International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2024)
S. Hasegawa, N. Sakata and V. Sundstedt (Editors)

The Influence of a Prop Mass on Task Performance in Virtual Reality

L. Thomesse[1,2], J. Cauquis[1,3], E. Peillard[1], L. Dominjon[3], T. Duval[1] & G. Moreau[1]

[1]IMT Atlantique, Lab-STICC, UMR CNRS 6285, France
[2]Université de Lille, France
[3]Clarté, France

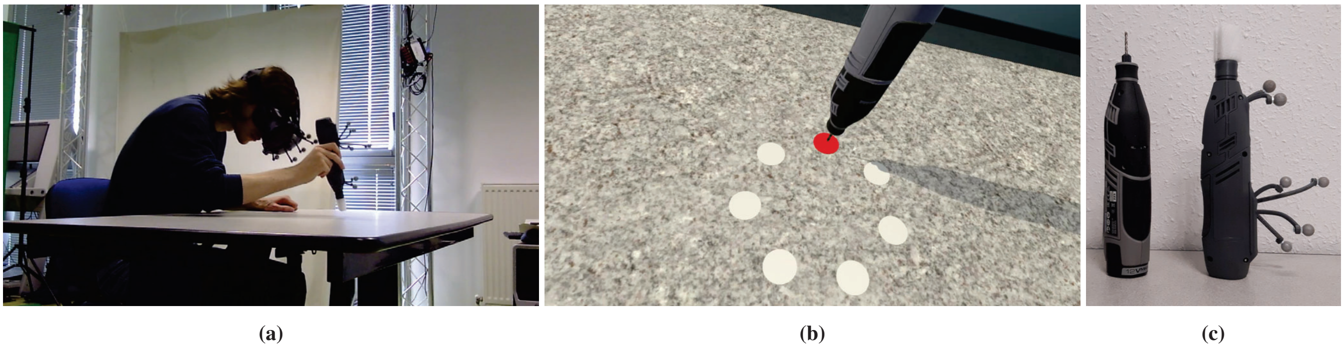

(a) (b) (c)

Figure 1: **(a)** *A participant during the experiment.* **(b)** *His point of view within the VR headset.* **(c)** *The original tool (left) and one of the three replicas used with markers attached to it (right).*

Abstract

As Virtual Reality (VR) applications continue to develop further, many questions persist regarding how to optimize user performance in virtual environments. Among the numerous variables that could influence performance, the mass of the props used within VR applications is particularly noteworthy. This paper thus proposes a user study to investigate the influence of the mass of a prop (a tool replica) on users' performance in a pointing task. A VR within-subject experiment was conducted, with three different weighted replicas, to collect objective and subjective data from participants. Results suggest that the mass of the prop can influence task performance in terms of error-free selection time, number of errors, and subjective perceptions such as perceived difficulty and cognitive load. Indeed, performance was significantly better when using a lighter replica than a heavier one, and subjective user-experience-related metrics were also significantly improved with a light replica. These results help pave the way for additional research on user performance within virtual environments.

CCS Concepts

• ***Human-centered computing*** → ***User studies; Virtual reality;***

1. Introduction

Virtual Reality (VR) has emerged as a transformative medium across a variety of fields, such as surgery [ASB*15] and industry [RKC21, NSRL*20], offering immersive experiences that can enhance skills acquisition.

By offering on-demand repetitions and minimizing risks through controlled environments, VR has demonstrated its effectiveness as a training medium for acquiring procedural knowledge, which involves understanding how to execute a task [APME21]. However, as VR continues to evolve, understanding the various factors influencing task performance within these virtual settings becomes increasingly important [TMN24].

Among those factors, the role of physical properties –such as the weight of props– remains under-investigated. Given that prop mass can potentially influence user dynamics, performance, and fatigue, this study aims to examine how varying masses of tool replicas affect task performance in a VR environment. Please be aware that

DOI: 10.2312/egve.20241354

this paper does not cover the issue of training outcomes transfer from virtual reality to the real world.

A user study is thus proposed to investigate whether the mass of a tool replica influences a user task performance in a virtual environment. In this study, participants performed a pointing task in a virtual environment, utilizing three different tool replicas. The only difference among these replicas was their mass. The performance metrics assessed in this study include task completion time and the frequency of errors made during the task. Questionnaires were also used to evaluate the overall user experience, thanks to subjective metrics like perceived performance, perceived difficulty, perceived tiredness, and perceived cognitive load.

Results from this study provide insights into the design and implementation of props for immersive applications to enhance user performance and potentially virtual training efficacy. Specifically, the results indicated that a lighter replica significantly reduced both the mean selection time and the number of errors compared to a heavier one, suggesting improved performance with the lightest prop. Additionally, participants reported lower perceived difficulty and fatigue, as well as a decreased perceived cognitive load when using the lightest replica. This indicates an overall better user experience when using a prop with a lighter mass.

The remainder of this paper is structured as follows: section 2 describes related work, section 3 presents the user study, while section 4 presents the results. The discussion and conclusion are found in section 5, and section 6 respectively.

2. Related work

2.1. Props in virtual reality

Over recent years, the quality of visual rendering in virtual environments has significantly advanced, allowing for the creation of highly realistic objects [GCM*22]. However, interactions within these environments frequently remain less realistic, confined to using controllers or mid-air gestures, potentially undermining the sense of presence experienced by users, and can lead to reduced user performance [AOB19].

To mitigate this issue, employing appropriate tools for specific tasks can enhance the experience, like using a hammer to drive a nail [SDN*20]. Utilizing tools enables users to interact in a manner that closely resembles real-world scenarios [MSH21]. However, physical tools can pose risks to users or damage their surroundings, and they also limit users to the functions defined by the base tool [MSH21]. Thus, the integration of proxies that mimic the characteristics of real objects is essential. We refer to these as haptic proxies—props that replicate the haptic properties of their virtual counterparts, such as shape, weight, and texture, to provide passive haptic feedback [NZS21]. They can pose less risk to users while offering more affordances than a dedicated tool, as a prop can serve multiple functions. These proxies facilitate more seamless interaction with the properties being manipulated [FBT*20], serving as a cost-effective and efficient alternative to enhance the sense of touch in virtual reality [FBT*20, SDN*20].

In designing these props, one of the most straightforward approaches is to create 3D-printed models that accurately replicate the shape and size of the original objects. Alternatively, Feick et al. in their work with TanGi [FBT*20] propose the use of basic geometric shapes such as cubes or spheres. These primitive shapes can be combined to assemble a form that closely resembles the virtual object the user will interact with. Both of these methods have their advantages and disadvantages, frequently leading to variations between the original tool and the produced replica, such as in the work of Takano et al. where a thin tooth drill is emulated by a controller attached to a stylus [TTF*21]. One of those discrepancies is the mass of the prop, which is often left aside during the conception process or arbitrarily chosen [LKS*20].

2.2. Mass of virtual reality props

Thanks to 3D printing techniques, props can replicate some properties of objects, such as their shape, affordances, or textures [FDHK23]. However, these techniques perfectly illustrate the significant mass differences that can arise between an object and its replica, due to the typically low-density materials employed in their manufacturing process, favoring the creation of lightweight replicas [JCW*22].

Thus, mass is often simulated. There are several techniques to mimic the perception of an object's weight. As our perception of mass is largely influenced by our sense of force, factors such as muscular fatigue can alter how we perceive lifting an object, leading to potential misjudgments about its mass [JH83]. Cognitive factors, such as conceptual expectations and social influences, also play a role in mass perception. For instance, a study manipulated the weight of training golf balls to match those of regular golf balls (golfers usually train with golf balls lighter than the regular ones) [EL98]. They discovered that experienced golfers perceived the training balls as heavier than standard ones, while novices did not notice any difference. On the social influence front, another study explored how dolls representing different genders, ages, and body types, all filled to the same weight, were perceived [Dij08]. The findings indicated that female dolls were perceived as heavier than larger male dolls, though this effect disappeared when participants closed their eyes. These results underscore how past experiences and social contexts can shape weight perception.

The way we grasp an object (using two fingers or an entire hand) along with our sensorimotor memory, also affects mass perception, as these can be influenced by prior physical exertion [QRPC03]. Additionally, weight perception may be altered by an object's perceived characteristics, including its size [CHA91], material [BCG09], mass distribution [SDN*20], shape [Dre94], temperature [KH20], color [DC17], or brightness [VRV19]. This implies that visual properties are crucial in our assessment of an object's mass, which is why larger objects are often assumed to be heavier [LYL*21]. It is thus feasible to simulate weight using only visual feedback. Such techniques are referred to as pseudo-haptic or vision-haptic illusions [LCK*00]. For example, modifying the control-display ratio can create a scenario where, while lifting a virtual bowling ball, the virtual hand movements do not match the real hand's movements, providing a sense of weight [RGGR18]. Similarly, varying the speed at which a virtual object is moved can affect weight perception. A user pulling an object that moves slowly will perceive it as heavier, while a faster-moving object will seem lighter [RHVW15]. Using such software simulation techniques to render the mass of objects enhances user comfort by eliminating the need for cumbersome devices that may restrict movement. Props on the other hand provide the user with real haptic feedback.

Haptic feedback allows for the simulation of weight by applying forces that distort skin or body parts and activate proprioceptive receptors. One method of enhancing the haptic experience is through electric pulsation, which can contract muscles when users are holding an object [LYC*17]. It is essential to calibrate the intensity of these signals to ensure they complement the intended haptic feedback without overshadowing it [LYC*17]. Another way to convey weight perception is through vibrations. Users typically expect stronger vibrations for heavier objects; however, in virtual reality environments, these vibrations may need to be intensified to accurately simulate specific weights [HS05]. Various physical devices are described in the literature to facilitate the simulation of changes in mass. For instance, props can replicate properties of virtual objects, such as air resistance and inertia, through systems like Drag:On [ZK19], liquid simulations for containers [CCC*18], or by adjusting an object's center of gravity to create varied sensations of weight [ZK17]. With a prop in their hands, users can perform tasks within virtual environments.

2.3. Task performance and user experience in virtual reality

In Virtual Reality, task performance is a crucial aspect that directly influences the effectiveness of simulations and applications. In the context of this study, the definition of a task refers to the specific objectives that users are expected to achieve within a virtual environment. Commonly evaluated tasks include those that align with Fitts' Law, such as pointing tasks [BS21], as well as various interactive activities like pick and place [RCO21, SFC*03], path following [FKT02], and assembly tasks [DTIS23]. These tasks serve as benchmarks for understanding how effectively users can interact with virtual elements. Nevertheless, it is noteworthy that VR allows the practice of more complex and industry-related tasks. Indeed, studies dealing with more specific tasks abound in the literature, such as the one of Cooper et al. where participants had to perform a wheel change [CMC*16].

Several metrics are typically employed to assess task performance in VR. The most prevalent metrics include task completion time and the number or frequency of errors made during the task execution, such as in [RCO21, SFC*03, CMC*16]. Task completion time provides insights into the efficiency of user interactions, while error metrics highlight areas where users may struggle, offering a comprehensive view of performance. In a training context, learning outcomes could also be evaluated. It has previously been shown that the mass of a tool replica does not impact such outcomes, under several conditions (industrial-like tasks, for masses ranging between 400g 900g) [CPD*24]. While objective metrics like these are vital for evaluating task efficiency and accuracy, they do not encompass the entire user experience. Thus, subjective metrics are also essential to consider.

User experience (UX) is defined as: *user's perceptions and responses that result from the use and/or anticipated use of a system, product or service* [Int12]. Within VR, UX encompasses users' overall satisfaction and emotional response as they engage with virtual environments. It reflects not only the perceived effectiveness of task performance but also user subjective appreciation, such as engagement and the feeling of presence. Metrics for assessing user experience can include user satisfaction surveys [Bro96], perceived ease of use, and various psychological scales that could measure engagement and immersion. Together, these metrics contribute to a more holistic understanding of how users interact with VR systems, illuminating both performance outcomes and the qualitative aspects of user engagement.

To the best of our knowledge, the influence of the mass of a replica on task performance and user experience in virtual reality has never been investigated.

2.4. Research question and hypotheses

Regarding these previous works, we proposed the following research question: *does the mass of a tool replica influence task performance in a virtual reality environment?* As in his paper Fitts [Fit54] implies that mass has no influence on speed and accuracy for its well-known pointing task, we suppose that we won't be able to invalidate this hypothesis:

- H1: There is no difference regarding task performance between each replica mass variant.

To investigate not only the user performance but also the overall experience within this virtual context, we made some other hypotheses regarding the user feeling of performance, and the user perceived difficulty for the pointing task. These hypotheses are formulated this way:

- H2: A lighter replica results in a higher performance feeling.
- H3: A heavier replica results in a higher perceived difficulty.

They are based on the fact that the heavier the mass, the greater the strength and effort to deploy, which according to us should lead to an increasing perceived difficulty and a lower feeling of performance.

To test these hypotheses, an immersive application dedicated to performing a pointing task in virtual reality was designed with several replicas of a tool with different masses.

3. User study

A user study was thus conducted to investigate the influence of the mass of a prop on performance for a pointing task in an immersive environment. To do so, a virtual environment dedicated to the pointing task was built including a training mode to prepare the participants for the task.

3.1. Participants

G*Power [FELB07] was used to perform an a priori sample analysis, which recommended the inclusion of at least 30 participants. 32 volunteers were thus recruited thanks to internal mailing lists. One was removed, for being an outlier taking too much time in task completion. Indeed, this participant had an error-free mean selection time with a Z-score of more than 3 standard deviations from the average ($Z - score = 4.09$). Among the remaining 31 participants (24 males, 7 females, Ages 20 to 54, $M = 27$, $SD = 8.0$), 14 wore glasses, 8 of whom used them with the VR headset and 1 participant had contact lenses. A Flanders laterality questionnaire [NTLG13] was used to determine participants' handedness. 26 participants were right-handed, 3 were left-handed and the last 2 were ambidextrous. Participants were also asked for their familiarity with VR, on a five-point Likert scale ranging from "1: I have never used it" to "5: I use it for a living" ($M = 2.6$, $SD = 1.0$).

3.2. Experience design

This study is a within-subject experiment, with the mass of a drill-like rotary tool replica as the main independent variable. This mass has three values: 50%, 100%, and 150% of the mass of the original tool (312g, 624g, 936g respectively), resulting in three tool replicas overall. The original tool and one of the replicas can be seen in Figure 1. Our dependent variables are related to the performance of a pointing task that participants were asked to undertake, they are detailed in subsection 3.4. In a virtual environment, they had to use the tip of a replica tool to select targets according to a seven-branch pattern defined by the ISO 9241-411 standard [Int12], which defines the order of appearance of the targets to be selected. These targets have three different sizes (0.01m, 0.02m, 0.04m) and a common distance to the pattern's center (0.1m, 0.15m, 0.25m). An example pattern is presented in Figure 2, as well as a representation of the three distances used from the center of the pattern. A VR view of a pattern during the experiment can be seen in Figure 1.

This leads to 3 (target sizes) × 3 (distances between targets) = 9 possible patterns. As stated, each pattern is made up of 7 targets and will be repeated 4 times. That is 28 targets to select. In all, a participant will have to sequentially select 3 (masses) × 9 (patterns) × 7 (targets per pattern) × 4 (number of repetitions) = 756 targets.

This study design is inspired by McAnally et al. [MWW23].

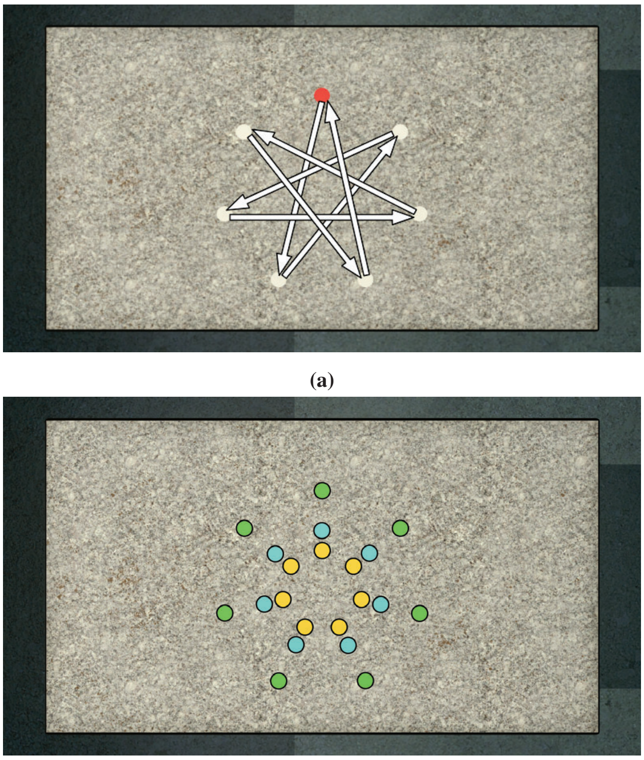
3.3. Procedure

At first, participants were asked to read the participant information sheet and then to read and sign a consent form. They were then briefed regarding the experiment.

Participants were seated in front of a table (Figure 1), where the targets to be selected were depicted. They were asked to select these targets as quickly and accurately as possible. The target to select was indicated thanks to a color change (all targets were in white, except the one to be selected which was in red). They were allowed to rest their elbow on the table during the task or use their available hand to hold their wrist.

The experiment started with a training mode, where participants had no time limit and were able to select as many targets as they wanted with the number of errors not being taken into account. In training mode, there were seven targets on the table as during the main experimental mode, but they were larger (0.1m) and farther away from the center of the table (0.3m). Furthermore, they did not change either in size or distance.

When the participants said they were ready, the experiment began. They used a first replica and then, once they had made the 252 associated selections (*3 (target sizes)* × *3 (distance between targets)* × *7 (targets per pattern)* × *4 (number of repetitions) = 252*), they proceeded to the next 252 selections with another replica and finally completed the last 252 selections with the third replica. The order in which replicas were used was defined by a Latin square of size 3. Similarly, the order in which the 9 patterns of each three selection phases were presented to each participant was randomized.



(a)

(b)

Figure 2: *Top view of the table inside the virtual scene, with* **(a)** *a representation of the seven-branch pattern used (the red target is the one to be selected) along with* **(b)** *a representation of the three distances used from the center of the table for the targets (yellow = 0.1m, cyan = 0.15m, green = 0.25m).*

For every 28 targets, participants were offered a non-mandatory pause of up to 30 seconds. These small pauses were limited to 30 seconds to not lengthen the duration of the study, and their duration was determined thanks to a pilot study. They were kept non-mandatory to prevent frustration from quick participants. A longer mandatory pause was offered every 252 targets, whenever the replica mass was changed. At all those times yellow circles appeared on both sides of the participants. Those circles indicated areas where replicas could be put down, to allow participants to rest. In addition, these areas served as a baseline zone for the study examiner to place the next replicas when they had to be changed. The replicas remained unseen by the participants until the experiment was completed.

During these longer breaks, participants were invited to remove the virtual reality headset and answer two questionnaires. The first was a user-experience-related questionnaire on how the user felt during the experience (5-point Likert scale), asking for their opinion on the perceived mass of the replica, their perceived performance, difficulty, and tiredness during the experience. The second questionnaire was a NASA-RTLX [HS88].

The experiment concluded after completing all the target selections with the three replicas. Participants were then required to fill out one additional demographic information form.

The study workflow is presented in Figure 3.

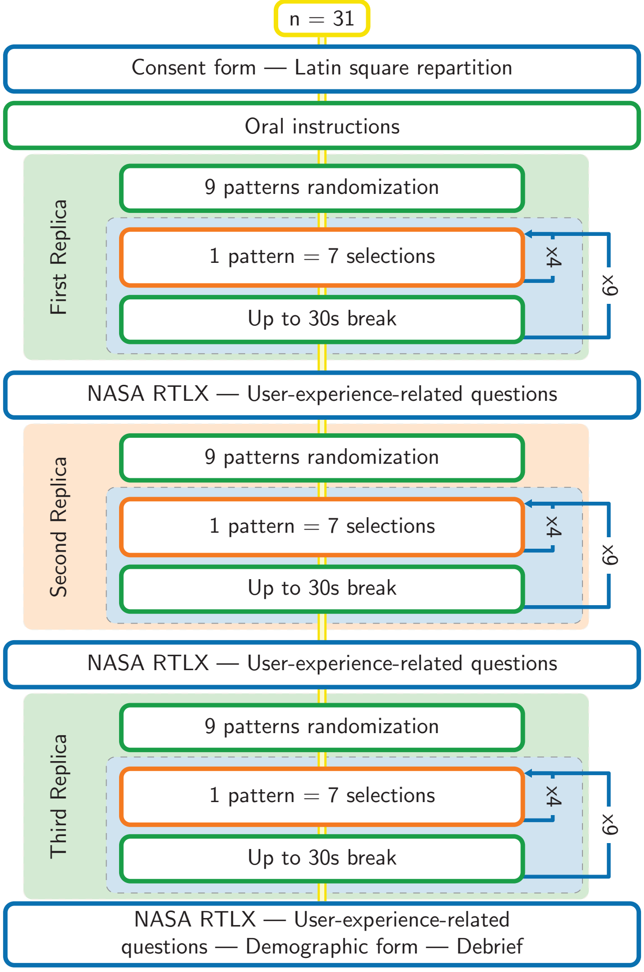


Figure 3: *Study workflow.*

3.4. Dependent variables

Objective variables. The data collected during the VR phases were the time between two target validations (Error-free Mean Selection Time) and the number of errors on each try (Number of Errors). This number refers to the count of collisions outside a target. These two variables encompass the notion of performance which is at the heart of H1. Another objective metric completes this set, the Index of Difficulty, which is defined as:

$$ID = \log_2\left(\frac{A}{W} + 1\right) \quad \text{[Mac92]}$$

where A is the distance between the targets and W is the size of the targets.

Subjective variables. Perceptions of the user were collected thanks to questionnaires. NASA-RTLX questionnaires were completed by each participant to evaluate their cognitive load [HS88]. Another brief 4-question survey was also answered by each participant. All these questions were rated on a 5-point Likert scale to measure Perceived Mass of the replica, Perceived Performance of the user, Perceived Difficulty of the task, and Perceived Tiredness of the user. The questions were:

- How did the replica feel? (1: Light, 5: Heavy)
- How do you judge your performance on this task? (1: Weak, 5: Excellent)
- Did this task seem difficult? (1: Very easy, 5: Very hard)
- Did this task seem tiring? (1: Not tiring, 5: Tiring)

All these subjective variables were used to investigate H2 and H3.

3.5. Apparatus

In this study, an HTC Vive VR headset coupled with an Optitrack system (8 Flex 13 cameras, 120FPS) was used. The tracking system was used to monitor the movements of the headset and the three replicas of the drill-like rotary tool. Replicas were fabricated using a 3D printer. Lead sheets served as ballast for weight distribution, a specific attention was put on balance. The replicas were loaded with three distinct masses: 312g (m=50%), 624g (m=100%, corresponding to the original tool mass), and 936g (m=150%). Each replica was equipped with six passive markers placed to avoid interference with a pen-like grip. One computer (Windows 7 Pro, 24GB RAM, Intel Xeon 2.80GHz) managed the Motive software (version 2.1.2) used to control the Optitrack system, while a second computer was used for participants to complete online surveys and to operate the VR software created with Unity (2022.3.22f1). The application ran at 90fps. The setup included a physical table (counterpart of the virtual one) for passive haptic feedback when the tool entered into contact with a target and a chair to allow participants to sit during the study. To prevent hard clashes between the table and the replicas, some pieces of foam were placed at their bits. The setup and one of the replicas can be seen in Figure 1.

4. Results

4.1. Method

Statistical analyses were performed with Python (Mathplotlib, Pandas, Numpy, SciPy, Modelstats) and a significance level $\alpha = 0.05$. One-way repeated ANOVAs were used when the normality assumption was met. When the normality assumption was not met, an attempt was made to normalize the data through the use of a Neperian logarithm transformation. If it failed, the data was processed using a Friedman test or a Kruskal-Wallis test when the measures were not repeated. Tukey's HSD post hoc tests were used when the normality assumption was met. If normality could not be established, Nemenyi tests followed Friedman tests, and Wilcoxon tests were utilized for Kruskal-Wallis test settings.

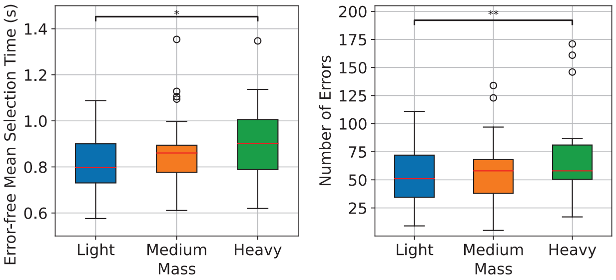
4.2. Objective results

4.2.1. Error-free mean selection time

Between the Error-free Mean Selection Time and the mass of the replicas, an ANOVA analysis showed a significant effect $F_{2,60} = 15,2$, $p = 5.0\times10^{-6} < 0.001$, $\eta^2_G = 0.06$. Post-hoc tests with Tukey's HSD revealed a significant effect between light and heavy-weighted replicas ($p = 0.043$), but not between light and medium ($p = 0.302$) or medium and heavy-weighted replicas ($p = 0.605$). Data are presented in Figure 4 and in Table 1.

4.2.2. Number of errors

Between the Number of Errors and the mass of the replicas, a Friedman test showed a significant effect with a small effect size ($Q = 11.95, p = 0.003, W = 0.19$). Post-hoc tests with Nemenyi's test revealed a significant effect between light and heavy-weighted replicas ($p = 0.002$), but not between light and medium ($p = 0.199$) or medium and heavy-weighted replicas ($p = 0.199$). Data are presented in Figure 4 and in Table 1.



(a) **(b)**

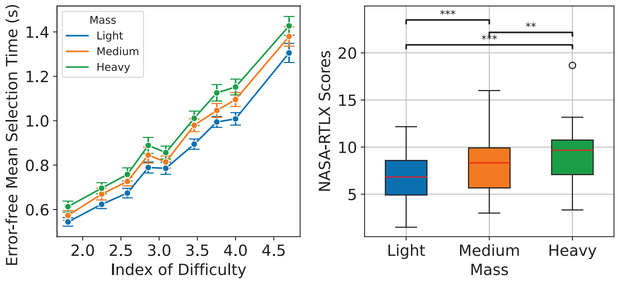
Figure 4: **(a)** *Participants'* Error-free Mean Selection Time *and* **(b)** *total* Number of Errors *according to the mass of the replica. '*' indicates significant difference with p < 0.05 and '**' indicates significant difference with p < 0.01.*

Table 1: Error-free Mean Selection Time *and* Number of Errors *summary statistics.*

Mass	Variable	Mean	SD	Median	IQR
Light	Time	0.816	0.124	0.797	0.170
Medium	Time	0.870	0.150	0.861	0.117
Heavy	Time	0.907	0.162	0.903	0.217
Light	Errors	54.355	28.096	51	37.5
Medium	Errors	56.903	28.923	58	30
Heavy	Errors	66.806	36.576	58	30.5

4.2.3. Index of difficulty

A two-way repeated measures ANOVA was conducted to see if there was an interaction between the Error-free Mean Selection Time and the Index of Difficulty according to the mass of the replica, but no interaction was found. However, the test revealed a significant effect between the Error-free Mean Selection Time and the Index of Difficulty $F_{8,240} = 518.7$, $p = 8.8\times10^{-147} < 0.001$, $\eta^2_G = 0.72$. Next, a linear mixed-effects model was developed to determine if an increase in the Index of Difficulty leads to higher Error-free Mean Selection Time, with results grouped by the replicas' mass. The model revealed that the Error-free Mean Selection Time increased linearly when the Index of Difficulty increased ($coef = 0.299, p < 0.001$). The comparison between replicas' mass showed that there is an offset between light and medium masses ($coef = -0.062, p < 0.001$) and between medium and heavy masses ($coef = 0.047, p < 0.001$). Data are presented in Figure 5 and in Table 2.



(a) **(b)**

Figure 5: **(a)** *Users'* Error-free Mean Selection Time *according to the* Index of Difficulty *grouped by replicas' weight and* **(b)** *users' answer on the* NASA-RTLX *according to the mass of the replicas. '***' indicates significant difference with p < 0.001 and '**' indicates significant difference with p < 0.01.*

Table 2: Error-free Mean Selection Time *according to the* Index of Difficulty *summary statistics.*

ID	Mass	Mean	SD	ID	Mass	Mean	SD
1.807	Light	0.544	0.019	3.459	Light	0.895	0.023
1.807	Medium	0.574	0.022	3.459	Medium	0.979	0.029
1.807	Heavy	0.613	0.025	3.459	Heavy	1.011	0.032
2.248	Light	0.624	0.019	3.755	Light	0.994	0.024
2.248	Medium	0.669	0.026	3.755	Medium	1.046	0.031
2.248	Heavy	0.696	0.025	3.755	Heavy	1.125	0.037
2.585	Light	0.674	0.021	4.000	Light	1.008	0.028
2.585	Medium	0.727	0.020	4.000	Medium	1.095	0.032
2.585	Heavy	0.758	0.030	4.000	Heavy	1.152	0.035
2.858	Light	0.789	0.025	4.700	Light	1.305	0.043
2.858	Medium	0.845	0.036	4.700	Medium	1.379	0.044
2.858	Heavy	0.889	0.036	4.700	Heavy	1.427	0.042
3.087	Light	0.786	0.027				
3.087	Medium	0.814	0.028				
3.087	Heavy	0.856	0.029				

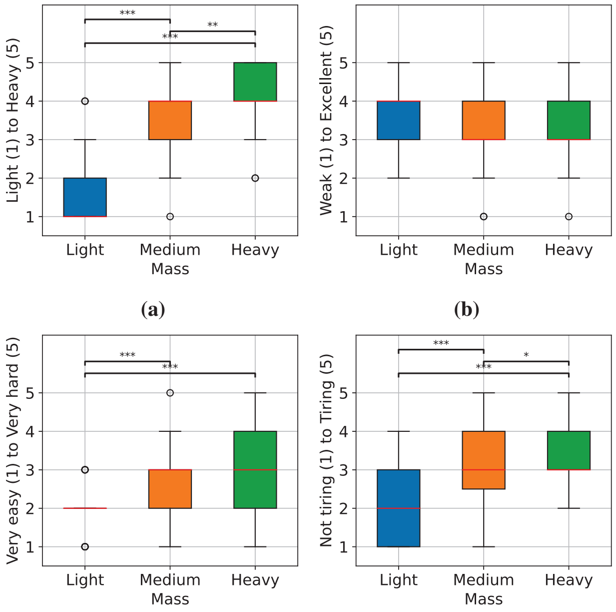
4.3. Subjective results

4.3.1. Likert's scale questionnaires

As normality assumptions were not met and measures not repeated, Kruskal-Wallis tests were performed on the participants' answers to the questionnaires.

For the Likert scales questionnaires, Kruskal-Wallis tests showed significant effects with large effect size on Perceived Mass

($H = 47.7, p < 0.001, \eta^2[H] = 0.51$), on Perceived Difficulty ($H = 18.2, p < 0.001, \eta^2[H] = 0.18$) and on Perceived Tiredness ($H = 23.6, p < 0.001, \eta^2[H] = 0.24$). No effect was observed regarding Perceived Performance. For Perceived Mass, a Wilcoxon's post hoc test revealed a significant effect between light and medium masses ($p < 0.001$), between medium and heavy masses ($p = 0.008$) and between light and heavy masses ($p < 0.001$). Regarding the Perceived Difficulty, Wilcoxon's post-hoc tests indicated a significant difference between light and medium masses ($p = 7.7e^{-05}$) and between light and heavy masses ($p < 0.001$). However, no significant difference was observed between medium and heavy masses ($p = 0.22$). Regarding the Perceived Tiredness, a post-hoc analysis using Wilcoxon's test indicated a significant difference between light and medium masses ($p < 0.001$), between medium and heavy masses ($p = 0.016$), and between light and heavy masses ($p < 0.001$). All Likert scale questionnaire data are presented in Figure 6 and in Table 3.



(a) (b)

(c) (d)

Figure 6: Perceived **(a)** Mass, **(b)** Performance, **(c)** Difficulty, *and* **(d)** Tiredness *according to the mass of the replicas. '*' indicates significant difference with* $p < 0.05$ *and '**' indicates significant difference with* $p < 0.01$ *while '***' indicates significant difference with* $p < 0.001$.

4.3.2. NASA-RTLX questionnaire

Between the answer to the NASA-RTLX questionnaire and the mass of the replicas, a Kruskal-Wallis test showed a significant effect with a moderate effect size ($H = 7.5, p = 0.023, \eta^2[H] = 0.062$). Subsequent analyses using Wilcoxon's tests demonstrated significant differences between light and medium masses ($p < 0.001$), medium and heavy masses ($p = 0.003$), and as well as between light and heavy masses ($p < 0.001$). Data are presented in Figure 5 and in Table 4.

Table 3: Perceived Mass, Performance, Difficulty, *and* Tiredness *summary statistics.*

Prop Mass	Perception	Mean	SD	Median	IQR
Light	Mass	1.742	0.999	1	1
Medium	Mass	3.613	0.919	4	1
Heavy	Mass	4.129	0.885	4	1
Light	Performance	3.806	0.792	4	1
Medium	Performance	3.355	0.950	3	1
Heavy	Performance	3.290	0.902	3	1
Light	Difficulty	1.968	0.657	2	0
Medium	Difficulty	2.742	0.815	3	1
Heavy	Difficulty	2.968	1.169	3	2
Light	Tiredness	2.129	0.957	2	2
Medium	Tiredness	3.097	0.908	3	1.5
Heavy	Tiredness	3.452	0.925	3	1

Table 4: NASA-RTLX *summary statistics.*

Mass	Mean	SD	Median	IQR
Light	6.753	2.923	6.833	3.667
Medium	8.124	3.250	8.333	4.250
Heavy	9.118	3.322	9.667	3.667

5. Discussion

The findings from the experiment present significant insights into the impact of prop mass on user performance in VR environments, highlighting the nuanced interplay between physical attributes of tools and cognitive perception as well as motor task execution. The study presented within this paper, framed around a Fitts-like task, provides evidence that the mass of a tool replica can indeed influence performance, specifically in terms of error-free mean selection time, number of errors, and subjective experiences related to perceived difficulty, fatigue, and cognitive load.

Results indicate an advantage in both speed and accuracy when participants utilized the lighter replica. The observed decline in error-free mean selection times and number of errors with the lighter replica, associated with the mass offsets regarding the positive linear relationship between the error-free mean selection time and the index of difficulty do not support hypothesis H1 that there is no difference regarding task performance between each replica mass variant. These results would rather suggest that the mass of a replica can influence task performance. This has to be nuanced, as it might be related to the task performed, and to the values of the masses used. This is supported by the results, as there is no significant difference between the medium-weighted replica and the two others regarding these performance metrics, suggesting that if a relationship does exist between mass and task performance, it would be non-linear and/or depending on other factors that were not used as variables within this study. Investigating additional variables such as prop shape, texture, and balance could provide more understanding of how these factors may interact with mass to influence performance outcomes. Also, further studies dealing with more complex or industry-specific tasks are needed to ensure the generalizability of those findings.

Moreover, the results indicate no significant impact on perceived performance. The second hypothesis H2 is thus invalidated. This indicates that while there were significant differences in error-free mean selection time and number of errors associated with the mass of the replicas, this did not translate into an overarching perception of improved ability or performance outcomes. This disassociation calls for further exploration into the psychological aspects of performance perception, specifically regarding how physical changes impact self-efficacy in task-oriented environments. Future studies could also investigate physiological data, in an attempt to correlate them to subjective and performance-related metrics.

It is also noteworthy that there were significant differences regarding the perceived mass of the replicas, indicating that participants were able to distinguish the three masses, regardless of their order of utilization. This is in line with just noticeable differences present within literature [LYL*21].

Further analysis revealed that lighter tools induced a diminished feeling of difficulty, partially validating the third hypothesis H3. This outcome may be correlated with the observed reductions in number of errors and mean error-free selection time. The implications of this finding suggest a fundamental relationship between the physical characteristics of tools and the psychological burden they impose on users. This is in line with previous studies suggesting that the physical learning environment (which includes tools used) could be a causal factor for cognitive load [CvP14]. When participants reported experiencing lower cognitive load with lighter replicas, it highlights the significance of ergonomic considerations in the design of virtual tools. Consequently, the design of replicas used in various applications could benefit from an emphasis on reduced mass to enhance user performance and comfort.

Conversely, the heavier replica was associated with a stronger perception of fatigability among participants. Several factors may have contributed to this result such as the nature of the task, the number of repetitions, and the mass itself. Additionally, the duration of the pauses may have been too short when using the heavier mass. This inadequacy may have prevented participants, especially the physically weakest ones, from resting sufficiently before continuing. Such findings are particularly consequential for applications requiring prolonged use of props, as the elevated feeling of tiredness may impair sustained attention and performance over time. This suggests that not only must we consider the effectiveness of replicas in terms of task performance, but also the longer-term implications on user well-being and engagement. Since VR applications typically require extended use, it would be advantageous to investigate how fatigue and performance evolve over time with the repeated use of heavier tools. Longitudinal studies examining the effects of repeated exposure to different prop weights could also shed light on the adaptation processes that users undergo, contributing to improved designs and user experiences.

6. Conclusion

The user study presented in this paper aims to investigate the influence of the mass of a prop on task performance within an immersive virtual application. 31 volunteers used three tool replicas of varying masses to perform a pointing task. Results indicate that the mass of the replica can influence objective metrics used to assess performance (error-free mean selection time and number of errors), as a lighter replica led to a significant increase of performance compared to a heavier replica. Furthermore, some subjective metrics such as perceived cognitive load and task difficulty were also found to be influenced by the mass of the prop used. Indeed, participants experienced a significantly reduced overall user experience when utilizing a heavier replica. Further research is needed to explore the nuanced relationships between prop characteristics and user outcomes, ultimately leading to the optimization of tool replica design and functionality.

Acknowledgments

The authors would like to thank all the participants for the time they spent in the completion of the user study. This work was supported by a French government funding managed by the National Research Agency under the Investments for the Future program (PIA) grant ANR-21-ESRE-0030 (CONTINUUM).

References

[AOB19] ABDI OSKOUIE M., BOULANGER P.: Using Proxy Haptic for a Pointing Task in the Virtual World: A Usability Study. In *Augmented Reality, Virtual Reality, and Computer Graphics*, De Paolis L. T., Bourdot P., (Eds.), vol. 11613. Springer International Publishing, Cham, 2019, pp. 292–299. Series Title: Lecture Notes in Computer Science. `doi:10.1007/978-3-030-25965-5_21`. 2

[APME21] ABICH J., PARKER J., MURPHY J. S., EUDY M.: A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality 25*, 4 (Dec. 2021), 919–933. `doi:10.1007/s10055-020-00498-8`. 1

[ASB*15] ABELSON J. S., SILVERMAN E., BANFELDER J., NAIDES A., COSTA R., DAKIN G.: Virtual operating room for team training in surgery. *The American Journal of Surgery 210*, 3 (Sept. 2015), 585–590. `doi:10.1016/j.amjsurg.2015.01.024`. 1

[BCG09] BUCKINGHAM G., CANT J. S., GOODALE M. A.: Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight. *Journal of neurophysiology 102*, 6 (Dec. 2009), 3111–3118. `doi:10.1152/jn.00515.2009`. 2

[Bro96] BROOKE J.: SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 1996. `doi:10.1201/9781498710411-35`. 3

[BS21] BATMAZ A. U., STUERZLINGER W.: The Effect of Pitch in Auditory Error Feedback for Fitts' Tasks in Virtual Reality Training Systems. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (Mar. 2021), pp. 85–94. `doi:10.1109/VR50410.2021.00029`. 3

[CCC*18] CHENG C.-H., CHANG C.-C., CHEN Y.-H., LIN Y.-L., HUANG J.-Y., HAN P.-H., KO J.-C., LEE L.-C.: GravityCup: A liquid-based haptics for simulating dynamic weight in virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, Nov. 2018), VRST '18, Association for Computing Machinery, pp. 1–2. `doi:10.1145/3281505.3281569`. 3

[CHA91] CHARPENTIER A.: Analyse experimentale de quelques elements de la sensation de poids [experimental analysis: On some of the elements of sensations of weight]. *Archive de Physiologie normale et pathologiques 3* (1891), 122–135. doi provided is referring to a traduction of the original paper. `doi:10.3758/BF03213127`. 2

[CMC*16] COOPER N., MILELLA F., CANT I., PINTO C., WHITE M., MEYER G.: Augmented Cues Facilitate Learning Transfer from Virtual to Real Environments. In *2016 IEEE International Symposium on*

Mixed and Augmented Reality (ISMAR-Adjunct) (Sept. 2016), pp. 194–198. doi:10.1109/ISMAR-Adjunct.2016.0075. 3

[CPD*24] CAUQUIS J., PEILLARD E., DOMINJON L., DUVAL T., MOREAU G.: Investigating Whether the Mass of a Tool Replica Influences Virtual Training Learning Outcomes. *Transactions on Visualization & Computer Graphics (TVCG)* (2024). doi:10.1109/TVCG.2024.3372041. 3

[CvP14] CHOI H.-H., VAN MERRIËNBOER J. J. G., PAAS F.: Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load. *Educational Psychology Review 26*, 2 (June 2014), 225–244. doi:10.1007/s10648-014-9262-6. 8

[DC17] DE CAMP J. E.: The influence of color on apparent weight. A preliminary study. *Journal of Experimental Psychology 2*, 5 (1917), 347–370. doi:10.1037/h0075903. 2

[Dij08] DIJKER A. J.: Why Barbie feels heavier than Ken: The influence of size-based expectancies and social cues on the illusory perception of weight. *Cognition 106*, 3 (Mar. 2008), 1109–1125. doi:10.1016/j.cognition.2007.05.009. 2

[Dre94] DRESSLAR F. B.: Studies in the Psychology of Touch. *The American Journal of Psychology 6*, 3 (1894), 313–368. doi:10.2307/1411644. 2

[DTIS23] DALING L. M., TENBROCK M., ISENHARDT I., SCHLITTMEIER S. J.: Assemble it like this! – Is AR- or VR-based training an effective alternative to video-based training in manual assembly? *Applied Ergonomics 110* (July 2023), 104021. doi:10.1016/j.apergo.2023.104021. 3

[EL98] ELLIS R. R., LEDERMAN S. J.: The Golf-Ball Illusion: Evidence for Top-down Processing in Weight Perception. *Perception 27*, 2 (Feb. 1998), 193–201. doi:10.1068/p270193. 2

[FBT*20] FEICK M., BATEMAN S., TANG A., MIEDE A., MARQUARDT N.: Tangi: Tangible Proxies For Embodied Object Exploration And Manipulation In Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Nov. 2020), pp. 195–206. doi:10.1109/ISMAR50242.2020.00042. 2

[FDHK23] FEICK M., DEGRAEN D., HUPPERICH F., KRÜGER A.: MetaReality: enhancing tactile experiences using actuated 3D-printed metamaterials in Virtual Reality. *Frontiers in Virtual Reality 4* (June 2023), 1172381. doi:10.3389/frvir.2023.1172381. 2

[FELB07] FAUL F., ERDFELDER E., LANG A.-G., BUCHNER A.: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 39*, 2 (May 2007), 175–191. doi:10.3758/BF03193146. 3

[Fit54] FITTS P. M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology 47*, 6 (1954), 381–391. doi:10.1037/h0055392. 3

[FKT02] FEYGIN D., KEEHNER M., TENDICK R.: Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002* (Mar. 2002), pp. 40–47. doi:10.1109/HAPTIC.2002.998939. 3

[GCM*22] GONÇALVES G., COELHO H., MONTEIRO P., MELO M., BESSA M.: Systematic review of comparative studies of the impact of realism in immersive virtual experiences. *ACM Computing Surveys* (Apr. 2022). doi:10.1145/3533377. 2

[HS88] HART S. G., STAVELAND L. E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Hancock P. A., Meshkati N., (Eds.), vol. 52 of *Human Mental Workload*. North-Holland, Jan. 1988, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9. 5

[HS05] HERBST I., STARK J.: Comparing force magnitudes by means of vibro-tactile, auditory, and visual feedback. In *IEEE International Workshop on Haptic Audio Visual Environments and Their Applications* (Oct. 2005), pp. 5 pp.–. doi:10.1109/HAVE.2005.1545654. 3

[Int12] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO/TS 9241-411:2012, 2012. URL: https://www.iso.org/standard/54106.html. 3, 4

[JCW*22] JANDYAL A., CHATURVEDI I., WAZIR I., RAINA A., UL HAQ M. I.: 3D printing – A review of processes, materials and applications in industry 4.0. *Sustainable Operations and Computers 3* (Jan. 2022), 33–42. doi:10.1016/j.susoc.2021.09.004. 2

[JH83] JONES L. A., HUNTER I. W.: Effect of fatigue on force sensation. *Experimental Neurology 81*, 3 (Sept. 1983), 640–650. doi:10.1016/0014-4886(83)90332-1. 2

[KH20] KUHTZ-BUSCHBECK J. P., HAGENKAMP J.: Cold and heavy: Grasping the temperature–weight illusion. *Experimental Brain Research 238*, 5 (May 2020), 1107–1117. doi:10.1007/s00221-020-05794-y. 2

[LCK*00] LECUYER A., COQUILLART S., KHEDDAR A., RICHARD P., COIFFET P.: Pseudo-haptic feedback: Can isometric input devices simulate force feedback? In *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)* (New Brunswick, NJ, USA, 2000), IEEE Comput. Soc, pp. 83–90. doi:10.1109/VR.2000.840369. 2

[LKS*20] LI N., KIM H.-J., SHEN L., TIAN F., HAN T., YANG X.-D., NAM T.-J.: HapLinkage: Prototyping Haptic Proxies for Virtual Hand Tools Using Linkage Mechanism. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, Oct. 2020), UIST '20, Association for Computing Machinery, pp. 1261–1274. doi:10.1145/3379337.3415812. 2

[LYC*17] LOPES P., YOU S., CHENG L.-P., MARWECKI S., BAUDISCH P.: Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2017), CHI '17, Association for Computing Machinery, pp. 1471–1482. doi:10.1145/3025453.3025600. 3

[LYL*21] LIM W. N., YAP K. M., LEE Y., WEE C., YEN C. C.: A Systematic Review of Weight Perception in Virtual Reality: Techniques, Challenges, and Road Ahead. *IEEE Access 9* (2021), 163253–163283. doi:10.1109/ACCESS.2021.3131525. 2, 8

[Mac92] MACKENZIE I. S.: Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact. 7*, 1 (Mar. 1992), 91–139. doi:10.1207/s15327051hci0701_3. 5

[MSH21] MORAN-LEDESMA M., SCHNEIDER O., HANCOCK M.: User-Defined Gestures with Physical Props in Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction 5*, ISS (Nov. 2021), 488:1–488:23. doi:10.1145/3486954. 2

[MWW23] MCANALLY K., WALLWORK K., WALLIS G.: The efficiency of visually guided movement in real and virtual space. *Virtual Reality 27*, 2 (June 2023), 1187–1197. doi:10.1007/s10055-022-00724-5. 4

[NSRL*20] NARANJO J. E., SANCHEZ D. G., ROBALINO-LOPEZ A., ROBALINO-LOPEZ P., ALARCON-ORTIZ A., GARCIA M. V.: A Scoping Review on Virtual Reality-Based Industrial Training. *Applied Sciences 10*, 22 (Nov. 2020), 8224. doi:10.3390/app10228224. 1

[NTLG13] NICHOLLS M. E. R., THOMAS N. A., LOETSCHER T., GRIMSHAW G. M.: The Flinders Handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex 49*, 10 (Nov. 2013), 2914–2926. doi:10.1016/j.cortex.2013.02.002. 4

[NZS21] NILSSON N. C., ZENNER A., SIMEONE A. L.: Propping Up Virtual Reality With Haptic Proxies. *IEEE Computer Graphics and Applications 41*, 5 (Sept. 2021), 104–112. doi:10.1109/MCG.2021.3097671. 2

[QRPC03] QUANEY B. M., ROTELLA D. L., PETERSON C., COLE K. J.: Sensorimotor Memory For Fingertip Forces: Evidence For A Task-Independent Motor Memory. *Journal of Neuroscience 23*, 5 (Mar. 2003), 1981–1986. doi:10.1523/JNEUROSCI.23-05-01981.2003. 2

[RCO21] RICCA A., CHELLALI A., OTRNANE S.: The influence of hand visualization in tool-based motor-skills training, a longitudinal study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (Lisboa, Portugal, Mar. 2021), IEEE, pp. 103–112. doi:10.1109/VR50410.2021.00031. 3

[RGGR18] Rietzler M., Geiselhart F., Gugenheimer J., Rukzio E.: Breaking the Tracking: Enabling Weight Perception using Perceivable Tracking Offsets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2018), CHI '18, Association for Computing Machinery, pp. 1–12. doi:10.1145/3173574.3173702. 2

[RHVW15] Rosa N., Hürst W., Vos W., Werkhoven P.: The Influence of Visual Cues on Passive Tactile Sensations in a Multimodal Immersive Virtual Environment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY, USA, Nov. 2015), ICMI '15, Association for Computing Machinery, pp. 327–334. doi:10.1145/2818346.2820744. 3

[RKC21] Radhakrishnan U., Koumaditis K., Chinello F.: A systematic review of immersive virtual reality for industrial skills training. *Behaviour & Information Technology 40*, 12 (Sept. 2021), 1310–1339. doi:10.1080/0144929X.2021.1954693. 1

[SDN*20] Strandholt P. L., Dogaru O. A., Nilsson N. C., Nordahl R., Serafin S.: Knock on Wood: Combining Redirected Touching and Physical Props for Tool-Based Interaction in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Apr. 2020, pp. 1–13. doi:10.1145/3313831.3376303. 2

[SFC*03] Shing C.-Y., Fung C.-P., Chuang T.-Y., Penn I.-W., Doong J.-L.: The study of auditory and haptic signals in a virtual reality-based hand rehabilitation system. *Robotica 21*, 2 (Mar. 2003), 211–218. doi:10.1017/S0263574702004708. 3

[TMN24] Tusher H. M., Mallam S., Nazir S.: A Systematic Review of Virtual Reality Features for Skill Training. *Technology, Knowledge and Learning* (Jan. 2024). doi:10.1007/s10758-023-09713-2. 1

[TTF*21] Takano T., Takashima K., Fujita K., Guang H., Ikematsu K., Kitamura Y.: A Compact and Low-cost VR Tooth Drill Training System using Mobile HMD and Stylus Smartphone. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka Japan, Dec. 2021), ACM, pp. 1–3. doi:10.1145/3489849.3489933. 2

[VRV19] Vicovaro M., Ruta K., Vidotto G.: Influence of visually perceived shape and brightness on perceived size, expected weight, and perceived weight of 3D objects. *PLOS ONE 14*, 8 (Aug. 2019), e0220149. doi:10.1371/journal.pone.0220149. 2

[ZK17] Zenner A., Krüger A.: Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics 23*, 4 (Apr. 2017), 1285–1294. doi:10.1109/TVCG.2017.2656978. 3

[ZK19] Zenner A., Krüger A.: Drag:on: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2019), CHI '19, Association for Computing Machinery, pp. 1–12. doi:10.1145/3290605.3300441. 3