# Within or Between? Comparing Experimental Designs for Virtual Embodiment Studies

Grégoire Richard[1*] Thomas Pietrzak[1] Ferran Argelaguet[2] Anatole Lécuyer[2] Géry Casiez[1,3]

[1]Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, Lille, France
[2]Inria Rennes - Bretagne Atlantique, Rennes, France [3]Institut Universitaire de France, Paris, France

## Abstract

When designing virtual embodiment studies, one of the key choices is the nature of the experimental factors, either between-subjects or within-subjects. However, it is well known that each design has advantages and disadvantages in terms of statistical power, sample size requirements and confounding factors. This paper reports a within-subjects experiment with 92 participants comparing self-reported embodiment scores under a visuomotor task with two conditions: synchronous motions and asynchronous motions with a latency of 300 ms. With the gathered data, using a Monte-Carlo method, we created numerous simulations of within- and between-subjects experiments by selecting subsets of the data. In particular, we explored the impact of the number of participants on the replicability of the results from the 92 within-subjects experiment. For the between-subjects simulations, only the first condition for each user was considered to create the simulations. The results showed that while the replicability of the results increased as the number of participants increased for the within-subjects simulations, no matter the number of participants, between-subjects simulations were not able to replicate the initial results. We discuss the potential reasons that could have led to this surprising result and potential methodological practices to mitigate them.

**Index Terms:** Virtual Embodiment—Methodology—Latency

## 1 Introduction

The design of a controlled experiment requires determining the number of independent variables to consider, and the number of levels taken by each of them. For any independent variable involved, the experimenter has to choose between a between-subjects design, where each participant is exposed to only one level, and a within-subjects design where each participant tests each level. A between-subjects design is required when investigating individual differences. For example, the effect of gender or age on a dependent variable, and a within-subjects design is required when studying learning or fatigue effects on an independent variable. However, most of the time the choice of an experimental design is not fixed and it is decided by the researcher.

Each design has its own advantages and disadvantages. Between-subjects designs prevent confounding factors such as learning, fatigue, and frustration since each participant is exposed to a single level. The time taken to complete the experiment is also shorter than a within-subjects design. However, between-subjects designs are based on the comparisons between different groups of participants to study the effect of an independent variable. The results can be affected by individual differences making it harder to detect significant differences, reducing the power of the statistical analysis. It also increases the chances to detect differences that can be attributed to differences between the groups instead of differences between the conditions, when the groups are too small and not homogeneous (*e.g.* typing speed when comparing different typing techniques [50]). To reduce the impact of individual differences, between-subjects designs comparatively require larger groups compared to within-subjects designs ($m$ times larger if $m$ is the number of levels of an independent variable). Mixed experiment designs can be used when considering two or more independent variables, where one or more independent variables are administered within-subjects and the others between-subjects, but it is beyond the scope of this work.

Controlled experiments in virtual reality (VR) to study user experience do not suffer an exception. Some contexts require to follow a within-subjects design, for example when studying the effect of VR therapy over different sessions on a phobia [73]. Other contexts, such as scenarios involving surprises/threats [25, 31], typically follow a between-subjects design as being aware of the effect can alter participants' reactions. Other contexts, such as the study of embodiment, offer more freedom in the choice of an experimental design. A key factor of user experience in VR is virtual embodiment, the "sense that emerges when [a body]'s properties are processed as if they were the properties of one's own biological body" [45]. While a majority of virtual embodiment studies are designed following a within-subjects design, a fair amount follows a between-subjects design (see Table 1). For each design, it remains unclear how many participants should be involved in a study to detect an effect if it exists. Furthermore, to get the same statistical power between the two designs, the number of participants required in a between-subjects design may not be directly related to the number of levels. The behavior of participants when answering successive questionnaires may also influence the results, as they may answer the questions depending on their previous answers, for example.

To answer these questions, we evaluate in this paper the influence of the experimental design and the sample size in a virtual embodiment experiment with a single independent variable with two levels. We first report an experiment ($n = 92$), in which participants had to perform a visuomotor task adapted from previous studies [12,46,59]. The experimental design had one within-subjects variable with two levels, synchronous and asynchronous integration. The task and design were chosen to ensure that a significant difference could be observed between the two levels with lower embodiment scores for the asynchronous condition, as it has been previously shown [12,46,59]. We then evaluated the influence of the number of participants and the type of experimental design on the effect size and the ability to find the expected significant difference between the conditions. We used a Monte-Carlo method on different subsets of the collected data to simulate a number of participants ranging from 10 to 92 in the within-subjects design and 12 to 46 per group in the between-subjects design. Our results show that we could find the expected result with more than 95% chance from 33 participants with the within-subjects design, in agreement with the results from the literature. However, we were not able to replicate the expected result with the between-subjects design. We discuss the implications of these findings and provide guidelines on the design of virtual embodiment studies and embodiment questionnaires.

*e-mail: gregoire.richard@inria.fr

## 2 Related Work

This section reviews a number of works defining the notion of embodiment and its different subcomponents before focusing on the impact of latency on embodiment. We follow with an analysis of the implementation of within-subjects and between-subjects designs in VR, and particularly when studying the notion of embodiment.

### 2.1 Virtual embodiment

The notion of embodiment usually refers to motor control and affective attachment over a body [22, 45]. It was historically laid out through the rubber-hand illusion (RHI) paradigm proposed by Botvinick & Cohen [11], and further researched in many studies [27,43,74,83]. Researchers use the concept of virtual embodiment because the sensations that enable embodiment are virtually simulated (i.e. visual and audio stimuli). Although in the original RHI experiment, Botvinick & Cohen showcased the illusion of ownership over a single limb, a number of studies have then explored the possibilities of full-body embodiment [25,46]. Embodiment can be studied in many contexts, through the visual appearance of users' virtual representations, also called *avatars* [2,7], or visuomotor and visuotactile integration [46,68,74,85]. The notion of embodiment has been shown to be correlated to the sense of presence [39], which has been defined by Lee *et al.* [51] as a psychological state where one cannot distinguish a virtual experience from reality.

Usually, the sense of embodiment is decomposed into multiple subdimensions [45, 71]. The main subdimension is body ownership, referring to affective attachment and the feeling of owning the body. The second subdimension is agency, referring to motor control. These two subdimensions are recognized to be the main subdimensions of embodiment. The study by Kilteni *et al.* proposed a third subdimension they called the sense of self-location, which corresponds to the space where one feels located. These three subdimensions form the working model of virtual embodiment that is generally used by researchers. Using questionnaires is often the main measure of the sense of embodiment [36, 71]. Recently, Peck & Gonzalez-Franco [63] proposed a new questionnaire to measure virtual embodiment in an absolute way, basing their model of embodiment on previous work from Kilteni *et al.*, and validating questionnaire items through controlled experiments. In the rest of this study, virtual embodiment and its subcomponents will be based on the work of Peck & Gonzalez-Franco [63].

### 2.2 Impact of latency on embodiment

Introducing discrepancy between an action and its visual feedback can impair and even prevent the emergence of embodiment [11,45, 83]. In such a case, embodiment decreases mainly through the sense of agency and the sense of ownership [33, 83]. The discrepancy can be implemented experimentally by prerecording animations for the avatar movements [46, 74], or by adding artificial latency [41, 85]. In the case of artificial latency, perceived agency decreases around the threshold of 100-150 ms, and does not emerge for delays around 500 ms [29, 54, 85]. Replications of the RHI [3, 67, 83] found that embodiment still emerged at 500 ms, but with sharp drops. A recent study by Caserman *et al.* [15] shows that 250 ms of latency did not impede agency but did reduce ownership. The study by Waltemate *et al.* [85] provides great insights to understand the impact of latency on motor performance and embodiment. They measured embodiment through agency and body ownership but both were assessed through one questionnaire item. They found that embodiment diminished slowly between delays of 45 and 150 ms, while it dropped more sharply afterward, reaching approximately 65% of its optimal value at around 350 ms. It appears that artificial latency has little to no impact on embodiment under 100 ms and that it impedes embodiment up to 500 ms, where it drops sharply.

### 2.3 Experimental protocol designs for embodiment studies

We reviewed 40 studies from the literature regarding virtual embodiment, to better understand when and how each experimental design is used (see Table 1). It is important to stress out that this is not meant to be a systematic review of literature, and more aimed at providing an overview of experimental designs when it comes to embodiment in VR. Besides a few that did not have the choice for their design, it is interesting to note that the others did not provide the motivations for the choice of their experimental design. In regards to embodiment measures, some of these studies implemented both objective and subjective measures, while others used only subjective means. For studies implementing threats [2, 30, 31], the objective measurement mainly consisted of participants' behavior and movement analyses. Virtual replications of the RHI were prone to use proprioceptive drift [11, 41, 46, 74], while others used other means, like the galvanic skin response (GSR) [46, 48]. Studies published before the recent propositions of standardized embodiment questionnaires [35, 63, 71] often proposed their own questionnaires, adapted from the presence questionnaire [87] for the context of their experiment, sometimes being only a few items [85]. More recent studies [1, 4, 13, 68] only implement a subjective measure of embodiment.

Embodiment questionnaires are either administered in the real world after conditions, after removing the VR gear [20,23,25,46,68], or inside the IVE [31, 47, 75, 85]. Some studies mention giving guidelines to participants to answer the questionnaires, mostly to answer based on what they felt and perceived [85], while others gave a shortened questionnaire before the experiment for the participants to assess embodiment under a baseline condition [12, 46, 47, 67]. Some other studies also implemented a familiarization phase [48, 65, 68]. In particular, Bovet *et al.* [13] had participants get familiar with two extreme conditions before the experiment. We noticed that there is no systematic report of effect size for embodiment studies. Some studies report effect sizes for objective measurement [13, 30, 41, 46], but only one study reported effect sizes for embodiment questionnaire answers [32].

From the 40 papers reviewed, 25 of these studies were conducted following a within-subjects design. Reported sample sizes range from 10 participants for Waltemate *et al.*'s study [85], to 99 participants in Peck *et al.* [64]. There were on average 30 participants in the reviewed within-subjects studies. Within-subjects design can lead participants to certain behaviors, such as acting a particular way they think experimenters expect from them. This effect is known as the "demand effect" [70, 86]. The demand effect is especially of concern for psychological experimental research and can induce a bias when administering questionnaires [18].

Among these studies, 12 of them were designed following a between-subjects design [4, 7, 8, 10, 12, 46, 48, 56, 59, 89]. For these studies, the number of participants ranged from 20 for the second experiment of Won *et al.* [89], to 69 participants in the study of Lugrin *et al.* [56]. The size of the groups ranged from 10 participants to 23, for an average of 15 participants per group. The sample size for between-subjects is of importance, because of possible bias caused by possible individual differences. When conducting between studies, if the homogeneity of groups has not been accounted for, it may be possible that results could be explained by differences between groups of participants. This is particularly relevant for studies regarding body ownership and embodiment. For example, Kalckert *et al.* [43] performed a systematic replication of the rubber-hand illusion (RHI) in order to compare visuomotor and visuotactile stimulations. In their review of literature, they noticed two studies [26, 84] that followed different designs and bore different results. Tsakiris *et al.* [84], in a within-subjects experiment, found no difference in ownership between visuomotor and visuotactile integration, while Dummer *et al.* [26], in a between-subjects, found that ownership was higher under visuotactile stimuli than visuomotor.

**Table 1: Summary of the reviewed virtual embodiment studies.**
***Effect is expressed as the difference in mean questionnaire answers between conditions. It is expressed in points, on a 7-point Likert scale.**
****Studies that reported effect sizes for embodiment questionnaire answers. The reported value here corresponds to Cohen's d.**

| Reference | Year | Design | Nb levels | Nb participants | Objective measurements | Subjective Measurements | | Effect (pt)* |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Number of questions | Reference questionnaire | |
| [74] | 2010 | Within | 2 | 14 | Proprioceptive drift | 11 | [11] | 1.29 |
| [80] | 2010 | Between | 2x2x2 | 24 | Heart rate | 8 | - | 1.1 |
| [60] | 2011 | Within | 2 | 22 | Behavior | 11 | [11] | 1.25 |
| [66] | 2011 | Within | 2 | 20 | GSR | 9 | - | 0.75 |
| [7] | 2013 | Between | 2 | 30 | Performance | 4 | - | 1.25 |
| [59] | 2013 | Between | 2x2 | 18 | Heart rate | 7 | [53, 80] | 1 |
| [46] | 2014 | Between | 2x2 | 60 | GSR | 8 | [87] | 1.25 |
| [61] | 2014 | Within | 2x2 | 15 | Proprioceptive drift | 12 | [11, 74] | 0.38 |
| [89] | 2015 | Between | 3 | 53 | Performance | 10 | - | 0.21 |
| [58] | 2015 | Within | 2x2x2 | 43 | Proprioceptive drift/GSR | 13 | [11, 43, 79] | 2 |
| [41] | 2015 | Within | 6 | 13 | Proprioceptive drift | 2 | - | 1.4 |
| [2] | 2016 | Within | 3 | 33 | Behavior and movement | 8 | - | 0.5 |
| [8] | 2016 | Between | 2x3 | 60 | Behavior | 5 | - | 0.75 |
| [40] | 2016 | Between | 2 | 24 | Performance | 12 | [52, 80] | 0.3 |
| [85] | 2016 | Within | 5 | 10 | Performance | 2 | [41] | - |
| [23] | 2017 | Mixed | 2x2 | 48 | GSR | 10 | - | 0.75 |
| [12] | 2017 | Between | 2 | 32 | Performance/Behavior | 7 | [52, 80] | 1 |
| [67] | 2017 | Within | 2x2x2 | 39 | Behavior | 14 | [55] | 0.9 |
| [21] | 2017 | Within | 3x3x2x1 | 23 | - | 6 | [2, 57, 80] | 1.4 |
| [37] | 2017 | Within | 2 | 28 | Performance | 10 | [2, 24] | 0.75 |
| [75] | 2017 | Within | 2x4 | 24 | - | 32 | [87] | - |
| [5] | 2017 | Within | 2x2 | 22 | Behavior | 12 | [55] | 2.5 |
| [30] | 2018 | Within | 3 | 20 | Performance | 9 | [43] | 1.5 |
| [1] | 2018 | Within | 4 | 25 | - | 18 | [2] | 3.75 |
| [32] | 2018 | Within | 3 | 32 | Proprioceptive drift | 10 | [55] | *0.8*** |
| [42] | 2018 | Within | 2 | 10 | Behavior | 16 | [6] | 1 |
| [56] | 2018 | Between | 3 | 75 | Performance | 13 | [72] | 0.35 |
| [77] | 2018 | Within | 5x2 | 41 | Performance | 14 | [11, 58] | 1.15 |
| [13] | 2018 | Within | 3x2 | 21 | - | 2 | - | - |
| [48] | 2019 | Between | 5 | 60 | GSR | 4 | [79] | 2 |
| [38] | 2019 | Within | 3x2 | 34 | Performance | 12 | [2, 24, 37] | 0.25 |
| [76] | 2019 | Within | 3x3 | 37 | Performance | 5 | [77] | 0.55 |
| [81] | 2019 | Within | 2 | 16 | - | 16 | [43] | 0.45 |
| [14] | 2019 | Within | 2 | 30 | Performance | 12 | - | 1.3 |
| [4] | 2019 | Between | 2x2 | 64 | Behavioral | 4 | [9] | 1.1 |
| [31] | 2020 | Mixed | 2x3 | 30 | Behavior | 14 | [35] | 1.5 |
| [65] | 2020 | Between | 2x2 | 60 | Performance | 11 | [35] | 1.5 |
| [64] | 2021 | Within | 2x2x2 | 99 | Performance | 24 | [35] | - |
| [68] | 2021 | Within | 3 | 24 | Performance | 14 | [35] | 0.35 |
| [78] | 2021 | Mixed | 2x2 | 74 | Performance | 4 | [17] | 0.6 |

In summary, even though the majority of virtual embodiment studies are conducted as within-subjects, a non-trivial amount of them implement a between-subjects design. Besides specific contexts, there is no clear motivation to choose either one. Effect sizes for virtual embodiment studies are not massively reported. There is no standard effect size in the literature, and as such, the number of participants varies across studies. It remains unclear how the experimental design can influence virtual embodiment results. The next section presents an experiment and a protocol to compare within-subjects and between-subjects and study the influence of the number of participants.

## 3 Experiment

The main objective of our experiment is to compare experimental designs to study virtual embodiment in VR. We want to study the ability of within-subjects and between-subjects designs to detect effects when they exist. As there exists a direct relationship between the number of participants in an experiment and the ability to detect an effect, we also want to study how results are impacted by the number of participants.

### 3.1 Experimental Design

Our experiment requires the use of a ground truth established from results validated and replicated in the literature. There are a few experiments in VR that have been replicated besides visuotactile synchrony with the RHI [11, 43, 82] and visuotactile/visuomotor synchrony with the study of the latency on embodiment [46, 74]. Previous studies have shown that there is no significant difference in embodiment between visuotactile and visuomotor synchrony [43]. We chose to replicate experiments based on visuomotor synchrony instead of visuotactile synchrony as their experimental setup and procedure are simpler to replicate, thus reducing potential bias. We also chose to test only two experimental conditions, one with minimal latency and the other with a level of latency known to significantly impact embodiment. It is the simplest experiment we can design and it simplifies the statistical analysis. Thus, it can be easily replicated in other studies to further strengthen results and methodological recommendations. Our experiment follows a within-subjects design where each participant experienced two different conditions: SYNCHRO and ASYNCHRO visuomotor (VM) integration. The two conditions are counterbalanced across participants to compensate for possible effects of presentation order and to be able to use the first condition tested by the participants as if we conducted a between-subject design with the two conditions.

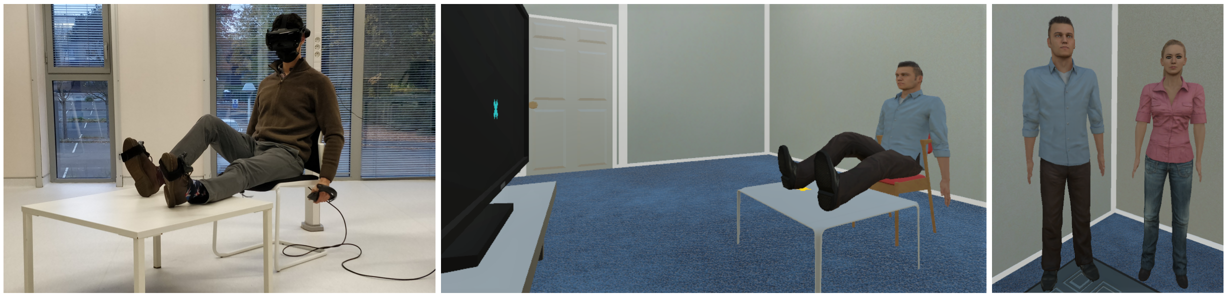

**Figure 1: Experimental setup, with headset, controllers and trackers fitted on a participant (left), the virtual environment with the male avatar (middle), and the two avatars (right).**

### 3.2 Conditions

The synchronous condition (SYNCHRO) integrates participants' movements with no additional delay between their actions in the physical and virtual environments.The asynchronous condition (ASYNCHRO) was designed to add a 300 ms delay compared to SYNCHRO.

Interactive systems have an inherent delay associated with the hardware, system, and processing. This end-to-end latency was the same in both conditions, we measured it with the LagMeter device [16] in the experiment application coded with Unity. It corresponds to the delay between a physical input action and a resulting change on the screen. To do so, the LagMeter first detects and measures when a vibration resulting from a tap on the tactile pad occurs, using a piezoelectric vibration sensor attached to the index finger. Second, it measures when the associated screen response occurs using a photodiode fitted on the left screen of the VR headset. The application switches the screen headset from black to white upon receiving the event from the pad. We collected 50 measures of latency and found an average end-to-end latency of 63 ms (SD=12 ms), similar to what has been found in the literature [16, 49, 85].

The latency for ASYNCHRO was artificially implemented by buffering all input events during a given time . This way ASYNCHRO had a total latency resulting in about 360 ms. According to the literature, the synchronous condition has a latency that is unlikely to impact embodiment (under 100 ms) while the latency in the asynchronous condition is in the 100-500 ms upper range where latency is known to affect embodiment.

### 3.3 Methodology

Participants were first instructed with information regarding the procedure and the experiment: what they were expected to do in the visuomotor task, how to complete the task, and what to focus on during the immersion. Participants were asked to look at the virtual legs when completing the task. Then, they were then asked to sign an informed consent form and answer basic demographic questions. Among those questions, they could express their gender expression, which was used to choose the corresponding avatar.

Participants were instructed to seat on a real chair and place their legs on a physical table (120 cm $\times$ 120 cm $\times$ 50 cm) located in front of them and keep their arm still on each side of the chair (Fig. 1). The chair and table were co-located with a virtual table and a virtual chair of similar shape within the IVE, in order to get coherent passive haptic feedback in the IVE.

The task consisted in replicating patterns with both heels, through leg movements. Instructions for leg movements were displayed on a virtual screen within the IVE, located in front of the participants. Each instruction required participants to repeat the pattern 10 times, and there were 12 different patterns, for a total of 120 movements per condition. The task lasted for about 5 minutes.

Participants went through both conditions within one session. The order of the conditions was counterbalanced. Each experimental condition was followed by an embodiment questionnaire to assess subjective embodiment (see Section 3.6). Participants were asked to remove their HMD and controllers and to complete the questionnaire on a laptop. Participants could not see their answers to the first condition when filling the questionnaire for the second one.

### 3.4 Apparatus

The participants were immersed in the IVE by wearing a *Valve Index* headset, which provided visual feedback and a 6-DOF head tracking. The feet of the participants were tracked with a *Vive Tracker* attached to each of them, and inverse kinematics [69] was used to match the avatar's movements with the actual movements of the participants. The virtual environment was implemented using *Unity3D* (2019.2.12f1) and the *Unity Steam VR* plugin. The experiment ran on an Intel(R) Core(TM) i7-9750H 2.60GHz CPU with 32GB of RAM and a 8GB 2070 RTX GeForce Nvidia GPU. The virtual representations of the participants were two avatars of both gender expressions from the *RocketBox* library [34].

### 3.5 Participants

92 participants volunteered to participate in the experiment, (64 male and 28 female), aged from 22 to 53 ($M = 30.2; SD = 7.7$). They were recruited by word of mouth. Participants had normal or corrected-to-normal vision. The experiment was approved by the local ethics board.

### 3.6 Embodiment Questionnaire

The sense of embodiment was self-reported by participants through an online questionnaire, using seven-point Likert scales. The questionnaire for embodiment was taken from the questionnaire proposed by Peck & Gonzalez-Franco [63]. This questionnaire was validated through a large number of observations and is the most recent. From the 16 items proposed in that questionnaire, one original item was removed (R8 in the original paper), as it could not be adapted for our scenario (Table 2).

## 4 Results

In this section, we first report the main scores for the within-subjects design ($n = 92$), then present an analysis as a between-subjects design, by only considering the first condition for each participant (two groups of $n = 46$). Then we present the simulations of within- and between-subjects designs when manipulating the number of participants[1].

Following the recommendations from [63], questionnaire responses were aggregated to compute the Appearance, Response, Ownership and Multi-sensory sub-scores, and a general score for embodiment and then re-scaled between -3 and +3.

[1]More can be found at ns.inria.fr/loki/WithinBetween

### 4.1 Results for the within-subjects design

As the normality assumptions were violated (Shapiro-Wilk normality test $p < 0.0001$, $W = 0.99$), in the following only non-parametric tests will be considered, with the exception of the following analysis to study interaction effects. Furthermore, to rule out order effects, we performed a repeated-measures ANOVA on Aligned Ranked Transformed data as the data did not follow a normal distribution [88], in which the *order of presentation* was treated as a between-subjects independent variable and the *VM condition* as a within-subjects variable. The analysis did not show any significant effect of the presentation order ($F(1,90) = 1.39$, $p = 0.24$) or any interaction between presentation order and VM condition ($F(1,90) = 0.51$, $p = 0.48$).

Table 3 and Fig. 2 summarize the results, in which Response, Ownership, Multi-sensory and overall embodiment showed significant differences (Wilcoxon signed-rank analysis). The computed effect size was 0.53. The effect size for the Wilcoxon signed rank test was computed using the method proposed by Pallant [62]. More precisely, we used the `wilcox_effsize` function from the `rstatix` R-package [44]. The analysis on Appearance did not reveal a significant difference ($p = 0.07$). These results replicate the previous results in the literature as we observe significantly higher scores of embodiment for the synchronous condition.

### 4.2 Results for the between-subjects design

For the between-subjects analysis, we used a Wilcoxon rank-sum test on the first condition participants were exposed to (Table 4). No significant effect was found between the SYNCHRONOUS and ASYNCHRONOUS conditions on **overall embodiment** ($p = 0.23$) nor any of the sub-components. This analysis failed to replicate the results from the literature, as no significant differences were found.

Table 2: Embodiment questionnaire adapted from Peck & Gonzalez-Franco.

| ID | Questions |
|---|---|
| R1 | I felt out of my body. |
| R2 | I felt as if my (real) body was drifting toward the virtual body or as if the virtual body was drifting toward my (real) body. |
| R3 | I felt as if the movements of the virtual body were influencing my own movements. |
| R4 | It felt as if my (real) body was turning into an "avatar" body. |
| R5 | At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw. |
| R6 | I felt like I was wearing different clothes from when I came to the laboratory. |
| R7 | I felt as if my body had changed. |
| R9 | I felt that my own body could be affected by the virtual body. |
| R10 | I felt as if the virtual body was my body. |
| R11 | At some point, it felt that the virtual body resembled my own (real) body, in terms of shape, skin tone or other visual features. |
| R12 | I felt as if my body was located where I saw the virtual body. |
| R13 | I felt I could control the virtual body as if it was my own body. |
| R14 | It seemed as if I felt the touch of the table in the location where I saw the virtual legs touched. |
| R15 | It seemed as if the touch I felt was caused by the table touching the virtual body. |
| R16 | It seemed as if my body was touching the table. |

Table 3: Summary of the results of the 92-participants within-subjects study.

| Metric | p-value | V | Sync. | | Async. | |
|---|---|---|---|---|---|---|
| | | | Mdn | IQ | Mdn | IQ |
| Response | $< 0.001$* | 743 | 0.0 | $[-0.8, 0.7]$ | $-0.2$ | $[-1.4, 0.6]$ |
| Ownership | $< 0.001$* | 623.5 | 0.83 | $[0.25, 1.42]$ | 0.16 | $[-0.5, 0.92]$ |
| Appearance | $= 0.07$ | 1528.5 | $-0.38$ | $[-1.13, 0.5]$ | $-0.69$ | $[-1.25, 0.38]$ |
| Multi-sensory | $< 0.001$* | 922 | 1.33 | $[1.0, 1.75]$ | 1.17 | $[0.34, 1.5]$ |
| Embodiment | $< 0.001$* | 959.5 | 0.42 | $[0.01, 1.00]$ | 0.08 | $[-0.56, 0.74]$ |

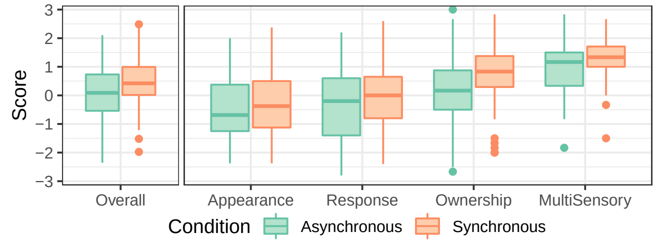


Figure 2: Overall and sub-components embodiment score in the within-subjects design for SYNCHRO and ASYNCHRO conditions.

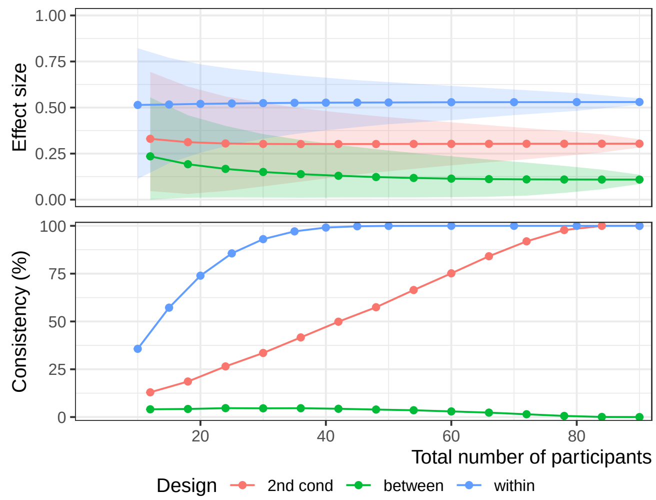


Figure 3: Mean effect size and consistency over the total number of participants in the within and between-subjects virtual experiments. The shaded regions for the effect size correspond to the 5th and 95th percentile of the values obtained with the simulations. "2nd cond" refers to the results when considering only the second condition of each participant, similar to the way we created the between-subjects design with the first condition.

### 4.3 Influence of the number of participants over the results for within-subjects studies

We analyzed the results from our 92-participants within-subjects study to evaluate the influence of the number of participants on the results regarding embodiment. Since the overall differences in embodiment scores were significant for the within-subject design, only the overall embodiment scores were considered in this analysis.

From the original 92-participants within-subjects study, we created virtual experiments by randomly selecting different subsets of participants according to a target sample size. As simulating all possible combinations is computationally too expensive, we used a Monte-Carlo method by creating several virtual experiments for each sample size, to get accurate estimations. To determine the right amount of simulations, we simulated between 10 and $1,000,000$ virtual experiments for 12-participant within-subjects and between-subjects. The 12-participants case is considered as a worst-case scenario as the variance is expected to be maximal. The results showed that $10,000$ simulations were enough to reach precise approximations for consistency and effect size. To keep a safety margin while remaining reasonable in terms of computational cost, we chose to run $100,000$ simulations for each sample size.

For each virtual experiment, a Wilcoxon signed-rank test was run between the SYNCHRO and ASYNCHRO conditions. As a significant effect was found in the original within-subjects study with the 92 participants, and according to the literature, we consider that any virtual experience with a significant difference ($p < 0.05$) between the two conditions, with the mean embodiment scores being higher in the SYNCHRO condition than the ASYNCHRO, is consistent with the original results. For each sample size, we computed the percentage of virtual experiments that were consistent with the original results and effect size (mean, 5$^{th}$ percentile and 95$^{th}$ percentile).

The blue line in Fig. 3 (bottom) shows the evolution of the consistency when varying the number of participants for the within-subjects virtual experiments. As can be seen from the figure, and as expected, the consistency of the results increases together with the sample size. Consistency reaches 95% at 33 participants and is virtually 100% beyond 40 participants. Moreover, the blue line in Fig. 3 (top) shows the evolution of the effect size as the sample size increases. While the mean effect size is almost constant (0.53) no matter the effect size, the 5$^{th}$ - 95$^{th}$ percentile range gets narrower as the sample size increases.

### 4.4 Influence of the number of participants over the results for between-subjects studies

To study the influence of the number of participants for the between-subjects study, we used the same method as described in the section above. We selected only the first condition of each participant to obtain a 92-participants between-subject study in which both groups had 46 participants.

The results for the full between-subjects design were not significant (*cf.* Sect. 4.2). We define *consistency* as based on the results of the 92-participants within-subjects study and on the literature, i.e. that there is a significant difference in embodiment between the SYNCHRONOUS and ASYNCHRONOUS conditions. Thus, in this case, the original 92-participants between-subjects study is considered inconsistent. The green lines in Fig. 3 show the evolution of the effect size and consistency respectively. As it can be expected from the results of the 92-participants between-subjects study, consistency values remain close to zero percent for all simulations. Similarly, the mean effect size is lower (converges to 0.11) with a larger interquartile range for small sample sizes.

During the data analysis, we observed an interesting behavior for the results when only considering the second condition of the original 92-participants study. When considered alone, the second condition can be considered as a between-subjects design, although the results could be strongly biased by the first condition. Nevertheless, we report them as they provide interesting insights and match the initial hypothesis we had for the between-subject designs. The red lines in Fig. 3 show the evolution of the effect size and consistency respectively when considering the data from the second condition. Interestingly, the consistency of the results almost increases linearly according to the number of participants and reaches the 100% of consistency at around $n = 84$. The evolution of the effect size also followed a similar pattern as the previous simulations converging towards 0.30.

**Table 4: Summary of the results of the 92-participants between-subjects study.**

| Metric | p-value | W | Sync. Mdn | Sync. IQ | Async. Mdn | Async. IQ |
|---|---|---|---|---|---|---|
| Response | $> 0.9$ | 1048.5 | 0.0 | $[-0.95, 0.4]$ | 0.0 | $[-1.34, 0.8]$ |
| Ownership | $> 0.2$ | 920 | 0.58 | $[0.04, 1.46]$ | 0.5 | $[-0.17, 1.17]$ |
| Appearance | $> 0.2$ | 909 | $-0.38$ | $[-1.09, 0.34]$ | $-0.63$ | $[-1.25, 0.22]$ |
| Multi-sensory | $> 0.1$ | 886.5 | 1.5 | $[0.83, 1.96]$ | 1.17 | $[0.67, 1.5]$ |
| Embodiment | $> 0.2$ | 905 | 0.38 | $[-0.02, 0.93]$ | 0.21 | $[-0.43, 0.84]$ |

## 5 DISCUSSION

There were several objectives in this study, the main one being the comparison of within-subjects and between-subjects designs.

### 5.1 Within-subjects designs

Our first objective was to replicate results regarding the impact of latency on virtual embodiment, in a visuomotor task inspired by the literature [12,46,59]. We considered synchronous and asynchronous integration as several works have shown embodiment is impaired when participants perceive visuomotor and/or visuotactile discrepancy [11,41,54,85]. Considering the results of the 92-participants within-subjects study, participants were significantly affected by visuomotor discrepancies during the asynchronous condition (latency around 360 ms), although the decrease of virtual embodiment scores compared to the synchronous condition was moderate compared to other works [41,85]. This effect was also observed by three of the four sub-components proposed by Peck and Gonzalez-Franco [63]. These results tend to validate both the protocol to elicit embodiment (the visuomotor task based on leg movements) and the value of delay to impede embodiment (360 ms).

The reported effect size for the 92-participants within-subjects study was 0.53. According to Cohen [19], the effect size for Wilcoxon signed ranked test can be interpreted the same way as Cohen's *d*, so an effect size of 0.53 is considered medium. This appears coherent when considering that latency starts to affect embodiment around 100 ms and severely impacts it around 500 ms. As a result we could expect to measure a small effect size around 100 ms and a large effect size around 500 ms and a medium effect-size for 360 ms. However, effect sizes are rarely reported in the literature for virtual embodiment studies. The effect size measures the magnitude of an effect but it usually requires results from different studies on a given phenomenon to be able to establish what values can be considered as low, medium, or high. In its simplest form, the effect size can be considered as the difference of the mean or median values between two conditions. In our case, the difference of embodiment score between the two conditions is equal to 0.34 points on a 7-point Likert scale, which can be considered as small considering the differences obtained in other studies (Table 1). However such approximation does not take into account the variance of the data as methods to compute actual effect sizes do. In addition the results reported in the literature did not necessarily use the same questionnaires. Future embodiment studies also reporting effect size can help characterize the magnitude of ours.

In order to account for the impact of the sample size in the potential outcome and the replicability of the results, we simulated virtual experiments with sample sizes varying from $n = 10$ to $n = 90$. The results, as hypothesized, showed an increase of the replicability as the sample size increased, reaching the 100% of consistency when $n = 40$. 75% of replicability was reached for virtual experiments with 20 participants and up to 33 participants were required to reach 95%. These results show, considering the within-subjects experimental design, that 40 participants would have ensured an optimal number of participants to observe the effect of latency on embodiment.

### 5.2 Between-subjects designs

Contrary to our initial hypothesis, the results for the 92-participants between-subjects designs failed to replicate the results obtained for the within-subjects design. Assuming that the effect exists, as the within-subjects virtual experiments were able to detect, even reaching a 100% of consistency around 40 participants, and considering that the data set was the same, we hypothesize that the between-group design failed to detect an effect that does exist. A power analysis was run post-experiment, to determine the sample size that would have been required to detect an effect. This was done using the G*Power software [28], and sample size was computed using the effect size found in the initial 92-participants within-subjects study. The analysis revealed that a sample size of 170 was needed for the between-subjects, whereas the within-subjects yielded a required sample size of 44. This puts our results in perspective: the sample size for a between-subjects design, even with a factor with two levels, might need four times the number of participants in a within-subjects design to detect an effect. It is important to note that these results are heavily dependent on the effect size. Previous studies about embodiment, conducted using between-subjects designs, have also shown it was possible to find significant differences for effect sizes equivalent or smaller than ours, with less participants [56, 89]. While the task and context are different, there might be factors that are overlooked when conducting between-subjects studies regarding embodiment in VR, such as individual differences, like gender or age. Considering the results for the 92-participants between-subjects design, it was not surprising that the virtual experiments with a smaller number of participants also failed to consistently detect the effect. These results are in line with the work of Kalckert *et al.* [43] which noted that results in the RHI could be different based on the experimental design.

Considering that no differences were found for the first condition between the group who started with the synchronous condition and the one who started with the asynchronous condition, the observed differences could only be explained by the second condition experienced by the participants and the way they answered the questionnaires. The analysis of the second condition between-group virtual experiments might reinforce this potential explanation, as consistency steadily increased as the number of participants increased. However, it is important to stress that the second condition cannot be considered as a pure between-subjects design as participants were exposed to both conditions.

A possible explanation for the discrepancies between the first and the second condition is that participants first answered the questions in an approximate way. They would then have a better understanding how to answer it, and to which aspects of the task they need to pay attention when performing the second condition. Then, when filling the questionnaire a second time for the other condition, they would answer the questions in a relative way, based on what they remember of their answers to the previous questions. For instance, previous studies regarding the RHI [3] have made the argument that it is the comparison among conditions that are of interest. This tends to be further supported by informal feedback from a few participants regarding the way they interpreted the questionnaire, and the strategy they came up with. Six participants mentioned at the end of the experiment having focused more on what was going on during the second task after having answered the embodiment questionnaire once and getting an "idea of what [they are] expected to do", as one of them put it. One participant, after asking for help understanding a particular questionnaire item for the first questionnaire, told the experimenter that it was "easier the second time, because [they] have a reference now". Further studies, in the form of semi-structured interviews after filling the questionnaires, would be required to better understand the strategy participants follow when answering. If this type of strategy is confirmed, it would mean protocols need to be refined. For between- and within-subjects studies, a solution would be to help participants answer questionnaires in an absolute manner, by presenting for instance examples of situations for the two extremes of the scale for each question. Another one would be to set up a training phase where the participants would discover a baseline condition, and the questionnaire before going through the experimental condition. The direct downside of such a solution is that participants would be aware of the questionnaire, just like in a within-subjects design. For within-subjects studies other solutions could be envisioned. For example, the answers to the questionnaire in the previous condition can be presented to the participants to help them answer to the current condition in a relative way. Another solution would be to provide questionnaires with scales that are relative to the previous condition.

The fact that a number of embodiment studies have found significant differences in between-group designs could be explained by a smaller effect size found in our study. Overall our results suggest that a within-subjects design should be preferred over a between-subjects design in embodiment studies when this is an option.

## 6 Limitations

The main limitation of this study concerns the generalization of our results. They are indeed related to the effect size that depends on the value of latency used in the asynchronous condition. Higher values of latency would increase the effect size and increase the chance of detecting the effect using a between-subjects design. There is also the possibility that the sample size for the between-subjects was too small, as post-experiment power analysis showed a 170-participants required sample size. Nevertheless, the differences found among within- and between-subjects designs might be related to the actual measure of embodiment.

Indeed, embodiment can be measured using subjective measures with questionnaires but also objective measures. We only used subjective measures to keep the experiment simple, and in particular, one embodiment questionnaires measuring absolute values of embodiment. Other embodiment questionnaires could have been used [71], which could have provided different results. Future work could combine subjective measures with objective ones to better characterize the ability to detect effects when they exist.

Furthermore, due to the highly heterogeneous experiments assessing embodiment, embodiment questionnaires need to be adapted to each experiment. We adapted an embodiment questionnaire to the best of our knowledge but the phrasing of the questions could introduce some noise in the answers to the questions. Further work comparing different adaptations of a questionnaire to a given experiment could help measure the impact of the adaptation of the questionnaires.

## 7 Conclusion

This paper presented a methodology to compare within-subjects and between-subjects designs. From a ground truth established by the literature, our study managed to show that depending on the experimental design and the effect size, it was possible to fail to detect an effect that exists. While we observed the expected effect of latency on embodiment with a within-subjects design, we did not detect it with a between-subjects design. Overall, within-subjects design is much more able to detect small to medium effect sizes, with a smaller required number of participants. We proposed explanations of this phenomenon. Participants most likely answered the second questionnaire relatively to the first one, and the measured difference in embodiment is due to this difference, not to the absolute values. Therefore our recommendation for embodiment studies is to use a within-subjects design. However, we acknowledge that some experimental tasks require a between-subjects design. Hence, we plan to conduct semi-structured interviews to better characterize the way participants answer embodiment questionnaires. Then we will study calibration methods that could enable participants to give better assessments of embodiment in between-subjects design studies. It could take the form of a baseline condition with an embodiment questionnaire, which could give participants the necessary perspective to answer less arbitrarily. Another possibility is to provide participants an idea of what extreme levels of virtual embodiment could feel like, similarly to Kokkinara *et al.* [46].

## Acknowledgments

This work was funded by the Inria Avatar project.

## References


[1] A. Andreasen, N. C. Nilsson, and S. Serafin. Agency enhances body ownership illusion of being a virtual bat. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 505–506. IEEE, 2018.

[2] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *Virtual Reality (VR), 2016 IEEE*, pp. 3–10. IEEE, 2016.

[3] T. Asai. Agency elicits body-ownership: proprioceptive drift toward a synchronously acting external proxy. *Experimental brain research*, 234(5):1163–1174, 2016.

[4] L. Aymerich-Franch, S. Kishore, and M. Slater. When your robot avatar misbehaves you are likely to apologize: an exploration of guilt during robot embodiment. *International Journal of Social Robotics*, pp. 1–10, 2019.

[5] L. Aymerich-Franch, D. Petit, G. Ganesh, and A. Kheddar. Object touch by a humanoid robot avatar induces haptic sensation in the real hand. *Journal of Computer-Mediated Communication*, 22(4):215–230, 2017.

[6] J. O. Bailey, J. N. Bailenson, and D. Casasanto. When does virtual embodiment change our minds? *Presence*, 25(3):222–233, 2016.

[7] D. Banakou, R. Groten, and M. Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31):12846–12851, 2013.

[8] D. Banakou, P. D. Hanumanthu, and M. Slater. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 10:601, 2016.

[9] D. Banakou and M. Slater. Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, 111(49):17678–17683, 2014.

[10] I. Barberia, R. Oliva, P. Bourdin, and M. Slater. Virtual mortality and near-death experience after a prolonged exposure in a shared virtual reality may lead to positive life-attitude changes. *PloS one*, 13(11):e0203358, 2018.

[11] M. Botvinick and J. Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756, 1998.

[12] P. Bourdin, I. Barberia, R. Oliva, and M. Slater. A virtual out-of-body experience reduces fear of death. *PloS one*, 12(1):e0169343, 2017.

[13] S. Bovet, H. G. Debarba, B. Herbelin, E. Molla, and R. Boulic. The critical role of self-contact for embodiment in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1428–1436, 2018.

[14] D. Burin, K. Kilteni, M. Rabuffetti, M. Slater, and L. Pia. Body ownership increases the interference between observed and executed movements. *PloS one*, 14(1):e0209899, 2019.

[15] P. Caserman, M. Martinussen, and S. Göbel. Effects of end-to-end latency on user experience and performance in immersive virtual reality applications. In *Joint International Conference on Entertainment Computing and Serious Games*, pp. 57–69. Springer, 2019.

[16] G. Casiez, T. Pietrzak, D. Marchal, S. Poulmane, M. Falce, and N. Roussel. Characterizing latency in touch and button-equipped interactive systems. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 29–39, 2017.

[17] E. A. Caspar, A. Cleeremans, and P. Haggard. The relationship between human agency and embodiment. *Consciousness and cognition*, 33:226–236, 2015.

[18] G. Charness, U. Gneezy, and M. A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8, 2012.

[19] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[20] F. Danieau, A. Lecuyer, P. Guillotel, J. Fleureau, N. Mollet, and M. Christie. Enhancing audiovisual experience with haptic feedback: a survey on hav. *IEEE transactions on haptics*, 6(2):193–205, 2012.

[21] F. Danieau, T. Lopez, N. Mollet, B. Leroy, O. Dumas, and J.-F. Vial. Enabling embodiment and interaction in omnidirectional videos. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 697–702. IEEE, 2017.

[22] F. De Vignemont. Embodiment, ownership and disownership. *Consciousness and cognition*, 20(1):82–93, 2011.

[23] H. G. Debarba, S. Bovet, R. Salomon, O. Blanke, B. Herbelin, and R. Boulic. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one*, 12(12):e0190109, 2017.

[24] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic. Characterizing embodied interaction in first and third person perspective viewpoints. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–72. IEEE, 2015.

[25] D. Dewez, R. Fribourg, F. Argelaguet, L. Hoyet, D. Mestre, M. Slater, and A. Lécuyer. Influence of personality traits and body awareness on the sense of embodiment in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 123–134. IEEE, 2019.

[26] T. Dummer, A. Picot-Annand, T. Neal, and C. Moore. Movement and the rubber hand illusion. *Perception*, 38(2):271–280, 2009.

[27] H. H. Ehrsson. 43 the concept of body ownership and its relation to multisensory integration. *The New Handbook of Multisensory Process*, 2012.

[28] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[29] N. Franck, C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, T. d'Amato, and M. Jeannerod. Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, 158(3):454–459, 2001.

[30] R. Fribourg, F. Argelaguet, L. Hoyet, and A. Lécuyer. Studying the sense of embodiment in vr shared experiences. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 273–280. IEEE, 2018.

[31] R. Fribourg, E. Blanpied, L. Hoyet, A. Lécuyer, and F. Argelaguet. Influence of threat occurrence and repeatability on the sense of embodiment and threat response in vr. In *International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments (ISMAR)*, p. 9p, 2020.

[32] J. Frohner, G. Salvietti, P. Beckerle, and D. Prattichizzo. Can wearable haptic devices foster the embodiment of virtual limbs? *IEEE transactions on haptics*, 2018.

[33] S. Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

[34] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano Gómez, S. Orts-Escolano, et al. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality*, 1(article 561558):1–23, 2020.

[35] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[36] M. Gonzalez-Franco, D. Perez-Marcos, B. Spanlang, and M. Slater. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *2010 IEEE virtual reality conference (VR)*, pp. 111–114. IEEE, 2010.

[37] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir. First-and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:33, 2017.

[38] G. Gorisse, O. Christmann, S. Houzangbe, and S. Richir. From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:8, 2019.

[39] G. Gorisse, O. Christmann, and S. Richir. De la présence à l'incarnation: proposition d'un méta-modèle pour la réalité virtuelle. *Interfaces numériques*, 7(1):114–114, 2018.

[40] L. Hoyet, F. Argelaguet, C. Nicole, and A. Lécuyer. "wow! i have six fingers!": Would you accept structural changes of your hand in vr? *Frontiers in Robotics and AI*, 3:27, 2016.

[41] S. Imaizumi and T. Asai. Dissociation of agency and body ownership following visuomotor temporal recalibration. *Frontiers in integrative neuroscience*, 9:35, 2015.

[42] S. Jung, P. J. Wisniewski, and C. E. Hughes. In limbo: The effect of gradual visual transition between real and virtual on virtual body ownership illusion and presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 267–272. IEEE, 2018.

[43] A. Kalckert and H. H. Ehrsson. The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and cognition*, 26:117–132, 2014.

[44] A. Kassambara. Finalik, 2021. last checked November 5, 2021.

[45] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[46] E. Kokkinara and M. Slater. Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43(1):43–58, 2014.

[47] E. Kokkinara, M. Slater, and J. López-Moliner. The effects of visuomotor calibration to the perceived space and body, through embodiment in immersive virtual reality. *ACM Transactions on Applied Perception (TAP)*, 13(1):3, 2015.

[48] C. Krogmeier, C. Mousas, and D. Whittinghill. Human–virtual character interaction: Toward understanding the influence of haptic feedback. *Computer Animation and Virtual Worlds*, p. e1883, 2019.

[49] M. Le Chénéchal and J. Chatel-Goldman. Htc vive pro time performance benchmark for scientific research. In *Icat-Egve 2018*, 2018.

[50] J. Lee, H. Cho, and R. B. McKay. A rapid screening and testing protocol for keyboard layout speed comparison. *IEEE Transactions on Human-Machine Systems*, 45(3):371–384, 2015.

[51] K. M. Lee. Presence, explicated. *Communication theory*, 14(1):27–50, 2004.

[52] B. Lenggenhager, M. Mouthon, and O. Blanke. Spatial aspects of bodily self-consciousness. *Consciousness and cognition*, 18(1):110–117, 2009.

[53] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke. Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 2007.

[54] M. R. Longo and P. Haggard. Sense of agency primes manual motor responses. *Perception*, 38(1):69–78, 2009.

[55] M. R. Longo, F. Schüür, M. P. Kammers, M. Tsakiris, and P. Haggard. What is embodiment? a psychometric approach. *Cognition*, 107(3):978–998, 2008.

[56] J.-L. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. Any "body" there? avatar visibility effects in a virtual reality game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 17–24. IEEE, 2018.

[57] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Avatar anthropomorphism and illusion of body ownership in vr. In *2015 IEEE Virtual Reality (VR)*, pp. 229–230. IEEE, 2015.

[58] K. Ma and B. Hommel. The role of agency for perceived ownership in the virtual hand illusion. *Consciousness and cognition*, 36:277–288, 2015.

[59] A. Maselli and M. Slater. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, 7:83, 2013.

[60] J.-M. Normand, E. Giannopoulos, B. Spanlang, and M. Slater. Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PloS one*, 6(1):e16128, 2011.

[61] M. A. Padilla-Castañeda, A. Frisoli, S. Pabon, and M. Bergamasco. The modulation of ownership and agency in the virtual hand illusion under visuotactile and visuomotor sensory feedback. *Presence: Teleoperators and Virtual Environments*, 23(2):209–225, 2014.

[62] J. Pallant. Survival manual. *A step by step guide to data analysis using SPSS*, 4, 2011.

[63] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:44, 2021. doi: 10.3389/frvir.2020.575943

[64] T. C. Peck, J. J. Good, and K. Seitz. Evidence of racial bias using immersive virtual reality: Analysis of head and hand motions during shooting decisions. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2502–2512, 2021.

[65] T. C. Peck and A. Tutar. The impact of a self-avatar, hand collocation, and hand proximity on embodiment and stroop interference. *IEEE transactions on visualization and computer graphics*, 26(5):1964–1971, 2020.

[66] V. I. Petkova, M. Khoshnevis, and H. H. Ehrsson. The perspective matters! multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in psychology*, 2:35, 2011.

[67] N. Ratcliffe and R. Newport. The effect of visual, spatial and temporal manipulations on embodiment and action. *Frontiers in human neuroscience*, 11:227, 2017.

[68] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez. Studying the role of haptic feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality*, 1:28, 2021. doi: 10.3389/frvir.2020.573167

[69] RootMotions. Finalik, 2014. last checked November 5, 2021.

[70] R. Rosenthal. *Experimenter effects in behavioral research*. Irvington, 1976.

[71] D. Roth and M. E. Latoschik. Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, 2020. doi: 10.1109/TVCG.2020.3023603

[72] D. Roth, J.-L. Lugrin, M. E. Latoschik, and S. Huber. Alpha ivbo-construction of a scale to measure the illusion of virtual body ownership. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2875–2883, 2017.

[73] B. O. Rothbaum and L. F. Hodges. The use of virtual reality exposure in the treatment of anxiety disorders. *Behavior Modification*, 23(4):507–525, 1999.

[74] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater. Virtual hand illusion induced by visuomotor correlations. *PloS one*, 5(4):e10381, 2010.

[75] V. Schwind, P. Knierim, L. Chuang, and N. Henze. " where's pinky?" the effects of a reduced number of fingers in virtual reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 507–515, 2017.

[76] V. Schwind, J. Leusmann, and N. Henze. Understanding visual-haptic integration of avatar hands using a fitts' law task in virtual reality. In *Proceedings of Mensch Und Computer 2019*, pp. 211–222. Association for Computing Machinery, 2019. doi: 10.1145/3340764.3340769

[77] V. Schwind, L. Lin, M. Di Luca, S. Jörg, and J. Hillis. Touch with foreign hands: The effect of virtual hand appearance on visual-haptic integration. In *Proceedings of the 15th ACM Symposium on Applied Perception*, pp. 1–8, 2018.

[78] M. Shin, S. Lee, S. W. Song, and D. Chung. Enhancement of perceived body ownership in virtual reality-based teleoperation may backfire in the execution of high-risk tasks. *Computers in Human Behavior*, 115:106605, 2021.

[79] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives. Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience*, 2:6, 2008.

[80] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke. First person experience of body transfer in virtual reality. *PloS one*, 5(5):e10564, 2010.

[81] A. M. Soccini, M. Grangetto, T. Inamura, and S. Shimada. Virtual hand illusion: The alien finger motion experiment. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1165–1166. IEEE, 2019.

[82] M. Tsakiris and P. Haggard. The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1):80, 2005.

[83] M. Tsakiris, M. R. Longo, and P. Haggard. Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia*, 48(9):2740–2749, 2010.

[84] M. Tsakiris, G. Prabhu, and P. Haggard. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition*, 15(2):423–432, 2006.

[85] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 27–35, 2016.

[86] R. White. The influence of experimenter motivation, attitudes, and methods of handling subjects on psi test results. *Handbook of parapsychology*, pp. 273–301, 1977.

[87] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[88] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[89] A. S. Won, J. Bailenson, J. Lee, and J. Lanier. Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20(3):241–259, 2015.