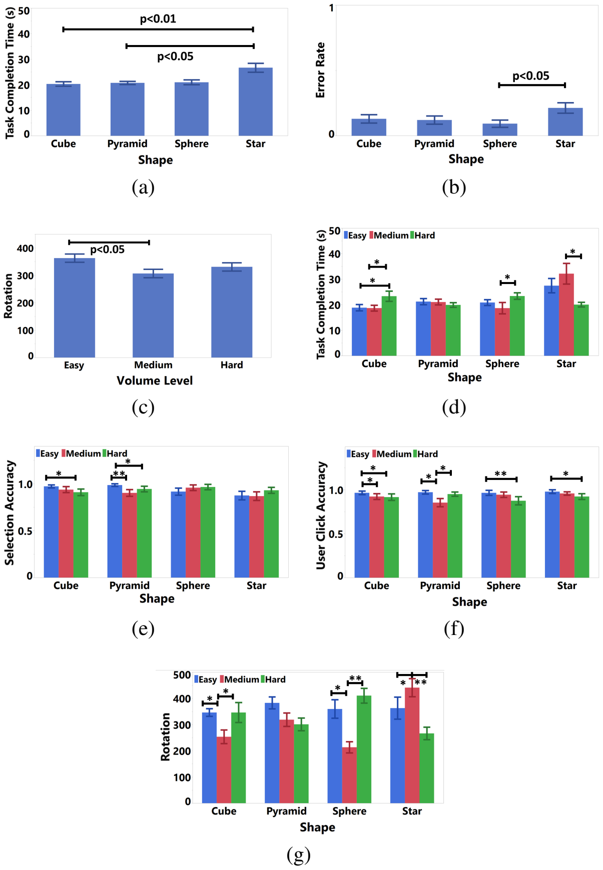


Figure 3: Two-way RM ANOVA results for study 1 for a) task completion time for shapes, b) error rate for shapes, and c) rotations for volume levels. Interactions between shape and volume level for d) task completion time, e) selection accuracy, f) user click accuracy, and g) rotations. * indicates p<0.05 significance level while ** indicates p<0.001 significance level.

User Experience: Four participants preferred the cube, four preferred the star, one preferred the pyramid, and three preferred the sphere. The participants who preferred cube commented *"Less complex as a shape."* and *"It is easy to focus on the cube."* Participant who preferred pyramid commented: *"...I was able to select given references faster"*. Participants who preferred sphere commented: *"...it was not hard as star and I turned it around easily"*, and *"...it was easy and comfortable to use..."*. Lastly, those who preferred star commented: *"...I would like to prefer hardest one to create strong password"* and *"...It seemed like a complex password for anyone to be able to crack into when I set it"*.]

4.5 Discussion

We evaluated different shapes and volume levels of 3D VGPs to better understand their usability and user preference.

For **usability**, we measured user performance in terms of time and rotations to gauge the ease of locating voxels. We found that simpler shapes (cube and pyramid) outperformed complex ones (star). This result aligns with prior research [13] and extends it by highlighting a non-linear interaction between volume and shape. The medium volume level was the slowest for some shapes (star), fastest for others (sphere), and showed no difference for specific shapes (pyramid). The rotation interaction results further verify the interaction between volume and shape. The position and accessibility of voxels within shapes likely explain these differences. For instance, increased volume in the pyramid added height, while in the star, it introduced more details to the sides. Participants executed more rotations in the easy volume, possibly due to its fewer salient features, hindering memorization, as noted in previous studies [1,14]. We also measured **user performance** in voxel selection using accuracy, error rate, and user click accuracy. Despite variations in voxel numbers across shapes and difficulty levels, no shape effect was observed. However, an interaction between shape and volume emerged. Easy volume exhibited higher user click accuracy than medium and difficult levels. These findings align with 3D selection, emphasizing the impact of object density on accuracy [3]. When looking at selection accuracy more in-depth, we found an interaction for the cube and the pyramid, where the easy level performed better than the medium level but no interaction for the sphere and the star. These results show the importance of using shapes with salient elements (sphere and star) as 3D VGPs, as the volume influences them less. Finally, for **user preference**, the cube and the star were the most preferred shapes. Interestingly, the participants who chose the cube focused on their user experience and how easy it was to select the voxels. On the other hand, the participants who selected the star focused on the feeling of security, as the shape made them believe it was more secure.

In conclusion, the cube was the best shape for 3D VGPs, which answer *RQ1*. Our results also show that when selecting the shape of 3D VGPs, it is important to consider how the number of voxels affects a shape's salient features. Yet, easy, symmetrical shapes like the cube offer the best trade-off between selection accuracy, speed, and user preference. On the other hand, more complex shapes like pyramids, spheres, and stars are more dependent on the interaction between volume and voxel positions.

5 User Study 2

5.1 Authentication Systems

We used the same VGP system as in study 1 (Sect. 4.1). For 3D VGPs, we used a cube and a star, as those were the shapes the user preferred most. For 2D VGPs, we used a square and a 2D star, as those are the same shapes as the 3D VGPs without depth. For the volume level, e.g., the number of voxels in shape, we chose the hard difficulty (64 voxels) because this condition provides high password variability without affecting user performance. Finally, we had two password difficulties based on password length: easy (4 voxels) and hard (6 voxels). We choose these lengths on spatial memory theory, where people cannot remember the location of seven objects at the same time [5].

The user selected the voxels in a similar way as in study 1. The only difference was that for the 3D VGPs, participants could rotate shapes using the controller thumbstick and reset them to their original pose by pushing down on the thumbstick, enabling exploration of all faces. For 2D VGPs, we turned off the rotation function.

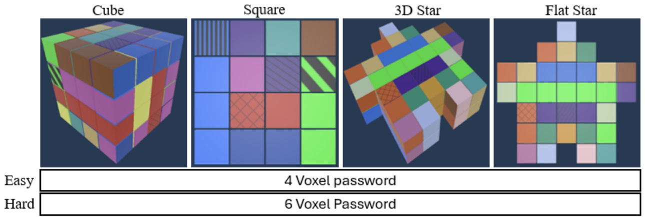


Figure 4: Experimental conditions for study 2.

5.2 Methodology

Participants: We recruited 20 participants (5 female, 15 male) from the local university. Their ages ranged from 19 to 47 (M=23.85, STD=5.98). Nineteen were right-handed, and one was left-handed. No participant suffers from color blindness. Regarding their experience with graphical passwords, seven reported as beginners, nine as novices, four as intermediates, and none reported as experts. No participant had used a VGP before.

Experimental Design We designed a two-factor within-subject study with four **Types** (4_T = 3D cube, square, 3D star, and 2D star), two **Difficulty Level** (2_{DF} = easy, hard). Each participant performed four types with two difficulty levels ($4_T \times 2_{DF}$ = 8 conditions (8_{Co}) each) with two repetitions (2_{rep}), which resulted in ($8_{Co} \times 20_{part} \times 2_{rep}$) 320 tasks in total. The order of conditions across within-subject dimensions was counter-balanced using a Latin Square.

Procedure The experiment procedure was similar to study 1 (Sect. 4.4). The difference lies in how the participant saw and memorized the passwords. For each condition, the participant inputs the password twice: the first time with the password shown in the reference screen (Fig. 2 (c)). After inputting the password from the reference, the participant had 30 seconds to memorize it. Then, they enter the password a second time from memory.

5.3 Evaluation Metrics

We logged task completion time, selection accuracy, error rate, user click accuracy, and user experience as described in Sect. 4.3. Additionally, to the open questions about the user experience, we collected the *System Usability Survey (SUS)* [4] to measure the usability of the system on a scale of 1 (strongly disagree) and 5 (strongly agree).

5.4 Results

The data were analyzed using three-way ANOVA in SPSS 27 for shape (cube and star), type (2D and 3D) and level (easy and hard). The data deviated from normal distribution except for the time variable, for which we applied a log transformation. The rest of the data was normalized by applying ART [20]. Statistical results are shown in Table 1. Results are shown in Table 2.

Task Completion Time We averaged time for each condition. The results show significant differences between types. A post hoc analysis shows that participants were faster with 2D VGPs than 3D VGPs ($p < 0.001$) (Fig. 5(a)). We also found that difficulty level significantly affects task completion time. The post hoc results indicate that participants were faster with the easy level than with the hard level ($p < 0.001$) (Fig. 5(b).

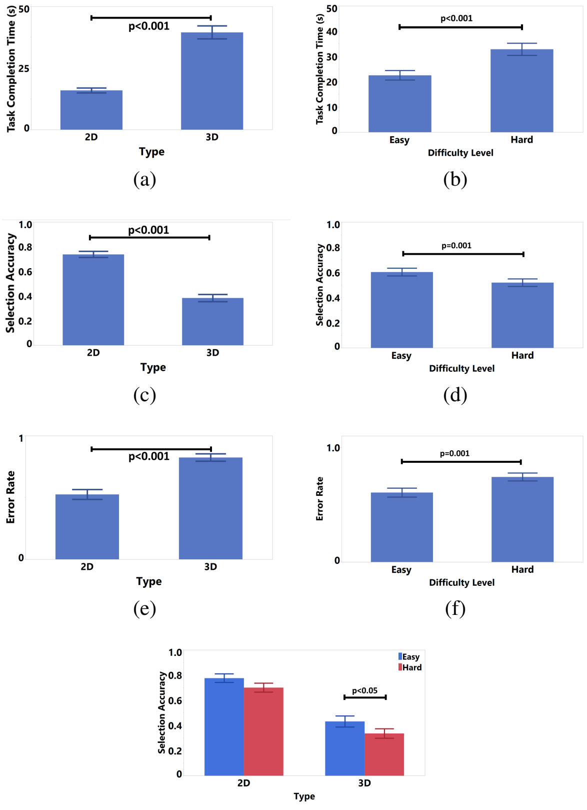
Selection Accuracy The results show a significant effect on type in terms of selection accuracy. A post hoc analysis shows that participants submitted more accurate passwords with 2D VGPs than 3D VGPs ($p < 0.001$)(Fig. 5(c)). The results also show a significant effect on difficulty level where the selection accuracy was higher with the easy level than the hard level ($p = 0.001$) (Fig. 5(d)). Lastly, results show a significant interaction between type and difficulty level. The post-hoc analysis indicates that participants submitted more accurate passwords with the easy level than the hard level in 3D VGPs ($p < 0.05$) (Fig. 5(g)).

Error Rate There is a significant difference between types in terms of error rate. The post hoc analysis shows that the error rate is less with 2D VGPs than with 3D VGPs ($p < 0.001$)(Fig. 5(e)). The results show a significant difference between difficulty levels. The error rate is less with the easy difficulty level than the hard one ($p < 0.001$) (Fig. 5(f)).

SUS: We asked participants to evaluate each method separately. We found that 2D VGPs received an excellent 'B' grade, and 3D VGPs received a good 'D' grade.

Table 2: Three-Way ANOVA Results for Shape, Type and Difficulty Level. Grey background shows statistically significant results

	Shape (S)	Type (T)	Difficulty Level (DL)	SxT	SxDL	TxDL	SxTxDL
Task Completion Time	F(1,19)=0.001, p = 0.983, $\eta^2 = 0.001$	F(1,19)=86.917, $p < 0.001$, $\eta^2 = 0.821$	F(1,19)=59.283, $p < 0.001$, $\eta^2 = 0.757$	F(1,19)=1.217, p = 0.284, $\eta^2 = 0.060$	F(1,19)=1.212, p = 0.311, $\eta^2 = 0.054$	F(1,19)=2.131, p = 0.161, $\eta^2 = 0.101$	F(1,19)=0.365, p = 0.553, $\eta^2 = 0.019$
Selection Accuracy	F(1,19)=1.312, p = 0.266, $\eta^2 = 0.065$	0F(1,19)=61.728, $p < 0.001$, $\eta^2 = 0.765$	F(1,19)=14.473, p = 0.001, $\eta^2 = 0.432$	F(1,19)=5.759, p = 0.083, $\eta^2 = 0.150$	F(1,19)=3.771, p = 0.067, $\eta^2 = 0.166$	F(1,19)=6.362, $p < 0.05$, $\eta^2 = 0.251$	F(1,19)=0.002, p = 0.968, $\eta^2 = 0.0$
Error Rate	F(1,19)=1.237, p = 0.280, $\eta^2 = 0.061$	F(1,19)=27.097, $p < 0.001$, $\eta^2 = 0.588$	F(1,19)=14.573, p = 0.001, $\eta^2 = 0.434$	F(1,19)=0.374, p = 0.548, $\eta^2 = 0.019$	F(1,19)=3.669, p = 0.071, $\eta^2 = 0.162$	F(1,19)=2.118, p = 0.162, $\eta^2 = 0.1$	F(1,19)=2.078, p = 0.166, $\eta^2 = 0.99$
User Click Accuracy	F(1, 19)=0.274, p = 0.607, $\eta^2 = 0.014$	F(1, 19)=0.557, p = 0.465, $\eta^2 = 0.028$	F(1, 19)=1.861, p = 0.188, $\eta^2 = 0.089$	F(1, 19)=0.840, p = 0.371, $\eta^2 = 0.042$	F(1,19)=7.331, p = 0.5, $\eta^2 = 0.0278$	F(1,19)=2.811, p = 0.11, $\eta^2 = 0.128$	F(1, 19)=0.588, p = 0.452, $\eta^2 = 0.03$



(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 5: Three-way RM ANOVA results for the shape, type and difficulty level analysis: a) task completion time results for type, b) task completion time results for difficulty level, c) selection accuracy results for type, d) selection accuracy results for difficulty level, e) error rate results for type, f) error rate results for difficulty level and g) selection accuracy interaction results between type and difficulty level.

User Experience: We asked participants which password method they would prefer and why. Fourteen preferred 2D VGPs, and the rest preferred 3D VGPs. Participants who preferred 2D VGPs commented: *"2D method was easy to understand, remember, and use..."* and *"...I was able to find the colors quickly."* Participants who preferred 3D VGPs commented: *"It was hard for anyone to guess as there are numerous possibilities."* and *"... It is hard to be broken."*. Lastly, we asked participants what they liked and disliked for each password method. The participants commented that they liked the 3D VPGs as *"difficult to crack"* and *"difficult for others to remember"*. They also commented: *"I didn't like the amount of time I was given to memorize the sequence"* as reasons for disliking the 3D VPGs. Participants commented: *"Simple to use"* as the reason for liking 2D VPGs. On the other hand, they stated: *"Not as detailed as 3D"* for what they disliked about 2D VPGs.

5.5 Discussion

Study 2 compares 3D and 2D VGPs to understand user performance, preference, and memorability. Next, we discuss the insight derived from this study:

When looking at the **user performance**, we analyzed task completion time and click accuracy. Our result shows that participants exhibited quicker password entry times when using 2D VGPs than 3D VGPs. Yet, the time difference (23.2 s) can be explained as the time needed to rotate the shape. Easier passwords were also faster and more accurate due to the time it takes to enter all characters.

We also focus on the **memorability**. We analyzed the password accuracy and the error rate. The results show higher accuracy and lower error rates when using the 2D VGPs than 3D VGPs. These results show that 2D VGPs do not suffer the steep learning curve compared to other graphical passwords [2], as users could have a higher performance and memorability with 2D VGPs than 3D VGPs. Yet, 3D VGPs are affected by the need to manipulate them to identify the correct voxel, which requires a higher spatial ability [7] and increases performance time.

Finally, we evaluated the **user experience**. 14 participants preferred 2D VGPs, rating them as excellent due to their strong memorability and usability. Specifically, 2D shapes are flat and exist on a single plane [17], which helps users avoid confusion when recalling previously entered passwords. In summary, the answer to *RQ2* is that 2D VGPs improve user performance over 3D VGPs, and the answer to *RQ3* is that 2D VGPs improve the user experience.

6 Recommendations for Graphical Passwords in VR Authentication

This work investigated various factors influencing usability, user preference, and memorability of VGPs, including shape, volume, and password length. Our results contribute insights to designing and implementing user-friendly authentication systems for VR applications with graphical passwords.

- **Graphical Password Shapes**: Study 1 showed that simpler graphical passwords are more practical and usable than complex ones (cube than the star). Yet, we also identified the importance of utilizing shapes with salient elements, as that might prevent some of the issues with complex shapes (larger volumes), e.g., reduced performance, and make it easy to remember which element to select. Future VR application designers should consider a wide range of shapes and volumes for VGPs, but if simplicity is required, the cube is a good shape.

- **Graphical Password Memorability**: Graphical passwords have demonstrated advantages that enhance memorability, especially when using personalized objects [15]. Yet, the results of Study 2 show that increasing the password length from 4 to 6 voxels decreased user performance, as demonstrated by a higher task completion time and error rate. Based on these findings, we recommend future VR application designers keep the password length as short as possible and add other elements to increase password complexity, like having to identify the shape.

- **Using 3D Graphical Passwords**: Study 2 found that 2D VGPs are overall better than 3D VGPs, regardless of the shape used. For example, 2D VGPS were faster, with higher selection accuracy and less error rate than 3D VGPs. Moreover, 2D VGPs got a B SUS grade, whereas 3D VGPs got a D. Finally, participants preferred 2D VGPs over 3D VGPS. These results, together with previous work on the importance of spatial abilities for 3D manipulations [7], show that even for VR authentication methods, 2D VGPs are better than 3D VGPs. We recommend future VR application designers use 2D graphical passwords for VR authentication. If using 3D graphical passwords, it is important to incorporate ways to reduce the spatial memory needs of the user.

- **Eye Gaze as an Input Method for Graphical Passwords in VR**: Our two studies identified that the user click accuracy, e.g., if they could point to the desired voxel, follow the same rules of 3D pointing, e.g., a trade-off between the size of an element and the selection time [3]. Yet, we also identified that user click accuracy is higher with 2D VGPs than with 3D VGPs due to the ability to have fixed voxels. Based on these results and the advantages of using eye-gaze for authentication methods in VR identified by previous work [13], we recommend using eye-gaze as an input method for graphical passwords in VR.

7 Limitations

This paper compared a limited number of shapes (cube, pyramid, sphere, and star). Moreover, each shape had a different number of voxels, as standardizing them made the shapes unrecognizable. Future work should evaluate more shapes and shapes with the same number of voxels to verify our results. Future work should also evaluate the security aspects of VGPs, as our study focused on user performance and usability. Lastly, the study primarily examined short-term user interactions and performance for authentication, while we did not explore long-term usability and adaptation. Future research may investigate how users adapt to VGPs over prolonged periods.

8 Conclusion

In this paper, we evaluated VGPs in VR using eye-gaze as the input method. Our results enable us to gain insights into user preferences for specific types of VGPs within VR environments; it also enhances our understanding of the types of VGPs that can significantly impact a user's ability to engage with VR-based authentication applications.

Acknowledgments

Thanks to Poppy Riddle for helping us design the user study.

References

[1] Y. Abdelrahman, F. Mathis, P. Knierim, A. Kettler, F. Alt, and M. Khamis. CueVR: Studying the Usability of Cue-based Authentication for Virtual Reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, pp. 1–9. ACM, Frascati, Rome Italy, June 2022.

[2] F. Alt, M. Mikusz, S. Schneegass, and A. Bulling. Memorability of cued-recall graphical passwords with saliency masks. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pp. 191–200, 2016.

[3] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers and Graphics*, 37(3):121–136, 2013.

[4] J. Brooke. SUS: a 'quick and dirty' usability scale. In *Usability Evaluation In Industry*. CRC Press, 1996.

[5] N. Cowan. George miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological review*, 122(3):536, 2015.

[6] A. De Luca, R. Weiss, and H. Drewes. Evaluation of eye-gaze interaction methods for security enhanced pin-entry. In *Proceedings of the 19th australasian conference on computer-human interaction: Entertaining user interfaces*, pp. 199–202, 2007.

[7] T. Drey, M. Montag, A. Vogt, N. Rixen, T. Seufert, S. Zander, M. Rietzler, and E. Rukzio. Investigating the effects of individual spatial abilities on virtual reality object manipulation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023.

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

[9] B. E. Fayyadh, K. Mansour, and K. W. Mahmoud. A new password authentication mechanism using 2d shapes. In *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, pp. 113–118. IEEE, 2018.

[10] J. F. Hair Jr, W. C. Black, B. J. Babin, and R. E. Anderson. Multivariate data analysis, 2014.

[11] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt. The role of eye gaze in security and privacy applications: Survey and future hci research directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–21, 2020.

[12] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–21. ACM, Honolulu HI USA, Apr. 2020.

[13] F. Mathis, J. Williamson, K. Vaniea, and M. Khamis. Rubikauth: Fast and secure authentication in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2020.

[14] 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)*, 28(1):1–44, 2021.

[15] M. Mohamed, T. Porterfield, and J. Chakraborty. Cross-cultural effects on graphical password memorability and design. *Journal of Systems and Information Technology*, 23(1):1328–7265, 2021.

[16] A. Murata and W. Karwowski. Automatic lock of cursor movement: Implications for an efficient eye-gaze input method for drag and menu selection. *IEEE Transactions on Human-Machine Systems*, 49(3):259–267, 2018.

[17] A. Sheffer and E. de Sturler. Parameterization of faceted surfaces for meshing using angle-based flattening. *Engineering with computers*, 17:326–337, 2001.

[18] R. Turkmen, C. Nwagu, P. Rawat, P. Riddle, K. Sunday, and M. B. Machuca. Put your glasses on: A voxel-based 3D authentication system in VR using eye-gaze. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 947–948. IEEE, Shanghai, China, Mar. 2023.

[19] J. Weaver, K. Mock, and B. Hoanca. Gaze-based password authentication through automatic clustering of gaze points. In *2011 IEEE international conference on systems, man, and cybernetics*, pp. 2749–2754. IEEE, 2011.

[20] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011.

[21] Z. Yu, H.-N. Liang, C. Fleming, and K. L. Man. An exploration of usable authentication mechanisms for virtual reality systems. In *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pp. 458–460. IEEE, 2016.

[22] H. Zhu, W. Jin, M. Xiao, S. Murali, and M. Li. Blinkey: A two-factor user authentication method for virtual reality devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–29, 2020.