Exploring the Design Space for Security Warnings in Immersive Environments

Andrea Mengascini*, Rebecca Weil *, Annabelle Walle*, Jürgen Steimle†, Giancarlo Pellegrino*
* *CISPA Helmholtz Center for Information Security Saarbrücken, Germany*
{andrea.mengascini, weil, walle, pellegrino}@cispa.de
†*Saarland University Saarbrücken, Germany*
steimle@cs.uni-saarland.de

***Abstract*—More and more immersive environments support third-party applications, leading to concerns about the trustworthiness of user interfaces (UIs), which attackers could exploit, endangering users. Although security warnings attempt to safeguard users by highlighting risks, most studies primarily target security indicators as usable intervention for app transitions in virtual reality. Our research broadens this focus by providing a systematic, data-driven investigation of security warnings for third-party applications in immersive environments. We analyzed a decade's worth of top VR interactions and security conference findings and assessed the top 10 free VR applications from two leading stores each. From our design process, we implemented four warnings. Through two user studies involving 61 participants, we measured their responses to these warnings during virtual object interactions. Our findings indicate that a red glow on an object was the most effective warning, frequently associated with danger, while pop-up warnings were the least effective.**

1. Introduction

The tremendous growth of immersive technology is reshaping the way users perceive and interact with application software: While traditional user interfaces (UIs) made of buttons and icons still exists, immersive environments increasingly leverage virtual reality objects that users interact via hand gestures in virtual or mixed reality environments. More immersive systems now support third-party applications [28], [65] that an attacker can use to superimpose visual elements or display harmful content [67], [71], [152], deceiving users into insecure interactions and resulting in, for example, password theft or malware download [152].

An existing approach to tackle UI trustworthiness concerns is using *security warnings*, which traditionally are trusted UI elements controlled by the system – that significantly reduce the incidence of insecure interactions by informing users about risks and introducing friction to encourage safer choices. In some cases, such as HTTPS warnings in browsers, blocking access outright is impractical; instead, warnings allow users to make informed decisions about proceeding despite potential risks. Prior work has extensively explored security warnings in web applications, investigating the design space, ranging from passive warnings like secret images embedded in HTML pages [29] to active interstitial messages breaking users' workflows [4], as well as their effectiveness when visiting malicious pages [4], [33] and those delivered over insecure connections [4], [42], [120], [136]. Only recently, the research community has started focusing on security warnings in immersive systems, proposing visual and haptic indicators to warn users about potential risks when interacting with static, well-defined, cross-applications UI elements, such as online ads [67] and Metaverse hyperlinks when moving between virtual applications [53], [158]. Unfortunately, recent works have shown that risks also exist when users interact with UI elements that are specific to the application logic, such as PIN pads, by placing a malicious object around a benign one to grab user input for click-jacking attacks [23], snooping on user interactions with invisible objects to steal secrets [23] or fingerprint users [98].

Research Questions In this paper, we focus on designing and testing security warnings to protect users when interacting with seemingly benign UI elements within an immersive environment. Our study operates within the VR space of the XR continuum [94], where the entire environment is recreated inside the headset. Since creating new warning is not a trivial task, we follow a data-driven approach to design warnings, drawing from academic literature and UI designs in real immersive applications, implementing those that can be supported by existing consumer-grade devices and cannot be easily bypassed by attackers. Then, we test the warnings' noticeability, comprehension (user study 1), and final effectiveness (user study 2). More specifically, in this paper, we address the following research questions:

- **RQ1: Design Space** What could be the designs of security warning systems in immersive environments? What designs are practically viable using consumer-grade devices?
- **RQ2: Initial User Response** How do users process and respond to immersive warnings?
- **RQ3: Behavioral Response** How effective are the considered designs in minimizing interactions with insecure objects?

Our Study Firstly, we conducted a comprehensive literature review to explore the design space of security warning systems in immersive environments (RQ1). We examined the past decade's literature from leading VR interaction conferences (ACM CHI and IEEE VR) and reviewed the top 20 free VR applications from the Oculus Quest Store [90] and Steam [133] to identify interaction modalities and feedback methods suitable for presenting warnings. We covered all works published in the top security and privacy and Human-Computer Interaction (i.e., IEEE S&P, USENIX Security, ACM CCS, NDSS, ACM CHI, SOUPS, and PETS) conferences of the last ten years to identify effective warnings for potential hazards. By combining VR interactions, security advice from conferences, and our threat model, we designed secure warnings for VR systems. Secondly, to understand the initial user response (RQ2) and behavioral effectiveness (RQ3) of these designs, we conducted two different user studies. We recruited 61 participants to evaluate how users perceive the warnings and how these warnings influence their hazard-related attitudes, beliefs, and behavior.

Insights From 99 papers, we identified the possible interaction modalities and warning design strategies, which led to the identification of 11 potential designs. However, only three of these designs are viable as they are not trivially vulnerable to attacks (e.g., spoofing or tampering) and can be implemented using consumer-grade VR devices. From these potential designs, we derived four concrete warnings: 1) a red glow decorating insecure objects, 2) blurring the visual appearance of insecure objects, 3) a pop-up system alert, and 4) altering the physics of potentially malicious objects by scaling them down. Our results show that red glow warnings were effective as many linked red to danger, but some mistook it for a guidance VR feature. Our physics-based alerts were noticeable but caused users stress by hastening their decision-making, reducing their effectiveness. Blurred objects reduced interaction yet were rarely seen as warnings. Pop-ups had mixed reactions, suggesting the need for clearer messaging. Our findings have implications for exploration and iteration in designing immersive security warnings.

Contributions We contribute the following:

1) A comprehensive **survey of the last decade's Human-Computer Interaction (HCI) and security research**, aimed at a) systematizing VR interaction and feedback modalities for warnings, and b) identifying best practices in warning design to enhance noticeability, comprehension, and effectiveness (§ 4.1).
2) Based on these properties, **the design of four distinct warnings** for both well-defined traditional UIs and more immersive UIs within the scene (§ 4.2)
3) Via two user studies with 61 participants, an **evaluation of noticeability, comprehension, and overall effectiveness** of the designed warnings (§ 5 and 6).
4) **Design strategies for warnings** in immersive systems based on our comprehensive survey, design, and analysis. For example, we encourage future work to explore both a combination of proposed warnings and the usage of context-specific warnings.

2. Background

Before we present our study, we briefly review the assumptions about the immersive system for this work.

2.1. System Model

We assume a system where multiple applications or sources can simultaneously output in the same scene. The output of these applications can be either a 2D canvas, an object, or a semi-transparent filter. Such a multi-object/application model is in line with prior works [66], [67]. In addition, we assume that our system contains a trusted output module that manages all output presented to the user. The output module is the only way an application can generate output. This model can deny and restrict the output or overlay feedback (audio or visual) to any application. This concept aligns with established frameworks where this output module resides at the system level, making it a part of the operating system [66], [67].

2.2. Threat Model and Attacks

We assume the attacker controls third-party code loaded into a benign app, such as malicious advertisements placed within the scene (see [67]). Alternatively, the attacker may control a malicious app that has been installed on the user's system or exploited a vulnerability in an already installed benign app. In either scenario, the attacker aims to deceive the user with 3D objects and induce them to perform unwanted operations or interactions. Such interactions can result in different type of attacks. For example, the attacker could mount *phishing attacks* by mimicking a system UI element, such as a PIN pad, and presenting it as part of a legitimate operation—like entering a PIN to confirm a transaction in another app. Deceived by its realistic appearance, the victim unknowingly enters their PIN, thereby compromising their sensitive data (see [23], [163]).

Phishing attacks are just one example; other attacks are also possible. For instance, the attacker can *fingerprint* and *(re)identify* users by inserting objects into scenes and recording victim interaction patterns (see [98], [99]). Another type of attack involves *malvertisement* and *ad fraud*, where the attacker exploits the system's inability to clearly signal the origin of third-party VR entities. This leads to the propagation of deceptive and malicious advertising (see [67]). Users who interact with these fraudulent ads inadvertently contribute to the financial gain of the attacker, resulting in personal data exploitation and potential monetary losses.

2.3. Security Warning Activation

Outright blocking content or applications may not always be practical, as it can hinder usability and limit user access to desired content. Instead, many existing systems employ cues to alert users to potential risks, enabling informed decision-making without fully restricting access. For instance, browsers use HTTPS warnings [33] to inform users about unsecure connections rather than blocking access outright. Similarly, warnings can indicate if an app is downloaded from unknown sources [47] or if it begins accessing new permissions [9], allowing users to stay vigilant. This paper assumes such an activation mechanism is in place and instead focuses on exploring the design space and effectiveness of the associated immersive security warnings.

In immersive environments, warnings play a crucial role in informing users about potential risks, enhancing awareness, and prompting reconsideration of risky actions [3]. They can mitigate threats from untrusted or malicious objects by alerting users when engaging with

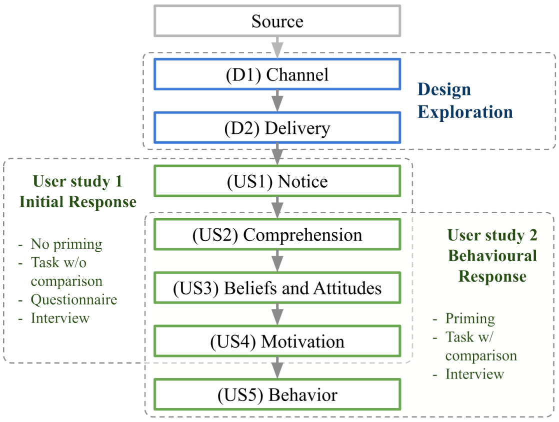


Figure 1: Our methodology mapped to the C-HIP model.

unverified interfaces, such as entering credentials into a spoofed UI or interacting with deceptive overlays.

3. Methodology

We address the challenges of designing and testing security warnings to protect users from seemingly benign UI elements within an immersive environment. The first challenge was identifying **suitable methods and output channels** for delivering warnings in these immersive interactions (RQ1). We conducted two surveys, initially analyzing literature and real-world VR applications to understand interaction and information delivery in VR. Subsequently, we drew upon security literature to identify best practices in warning presentations. We then set to resolve other challenges with the design of our warnings. As the VR experiences are full of different stimuli, a key challenge was ensuring the **noticeability** of the warnings among all the other stimuli. Another critical aspect was ensuring users **comprehended** the risks when presented with the warnings. Thus, our design efforts concentrated on creating noticeable and comprehensible warnings. To check the result of our design process, we tested the designed warnings' noticeability, comprehension (RQ2), and final effectiveness (RQ3) with two user studies. We structure our methodology using the Communication-Human Information Processing (C-HIP) model [160], a multi-step model to structure the warning research [25], [159], and map the steps to one of our research questions.

C-HIP Model The C-HIP model, sketched in Fig. 1, offers a multi-stage, structured approach to understand how users perceive and respond to warnings. The first stage of the model focuses on the delivery of a warning from a *source*, e.g., an operating system or browser, to the user, using a *channel*, e.g., an image or a text message, that the user receives through an interface, i.e., *delivery*. The second stage starts once the user receives the warning and is divided in five steps. It starts with the warning attracting attention, i.e., *notice*, as it is competing with other elements on the same UI. Then, the message of a warning needs to be understood, i.e., *comprehension*, and, to be successful, it needs to concur with the user's *beliefs and attitudes*. Another key step for successful warnings is about the trade-off between the compliance and non-compliance costs (e.g., time, effort, and stress), defining the user's *motivation*. The last step of the warning processing is to alter the user's *behavior*, e.g., executing a secure action or avoiding an insecure one.

Methodology Overview Overall, we organize our study as follows. First, we identify possible designs of immersive warnings, covering the *channel* and *delivery* steps of the model. Here, we follow a data-driven approach by looking at prior work in VR interactions, modalities, feedback methods, and security warnings (§ 3.1), answering RQ1. Then, we evaluate the identified warnings, covering the individual's information processing pipeline with two user studies (§ 3.2), answering respectively RQ2 and RQ3. The studies are designed to be independent and non-sequential, and we present them in the same order as the C-HIP model for coherence.

3.1. RQ1: Design Exploration

In our methodology's initial phase, we identified suitable channels and delivery methods for VR warnings, analyzing VR interaction modalities based on the C-HIP model's "D1 – Channel" component and existing feedback responses ("D2 – Delivery"). We then reviewed past security warning literature for guidelines on noticeability, comprehension and other aspect to increase the overall effectiveness. Lastly, we integrated potential information presentation methods with insights from existing warnings, aiming to create effective, feasible warnings for current head-mounted displays (HMDs) that are resistant to spoofing by attackers.

Systematization of VR Interactions We conducted a survey of ACM CHI and IEEE Virtual Reality publications from the past decade, focusing on VR interaction technologies, excluding papers unrelated to interaction methods (e.g., locomotion, surgery). Using a collaboratively developed codebook [74], two researchers categorized interaction modalities and feedback designs across academic research and popular VR applications from the Oculus Quest Store [90] and Steam [133]. We coded interaction modalities from literature and applications based on input and output modalities (D1). One researcher categorized studies, while a second conducted quality assurance to refine and group similar codes with disagreements resolved through discussions to reach a consensus. The same approach was applied across both academic literature and VR applications, ensuring consistency. Core application experiences were observed for 15 minutes (excluding setup and menus) using the same codebook (§ A.2).

Systematization of Prior Warnings We reviewed the proceedings of the top Security and HCI conferences in the last ten years (including USENIX Security, NDSS, IEEE S&P, ACM CCS, CHI, PETS, and SOUPS) for effective warning design ideas and guidelines, selecting them as key venues at the intersection of security and usability. Using relevant keywords (full queries listed in § A.1), we focused on papers specifically discussing and studying warning designs. We included only work that presented new warning designs, evaluated existing warnings, proposed improvements, identified flaws, or provided design recommendations. Works that did not focus on warnings, e.g., addressing only the underlying systems managing warnings or focusing solely on user training, were excluded. A codebook [74] guided the extraction and categorization of warning designs into recommendations that designers can use to provide better warning delivery (D2). This involved a two-step process where one researcher initially identified relevant papers and extracted key design features and their impacts, categorizing them based on their implementation, with additional categories for discrete warning attributes like color or tone. A second researcher then conducted quality assurance checks to ensure consistency in categorization and relevance to warning efficacy, with disagreements resolved through discussions to reach a consensus. The result can be seen in Tab. 3.

Designs of Secure Immersive Warnings After a comprehensive review of existing interaction modalities and security warnings, we formulated guiding principles for designing warnings in VR, ensuring the utilization of all available feedback methods. These rules directed us to investigate immersivity levels, feedback mechanisms, and to ensure universal applicability across VR devices. Lastly, our rules guided us to avoid known design pitfalls from past warnings. Details on these principles are in § 4.2.

3.2. RQ2-3: Initial and Behavioral User Response

We conducted two user studies to evaluate users' initial and behavioral responses to warnings in a VR environment when interacting with **application-logic specific UI elements**, with each study addressing different stages of the C-HIP model. The distinct objectives of each study required different methodological approaches, allowing us to explore separate aspects of warning perception and decision-making while maintaining consistency in objects and tasks for comparison.

3.2.1. User Study 1: Initial User Response. In the first study, we evaluated the first four stages of the C-HIP model —from "(US1) Notice" to "(US4) Motivation"— using a user task, followed by a questionnaire and an interview. In the user task, participants interacted with VR objects with and without the warnings to assess their impact on task performance and noticeability. Specifically, we examined how warning condition influenced interaction times, with longer interactions possibly indicating friction or hesitation. As it is described in the subsequent part of the paper, due to a smaller-than-expected participant size, we relied on general trends rather than statistical tests for this analysis. The questionnaire, inspired by SUS [20] and NASA-TLX [49], assessed workload, usability, and perceived risk. We investigated whether our warnings increase workload and usability friction and decrease the users' sense of security and motivation to interact with objects, as a successful warning should, compared to the no-warning condition. To analyze these effects, we used linear mixed models with warning as a fixed effect and a maximal random structure to account for dependencies across participants, items, and objects. The interview explored warning noticeability, user reactions, and beliefs about VR security and assessed participants' perceptions of malicious intent behind warnings and their possible motivation to avoid interaction.

3.2.2. User Study 2: Behavioral Response. The second study explored the latter stages of the C-HIP model, from "(US2) Comprehension" to "(US5) Behaviour". It investigated users' decision-making in assessing VR object security and collected feedback on designed warnings. The study was divided into a task and an interview. Participants were exposed to all warnings individually, using the same objects and tasks from the first study. This setup allowed us to assess their effectiveness in minimizing interactions with insecure objects while also understanding why certain warnings succeeded or failed. To limit free exploration and prevent potential confounds from repeated exposure to the same scene, we presented the same tasks as in the first study but placed them individually in distinct environments. We examined differences in object interactions (yes/no) as a function of warning and object. To determine whether warnings reduced interaction and whether this effect varied by object, we conducted linear mixed model analyses, including warning, object, and their interaction as fixed effects, with object interaction as the outcome variable. Estimated marginal means were used to further analyze these interactions. Post-experiment interview let participants elaborate on their experiences with each warning, focusing on comprehension, motivation, and behavior. Additionally, it assessed participants' perceptions of malicious intent, as they were explicitly aware that some objects posed security risks.

3.2.3. Data Collection and Metrics. To determine the success of our warning system in the user study, we focus on several essential metrics. These include evaluating the system's noticeability, the caution it instills in users, its impact on interactions, and overall behavioral influence. We aimed to thoroughly evaluate the warning's performance against its goals through quantitative and qualitative data from interactions and feedback.

Interaction Data Throughout the VR experience, we collected timestamped events such as when the user interacts with objects and state changes (e.g., warning on/off, start/stop action, and objects' position). We validated the experience by re-enacting users' actions with timestamped head and hand positions and rotations.

Interviews We audio-recorded interviews under the explicit consent of the participants for the sole purpose of transcribing and analyzing them.

3.2.4. Recruitment. We used both remote and in-person recruitment. Study 1 was conducted remotely via Prolific due to pandemic-related restrictions on laboratory access, while Study 2 was conducted in person once access was restored. Additionally, the Prolific participant pool was exhausted after Study 1, requiring in-person recruitment for Study 2. Participants were screened to meet the study requirements, and compensation was set above the local minimum wage for in-person participants. For remote participants via Prolific, we followed Prolific's payment principles guidelines, using their recommended payment [112]. Adjustments were made for commuting costs (in-person) and potential technical issues (remote). Tab. 15 and Tab. 16 in § C.3 provide a summary of participants' demographics. The full dataset is in the public repository.

3.2.5. Data Availability. The data collection script and processed data for statistical analysis, including interaction times, are available in an anonymized repository and will be permanently hosted on Zenodo. Raw data, such as controller and head movements, have been excluded to protect participant privacy and mitigate fingerprinting risks. The repository includes detailed instructions for re-running the analysis. The shared files can be accessed at: https://doi.org/10.5281/zenodo.15210478

3.3. Ethical Considerations

Our study involved human participants and was approved by our institute's ethical review board. To ensure privacy and security, all data were collected over a TLS-secured connection, anonymized for processing, and audio recordings of interviews were deleted after transcription, with any personally identifying information removed. To minimize safety risks, we ensured a controlled environment to reduce injury risks due to the VR headset covering both eyes and for seizures, we relied on the Quest's own warnings. For remote participants, the application was designed to be self-contained, requiring no installation, only internet access, and leaving no traces on personal devices. It could not be reused after the experiment. All participants provided informed consent before the study, with a post-study debriefing when full disclosure was initially limited. They could withdraw participation at any time, even post-debriefing, while still receiving compensation (via Amazon voucher or bank transfer).

4. RQ1: Design Exploration

In this section, we instantiate our methodology and present the designs we identified (RQ1). By examining interaction and feedback modalities, we assess potential warning channels (D1) and, drawing on insights from prior warnings, we develop approaches for VR warning delivery (D2). The output of this section is the list of designs to consider for the user studies.

4.1. Analysis of Prior and Existing Systems

4.1.1. Interaction Modalities and Feedback Mechanisms. From the past decade of ACM CHI and IEEE Virtual Reality publications (9,896 papers), we identified 243 relevant papers using keywords like `VR`, `virtual`, and `haptic`. After excluding non-relevant papers, we analyzed 73 papers. We also reviewed a total of 20 VR applications for 15 minutes each to enumerate the observed interaction modes with the objects in the scene and the feedback responses. We focused only on the core experience of the applications and did not explore the interstitial menus or setup needed to enter the core experiences. A researcher coded papers and applications with the same codebook, resulting in 17 distinct designs for interaction modalities and feedback mechanisms with virtual objects, as illustrated in Tab. 1.

We identified two sets of object interaction modalities. The first category encompasses interactions via input devices that do not realistically model the intended interaction modes of the virtual object, e.g., VR controller triggers that are used to grab objects [103]. The other category captures more natural interactions, e.g., via custom

TABLE 1: Systematization of modalities (Input) and feedback (Output Ch.): *Phy* are physical input dev., *Vir* are virtual dev., *A/V* are audio/video ch., *UI* are 2D canvas , *Sens.* are sensorial ch., and *Beyond* are beyond-real ch.

Immersivness Level	**Input**		**Output Channel**				**References**
	Phy	*Vir*	*A/V*	*Sens.*	*UI*	*Beyond*	
Interrupt the experience	●	●	●		●		1 paper; 5 apps
		●			●		2 papers
	●				●		2 papers
	●		●		●		4 apps
Enhance the experience		●		●			1 paper
		●	●				19 papers
		●	●	●			2 papers
	●		●	●			4 papers; 4 apps
	●		●				3 papers; 1 app
	●			●			23 papers
	●	●	●				4 papers
	●	●	●	●			1 paper; 3 apps
		●					1 paper
	●	●	●	●	●		3 apps
Beyond-real experience		●				●	10 papers
	●					●	1 paper
	●			●		●	2 papers

controllers able to actuate forces [135] or computer vision to track fingers' movements [149] to reproduce the act of grabbing an object.

The output coding identified four broad feedback mechanism categories. The first category (i.e., *A/V clues*) uses visual or auditory cues to give feedback to the users, which is the most common category identified in our survey. Examples are sketching [103] and designing tools [56] that help create interactive spaces giving a visual feedback to users. The second category relies on sensory reconstruction such as output simulating the tactile experience when interacting with real-world objects. For example, haptic [111] or electrovibration [169] to enhance tactile perception in virtual manipulation. Many approaches also rely on traditional two-dimensional UI controls. For example, applications showing video [92] or Social Networks [88] to users. The last category is for output which relies on beyond-real experiences by manipulating spatial-temporal perception, e.g., extending virtual spaces by amplifying the mapping of physical to virtual movement [157], or by altering the physics of the virtual environment, e.g., adding visuo-haptic illusion for stretching distances [39], or by changing body perception, e.g., by deviating the mapping with their avatar limbs [36].

Moreover, we categorized the various combinations of interaction and feedback into distinct levels of immersiveness, as detailed in Tab. 1. In four designs, the feedback mechanism places the output outside of the VR world, *interrupting* the experience. For example, the user must use an external device like a mobile phone to interact with an object [80]. In another case, interacting with a class of objects, the user is teleported to another VR scene, interrupting the current experience, e.g. in Social VR App [131] to enter different worlds. In ten designs, the feedback mechanisms *enhance* the immersive experience with visual, auditory, or other sensory augmentations, e.g., the object glows, vibrates, or emits a sound when the user grabs it. Enhancing the experience is by far the most common type of design in our survey. The final type of design relies on *beyond-real experiences*, where the feedback mechanism is implemented by bending physic rules, e.g., the user's virtual hand is moving slower than the real hand to simulate force when moving an object [166]. The complete table mapping articles, interaction modalities, and feedback mechanism is in § A.2.

4.1.2. Prior Security Warnings. From the last decade of Security and HCI conferences (15,214 papers), 68 articles matched our keywords, and we reviewed their abstracts. We then excluded 42 as out of scope, i.e., not presenting or discussing warning designs. One researcher then reviewed the remaining papers (26) utilizing a codebook to identify design principles to reach effective warnings, focusing on our goal of noticeability and comprehension. Tab. 2 shows our results. Our analysis identified two main design

TABLE 2: Prior security warning design strategies. Expanded advices can be seen in Tab. 3

	Design		**Notice & Comprehension**							**General**			
Ref.	*Act.*	*Pas.*	A	B	C	D	E	F	G	H	I	J	K
[101]												●	
[4]	●								●		●		
[43]	●							●					
[6]	●						●				●		
[8]	●		●							●			
[110]		●	●	●	●								
[109], [145]		●		●									
[148]										●			
[155]		●		●	●								
[54]		●	●										
[114]		●									●		
[142]	●				●						●		
[150]		●				●							
[108]	●				●			●			●		
[115], [161]	●										●		
[147]	●	●					●			●			
[55]			●										
[75]								●					
[64], [167]													●
[57]	●	●							●		●		
[19], [41]	●		●					●					
[100]				●							●		●

strategies in security warnings that we call *active* and *passive*. Active strategies intend to interrupt or alter the user workflow. They are typically implemented as intermediate pages or pop-up windows that prompt user interaction before proceeding. For example, Chrome may display an alert page before entering a suspicious site, requiring the user to click a button to continue to the page. Passive strategies do not interrupt users' activity but present supplemental information or cues about potentially dangerous actions. Examples of such warnings are non-modal, non-blocking pop-ups in older browsers (e.g., Internet Explorer 7.0 and Firefox 2.0 [33]) informing users if the page being visited is known to be malicious. Tab. 3 presents the specific codes we created for advice and techniques aimed at enhancing both the noticeability of warnings and promoting user adherence. Our findings highlight several effective strategies. For noticeability, employing iconography and distinct color schemes [19], [41], [54] (A), integrating multi-modal elements [55], [109], [110] (B), using warnings within the scene [108], [142], [145], [155] (C), and implementing physics-based designs [150] (D) are all shown to significantly capture users' attention. However,

TABLE 3: Enumeration of prior security warning design recommendations from the literature review.

Notice	A	Employ iconography and color scheme
	B	Utilize multi-modal elements
	C	Integrate warning within the scene
	D	Implement Physics based warning
	E	Develop a distinct UI style
Comprehension	F	Opinionated design to engage with the safe option
	G	Increase task friction
General	H	Avoid habituation
	I	Provide comprehensive risk information
	J	Training to (re)form beliefs and attitudes
	K	Ensure consistency and accessibility of warning

while ensuring visibility, warnings should also be distinct from existing UI elements to avoid confusion [6], [147] (E). In terms of promoting comprehension, incorporating opinionated design elements [19], [43], [75], creating a "sense of fear" [57], and utilizing prohibitive design techniques [75] (F) have proven effective in deterring users from engaging with unsafe options. Additionally, increasing task friction [4] (G), while maintaining usability, can further guide users towards safer interactions.

Additionally, we coded and followed advice from the C-HIP scale, informing our development of effective warnings and providing valuable guidance for others designing and refining warning systems. Avoiding habituation (H), through polymorphic warnings [8], [148] or avoiding unneccesary or too frequent warning [147] is essential to have a functional warning. Providing comprehensive risk information [4], [114], [115], [161] (I) and providing training [101] (J) is crucial for enhancing user comprehension and motivation. Moreover, ensuring consistency and accessibility (K) in warnings is vital for supporting users with impairments, making them inclusive and effective for a wider audience [19], [64], [167]. However, since VR attacks and warnings are still emerging and not widely recognized, habituation and rational rejection were less decisive in our design choices than factors in the Notice and Comprehension categories, though their importance may grow as threats evolve.

4.2. Design Immersive Warnings

Guided by our overarching goals for effective warning communication, our design exploration identified three out of 11 potential immersive non-spoofable design ideas that can be implemented with readily-available technology.

4.2.1. Design Principles. We faced the challenge of designing warnings by defining a set of principles. The first and second principles are based on our systematization which identified three levels of immersiveness implemented via various feedback mechanisms (see Tab. 1): First, as we focused on immersive warnings, we deliberately excluded those that significantly interrupted the immersive experience, as maintaining immersiveness was critical to ensure a continuous and cohesive user experience within a virtual reality environment. Second, we addressed various feedback mechanisms that define levels of immersiveness, including warnings through audio, visual cues, sensor-based feedback, UIs, and beyond-real interactions, since noticeability and comprehension strategies [159] effective in 2D may not seamlessly translate to 3D, where attention is distributed across multiple sensory inputs. Third, our literature survey identified two broad design categories—active and passive warnings—and we ensured coverage of both categories.

4.2.2. Feasibility and Channels Exclusion. A fundamental property of a warning sign is to be accepted as a *universal* danger sign [159], which can be undermined by variable capabilities of existing VR devices (i.e., HMDs). For example, warnings relying on haptic feedback will not work on HMDs not supporting haptic feedback. Accordingly, our design exploration considered the input/output capabilities of current HDMs. We looked for the input/output channels and technical specifications (e.g., hand controllers, gaze tracking, and haptic feedback) of the top four best-selling consumer-grade HMDs [123], i.e., Meta Quest 2, Playstation VR, HTC Vive Pro 2, and Valve Index. We considered all the different modes in which each device can be used, e.g., the Quest 2 works either linked to a PC or in standalone mode, and it can use either the controller or the hand tracking only. With this separation applied to each device, we could not design warnings relying on external devices, such as smartwatches and smartphones, and external controllers, such as vibro-haptic feedback devices.

We avoided problematic color distinctions for color vision deficiencies and opted not to use audio warnings, recognizing their limitations for individuals who are hard-of-hearing. This decision was informed by the critical importance of accurately locating the spatial source of warnings for their effectiveness [159]. Acknowledging the importance of accessibility and inclusivity, our designs prioritize visual warnings that work across all VR hardware, not limited by specific devices or sensory impairments, ensuring universal applicability and effectiveness.

4.2.3. Security Considerations. As a final step, we compared each design idea against our threat model (Section § 2), to determine if an attacker can tamper or spoof a security warning. First, we avoided designs whose channels are UI elements that are part of the scene, including, for example, objects that a user can grab to inspect the security of other objects. For example, when brainstorming for ideas, we considered a magnifier that a user can grab and use to inspect other objects in the environment. However, an attacker could place in the scene a similar object with identical functionality, making it hard for the user to tell apart the original from the forged magnifier. Then, we also decided to consider designs using only negative feedback, i.e., a warning will identify objects that are insecure, because positive feedback, e.g., decorations for secure objects, can be recreated by an attacker around malicious objects. Similarly, we did not place warnings around or inside objects such as labels or images, e.g., locks, as these can be forged, too.

4.2.4. Warnings. After the security review of our warnings, we moved from 11 to five channels for warnings. However, since our focus is on immersive designs, we removed two warning channels (*Controller (LED)* and *Smartphone/watch*) interrupting the experience (see "Out of Scope" in Tab. 4). Our warning design process was informed by principles from past usable security research in 2D environments and adapted for immersive 3D VR settings. Starting from the remaining three categories, we derived four designs (see Fig. 2):

Virtual Object This design relies on visual decorations to highlight insecure objects. Drawing from 2D visual boundary indicators [164], we translated this concept into 3D. Several possible decorations can be implemented in immersive environments, e.g., the visual sandbox of AdCube [67]. Here, we took a different design direction, and instead of containing the entire third-party content in a wireframe, we created a red glow effect (*Red Glow*, hereafter) around the object to indicate maliciousness.

System UI These designs rely on system UIs to warn the user. We explored two. The first intends to change the visual appearance of an object to make it less desirable to interact with. The second one is an active warning system, which appears as soon as the user interacts with the object:

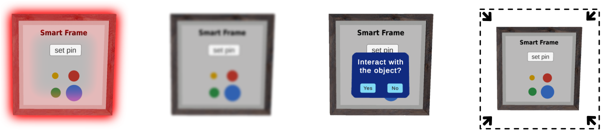
Visual alteration. A visual alteration changes the rendering of an object. We focused only on an alteration that results in visual degradation [159] to create a warning for insecure objects, similar to the visual decoration. Visual degradation can be achieved in multiple ways, such as rendering an object in grayscale. In 2D UIs, grayscale is commonly used to signal inactive elements [146]. We focused on an implementation that can render the object less attractive [12], [62] when insecure and selected the blur effect (*Blur*, hereafter), by adding a blurry layer on top of the object until the user interacts with it.

Interstitial UI. The second design explored an active warning system, which engages the user through a pop-up alert window (*Pop-Up*, hereafter), a typical direct attention-grabbing mechanism to prevent unsafe actions in 2D interfaces [33], [136]. The presented *Pop-Up* shows the "Interact with the object?" message when the user interacts with the insecure object. The user must dismiss the *Pop-Up* by answering yes to resume the interaction.

Alteration of Physics Unlike the other three warnings, which adapt existing 2D security principles, the alteration of physics introduces a beyond-real design unique to immersive environments by modifying the interaction dynamics of insecure objects [2]. We considered different options, e.g., via magnetically repulsing objects. We decided to implement negative feedback in the form of scaling down the dimension of the object (*Scale Down*, hereafter) as soon as the user starts the interaction.

5. RQ2: Initial User Responses

In the first study, we explored participants' initial responses to immersive warnings (RQ2) using a task, a

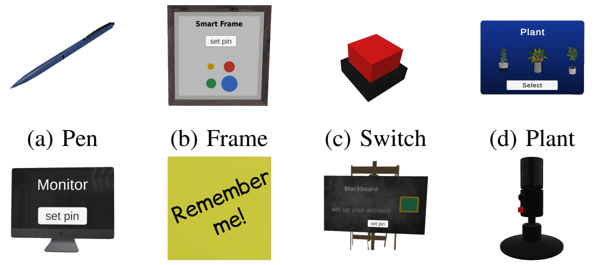


(a) *Red Glow* (b) *Blur* (c) *Pop-Up* (d) *Scale Down*

Figure 2: Examples of the four warnings implemented and evaluated with one of the task primary object.

TABLE 4: In grey, candidate designs for the user study.

Output: A/V	Sens.	UI	Beyond	Device: Phy HMD	Phy Ext	Virt	Warning channel	Immersiv.: Interrupt	Enhance	Beyond	Infeasible	Out of Sc.
●				●			Controller (LED)	●			●	●
●					●		Smartphone/watch	●			●	●
●						●	Virt. obj and sound	●				●
●						●	Virt. obj and sound		●			
	●			●			Controller (vibr)		●	●	●	
	●				●		Smartphone/watch (vibr)		●	●	●	
		●			●		Smartphone/watch	●			●	●
		●				●	System UI	●				●
		●				●	System UI		●			
			●	●			Controller (tact./ctrl. exp)		●	●	●	
			●			●	Alteration of physics			●		



(a) Pen (b) Frame (c) Switch (d) Plant
(e) Monitor (f) Sticky Note (g) Blackboard (h) Microphone

Figure 3: The eight objects of our user studies. (a-d) are primary objects and (e-h) are decoy.

questionnaire, and an interview. To get an unbiased initial response, participants were not aware some objects had warnings up until the end. During the task, we measured performance during object interactions under varying conditions to observe the impact of warnings on task execution. The questionnaire evaluated the effects of warnings on user interactions, mental workload, and perceived risk, while the interview probed noticeability, comprehension of warnings, and participants' motivation to not interact with warnings. The study's procedure, recruitment, and results are discussed in § 5.1 to 5.3, respectively.

5.1. Procedure

Prep Phase Participants were recruited via Prolific [1] and provided consent before joining the study. They then received a PDF with instructions, a binary link, and execution steps for their HMDs. The task began with a mandatory tutorial to familiarize them with interactions. The tutorial scene and objects resembled but were not identical to the main task to prevent habituation. To simulate real-world high-stress conditions in warning processing, participants were told the task had a time limit [24], though none existed. To avoid priming, we omitted security-related terms in the study description and tutorial and gave no instructions on how to behave toward warnings.

Experiment Phase After completing the tutorial, participants undertook the main task, which took place at a virtual office. Virtual offices are an increasingly important use case for immersive technology as witnessed by recent ideas by Microsoft [93] and Meta [82]. On a desk, participants found eight common office items, implementing specific functions. These objects included interactive items like a smart picture frame, which was activated by typing a secret PIN, and a pen that could be moved around. The eight objects are in Fig. 3 and the list of task is in Tab. 7 in § B.3. We divided those objects evenly into two groups: four primary and four decoy objects. Primary objects were the focus of our study, therefore each participant was presented with one of our warnings in two of those objects (i.e., between-participant design), one on a desk object and another on a canvas over the desk. Decoy objects, 'e-h' in Fig. 3, were interactable but never showed warnings. Participants engaged in diverse interactions with the objects, like moving a pen into a mug or typing a PIN to enable a monitor (details in § B.3).

Participants could track progress with an always visible to-do list, randomized per user to reduce bias. The task was developed for Meta Quest HMDs, chosen for its market dominance [123], using C# and Unity.

Post-Experiment Questionnaire and Interview After completing the task, participants answered four questionnaires—one per primary object, two of which had warnings. These assessed whether and how warnings impacted usability, workload, and perceived security risks. We designed the questionnaire by adapting the SUS [20] and NASA-TLX [49] scales, adding two custom questions on motivation and comprehension of the C-HIP model. Participants indicated whether they wanted (*want* item) to interact with the object and if they felt safe (*secure* item) doing so. The term safe was used to capture a general sense of risk, as participants were unaware of insecure objects in the scene. The complete questionnaire can be seen in § B.4. It was administered within the VR environment, following established methodology [5], to avoid potential bias from exiting the virtual environment before responding [113], [121].

The interview, detailed in § B.5, aimed to gather qualitative feedback on notice and initial processing of warnings, structured per C-HIP stages for analyzing warnings. After onboarding, which covered IRB procedures and welcomed participants, we asked if they noticed (US1) the warnings ('unexpected behavior or appearance'). We then explored participants' perspectives and comprehension (US2). To understand their beliefs and attitudes (US3), we examined their perceptions of VR objects as security risks. Next, we assessed motivation (US4) by asking whether the warnings influenced their willingness to complete tasks. The interview concluded with a debriefing, where participants learned the study's true purpose—examining responses to security indicators. They were given options to withdraw and the opportunity to suggest warning designs.

5.2. Recruitment and Demographics

We recruited remote participants as the study started towards the end of the COVID-19 pandemic, and used Prolific, an online platform [37] to find participants for user studies. As we developed our experiment for Meta Quest, we first screened Prolific users for Meta Quest owners with a short survey asking if (i) the survey participant owns a Meta Quest headset and (ii) if they were willing to participate in a user study on VR with or without the interview. We invited to participate in our short survey all Prolific users tagged by Prolific as VR owners and selected the first 1,000 respondents. We compensated each survey participant with £0.15[1]. We estimated 60 minutes for the preparatory phase, user task execution, and questionnaires, including a 10 minutes buffer for troubleshooting. We also estimated additional 40 minutes for the online interview, including 10 minutes breaks before the interview. We applied an hourly rate of £14/hour (£14/£24 without/with the interview) against the £9/hour recommended by Prolific, to account for the added challenge of using a VR headset.

A total of 212 respondents owned a Meta Quest device and expressed interest in participating in follow-up studies. Among them, 43 preferred to participate without an interview, while 169 were willing to participate with an interview included. Of these 212, 172 did not join our study, and two joined but did not start our study for technical reasons. The remaining 38 participants finished the user study. After reviewing the submitted data points, one participant concluded the user task and the questionnaire twice; we, therefore, excluded this participant from the analysis. Of the remaining 37 participants, 30 submitted valid data points for the objects interaction (24 men, six women; age: M = 31.37, range = 18-72), 37 had valid questionnaire data (31 men, six women; age: M = 30.81, range = 18-72), and ten successfully participated in the interview (six men, four women; age: M = 32.20, range = 22-42). The total observed participant drop-off could be attributed to the need for our own screening process to find Meta Quest headset owners, reluctance to join audio-recorded interviews and other unknown factors causing a change of heart in participants in between agreeing to take part and actually taking part in the study.

5.3. Results

We report task metrics as general trends to assess differences in task performance. The questionnaire responses were inferential statistically interpreted to measure the impact on workload and usability, and to gauge participants' comprehension of potential security risks. Insights from the interviews were extracted via thematic analysis, further contributing to our understanding of noticeability and initial comprehension.

5.3.1. Initial Response–Interaction.

Insights. We examined interaction times (time from the first until the last moment of interaction with any object in the scene) across various warning conditions. The rationale behind this approach is the assumption that warnings might have caused disruptions in the interactions with objects (e.g., due to hesitation to interact with or longer inspection of objects), leading to longer total interaction times. With this approach it is not possible to compare each warning condition with a no warning condition since this was manipulated as a within participant factor and, thus, potential differences between warnings would have to be interpreted as relative in nature.

The examination revealed intriguing behavioral patterns. As can be seen in the inter quartile range the *Pop-Up* condition (Mdn = 121.61, IQR = 78.63, Min = 81.57, Max = 320.62) created the largest variation in interaction times, indicating a differential effect of the warning on participants. Conversely, the *Scale Down* condition (Mdn = 98.72, IQR = 18.13, Min = 69.15, Max = 246.1) demonstrated the most uniform interaction times, indicating that the warning had a similar effect on participants. Moreover, *Scale Down* showed the lowest minimum and maximum interaction times, possibly, as we will see in § 6.2.2, due to a sense of urgency caused by reducing object size. The *Blur* (Mdn = 124.11, IQR = 46.24, Min = 105.21, Max = 249.31) and *Red Glow* condition (Mdn = 101.41, IQR = 52.54, Min = 76.34, Max = 383.90) ranged somewhere in between the *Pop-Up* and *Scale Down* condition, showing some variation in interaction times, both showing more differences between participants as compared to the *Scale Down* condition but less differences compared to the *Pop-Up* condition.

Participant Pool. The design for the assessment of task metrics would have required a larger number of participants per experimental cell to conduct an inferential statistical analysis (we exhausted the available participant pool). Consequently, we reported general trends for these metrics. However, questionnaire responses, for which the design was unaffected by these sample size considerations, were analyzed inferentially to assess workload, usability, and participants' comprehension of potential security risks. Additionally, insights from the interviews were analyzed thematically, enhancing our understanding of noticeability and initial comprehension of warnings.

5.3.2. Decomposing the Initial Response–Questionnaire.

We ensured suitability of the data for the analysis calculating the Cronbach's alpha [26], both indicating good reliability for both scales (α = .83 for SUS, and α = .82 for NASA-TLX), then we conducted linear mixed model analyses and report the maximal random-effect structures that converged or allowed a non-singular fit [14]. Detailed description of the models, with parameters and estimates can be found in § C.1, from Tab. 9 to Tab. 12. The conditional design of the questionnaire enabled statistical analyses, despite the limited number of participants.

As a first step, we assessed whether we observe any difference between warning and no warning. The analysis revealed a significant main effect of 'warning' for SUS items, $\chi^2(4)$ = 41.58, $p < .001$ and NASA items, $\chi^2(4)$ = 18.37, p = .001, indicating that the warnings differed with respect to their perceived usability and workload from the control condition (i.e., no warning attached to the objects).

Moreover, the analysis revealed a significant main effect for both of our own items, for the *want* item (Q10 in Tab. 8), $\chi^2(4)$ = 11.03, p = .026 and for the *secure* item (Q11 in Tab. 8), $\chi^2(4)$ = 14.02, $p < .007$, indicating that the warning had an effect on whether the object was perceived as *secure* and desirable to interact with.

(2) Comprehension of Warnings We assessed comprehension with the *'secure'* item. Results show that objects that *Scale Down* (M = 4.22, SD = 0.81) were perceived as significantly less secure to interact with and *Red Glow* objects (M = 4.82, SD = 0.39) were perceived as significantly more secure as compared to objects with no warning attached (M = 4.46, SD = 0.67).

1. This is the Prolific recommended compensation for a 60 seconds short survey.

(3) Motivation for not Interacting with Warnings We then analyzed users' motivation with the SUS items (usability), NASA-TLX items (workload), and the *want* item. The friction of object interaction influences motivation (not) to complete a task [159]. We measured friction by analyzing the reduced usability and mental workload associated with using the object if the warning is present. Additionally, we measured the *want* factor to interact with the object despite the warning.

Usability. Results show that the *Blur* (M = 3.93, SD = 1.04) and *Scale Down* (M = 3.56, SD = 1.18) warnings were perceived as significantly less usable as opposed to no warning attached (M = 4.3, SD = 0.98). No other effect reached statistical significance for usability.

Mental Workload. The NASA items show that *Scale Down* (M = 3.99, SD = 1.05) was perceived as significantly more mentally taxing as opposed to no warning attached (M = 4.47, SD = 0.87). No other effect reached statistical significance for mental workload.

Want. The analysis of the *want* item shows that blurring objects significantly reduced users' desire to interact with the objects (M = 3.44, SD = 1.26) compared to objects without any warning attached (M = 4.48, SD = 0.90). No other effect reached statistical significance for the motivation to interact with the object. More details about the statistical analyses are given in § C. We performed sensitivity checks due to the small sample size, using a significance level of $p = .05$ and a power of .80 to detect warning effects. These checks established a minimal detectable effect of .44, suggesting cautious interpretation of our results, as most estimates were below this threshold.

5.3.3. Decomposing the Initial Response–Interviews.

From our analysis of the interviews, we identified several insights and participants' reflections that we grouped following the C-HIP stages: (1) Noticeability and Perception of Warnings, (2) Comprehension of Warnings, (3) Beliefs and Attitudes towards VR Security and Privacy, (4) Motivation for Interacting with Objects with Warnings. While participating in these interviews, the participants were not yet aware that the study focused on security warnings. We slightly edited some quotes for readability.

(1) Noticeability and Perception of Warnings We assessed noticeability and perception, asking participants if they noticed unexpected behavior or appearance. Seven out of ten participants noticed the warnings, while three did not (2 *Red Glow* and 1 *Pop-Up*). Those who noticed had mixed impressions: *Scale Down* and *Blur* were consistently described as out of place with the experience, while *Red Glow* and *Pop-Up* felt part of the experience.

(2) Comprehension of Warnings Not all warnings were associated with negative feelings. For example, P8, one of the three who interacted with *Red Glow*, associated it with danger but confused it for an object highlighting feature:

> *I know generally red is perceived as danger. But just in that environment, I think it was there just because it was pointing my direction to something.* (P8)

P9 made a similar confusion about *Red Glow*, but did not associate it with danger. The last participant pointed out that, even if it was unnoticed, *Red Glow* suggested caution:

> *I would definitely think that it was some kind of caution. I would maybe think, if it looked particularly out of place, I might think something was going wrong with the actual VR program itself, that that's not meant to be there. But red is generally not a good sign.* (P1)

P2, like P8 and P9, associated *Pop-Up* with a highlighting feature:

> *You want to interact with one object, and you want to make sure it's this one and not the one next to it, then [the pop-up] would make sense.* (P2)

Another contradiction in perception came from P7, who misinterpreted *Blur* as a privacy shield despite it causing motion sickness by hindering depth perception.

Other participants associated negative feelings with *Scale Down* and *Pop-Up*. For example, P3, P4, and P10 noted they were either distracting or disruptive, indicating an increase of friction in the interaction, which is associated with higher warning compliance [159]. The remaining four shared mixed feedback. For example, P5 acknowledged that *Pop-Up* adds friction, but it may cause habituation over time. P6 shared that *Scale Down* was confusing as it may signal some wrong action and was unsure if the task, i.e., typing a PIN, would have been completed on time.

(3) Beliefs and Attitudes Towards Security and Privacy We asked participants about general risks affecting virtual worlds, and they identified risks from four categories, i.e., personal data theft (6/10), social engineering attacks such as phishing (5/10), malware (3/10), and physical harm (1/10). Only one participant, P6, did not mention any security risks. When we asked participants about the potential consequences of interacting with a generic malicious object, most answers pointed out that social engineering attacks are the first security consequence (5/10), followed by malware (3/10), personal data theft (2/10), and physical harm (1/10). We note that at this point, we did not inform the participant that the objects with warnings were malicious.

(4) Motivation for Not Interacting with Warnings Participants demonstrated caution towards potential security threats, with 7/10 stating they would avoid interacting with warning-marked objects. Pre-existing awareness of security risks significantly influenced participant behavior, whereas unfamiliar threats were often treated with less gravity. For instance, when discussing a PIN entry object, P1 expressed trust in a more interactive method, stating:

> *I think I'd have more trust in a pen that you can pick up and draw numbers... maybe just the idea that it would be slightly harder for a machine to read it if I'm drawing out on a whiteboard* (P1)

Finally, when we asked participants if the displayed warning acted as a deterrent for interaction, 6/10 answered yes, and 4/10 answered no. The merged codebook for study 1 can be found in § B.7.

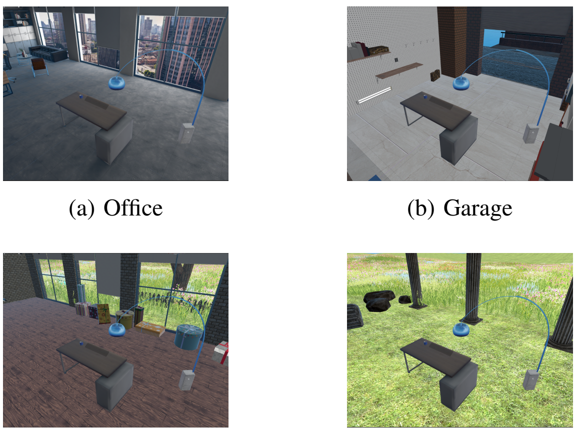
6. RQ3: Behavioral Responses

Our second study measured the second part of the C-HIP scale, the behavioral compliance towards warnings, through a behavioral task and user interviews. This approach evaluated the C-HIP model's latter stages, focusing on comprehension and behavior towards warnings. We utilized quantitative measures like object interaction and qualitative analyses to measure warning effectiveness. The study's procedure, recruitment, and results are discussed in sections § 6.1, § 6.1.1, and § 6.2, respectively.

6.1. Procedure

Prep Phase Participants were instructed to interact with secure objects and avoid non-secure ones, omitting terms like 'warning' to prevent bias. This ensured we measured warning effectiveness while making participants aware that some objects could be malicious without specifying which or how to identify them. Full context and threat model details were provided in the interview (§ B.6). Next, participants received the consent form, had time to read it, and were informed about data collection. After signing, they completed a demographics and prior VR experience questionnaire. They then wore the Meta Quest 2 headset, seated with ample space to prevent injury, with the HMD cable secured to avoid tangling. The experiment began with the same tutorial as the first user study.

Experiment Phase Instead of displaying all eight objects at once as in study 1, we sequentially showed only the four primary ones, with and without warning (i.e., within-participant design). Accordingly, each participant saw all combinations of objects with warnings. To simulate real-world conditions, we set a 20-minute time limit, which was sufficient, and no participant exceeded it. At any time, users could interact with the displayed object or skip it via the controller A button. We minimized potential confounds between the office scene and the four primary objects by using three additional environments, i.e., a festive room, a garage workshop, and an outdoor garden (see Fig. 4). For the quantitative analysis, we treated warning as a five-level condition (i.e., *No Warning*, *Red Glow*, *Blur*, *Scale Down*, *Pop-Up*), object as a four-level condition (i.e., Pen, Frame, Switch, Plant) and collapsed over all scenes, for a total of 80 interactions. We randomly ordered the sequence of interactions per participant.



(a) Office (b) Garage (c) Party (d) Outdoor

Figure 4: We used scene (a) as the main study scene (RQ2) and added scenes (b-d) to minimize potential confounds between the office setting and the four primary objects.

Post-Experiment Interview Before the interview, participants saw each object individually in five warning configurations, and we asked them to describe the thought process when interacting with objects. In the interview, we explored participants' criteria for object security, focusing on understanding and emotional responses (US2), followed by security beliefs and motivations (US3-4) in VR. The session ended with a debrief on the threat model, reflections on warning effectiveness, and design ideas.

6.1.1. Recruitment and Demographics. We conducted our second user study in person, and recruited participants on the University campus via flyers and student mailing lists. Participants needed prior VR HMD experience. We estimated the experiment and interview would take about two hours, and compensated participants with approx. £13/hour in local currency. In total, we recruited 23 participants, used the first three as pilot and excluded them from our analysis. We removed two more participants for not meeting the study's requirements, i.e., prior VR HMD use. In total, we analyzed data from 18 (11 men, 7 women; age: M = 25.33, range = 19-36) participants.

6.2. Results

We analyzed task metrics, and followed the same method for extracting insights from interviews via thematic analysis.

6.2.1. Behavioral Response. We investigated warning effectiveness based on participant interaction, fitting a model with 'warning' (i.e., *No Warning* vs. *Blur* vs. *Red Glow* vs. *Scale Down* vs. *Pop-Up*), 'object' (i.e., frame vs. pen vs. switch vs. plant) and the two-way interaction of 'warning' and 'object' as fixed effects. The model had a by-subject (i.e., participants) random intercept. We report the maximal random-effect structures that converged or allowed a non-singular fit [14]. A detailed description of the following analyses, including p-value adjustments for multiple comparisons, is in § C.2, Tab. 13, and Tab. 14.

The analysis revealed a main effect of 'warning', $\chi^2(4)$ = 286.33, $p < .001$, and a two-way interaction of 'warning' and 'object' $\chi^2(12)$ = 28.21, $p = .005$. We calculated simple contrasts to unpack the two-way interaction, which showed no differences between different objects per warning condition, indicating that each warning had the same effect on all objects, with the exception of the *Red Glow* warning on the frame object. More specifically, the *Red Glow* warning worked significantly better in preventing participants from interacting with objects for the frame as compared to the pen, the switch, and the plant. Sensitivity checks, conducted with a significance level of $p = .05$ and a power of .80, confirmed that our study's estimates exceeded the minimal detectable effect of .49.

Most Effective Warning We determined the most effective warning by calculating simple contrasts between warnings, split by the different objects. Results are in Fig. 5. A consistent pattern emerged for all objects with respect to the *Red Glow* warning. The *Red Glow* warning always led to more rejections compared to the no warning condition, irrespective of the object it was attached to. The *Red Glow* warning also led to more rejections than all other warning conditions for all objects except for the switch object. A consistent pattern also emerged for all objects concerning the *Pop-Up* warning. The *Pop-Up* warning did not differ from the no warning condition

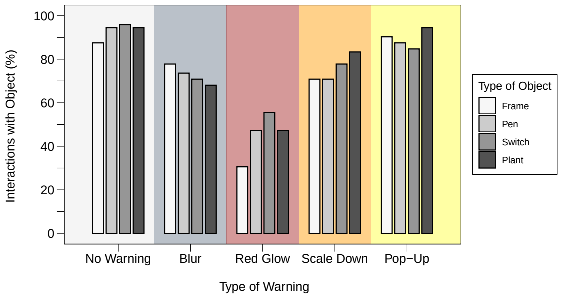


Figure 5: Percentage of object interactions as a function of object and warning.

irrespective of the object it was attached to, both leading to most object interactions. However, the effectiveness of *Blur* and *Scale Down* warnings differed between different objects. *Blur* was effective for all objects, as indicated by more rejections as compared to objects with no warnings attached, except for the frame. *Scale Down* was effective for all objects, as indicated by more rejections than objects with no warnings attached, except for the plant selector.

Overall, *Red Glow* was the most effective warning in preventing participants from interacting with objects, leading to most rejections. This effect was particularly pronounced for the frame object. *Pop-Up* were no more effective than objects without warnings, rendering *Pop-Up*s as ineffective in preventing interaction. *Blur* and *Scale Down* might be effective warnings for some objects, although they were not as effective as *Red Glow*.

6.2.2. Decomposing the Behavioral Response–Interview. We now present the results from our interview, following the C-HIP decomposition, focusing on (2) comprehension, (3) beliefs and attitudes, (4) motivation, and (5) behavior. We coded the interview following the methodology presented in § 3.2.3. Unlike in the first study, participants in this phase were aware that some modifications to objects could indicate security features. We slightly edited some quotes for readability.

(2) Comprehension of Warnings *Red Glow* was the warning most associated with danger. However, in line with the first study, 4/18 participants interpreted *Red Glow* as helpful in selecting the object, suggesting a transfer of understanding from existing VR experiences. P20 said:

You chose red probably because it was a bad color. But maybe it was just a targeting thing, because I know many VR games do... You mouse over something and it glows red (P20)

Participants found *Blur* and *Scale Down* confusing and did not associate them with danger or malice. The *Blur* warning, seen as inconsistent, evoked curiosity in some participants (3/18, e.g., P11) while others (3/18, e.g., P12) attributed it to vision issues:

Most likely, I'm just going to press it faster than the others, just to investigate (P11)

The blur gave me the sensation that there was something wrong with me, not wrong with the object (P12)

The *Scale Down* warning did not evoke a sense of danger and was not associated with security risks. Most (7/18) participants found it annoying and identified it as a glitch or weird interaction. For example, P25 said:

It was just weird. I didn't feel unsafe (P25)

In addition, 2/18 participants rushed to complete the task in response to the object scaling down.

Then, 8/18 participants expressed habituation to *Pop-Up* because these elements are common in other systems, leading people to overlook their significance.

Nowadays there are pop-up windows everywhere. I'm just used to clicking through them (P25)

Also, 5/18 participants felt they were faced with a critical security decision and made an additional effort to read them.

You are changing something in the world, so maybe it needs permission. (P20)

(3) Beliefs and Attitudes Towards Security and Privacy We asked participants about general threats in VR environments. By exploring their understanding of maliciousness, we aimed to determine whether they recognized or were aware of potential security and privacy risks associated with objects (e.g., a malicious pin-pad placed to steal the PIN). However, they identified an object as a possible threat only after the interviewer' explanation of the threat landscape at the end of the interview. The identified potential consequences are from the same four categories as in the previous study, with the same ranking. Social engineering attacks are by far the first concerns in VR environments (13/18), followed by malware (8/18), personal data theft (6/18), and physical harm (3/18).

(4) Motivation for Not Interacting with Warnings When questioned on warning effectiveness in influencing their decision, most participants, i.e., 13/18, indicated that the *Red Glow* warning was the most effective versus 4/18 saying it was not. Many participants, i.e., 12/18, indicated that *Scale Down* is also effective; however, 3/18 had conflicting opinion about effectiveness while 5/18 considered it ineffective. Lastly, many participants found *Pop-Up* (9/18) and *Blur* (7/18) to be ineffective. The merged codebook for study 2 can be found in § B.8.

(5) Behaviour toward Warnings While we measured behavior change in the user task (results in Fig. 5), the interview explored participants' intention to comply. At the end of the interview, they ranked the warnings and the intention to comply. Results are in Fig. 6. Most participants (13/18) perceived the *Red Glow* warning as effective, suggesting an inclination to avoid the objects. We examined if some warnings were ranked as more effective than others by calculating an ordinal logistic regression model over the rankings for all warning types. The overall analysis revealed a significant effect of type of warning $\chi^2(6) = 21.57$, $p = .001$, indicating a difference in the ranked effectiveness between the warnings. Pairwise comparisons showed that the *Red Glow* warning was rated as significantly more effective than the *Blur*, $\chi^2(66) = 12.52$, $p < .001$, the *Pop-Up*, $\chi^2(66) = 12.99$, $p < .001$, and the *Scale Down* warning, $\chi^2(66) = 16.03$, $p < .001$. No other comparisons were significant, all $ps > .36$.

7. Discussion

We explored security warning designs for VR interactions, suggesting 11 potential warning delivery channels,

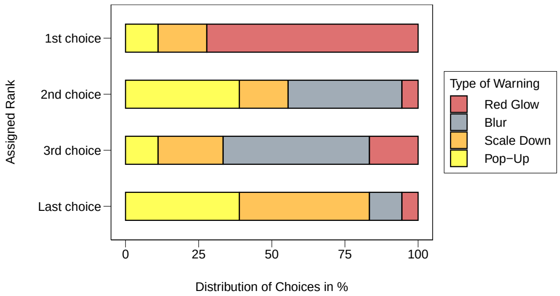


Figure 6: Distribution of rank choices.

three viable with current HMDs. We then developed and tested four warnings in two user studies. Below, we summarize our findings from our user studies.

7.1. Warning Effectiveness

Red Glow Our results indicate *Red Glow* as the most effective warning design, significantly reducing interactions with insecure objects. It was noticed among other visual stimuli and associated with danger and caution. While the *Red Glow* warning effectively influenced users' decisions and triggered compliance some feedback from the initial non-primed study suggested potential confusion with navigational aids in immersive environments. Moreover, the questionnaire results suggest it makes users feel more secure, requiring investigation into whether it might inadvertently encourage interaction with insecure objects.

Red Glow is highly effective but there are indications that it may be misinterpreted as a navigational aid, potentially reinforcing a false sense of security.

Scale Down Our results show that *Scale Down* reduced interactions when compared with *No Warning*; however, it is far from the reduction observed with the best performing warning. *Scale Down* design has desirable features of a warning, such as noticeability, usability degradation, and mental workload increase. However, participants reported less motivation to comply, and it was not associated with danger or security risks. In addition, a small number of participants in both studies pointed out that this warning caused a sense of urgency, fearing the object may disappear, suggesting to complete the interaction as soon as possible. While *Scale Down* may not be effective, altering virtual physics shows promise and should be further explored, e.g., magnetically repulsive objects.

Scale Down's approach of altering virtual physics increases noticeability, which is promising, but this specific design was not effective.

Blur *Blur* reduced interactions for almost all objects, still, like *Scale Down*, its effectiveness is well below the top performing warning. This design has both desirable and undesirable features. Among the desirable, this warning was noticeable, increased friction, and gave less desire to interact. Among the undesirable, *Blur* is not associated with danger and most of the participants said that it did not influence their decision to interact with the objects.

While Blur increases friction and discourages interaction, it lacks the necessary association with danger, limiting its effectiveness as a security warning.

Pop-Up Our results indicate that *Pop-Up* is likely unsuitable as an effective warning. More specifically, our results show that this warning is ineffective, as interactions with insecure objects did not differ from those with secure objects. About half of the participants said they overlooked its significance (comprehension) and it did not influence the decision to interact with the object (motivation).

Pop-Up failed to impact behavior, suggesting that traditional 2D-style pop-ups may be less effective in immersive environments.

7.2. Additional Insights from the Interviews

Warning Customization Many participants (16/28) proposed alternatives ideas, with varied visual elements (exclamation marks, skull icons) or interactions (unskippable popups for ten seconds). Some (9/28) suggested major changes, like informing users about sensitive data or pausing the experience. Others (6/28) recommended different channels, like sensory feedback and virtual assistants. These diverse suggestions show a preference for customizable warnings, a concept needing further investigation.
Assess Warnings in Mixed Environments Few participants noted that certain warnings could be more alarming in real-world contexts than in VR, implying our design effectiveness might vary in mixed scenarios with video overlays and should be explored in future studies.

7.3. Traditional and Immersive warning

Traditional security warnings are designed for 2D interfaces, but immersive environments require a different approach. Rather than directly adapting existing warnings, we started from warning design principles and built new security cues around immersive interaction. Our results show that more conventional warnings, like *Pop-Up*, were often ignored, while immersive warnings, like *Red Glow*, were more effective but introduced new challenges, such as navigational confusion. These findings suggest that emerging security hazards in VR demand warnings that are not just adaptations of 2D designs but are fully integrated into an immersive human-computer interaction.

8. Limitations

Demographics We recruited VR-experienced remote and in-person participants until platform saturation. Remote participants came from diverse backgrounds and countries (UK, USA, Portugal, Italy, Poland), while in-person participants, mostly computer science students, were based in Germany. The sample was predominantly composed of younger participants, particularly in the in-person group, where most were university students. This age distribution may limit generalizability to older professionals more likely to use VR for productivity. No participants reported color vision issues. The Meta Quest 2's spacer mitigated any impact to participants with corrective glasses. To reflect real-world constraints, time limits were used.

Apparent Discrepancies At first glance, the two studies may appear to conflict on the *Red Glow* warning, but the differing methodologies hinder direct comparison. The first study, which did not prime participants about security, used self-reports and interviews to assess independent interpretation of warnings. The second study, with primed participants, employed behavioral observations to evaluate effectiveness. These differences make it challenging to compare the studies directly. Priming in the second study shaped participant responses, unlike the non-primed first study. Additionally, the first study's reliance on self-reports introduces inherent biases, which the behavior-based measurements of the second study mitigate. While the first study establishes the warning's noticeability, the second reveals how context influences compliance. Together, they provide a more comprehensive understanding of the *Red Glow* warning's effectiveness.

Risk perception. Since immersive tasks are not yet central to productivity or frequent attack targets, users may perceive VR risks differently. While entering a PIN felt as high-risk, handling a virtual pen seemed low-risk, despite research highlighting potential privacy threats [99]. Interviews suggest participants relied on traditional security models, overlooking VR-specific risks. This discrepancy in mental models highlights the need to better understand security awareness in immersive environments better. Additionally, most participants primarily used VR for gaming, which may have influenced their perception of security warnings. Red highlights, often signaling interactivity in games rather than security risks, may have reduced the effectiveness of *Red Glow* in productivity settings. This suggests that findings may not fully generalize to office-productivity settings, particularly in high-stakes scenarios, and future research should explore these contexts further.

8.1. Further Improvement

Warning Alternatives and Improvements Future research should explore refinements and alternative warning designs to enhance effectiveness. Physics-based warnings could be strengthened as deterrents, while visual degradation methods may require clearer security associations than simple blurring. Additionally, *Red Glow*'s effectiveness could be improved with iconography to prevent misinterpretation as a navigational aid.

Warning Channel Combination Warning Science [159] suggest using multiple sensory channels can be effective in conveying warnings. However, we designed each warning for a distinct channel, to isolate the variables we desired to study. Future studies might explore combining warning channels to capture attention and convey messages.

Explicitness We omitted explicit explanations of potential risks to avoid priming in the first study and keep consistency with the second study. However, informing users about the consequences of actions increases warning effectiveness. Investigating the extent to which more verbose warning messages can influence the effectiveness of immersive warnings requires additional research.

Context-Specific Warnings Our research shows warning effectiveness in VR varies, with no single warning being ideal for all contexts. Customizing warnings to threats may improve effectiveness. *Red Glow* suits high-visibility, danger-associated scenarios, while *Scale Down* and *Blur* could suit subtler situations. The *Pop-Up* warning could be good for clear messages. Tailoring warnings to VR risks and experiences could enhance effectiveness by aligning them with scenario-specific needs.

9. Related Work

Our review identified relevant works in security warnings, VR interactions, and feedback methods. Here, we focus on studies concerning warnings in immersive environments. Recent studies [53], [158] explored security indicators for hyperlinking within virtual reality. Windl et al. [158] employed a design-centric approach, primarily informed by expert interviews instead of our systematic surveys. Their findings align with ours, demonstrating that a red visual warning captures user attention most effectively. While they explored multiple sensory channels, their results indicate that visual indicators—particularly red—were the most effective nudge, irrespective of placement in the VR scene. This supports our conclusion that a red glow is the most effective warning, reinforcing its association with danger and caution. Hosfelt et al. [53] investigated warnings using a traditional 2D interface and agent-assistant. Their interviews revealed that customization played a crucial role in user recall and engagement, with participants expressing a preference for adaptable security indicators. Our study similarly found that customization was a recurring theme in the user's feedback, where participants suggested custom visual elements and interaction-based security cues. These insights align with our broader interest, which extends beyond hyperlinking to explore warnings applicable to various metaverse interactions, from grabbable objects to canvases. Other works [66], [67] focused on mechanisms to separate third-party objects. Lee et al. [67] developed a module that isolates third-party content using visual wireframes. Our research extends this by systematically analyzing user responses to security warnings, assessing their efficacy in preventing interaction with malicious objects.

10. Conclusion

This paper presented a systematic design exploration of immersive security warnings in VR environments. Our systematic literature review of prior work and analysis of popular VR applications identified two input modalities, four output channels, and two warning design strategies from which we extracted eleven warning channels as potential design ideas. We implemented four warnings and assessed their impact and effectiveness with two user studies. Our results revealed different outcomes for effectiveness, with the *Red Glow* warning being the most effective for preventing interaction, while others showed potential for improvement. Our findings emphasize the need for a continued exploration and iteration in designing immersive security warnings.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft Grant No. 521601028 within SPP 2199.

References

[1] Prolific. https://www.prolific.com/, 2014. (Accessed on 06/05/2024).

[2] Parastoo Abtahi, Sidney Q. Hough, James A. Landay, and Sean Follmer. Beyond being real: A sensorimotor control perspective on interactions in virtual reality. In *CHI*, Apr 2022.

[3] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Comput. Surv.*, August 2017.

[4] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *USENIX*, 2013.

[5] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. Examining design choices of questionnaires in VR user studies. In *CHI*, 2020.

[6] Hazim Almuhimedi, Adrienne Porter Felt, Robert W. Reeder, and Sunny Consolvo. Your reputation precedes you: History, reputation, and the chrome malware warning. In *SOUPS @ USENIX*, 2014.

[7] Ayman Alzayat, Mark Hancock, and Miguel A. Nacenta. Quantitative measurement of tool embodiment for virtual reality input alternatives. In *CHI*, May 2019.

[8] Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. *33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[9] Apple. Adopting best practices for privacy and user preferences — Apple Developer Documentation. https://developer.apple.com/documentation/visionos/adopting-best-practices-for-privacy. (Accessed on 10/17/2024).

[10] Emilie Maria Nybo Arendttorp, Heike Winschiers-Theophilus, Kasper Rodil, Freja B. K. Johansen, Mads Rosengreen Jørgensen, Thomas K. K. Kjeldsen, and Samkao Magot. Grab it, while you can: a VR gesture evaluation of a co-designed traditional narrative by indigenous people. In *CHI*, Apr 2023.

[11] Oscar Ariza, Gerd Bruder, Nicholas Katzakis, and Frank Steinicke. Analysis of proximity-based multimodal feedback for 3D selection in immersive virtual environments. In *IEEE VR*, Mar 2018.

[12] Maansi Bansal-Travers, David Hammond, Philip Smith, and K. Michael Cummings. The impact of cigarette pack design, descriptors, and warning labels on risk perception in the U.S. *American Journal of Preventive Medicine*, 40(6):674–682, 2011.

[13] Nicolas Barbotin, James Baumeister, Andrew Cunningham, Thierry Duval, Olivier Grisvard, and Bruce H. Thomas. Evaluating visual cues for future airborne surveillance using simulated augmented reality displays. In *IEEE VR*, Mar 2022.

[14] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 2013.

[15] Mayra Donaji Barrera Machuca, Junwei Sun, Due-Minh Pham, and Wolfgang Stuerzlinger. Fluid VR: extended object associations for automatic mode switching in virtual reality. In *IEEE VR*, Mar 2018.

[16] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 2015.

[17] Andreea Dalia Blaga, Maite Frutos-Pascual, Chris Creed, and Ian Williams. Too hot to handle: An evaluation of the effect of thermal visual representation on user grasping interaction in virtual reality. In *CHI*, Apr 2020.

[18] Alberto Boem, Yuuki Enzaki, Hiroaki Yano, and Hiroo Iwata. Human perception of a haptic shape-changing interface with variable rigidity and size. In *IEEE VR*, Mar 2019.

[19] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie S. Downs, and Stuart E. Schechter. Your attention please: designing security-decision UIs to make genuine risks harder to ignore. In *SOUPS @ USENIX*, 2013.

[20] John Brooke et al. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 1996.

[21] Simon Cailhol, Philippe Fillatreau, Jean-Yves Fourquet, and Yingshen Zhao. A multi-layer approach of interactive path planning for assisted manipulation in virtual reality. In *2015 IEEE VR*, Mar 2015.

[22] Juan Sebastian Casallas, James H. Oliver, Jonathan W. Kelly, Frederic Merienne, and Samir Garbaya. Using relative head and hand-target features to predict intention in 3D moving-target selection. In *2014 IEEE VR*, Mar 2014.

[23] Kaiming Cheng, Arkaprabha Bhattacharya, Michelle Lin, Jaewook Lee, Aroosh Kumar, Jeffery F Tian, Tadayoshi Kohno, and Franziska Roesner. When the user is inside the user interface: An empirical study of UI security properties in augmented reality. In *USENIX*, 2024.

[24] Kaiming Cheng, Jeffery F. Tian, Tadayoshi Kohno, and Franziska Roesner. Exploring user reactions and mental models towards perceptual manipulation attacks in mixed reality. In *USENIX*, August 2023.

[25] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *UPSEC*, 2008.

[26] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 1951.

[27] Andreea Dalia Blaga, Maite Frutos-Pascual, Chris Creed, and Ian Williams. Freehand grasping: An analysis of grasping for docking tasks in virtual reality. In *IEEE VR*, Mar 2021.

[28] Apple Developer. VisionOS. https://developer.apple.com/visionos/, 2023. (Accessed on 07/28/2023).

[29] Rachna Dhamija and Adrian Perrig. Deja vu-a user study: using images for authentication. In *USENIX*, 2000.

[30] Patrick Dickinson, Andrew Cardwell, Adrian Parke, Kathrin Gerling, and John Murray. Diegetic tool management in a virtual reality training simulation. In *IEEE VR*, Mar 2021.

[31] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. VRsketchin: Exploring the design space of pen and tablet interaction for 3D sketching in virtual reality. In *CHI*, Apr 2020.

[32] Tobias Drey, Michael Montag, Andrea Vogt, Nico Rixen, Tina Seufert, Steffi Zander, Michael Rietzler, and Enrico Rukzio. Investigating the effects of individual spatial abilities on virtual reality object manipulation. In *CHI*, Apr 2023.

[33] Serge Egelman, Lorrie Faith Cranor, and Jason I. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. *SIGCHI*, 2008.

[34] David Englmeier, Isabel Schonewald, Andreas Butz, and Tobias Hollerer. Feel the globe: Enhancing the perception of immersive spherical visualizations with tangible proxies. In *IEEE VR*, Mar 2019.

[35] David Englmeier, Isabel Schonewald, Andreas Butz, and Tobias Hollerer. Sphere in hand: Exploring tangible interaction with immersive spherical visualizations. In *IEEE VR*, Mar 2019.

[36] Shaghayegh Esmaeili, Brett Benda, and Eric D. Ragan. Detection of scaled hand interactions in virtual reality: The effects of motion direction and task complexity. In *IEEE VR*, Mar 2020.

[37] Peer Eyal, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 2021.

[38] Cathy Fang, Yang Zhang, Matthew Dworman, and Chris Harrison. Wireality: Enabling complex tangible geometries in virtual reality with worn multi-string haptics. In *CHI*, Apr 2020.

[39] Martin Feick, Niko Kleer, André Zenner, Anthony Tang, and Antonio Krüger. Visuo-haptic illusions for linear translation and stretching using physical proxies in virtual reality. In *CHI*, May 2021.

[40] Martin Feick, Kora Persephone Regitz, Anthony Tang, and Antonio Krüger. Designing visuo-haptic illusions with proxies in virtual reality: Exploration of grasp, movement trajectory and object mass. In *CHI*, Apr 2022.

[41] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL warnings: Comprehension and adherence. *33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[42] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Emre Acer, Elisabeth Morant, and Sunny Consolvo. Rethinking connection security indicators. In *SOUPS @ USENIX*, 2016.

[43] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. Experimenting at scale with google chrome's SSL warning. *SIGCHI*, 2014.

[44] Lukas Gehrke, Sezen Akman, Pedro Lopes, Albert Chen, Avinash Kumar Singh, Hsiang-Ting Chen, Chin-Teng Lin, and Klaus Gramann. Detecting visuo-haptic mismatches in virtual reality using the prediction error negativity of event-related brain potentials. In *CHI*, May 2019.

[45] Eric J Gonzalez, Elyse D. Z. Chase, Pramod Kotipalli, and Sean Follmer. A model predictive control approach for reach redirection in virtual reality. In *CHI*, Apr 2022.

[46] Eric J Gonzalez and Sean Follmer. Sensorimotor simulation of redirected reaching using stochastic optimal feedback control. In *CHI*, Apr 2023.

[47] Google. Use Google Play Protect to help keep your apps safe & your data private - Nexus Help. https://support.google.com/nexus/answer/2812853?hl=en. (Accessed on 10/17/2024).

[48] Tobias Gunther, Lars Engeln, Sally J. Busch, and Rainer Groh. The effect of elastic feedback on the perceived user experience and presence of travel methods in immersive environments. In *IEEE VR*, Mar 2019.

[49] S. G. Hart and Lowell E. Staveland. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Advances in psychology*, 52, 1988.

[50] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. A design space for gaze interaction on head-mounted displays. In *CHI*, May 2019.

[51] Markus Holl, Markus Oberweger, Clemens Arth, and Vincent Lepetit. Efficient physics-based implementation for realistic hand-object interaction in virtual reality. In *IEEE VR*, Mar 2018.

[52] Adrian H. Hoppe, Felix Marek, Florian Van De Camp, and Rainer Stiefelhagen. VirtualTablet: Extending movable surfaces with touch interaction. In *IEEE VR*, Mar 2019.

[53] Diane Duros Hosfelt, Jessica Outlaw, Tyesha Snow, and Sara H. Carbonneau. Look before you leap: Trusted user interfaces for the immersive web. *unpublished*, abs/2011.03570, 2020.

[54] Hang Hu and G. Wang. End-to-end measurements of email spoofing attacks. In *USENIX*, 2018.

[55] Rebecca Jeong and Sonia Chiasson. 'Lime', 'Open lock', and 'Blocked': Children's perception of colors, symbols, and words in cybersecurity warnings. *CHI*, 2020.

[56] Hans-Christian Jetter, Roman Rädle, Tiare Feuchtner, Christoph Anthes, Judith Friedl, and Clemens Nylandsted Klokmose. "in VR, everything is possible!": Sketching and simulating spatially-aware interactive spaces in virtual reality. In *CHI*, Apr 2020.

[57] Ben Kaiser, Jerry W. Wei, Elena Lucherini, Kevin Lee, Jorge Nathan Matias, and Jonathan R. Mayer. Adapting security warnings to counter online disinformation. In *USENIX*, 2020.

[58] Hyo Jeong Kang. Designing next generation marketplace: The effect of 3D VR store interface design on shopping behavior. In *2017 IEEE VR*, 2017.

[59] Mohamed Kari and Christian Holz. HandyCast: Phone-based bimanual input for virtual reality in mobile and space-constrained settings via pose-and-touch transfer. In *CHI*, Apr 2023.

[60] Myung Jin Kim, Neung Ryu, Wooje Chang, Michel Pahud, Mike Sinclair, and Andrea Bianchi. SpinOcchio: Understanding haptic-visual congruency of skin-slip in VR with a dynamic grip controller. In *CHI*, Apr 2022.

[61] Yusuke Kishishita, Swagata Das, Antonio Vega Ramirez, Chetan Thakur, Ramin Tadayon, and Yuichi Kurita. Muscleblazer: Force-feedback suit for immersive experience. In *IEEE VR*, Mar 2019.

[62] Jan Andre Koch, Jan Willem Bolderdijk, and Koert van Ittersum. Can graphic warning labels reduce the consumption of meat? *Appetite*, 168:105690, 2022.

[63] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 2017.

[64] Elaine Lau and Zachary N. J. Peterson. A research framework and initial study of browser security for the visually impaired. In *USENIX*, 2023.

[65] Magic Leap. Landscape. https://ml1-developer.magicleap.com/en-us/learn/guides/design-landscape, 2023. (Accessed on 05/31/2023).

[66] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. Arya: Operating system support for securely augmenting reality. *IEEE Security & Privacy*, 16, 2018.

[67] Hyunjoo Lee, Jiyeon Lee, Daejun Kim, Suman Sekhar Jana, Insik Shin, and Sooel Son. AdCube: WebVR ad fraud and practical confinement of third-party ads. In *USENIX*, 2021.

[68] Jaeyeon Lee, Mike Sinclair, Mar Gonzalez-Franco, Eyal Ofek, and Christian Holz. TORC: A virtual reality controller for in-hand high-dexterity finger interaction. In *CHI*, May 2019.

[69] Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, and Seok-Hyung Bae. Projective windows: Bringing windows in space to the fingertip. In *CHI*, Apr 2018.

[70] Russell V Lenth. Least-squares means: The R package lsmeans. *Journal of statistical software*, 69, 2016.

[71] Jingjie Li, Sunpreet Singh Arora, Kassem Fawaz, Younghyun Kim, Can Liu, Sebastian Meiser, Mohsen Minaei, Maliheh Shirvanian, and Kim Wagner. How interactions influence users' security perception of virtual reality authentication? *unpublished*.

[72] Zhen Li, Joannes Chan, Joshua Walton, Hrvoje Benko, Daniel Wigdor, and Michael Glueck. Armstrong: An empirical examination of pointing at non-dominant arm-anchored UIs in virtual reality. In *CHI*, May 2021.

[73] Christos Lougiakis, Akrivi Katifori, Maria Roussou, and Ioannis-Panagiotis Ioannidis. Effects of virtual hand representation on interaction and embodiment in HMD-based virtual environments using controllers. In *IEEE VR*, Mar 2020.

[74] Kathleen M. MacQueen, Eleanor McLellan, Kelly Kay, and Bobby Milstein. Codebook development for team-based qualitative analysis. *Field Methods*, 10, 1998.

[75] Jeehan Malik, Morgan N D Parr, Jessica Flathau, Hanxi Tang, Joseph K. Kearney, Jodie M. Plumert, and Kyle Rector. Determining the effect of smartphone alerts and warnings on street-crossing behavior in non-mobility-impaired older and younger adults. *CHI*, 2021.

[76] Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. EyeSeeThrough: Unifying tool selection and application in virtual environments. In *IEEE VR*, Mar 2019.

[77] Jonatan Martinez, Daniel Griffiths, Valerio Biscione, Orestis Georgiou, and Tom Carter. Touchless haptic feedback for supernatural VR experiences. In *IEEE VR*, Mar 2018.

[78] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *33rd Annual ACM Conference on Human Factors in Computing Systems*, Apr 2015.

[79] Daniel Medeiros, Rafael Dos Anjos, Nadia Pantidi, Kun Huang, Mauricio Sousa, Craig Anslow, and Joaquim Jorge. Promoting reality awareness in virtual reality through proxemics. In *IEEE VR*, Mar 2021.

[80] Manuel Meier, Paul Streli, Andreas Fender, and Christian Holz. TapID: Rapid touch interaction in virtual reality using wearable sensing. In *IEEE VR*, Mar 2021.

[81] Tim Menzner, Alexander Otte, Travis Gesslein, Jens Grubert, Philipp Gagel, and Daniel Schneider. A capacitive-sensing physical keyboard for VR text entry. In *IEEE VR*, Mar 2019.

[82] Meta. Horizon Workrooms virtual office and meetings — Meta for Work. https://forwork.meta.com/de/en/horizon-workrooms/. (Accessed on 01/29/2024).

[83] Meta. Bigscreen Beta on Oculus Quest 2. https://www.oculus.com/experiences/quest/2497738113633933/, 2023. (Accessed on 07/27/2023).

[84] Meta. DeoVR Video Streaming on Oculus Quest 2. https://www.oculus.com/experiences/quest/2382576078453818/, 2023. (Accessed on 07/27/2023).

[85] Meta. First Steps for Quest 2 on Oculus Quest 2. https://www.oculus.com/experiences/quest/3675568169182204/, 2023. (Accessed on 07/27/2023).

[86] Meta. FitXR - Box. HIIT. Dance. Sculpt. Combat. on Oculus Quest 2. https://www.oculus.com/experiences/quest/2327205800645550/, 2023. (Accessed on 07/27/2023).

[87] Meta. GYM CLASS - BASKETBALL VR on Oculus Quest 2. https://www.oculus.com/experiences/quest/3661420607275144/, 2023. (Accessed on 07/27/2023).

[88] Meta. Instagram (Beta) on Oculus Quest 2. https://www.oculus.com/experiences/quest/6102857836422862/, 2023. (Accessed on 07/27/2023).

[89] Meta. Netflix on Oculus Quest 2. https://www.oculus.com/experiences/quest/2184912004923042/, 2023. (Accessed on 07/27/2023).

[90] Meta. Oculus Store: VR Games, Apps, & More. https://www.oculus.com/experiences/quest/, 2023. (Accessed on 08/01/2023).

[91] Meta. VR Animation Player on Oculus Quest 2. https://www.oculus.com/experiences/quest/2515021945210953/, 2023. (Accessed on 07/27/2023).

[92] Meta. YouTube VR on Oculus Quest 2. https://www.oculus.com/experiences/quest/2002317119880945/, 2023. (Accessed on 07/27/2023).

[93] Microsoft. Introducing Microsoft Mesh — Connect like never before. https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh. (Accessed on 01/29/2024).

[94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[95] Roberto A. Montano-Murillo, Cuong Nguyen, Rubaiat Habib Kazi, Sriram Subramanian, Stephen DiVerdi, and Diego Martinez-Plasencia. Slicing-Volume: Hybrid 3D/2D multi-target selection technique for dense virtual environments. In *IEEE VR*, Mar 2020.

[96] Soroosh Mortezapoor, Khrystyna Vasylevska, Emanuel Vonach, and Hannes Kaufmann. CoboDeck: A large-scale haptic VR system using a collaborative mobile robot. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, Mar 2023.

[97] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. Does it feel real?: Using tangibles with different fidelities to build and explore scenes in virtual reality. In *CHI*, May 2019.

[98] Vivek Nair, Gonzalo Munilla Garrido, Dawn Song, and James O'Brien. Exploring the privacy risks of adversarial VR game design. In *23rd Privacy Enhancing Technologies Symposium (PETS 23)*, 2023.

[99] Vivek Nair, Wenbo Guo, Justus Mattern, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song. Unique identification of 50,000+ virtual reality users from head & hand motion data. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 895–910, Anaheim, CA, August 2023. USENIX Association.

[100] Daniela Napoli and Sonia Chiasson. Assessing non-visual SSL certificates with desktop and mobile screen readers. *2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018.

[101] Ajaya Neupane, Md. Lutfor Rahman, Nitesh Saxena, and Leanne M. Hirshfield. A multi-modal neuro-physiological study of phishing detection and malware warnings. *22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015.

[102] Francisco Nicolau, Johan Gielis, Adalberto L. Simeone, and Daniel Simoes Lopes. Exploring and selecting supershapes in virtual reality with line, quad, and cube shaped widgets. In *IEEE VR*, Mar 2022.

[103] Oculus. Gravity Sketch on Oculus Quest 2. https://www.oculus.com/experiences/quest/1587090851394426/, 2023. (Accessed on 07/27/2023).

[104] Yuki Onishi, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. BouncyScreen: Physical enhancement of pseudo-force feedback. In *IEEE VR*, Mar 2021.

[105] Alexander Otte, Tim Menzner, Travis Gesslein, Philipp Gagel, Daniel Schneider, and Jens Grubert. Towards utilizing touch-sensitive physical keyboards for text entry in virtual reality. In *IEEE VR*, Mar 2019.

[106] Toni Pakkanen, Jaakko Hakulinen, Tero Jokela, Ismo Rakkolainen, Jari Kangas, Petri Piippo, Roope Raisamo, and Marja Salmimaa. Interaction with WebVR 360° video player: Comparing three interaction paradigms. In *2017 IEEE VR*, 2017.

[107] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. Hand interfaces: Using hands to imitate objects in AR/VR for expressive interactions. In *CHI*, Apr 2022.

[108] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. *CHI*, 2019.

[109] Ioannis Politis, Stephen Anthony Brewster, and Frank Earl Pollick. Evaluating multimodal driver displays under varying situational urgency. *SIGCHI*, 2014.

[110] Ioannis Politis, Stephen Anthony Brewster, and Frank Earl Pollick. To beep or not to beep?: Comparing abstract versus language-based multimodal driver displays. *33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[111] Pornthep Preechayasomboon, Ali Israr, and Majed Samad. Chasm: A screw based expressive compact haptic actuator. In *CHI*, Apr 2020.

[112] Prolific. Prolific's payment principles. [Online; accessed 2025-04-16].

[113] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. Breaking the experience: Effects of questionnaires in VR user studies. *CHI*, 2020.

[114] Elissa M. Redmiles. "Should I worry?" a cross-cultural examination of account security incident response. *IEEE SP*, 2018.

[115] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. An experience sampling study of user reactions to browser warnings in the field. *CHI*, 2018.

[116] Michael Rietzler, Florian Geiselhart, Julia Brich, and Enrico Rukzio. Demo of the matrix has you: Realizing slow motion in full-body virtual reality. In *IEEE VR*, Mar 2018.

[117] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. Exploring interaction fidelity in virtual reality: Object manipulation and whole-body movements. In *CHI*, May 2019.

[118] Hugo Romat, Andreas Fender, Manuel Meier, and Christian Holz. Flashpen: A high-fidelity and high-precision multi-surface pen for virtual reality. In *IEEE VR*, Mar 2021.

[119] Sylvia Rothe, Pascal Pothmann, Heiko Drewe, and Heinrich Hussmann. Interaction techniques for cinematic virtual reality. In *IEEE VR*, Mar 2019.

[120] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian S. Fischer. The emperor's new security indicators. *2007 IEEE Symposium on Security and Privacy (SP '07)*, 2007.

[121] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. Using presence questionnaires in virtual reality. *CHI*, 2019.

[122] Mike Sinclair, Eyal Ofek, Christian Holz, Inrak Choi, Eric Whitmire, Evan Strasnick, and Hrvoje Benko. Three haptic shape-feedback controllers for virtual reality. In *IEEE VR*, Mar 2018.

[123] Statista. Chart: Meta Leads the way in VR Headsets. https://www.statista.com/chart/29398/vr-headset-kpis/, 2023. (Accessed on 08/03/2023).

[124] Steam. Aces High III. https://store.steampowered.com/app/651090/Aces_High_III/, 2023. (Accessed on 07/27/2023).

[125] Steam. ChilloutVR. https://store.steampowered.com/app/661130/ChilloutVR/, 2023. (Accessed on 07/27/2023).

[126] Steam. DCS World. https://store.steampowered.com/app/223750/DCS_World_Steam_Edition/, 2023. (Accessed on 07/27/2023).

[127] Steam. Epic roller Coasters. https://store.steampowered.com/app/787790/Epic_Roller_Coasters/, 2023. (Accessed on 07/27/2023).

[128] Steam. PokerStars VR. https://store.steampowered.com/agecheck/app/886250/, 2023. (Accessed on 07/27/2023).

[129] Steam. RaceRoom Racing Experience. https://store.steampowered.com/app/211500/RaceRoom_Racing_Experience/, 2023. (Accessed on 07/27/2023).

[130] Steam. Rec Room. https://store.steampowered.com/app/471710/Rec_Room/, 2023. (Accessed on 07/27/2023).

[131] Steam. VRChat. https://store.steampowered.com/app/438100/VRChat/, 2023. (Accessed on 07/27/2023).

[132] Steam. War Thunder. https://store.steampowered.com/app/236390/War_Thunder/, 2023. (Accessed on 07/27/2023).

[133] Steam. Welcome to Steam. https://store.steampowered.com/, 2023. (Accessed on 08/01/2023).

[134] Steam. Zaccaria Pinball. https://store.steampowered.com/app/444930/Zaccaria_Pinball/, 2023. (Accessed on 07/27/2023).

[135] Evan Strasnick, Christian Holz, Eyal Ofek, Mike Sinclair, and Hrvoje Benko. Haptic links: Bimanual haptics for virtual reality using variable stiffness actuation. In *CHI*, Apr 2018.

[136] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying wolf: an empirical study of SSL warning effectiveness. In *USENIX*, 2009.

[137] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. TabletInVR: Exploring the design space for using a multi-touch tablet in virtual reality. In *CHI*, May 2019.

[138] Yudai Tanaka, Alan Shen, Andy Kong, and Pedro Lopes. Full-hand electro-tactile feedback without obstructing palmar side of hand. In *CHI*, Apr 2023.

[139] Markus Tatzgern and Christoph Birgmann. Exploring input approximations for control panels in virtual reality. In *IEEE VR*, Mar 2021.

[140] Shan-Yuan Teng, Da-Yuan Huang, Chi Wang, Jun Gong, Teddy Seyed, Xing-Dong Yang, and Bing-Yu Chen. Aarnio: Passive kinesthetic force output for foreground interactions on an interactive chair. In *CHI*, May 2019.

[141] Theophilus Teo, Kuniharu Sakurada, and Maki Sugimoto. Exploring enhancements towards gaze oriented parallel views in immersive tasks. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, Mar 2023.

[142] Kurt Thomas, Jennifer Pullman, Kevin Yeo, Ananth Raghunathan, Patrick Gage Kelley, Luca Invernizzi, Borbala Benko, Tadek Pietraszek, Sarvar Patel, Dan Boneh, and Elie Bursztein. Protecting accounts from credential stuffing with password breach alerting. In *USENIX*, 2019.
[143] Hao Tian, Chanabo Wana, and Xinvu Zhana. A realtime virtual grasping system for manipulating complex objects. In *IEEE VR*, Mar 2018.
[144] Hsin-Ruey Tsai, Chieh Tsai, Yu-So Liao, Yi-Ting Chiang, and Zhong-Yi Zhang. FingerX: Rendering haptic shapes of virtual objects augmented by real objects using extendable and withdrawable supports on fingers. In *CHI*, Apr 2022.
[145] Remo M. A. van der Heiden, Shamsi T. Iqbal, and Christian P. Janssen. Priming drivers before handover in semi-autonomous cars. *CHI*, 2017.
[146] Christof van Nimwegen, Herre van Oostendorp, and Hermina (Tabachneck-) Schijf. Can more help be worse? the over-assisting interface. In *Proceedings of the Conference on Dutch Directions in HCI*, Dutch HCI '04, page 4, New York, NY, USA, 2004. Association for Computing Machinery.
[147] Anthony Vance, David Eargle, Jeffrey L. Jenkins, C. Brock Kirwan, and Bonnie Brinton Anderson. The fog of warnings: How non-essential notifications blur with security warnings. In *SOUPS @ USENIX*, 2019.
[148] Anthony Vance, Brock Kirwan, Daniel Bjornn, Jeffrey L. Jenkins, and Bonnie Brinton Anderson. What do we really know about how habituation to warnings occurs over time?: A longitudinal fMRI study of habituation and polymorphic warnings. *CHI*, 2017.
[149] Mickeal Verschoor, Daniel Lobo, and Miguel A. Otaduy. Soft hand simulation for smooth and robust natural interaction. In *IEEE VR*, Mar 2018.
[150] Patrizia Di Campli San Vito, Gözel Shakeri, Stephen Anthony Brewster, Frank Earl Pollick, Edward Brown, Lee Skrypchuk, and Alexandros Mouzakitis. Haptic navigation cues on the steering wheel. *CHI*, 2019.
[151] Emanuel Vonach, Clemens Gatterer, and Hannes Kaufmann. VRRobot: Robot actuated props in an infinite virtual environment. In *2017 IEEE VR*, 2017.
[152] Martin Vondrácek, Ibrahim M. Baggili, Peter Casey, and Mehdi Mekni. Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses. *CoSe*, 127, 2023.
[153] Lili Wang, Jianjun Chen, Qixiang Ma, and Voicu Popescu. Disocclusion headlight for selection assistance in VR. In *IEEE VR*, Mar 2021.
[154] Miao Wang, Zi-Ming Ye, Jin-Chuan Shi, and Yang-Liang Yang. Scene-Context-Aware indoor object selection and movement in VR. In *IEEE VR*, Mar 2021.
[155] Min Juan Wang, Sus Lundgren Lyckvi, Chenhui Chen, Palle Dahlstedt, and Fang Chen. Using advisory 3D sound cues to improve drivers' performance and situation awareness. *CHI*, 2017.
[156] William Revelle. *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, 2023. R package version 2.3.3.
[157] Graham Wilson, Mark McGill, Matthew Jamieson, Julie R. Williamson, and Stephen A. Brewster. Object manipulation in virtual reality under increasing levels of translational gain. In *CHI*, Apr 2018.
[158] Maximiliane Windl, Anna Scheidle, Ceenu George, and Sven Mayer. Investigating security indicators for hyperlinking within the metaverse. In *SOUPS @ USENIX*, 2023.
[159] Michael S. Wogalter. Handbook of warnings. 2006.
[160] Michael S Wogalter, Dave DeJoy, and Kenneth R Laughery. Organizing theoretical framework: a consolidated communication-human information processing (C-HIP) model. In *Warnings and risk communication*. 1999.
[161] Justin Wu, Cyrus Gattrell, Devon Howard, Jake Tyler, Elham Vaziripour, Daniel Zappala, and Kent E. Seamons. "Something isn't secure, but I'm not sure how that translates into a problem": Promoting autonomy by designing for understanding in Signal. In *SOUPS @ USENIX*, 2019.
[162] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. Eyes-free target acquisition in interaction space around the body for virtual reality. In *CHI*, Apr 2018.
[163] Zhuolin Yang, Cathy Yuanchen Li, Arman Bhalla, Ben Y. Zhao, and Haitao Zheng. Inception attacks: Immersive hijacking in virtual reality systems.
[164] Zishuang Ye. Building trusted paths for web browsers. 2002.
[165] Xin Yi, Yiqin Lu, Ziyin Cai, Zihan Wu, Yuntao Wang, and Yuanchun Shi. GazeDock: Gaze-only menu selection in virtual reality using auto-triggering peripheral menu. In *IEEE VR*, Mar 2022.
[166] Run Yu and Doug A. Bowman. Force push: Exploring expressive gesture-to-force mappings for indirect 3D object manipulation. In *IEEE VR*, Mar 2018.
[167] Yaman Yu, Saidivya Ashok, Smirity Kaushik, Yang Wang, and G. Wang. Design and evaluation of inclusive email security indicators for people with visual impairments. *IEEE SP*, 2023.
[168] André Zenner and Antonio Krüger. Drag:on: A virtual reality controller providing haptic feedback based on drag and weight shift. In *CHI*, May 2019.
[169] Lu Zhao, Yue Liu, Dejiang Ye, Zhuoluo Ma, and Weitao Song. Implementation and evaluation of touch-based interaction using electrovibration haptic feedback in virtual environments. In *IEEE VR*, Mar 2020.
[170] Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *CHI*, Apr 2018.
[171] Yuqi Zhou and Voicu Popescu. Tapping with a handheld stick in VR: Redirection detection thresholds for passive haptic feedback. In *IEEE VR*, Mar 2022.

Appendix A. Warning Design

A.1. Literature Review Search Queries

This section of the appendix presents the query we used in our literature survey to filter the relevant paper.

A.1.1. Query for VR Interaction. `(VR OR virtual) AND (interact*) AND (technique* OR feedback* OR controller* OR experience OR UX OR user interface* OR UI OR select* OR manipulat* OR grasp* OR pointing OR gestures OR sensor* OR design OR haptic OR vibration OR auditory OR visual OR usability)`

A.1.2. Query for Warnings. `(Warning* AND (securit* OR web OR design* OR effectiveness OR usability OR user* OR message* OR dialogue* OR phishing OR SSL OR alert*))`

A.2. Design Exploration References

This section of the appendix includes references to the papers and applications of our survey:

- Tab. 5 lists our full literature survey. The literature survey maps the coding of our design exploration with prior work and VR applications.
- Tab. 6 lists the VR applications analyzed in this work.

Appendix B. User studies

In this section we describe the tools, e.g. tasks, questions, interview scripts used during the user studies.

B.1. Initial response: Remote Screening Questionnaire

Here we report the questionnaire we used for the online screening in the first user study.

TABLE 5: Exploration of the interaction modalities and feedback mechanisms including external references.

	Input		Output				
Immers.	*Phy*	*Vir*	*A/V*	*Sens.*	*UI*	*Beyond*	**References**
Interrupt the experience	●	●	●	○	●	○	[31], Q3, Q7, Q10, S5-6
	○	●	○	○	●	○	[69], [72]
	●	○	○	○	●	○	[80], [165]
	●	○	●	○	●	○	Q1, Q5, Q6, Q8
Enhance the experience	○	●	○	●	○	○	[7]
	○	●	●	○	○	○	[10], [13], [15], [17], [27], [30], [32], [51], [56], [58], [73], [79], [102], [107], [117], [143], [149], [153], [154]
	○	●	●	●	○	○	[11], [139]
	●	○	●	●	○	○	[34], [60], [105], [137], S1, S4, S7, S10
	●	○	●	○	○	○	[50], [59], [141], S9
	●	○	○	●	○	○	[18], [35], [38], [48], [52], [61], [68], [78], [81], [96], [97], [104], [111], [118], [122], [135], [138], [140], [144], [151], [169]–[171]
	●	●	●	○	○	○	[76], [95], [106], [119]
	●	●	●	●	○	○	[77], Q9, S2, S8
	○	●	○	○	○	○	[162]
	●	●	●	●	●	○	Q2, Q4, S3
Part of the experience	○	●	○	○	○	●	[2], [21], [22], [36], [40], [45], [46], [116], [157], [166]
	●	○	○	○	○	●	[39]
	●	○	○	●	○	●	[44], [168]

TABLE 6: Top 10 applications from Quest Store (Q) and SteamVR (S).

Quest top 10	Steam top 10
Q1. YoutubeVR [92]	S1. War Thunder [132]
Q2. GymClass [87]	S2. DCS World [126]
Q3. Instagram [88]	S3. VRchat [131]
Q4. First Steps for Quest2 [85]	S4. Raceroom [129]
Q5. Netflix [89]	S5. Recroom [130]
Q6. Deovr [84]	S6. PokerStarVR [128]
Q7. Bigscreen beta [83]	S7. ChilloutVR [125]
Q8. VR Animation Player [91]	S8. Aces High 3 [124]
Q9. Fitxr [86]	S9. Zaccaria pinball [134]
Q10. Gravity sketch [103]	S10. Epic Roller Coaster [127]

1) Do you have an Oculus Quest? [**a**) Yes **b**) No] **2)** Do you play PC video games on your Quest (via Link Cable or Air Link)? [**a**) Yes **b**) No **c**) Not sure] **3)** Would you like to participate in a virtual reality user study with your Oculus device? [**a**) Yes, I want to participate in the VR study, including the Zoom video interview after the study. We will contact you on Prolific to schedule an appropriate time slot for you and us. The estimated total time is between 60 and 90 minutes. The reward is £24. **b**) Yes, I want to participate in the VR study only. The estimated time is between 45 and 60 minutes. The reward is £14. **c**) No, I do not want to participate in the study.]

B.2. Behavioural response: Demographic Questionnaire

In this section we report the demographic and VR usage questionnaire administered in the prep-phase of the second user study.

1) Age **2)** Gender **3)** Country of Residence **4)** Do you own a VR device? [**a**) Yes **b**) No] **5)** If yes, which one? **6)** How often do you use your VR device? [**a**) Daily **b**) Weekly **c**) Monthly **d**) Rarely **e**) Never] **7)** If not, have you ever used a VR device before? [**a**) Yes **b**) No] **8)** What type of content do you think is most engaging in VR? [**a**) Games **b**) Educational Content **c**) Fitness **d**) Social Experiences **e**) Others] **9)** Have you ever experienced discomfort from using a VR device? [**a**) Yes **b**) No]

B.3. Behavioural and Initial Response: Objects Task

In Tab. 7 we report the explanation of the task we asked participants to complete during the task in both user studies. Grabbable objects included a pen to place in a mug, a sticky note to attach to surfaces or mid-air, a microphone with an on/off button to press, and a light switch to turn on a lamp. Pointer-based interactions involved selecting a plant from a scrollable UI, entering a PIN to log into a smart frame, and choosing a PIN for the blackboard and monitor.

B.4. Initial Response: Questionnaire

In Tab. 8 we present the first study questionnaire. questions. Response need to be selected for each question with a 5 point likert scale displayed in the following order: Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree.

B.5. Initial Response: Interview

In our research, we utilized a semi-structured interview method, detailed here. However, please note

TABLE 7: Task per object.

Object	Instructions
Pen	Put the pen in the mug
Sticky Note	Put the sticky note where you will remember it
Microphone	Turn on the microphone
Light Switch	Light up the lamp with the switch on the desk
Plant Selector	Select a plant
Smart Frame	Use 3494 to login into the frame
Blackboard	Select a pin for the blackboard
Monitor	Select a pin for the monitor

TABLE 8: Questionnaire Questions grouped by category: SUS, NASA-TLX, and Motivation and Comprehension. Participants responded on a 1-10 scale.

Questionnaire Questions
SUS
It was complicated to interact with this object.
I thought this object was easy to use.
I found the interaction with this object was well integrated.
I expected the object to react in a different way.
NASA-TLX
The object reacted how I expected it would.
It was physically draining to interact with this object.
I was satisfied with the interaction with this object.
I worked mentally hard to interact with this object.
I was irritated interacting with this object.
Motivation and Comprehension
I wanted to interact with this object.
I felt safe interacting with this object.

that the flow of conversation in each interview could vary, with some not strictly following the provided structure due to participant responses and context. Here we present the main questions and procedure of the interview we adhered that acted like guidelines during all the interviews.

Warm Up Phase. **(a)** Greetings **(b)** Explanation of the the interview and what will be collected **(c)** Explanation of the IRB and the possibility of withdrawing at any moment.

Broad Task Question. **(a)** What are your experiences with VR applications so far? **(b)** Tell me about the experience you just played. How did you perceive this experience? **(c)** Did you notice some unexpected behaviour or appearance in some objects?

Specific Task Question. **(a)** What do you think that [Warning] meant? **(b)** Would you describe the [Warning] positively or negatively? And Why?

VR Privacy Question. **(a)** What kind of security risks can you imagine in a virtual world? **(b)** And what about risk from a malicious object?

Motivation. **(a)** Do you think the [Warning] you experienced will deter users from interacting with them? Why?

New Warning Design. **(a)** What type of warning do you think will catch a user's attention about a security risk?

Concluding. **(a)** Debrief about the interview motives **(b)** Answering participant question

B.6. Behavioural Response: Interview

Security priming. At the start of the study, i.e. when explaining the study structure and data collection, and during the tutorial we state: "Interact with the secure object and pass the non secure ones within the time limit." In the talk aloud part we let the user explore the objects freely.

Warm Up Phase. **1)** Welcome **2)** Explanation (again) of the audio recording and processing **3)** Explanation of the IRB and the possibility of withdrawing at any moment

Broad Task Question. **1)** Tell me about the experience you just played. How did you perceive this experience? **2)** What were your thoughts when you were asked to interact only with secure objects and avoid insecure ones?

Decision criteria. **1)** How did you identify which objects were secure and which were insecure? **2)** Do you recall some unexpected interaction during the task with some objects? Which one do you recall? **3)** How did you interpret these unexpected interactions? What impact did they have on your decisions to interact or not interact with the objects? **4)** How did the different behaviors make you feel? Did these feelings influence your decisions to use the object or not? **5)** When you think about your decisions to (not) react over the course of the task, did anything change over time?

Metaverse Privacy Question. **1)** What kind of security risks can you imagine in a virtual world? **2)** What do you think can happen interacting with a nonsecure object? **3)** Do you think the [Warning] you experienced will deter users from interacting with them? And why?

Warning Reflection. **1)** Did you feel that some warnings were more effective than others in influencing your decisions? If yes, could you explain why and how? **2)** Now, we will revisit some of the warnings you encountered during the task. I would like you to order them from the one you perceived to be most effective to the one you perceived to be least effective. Can you explain your reasoning behind the order? **3)** Thinking back, would there have been a different type of warning that might have influenced your actions more effectively? Can you describe it?

Concluding. **1)** Debrief about the interview motives **2)** Answering participant question

B.7. Initial Response: Interview Codebook

For the first study this is the merged codebook the researcher based the qualitative analyses on.

VR Experience User's previous encounters or familiarity with virtual worlds or virtual environments. [• Work • Spare time]**In-study feeling** User's emotional response or subjective feelings towards the VR experience in the conducted user study. [• Fine • Issue]**Noticing** User's attention towards the warning signal or users' identification of elements that would have attracted their attention. [• No notice • Noticed as weird • Noticed as normal • Noticed on canvas UIs • Noticed on grabbable objects]**Perception of warning** User's interpretation or understanding of the meaning behind the specific warning signal or visual cue. [• Warning – Not Working: Perceived as ineffective or non-functional, failing to deter interaction with potentially risky objects. • Warning – Working: Considered effective and appropriate, successfully reducing interaction with potentially risky objects.]**Privacy and security risks** User's perception or awareness of the privacy and security risks present in virtual worlds. [• None • Malware • Social Engineering • Profiling • Physical Harm]**Deterrence of proposed warning** Users' opinions on whether indicators or warnings for malicious objects would discourage them from interacting with such objects. [• Yes • No]**Different Design** User-generated ideas or suggestions for alternative designs or formats of warning signals or indicators. [• Enhanced Version • Before Entering • Pause Experience • Sensory Feedback • Virtual Assistant]

B.8. Behaviour Response: Interview Codebook

For the second study this is the merged codebook the researcher based the qualitative analyses on.
VR Expertice Possible expertise levels of participants using VR. [• Novice • Medium • Expert] **Expected unsecure object** User's expectation before entering the study about what an insecure object could be. [• Different interaction: Participants view insecure objects as the ones that behaving abnormally. • Scary: Objects that evoke fear or concern in the user. • Physical Harm: Objects that could potentially cause physical damage.] **Criteria** Criteria users used to identify insecure objects. [• Different interaction from the "normal" one: Avoidance of objects with behavior different from the *No Warning* one. • Strange behavior: Refraining from interacting with oddly behaving objects. • Object based: Non-interaction based on a class of objects perceived as insecure.] **Per-warning feedback** For each warning the perception and motivation used to categorize the warning as secure or insecure. [• Secure: Perceived as safe and trustworthy. • Insecure: Perceived as risky or dangerous. • Effective: Considered useful in indicating risks. • Ineffective: Seen as failing to effectively indicate risks. • Feelings: Which emotional responses the warnings evoked.] **Privacy and security risks** User's perception or awareness of the privacy and security risks present in virtual worlds. [• None • Malware • Social Engineering • Profiling • Physical Harm]

Appendix C. Additional Numeric Details of our Analysis

For comprehensive insights into our research, here, we report the statistical evaluations from both user studies and the codebooks used.

C.1. User Study 1: Model Parameters and Estimates Questionnaire

To examine whether the SUS and NASA items measured the underlying constructs of usability and mental load reliably and were thus, suitable to be used in further analyses, we calculated Cronbach's alpha [26] with the help of the psych package [156], determining internal consistency for the items of each scale. For SUS items Cronbach's alpha was α = .83, and for NASA items Cronbach's alpha was α = .82, indicating good reliability for both scales.

To investigate the effect of warnings, we conducted linear mixed model analyses using the lme4 package [16] and lmerTest package [63]. In the following, we report the maximal random-effect structures that converged or allowed a non-singular fit [14]. We fitted models with 'warning' (i.e. *No Warning* vs. *Blur* vs. *Red Glow* vs. *Scale Down* vs. *Pop-Up*) as fixed effect. We fitted separate models for SUS, NASA, as well as our own two items (i.e. "I wanted to interact with this object"; "I felt safe interacting with this object") as outcome variable, resulting in four different models. To eliminate potential dependencies nested within participants, items, and objects, the SUS and NASA models had by-subject (i.e. participants), by-item (i.e. items of the scale), and by-object random intercepts. The models for our own two items had by-subject (i.e. participants) and by-object random intercepts.
Warnings vs *No Warning* The analysis revealed a significant main effect of 'warning' for SUS items, $\chi^2(4)$ = 41.58, p < .001 and NASA items, $\chi^2(4)$ = 18.37, p = .001, indicating that the warnings differed with respect to their perceived usability and workload from the control condition (i.e. *No Warning* attached to the objects). Moreover, the analysis revealed a significant main effect for the 'want' item, $\chi^2(4)$ = 11.03, p = .026 and the 'secure' item, $\chi^2(4)$ = 14.02, $p < .007$, indicating that the warning had an effect on whether the object was perceived as secure and desirable to interact with. Model parameters and estimates for the models are given in Tab. 9 to Tab. 12.
SUS For usability, the model shows that when objects had the *Blur* (M = 3.93, SD = 1.04) or the *Scale Down* warning (M = 3.56, SD = 1.18) they were perceived as significantly less usable as opposed to *No Warning* attached (M = 4.3, SD = 0.98, see Tab. 9). No other effect reached statistical significance for usability.

TABLE 9: Model Parameters and Estimates for SUS Model. A significance level of p < .0001 is indicated by †, a significance level of $p < .01$ is indicated by ‡.

Parameter	Estimate	SE	t value	95% CI
Intercept	4.30	0.15	29.43	[4.01, 4.59]†
Blur vs. control	-0.34	0.13	-2.53	[-0.60, -0.08]‡
Red Glow vs. control	0.02	0.12	0.21	[-0.20, 0.25]
Scale Down vs. control	-0.77	0.13	-6.02	[-1.02, -0.52]†
Pop-Up vs. control	-0.10	0.13	-0.81	[-0.35, 0.15]

NASA-TLX For mental workload, the model shows that when objects had the *Scale Down* warning (M = 3.99, SD = 1.05), they were perceived as significantly more mentally taxing as opposed to *No Warning* attached (M = 4.47, SD = 0.87, see Tab. 10). No other effect reached statistical significance for mental workload.
Want item For participants' motivation to interact with the object, the model shows that when objects had the *Blur* warning, users reported significantly less desire to interact with the objects (M = 3.44, SD = 1.26) compared to objects with *No Warning* attached (M = 4.48, SD = 0.90, see Tab. 11). No other effect reached statistical significance for the motivation to interact with the object.

Secure item For perceived security, the model shows that objects that *Scale Down* (M = 4.22, SD = 0.81)

TABLE 10: Model Parameters and Estimates for NASA Model. A significance level of $p < .0001$ is indicated by †.

Parameter	Estimate	SE	t value	95% CI
Intercept	4.47	0.18	24.79	[4.12, 4.83]†
Blur vs. control	-0.15	0.13	-1.10	[-0.41, 0.12]
Red Glow vs. control	0.13	0.12	1.12	[-0.10, 0.35]
Scale Down vs. control	-0.50	0.13	-3.90	[-0.75, -0.25]†
Pop-Up vs. control	-0.12	0.13	-0.96	[-0.37, 0.13]

TABLE 11: Model Parameters and Estimates for 'want' item. A significance level of $p < .0001$ is indicated by †, a significance level of $p < .001$ is indicated by ‡

Parameter	Estimate	SE	t value	95% CI
Intercept	4.11	0.21	19.90	[3.70, 4.51]†
Blur vs. control	-0.71	0.23	-3.03	[-1.17, -0.25]‡
Red Glow vs. control	-0.11	0.20	-0.53	[-0.50, 0.29]
Scale Down vs. control	-0.11	0.22	-0.51	[-0.55, 0.32]
Pop-Up vs. control	-0.35	0.22	-1.56	[-0.78, 0.09]

where perceived as significantly less secure and *Red Glow* objects (M = 4.82, SD = 0.39) were perceived as significantly more secure as compared to objects with *No Warning* attached (M = 4.46, SD = 0.67, see Table 12).

TABLE 12: Model Parameters and Estimates for 'secure' item. A significance level of $p < .0001$ is indicated by †, a significance level of $p < .01$ is indicated by ‡

Parameter	Estimate	SE	t value	95% CI
Intercept	4.46	0.12	38.59	[4.23, 4.69]†
Blur vs. control	-0.15	0.14	-1.11	[-0.42, 0.12]
Red Glow vs. control	0.26	0.12	2.17	[0.02, 0.49]‡
Scale Down vs. control	-0.32	0.13	-2.41	[-0.57, -0.06]‡
Pop-Up vs. control	-0.19	0.13	-1.48	[-0.45, 0.06]

Take-away Taken together the results from the questionnaire revealed that both *Blur* and *Scale Down* reduce the perceived usability of an object, while *Scale Down* is in addition also perceived as mentally taxing and not secure. Only *Blur* seemed to have the effect to reduce the motivation to interact with an object. Interestingly, *Red Glow* made objects seem more rather than less secure, contrary to the intention of the manipulation.

C.2. User Study 2: Log Odds, SE, and p-values Warning Effectiveness

We investigate the effectiveness of warnings as measured by whether participants interacted or not by fitting a model, including 'warning' (i.e. *No Warning* vs. *Blur* vs. *Red Glow* vs. *Scale Down* vs. *Pop-Up*), 'object' (i.e. frame vs. pen vs. switch vs. plant) and the two-way interaction of 'warning' and 'object' as fixed effects, using the lme4 package [16] and lmerTest package [63]. The model had a by-subject (i.e. participants) random intercept. We report the maximal random-effect structures that converged or allowed a non-singular fit [14].

The analysis revealed a main effect of 'warning', $\chi^2(4)$ = 286.33, $p < .001$, and a two-way interaction of 'warning' and 'object' $\chi^2(12)$ = 28.21, $p = .005$. To unpack the two-way interaction, simple contrasts were calculated using the R emmeans package based on the R lsmeans package [70]. There were no differences between different objects per warning condition, indicating that each warnings had the same effect on all objects, with the exception of the *Red Glow* warning on the frame object. More specifically, the *Red Glow* warning worked significantly better in preventing participants from interacting with objects for the frame as compared to the pen, the switch, and the plant. Log odds ratio, p-values, and standard error are reported in Tab. 13.

TABLE 13: Comparisons of log odds ratio of participants' interaction frequency split by warnings, p-values are adjusted by Tukey-method.

Warning - Contrast		Log-Odds	SE	p-value
No Warning	Frame vs. Pen	-1.04	0.67	0.41
	Frame vs. Switch	-1.38	0.73	0.23
	Frame vs. Plant	-1.04	0.67	0.41
	Pen vs. Switch	-0.34	0.82	0.98
	Pen vs. Plant	0.00	0.77	1.00
	Switch vs. Plant	0.34	0.82	0.98
Blur	Frame vs. Pen	0.31	0.46	0.90
	Frame vs. Switch	0.51	0.45	0.67
	Frame vs. Plant	0.70	0.45	0.40
	Pen vs. Switch	0.20	0.44	0.97
	Pen vs. Plant	0.38	0.44	0.82
	Switch vs. Plant	0.19	0.43	0.97
Red Glow	Frame vs. Pen	-1.32	0.48	0.03
	Frame vs. Switch	-1.86	0.48	< 0.001
	Frame vs. Plant	-1.32	0.48	0.03
	Pen vs. Switch	-0.54	0.43	0.58
	Pen vs. Plant	0.00	0.43	1.00
	Switch vs. Plant	0.54	0.43	0.58
Scale Down	Frame vs. Pen	0.00	0.44	1.00
	Frame vs. Switch	-0.51	0.45	0.67
	Frame vs. Plant	-0.98	0.47	0.16
	Pen vs. Switch	-0.51	0.45	0.67
	Pen vs. Plant	-0.98	0.47	0.16
	Switch vs. Plant	-0.47	0.49	0.77
Pop-Up	Frame vs. Pen	0.34	0.59	0.94
	Frame vs. Switch	0.64	0.57	0.68
	Frame vs. Plant	-0.70	0.70	0.75
	Pen vs. Switch	0.29	0.54	0.95
	Pen vs. Plant	-1.04	0.67	0.41
	Switch vs. Plant	-1.34	0.66	0.18

Most Effective Warning We determine the most effective warning by calculating simple contrasts between warnings, split by the different objects. Analysis parameters are in Tab. 14.

A consistent pattern emerged for all objects with respect to the *Red Glow* warning. The *Red Glow* warning always led to more rejections compared to the *No Warning* condition, irrespective of the object it was attached to. The *Red Glow* warning also led to more rejections compared to all other warning conditions for all objects, except for the switch object.

A consistent pattern also emerged for all objects with respect to the *Pop-Up* warning. The *Pop-Up* warning did not differ from the *No Warning* condition irrespective of the object it was attached to, both leading to most object interactions. The effectiveness of *Blur* and *Scale Down* warnings however, differed between different objects. The *Blur* warning was effective for all objects, as indicated by more rejections as compared to objects with *No Warning* attached, except for the frame.

The *Scale Down* warning was effective for all objects, as indicated by more rejections as compared to objects with *No Warning* attached, except for the plant selector.

Taken together the results show that the *Red Glow* warning is the most effective warning in preventing participants to interact with objects, leading to most rejections. This effect is particularly pronounced for the frame object. The *Pop-Up* warnings do not differ in their effectiveness from objects without warnings attached, rendering *Pop-Up* warnings as ineffective to prevent interaction. The *Blur* and *Scale Down* warnings might be effective warnings for some objects, although they were not as effective as the *Red Glow* warning.

TABLE 14: Comparisons of log odds ratio of participants' interaction frequency split by objects, p-values are adjusted by Tukey-method.

Object -	**Contrast**	**Log-Odds**	**SE**	**p-value**
Frame	*No Warning* vs. *Blur*	0.90	0.51	0.41
	No Warning vs. *Red Glow*	4.26	0.56	< 0.001
	No Warning vs. *Scale Down*	1.40	0.50	0.04
	No Warning vs. *Pop-Up*	-0.34	0.59	0.98
	Blur vs. *Red Glow*	3.36	0.51	< 0.001
	Blur vs. *Scale Down*	0.51	0.45	0.79
	Blur vs. *Pop-Up*	-1.24	0.55	0.15
	Red Glow vs. *Scale Down*	-2.86	0.49	< 0.001
	Red Glow vs. *Pop-Up*	-4.61	0.59	< 0.001
	Scale Down vs. *Pop-Up*	-1.75	0.53	0.01
Pen	*No Warning* vs. *Blur*	2.25	0.63	0.003
	No Warning vs. *Red Glow*	3.99	0.63	< 0.001
	No Warning vs. *Scale Down*	2.45	0.63	< 0.001
	No Warning vs. *Pop-Up*	1.04	0.67	0.53
	Blur vs. *Red Glow*	1.73	0.44	< 0.001
	Blur vs. *Scale Down*	0.20	0.44	0.99
	Blur vs. *Pop-Up*	-1.21	0.51	0.12
	Red Glow vs. *Scale Down*	-1.54	0.44	0.004
	Red Glow vs. *Pop-Up*	-2.94	0.51	< 0.001
	Scale Down vs. *Pop-Up*	-1.40	0.50	0.04
Switch	*No Warning* vs. *Blur*	2.78	0.69	< 0.001
	No Warning vs. *Red Glow*	3.78	0.69	< 0.001
	No Warning vs. *Scale Down*	2.27	0.70	0.01
	No Warning vs. *Pop-Up*	1.67	0.72	0.14
	Blur vs. *Red Glow*	0.99	0.43	0.14
	Blur vs. *Scale Down*	-0.51	0.45	0.79
	Blur vs. *Pop-Up*	-1.11	0.48	0.14
	Red Glow vs. *Scale Down*	-1.50	0.45	0.007
	Red Glow vs. *Pop-Up*	-2.11	0.48	< 0.001
	Scale Down vs. *Pop-Up*	-0.60	0.49	0.74
Plant	*No Warning* vs. *Blur*	2.64	0.63	< 0.001
	No Warning vs. *Red Glow*	3.99	0.63	< 0.001
	No Warning vs. *Scale Down*	1.47	0.65	0.16
	No Warning vs. *Pop-Up*	0.00	0.77	1.00
	Blur vs. *Red Glow*	1.35	0.43	0.02
	Blur vs. *Scale Down*	-1.17	0.47	0.09
	Blur vs. *Pop-Up*	-2.64	0.63	< 0.001
	Red Glow vs. *Scale Down*	-2.52	0.48	< 0.001
	Red Glow vs. *Pop-Up*	-3.99	0.63	< 0.001
	Scale Down vs. *Pop-Up*	-1.47	0.65	0.16

C.3. Demographics

Table 15 provides an overview of participants' age and sex distributions across both studies. The left section of the Table 16 corresponds to User Study 1, covering participants P1 to P10, while the right section corresponds to User Study 2, listing participants labeled from P11 to P28. For transparency, the full dataset containing additional demographic attributes is available in the public repository referenced in the main text.

TABLE 15: Participants' demographics for User Study 1 and User Study 2.

User Study 1

Category	**N**	**%**
Gender		
Male	32	84.2
Female	6	15.8
Age		
18–24	15	39.5
25–34	11	28.9
35–44	9	23.7
45–54	1	2.6
55–64	1	2.6
>64	1	2.6
Country		
United Kingdom	9	23.7
United States	7	18.4
Portugal	6	15.8
Italy	4	10.5
Poland	4	10.5
Turkey	1	2.6
South Africa	1	2.6
Austria	1	2.6
Vietnam	1	2.6
Canada	1	2.6
Australia	1	2.6
Latvia	1	2.6
Netherlands	1	2.6

User Study 2

Category	**N**	**%**
Gender		
Male	11	61.1
Female	7	38.9
Age		
18–24	9	50
25–34	8	44.4
35–44	1	5.6
>44	0	0
Country		
Germany	18	100.0

TABLE 16: Demographics of interview participants from User Study 1 and User Study 2, including age, sex, and country.

#	Age	Sex	Country	Study
P1	31	M	United Kingdom	1
P2	25	M	Portugal	1
P3	38	M	Italy	1
P4	34	F	United States	1
P5	42	M	United Kingdom	1
P6	32	F	Spain	1
P7	23	F	South Africa	1
P8	22	M	United Kingdom	1
P9	42	F	United Kingdom	1
P10	33	M	United Kingdom	1
P11	27	M	Germany	2
P12	36	F	Germany	2
P13	34	M	Germany	2
P14	22	M	Germany	2
P15	25	F	Germany	2
P16	24	F	Germany	2
P17	20	F	Germany	2
P18	34	M	Germany	2
P19	26	M	Germany	2
P20	20	M	Germany	2
P21	24	M	Germany	2
P22	19	M	Germany	2
P23	24	M	Germany	2
P24	20	M	Germany	2
P25	29	F	Germany	2
P26	22	F	Germany	2
P27	25	M	Germany	2
P28	25	F	Germany	2