

## Article

# Virtual Reality as an Innovative Tool for Numerosity Perception

Beatrice Aruanno <sup>1,\*</sup>, Giovanni Anobile <sup>2</sup>, Armando Viviano Razionale <sup>1</sup>, Monica Bordegoni <sup>3</sup>  
and Guido Marco Cicchini <sup>4</sup>

<sup>1</sup> Department of Civil and Industrial Engineering, University of Pisa, 56122 Pisa, Italy

<sup>2</sup> Department of Neuroscience, Psychology, Pharmacology and Child Health, University of Florence, 50121 Florence, Italy

<sup>3</sup> Department of Mechanical Engineering, Politecnico di Milano, 20156 Milano, Italy

<sup>4</sup> Institute of Neuroscience, CNR-Via Moruzzi 1, 56124 Pisa, Italy

\* Correspondence: [beatrice.aruanno@unipi.it](mailto:beatrice.aruanno@unipi.it)

**Featured Application:** Use of virtual reality technology in numerosity estimation tasks.

**Abstract:** Numeracy, the ability to use basic mathematical skills in everyday life, is essential in modern society. Recent studies have shown a connection between numeracy and visual numerosity perception, yet traditional 2D screen-based assessment methods often lack ecological validity and participant engagement. This study evaluates the viability of conducting numerosity estimation tasks in virtual reality (VR) and to determine whether hallmarks of numerosity processing, typically observed in laboratory settings, can be replicated in immersive environments. Six participants completed a psychophysical evaluation in VR, comparing the numerosity of visual stimuli consisting of two sets of spheres. The VR experiment successfully replicated two distinctive patterns found in traditional psychophysical studies: increased precision and decreased response times at high numerosities. Specifically, Weber fractions drop by approximately a factor of two, with values ranging from ~15% for low and intermediate conditions to ~8% in high numerosities, and response times decreases from ~663 ms for low numerosities to ~593 ms for high numerosities. These findings highlight that VR can be effectively used for numerosity estimation tasks, providing a controlled and immersive environment that traditional methods cannot achieve, while significantly expanding methodological possibilities in psychophysical research.

**Keywords:** numerosity estimation; virtual reality; perception; human–computer interaction



Academic Editor: Zhonghua Sun

Received: 14 February 2025

Revised: 1 April 2025

Accepted: 2 April 2025

Published: 3 April 2025

**Citation:** Aruanno, B.; Anobile, G.; Razionale, A.V.; Bordegoni, M.; Cicchini, G.M. Virtual Reality as an Innovative Tool for Numerosity Perception. *Appl. Sci.* **2025**, *15*, 3976. <https://doi.org/10.3390/app15073976>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Math is crucial in modern society. During our daily activities, we often base decisions on numerical information, for example, to organize our diary commitments, to check that we have received the right change, to manage diet and nutrition, to measure medicine doses, to estimate time until our train, and much more. Low numeracy, defined as the inability to understand and use numbers effectively, is a well-known limiting factor for both education (particularly for STEM disciplines) and employment choices, with huge economical costs [1,2]. Moreover, low numeracy characterizes developmental dyscalculia, a common but still poorly understood learning disorder, impairing children mathematical learning with negative repercussions on quality of life [3]. Numeracy is also often linked to math anxiety (excessive feelings of fear when exposed to mathematical tasks [4]), a relevant clinical condition, which is recently attracting much scientific interest. For these and other reasons, many governments around the world have established activities to continuously track the quality of mathematical learning in children and adolescents, on a large scale.

OECD (Organization for Economic Co-operation and Development) director for education and skills defined good numeracy as “the best protection against unemployment, low wages and poor health” [5]. Neuroscience can provide significant contributions to this goal. In this regard, it is widely accepted that understanding the basic neurocognitive mechanisms governing the early precursors of mathematical abilities represent a key step to contrast low numeracy, and help dyscalculia and math anxiety [6,7].

Where does mental math come from? Historically, mathematical skills were considered a paradigmatic example of high-level (human specific) verbal cognitive ability. In recent decades, this idea has radically changed, grounding mental math into a basic sensory mechanism: a visual number sense [8–10]. This fascinating idea triggered a huge burst of research, also for the implications sketched above, and there is now a huge amount of evidence, coming from different scientific fields, demonstrating that humans have a brain mechanism able to roughly but very quickly estimate the numerosity of objects in the environment [10–16]. This mechanism, often referred as the “Approximate Number System (ANS)” differs from serial counting in that it is faster, much more error-prone [9] and independent of mathematical language [17]. Moreover, whereas symbolic mathematics is a uniquely human function, we share the non-symbolic ANS with many animals’ species, including non-human primates [18] but even birds [19,20], fishes [21], and insects (e.g., bees [22,23]). Going beyond behavioral results, there is now also clear evidence for the existence of numerosity selective neurons, mainly located in parietal and frontal cortices of both humans and non-human primates [18].

How does math emerge from this sensory mechanism in humans? The rationale is that numerical and mathematical meaning is mapped onto this pre-existing sensory number system, in charge of encoding the numerosity of visual objects. Atypical development of this basic sensory mechanism, should compromise a meaningful mapping between numerical symbols (digits) and their non-symbolic counterpart (the associated numerical quantity), limiting the development of mathematical learning [7]. In line with this idea, a seminal paper by Halberda and collaborators [24] reported positive correlations between visual numerosity estimation precision (more blue or yellow dots?) and math performance (e.g., mental calculation) in adolescents and in previous years, back to kindergarten. Moreover, children with dyscalculia often show deficits in visual numerosity tasks, even when the verbal component is eliminated, such as by indicating which of two visual ensembles contains more objects [25–28]. The link between math and visual numerosity abilities is also in line with recent imaging studies describing areas in the parietal cortex responding to numerosity [11], digit perception [29], as well as mathematical reasoning [12].

These results highlight the potential of this sensory system to support mathematical learning. With effective training, it could help mitigate the negative effects of low numeracy, including those associated with dyscalculia and math anxiety. Moreover, as numerosity tasks do not require sophisticated language skills, these training activities could be potentially performed relatively early in life, also as a pedagogical strategy for the empowerment of basic non-symbolic prerequisites for later symbolic mathematical learning (for example in pre-school children).

Given the potential link between the ANS and math, some studies measured whether and how much this perceptual system can be improved by perceptual training procedures and whether the effect of the training generalizes over mathematical abilities (e.g., improving mental calculation proficiency). Although there are several methodological differences between studies, in general, the results show a relatively high level of plasticity of the ANS, even in adults [30,31]. At odds with this, the evidence for a potential transfer of perceptual improvements to symbolic mathematical abilities (e.g., calculation) is less clear and still highly debated. For example, Cochrane and colleagues [30] trained adults to

identify whether a set of elements (dots) briefly (to avoid counting, 500 ms) presented on a screen contained more white or black dots. By changing (trial-by-trial) the ratio between black and white dots, they measure sensory thresholds as the difference in numbers of black and white dots necessary to reach a pre-defined correct response rate (e.g., 75%). The results showed a clear improvement in thresholds, with performance continuing to improve even after thousands of trials. Despite the clear sensory improvement, the learning did not transfer to math abilities (arithmetic operations, including addition, subtraction, and multiplication), leaving the performance obtained before and after the perceptual training virtually unchanged.

The traditional approach to numerosity estimation typically involves presenting participants with arrays of visual stimuli, such as dots or other shapes, arranged on a 2D display [32]. These arrays can vary in configuration, including regular grids, random distributions, or clustered patterns, with key variables such as item density, size, spacing, and overall arrangement being manipulated [33]. Participants are tasked with estimating the number of items in the display, typically through direct estimation, comparison with other arrays, or magnitude estimation techniques. In addition to accuracy, response time is also often measured [34]. These results can also shed light on perceptual phenomena such as the Weber fraction, which quantifies sensitivity to numerical differences [35].

Other attempts have been made, with different outcomes, sometime showing transfer to math [31,36,37], sometimes not [30,38–40]. A relatively recent review critically evaluated all published studies that aimed to train the ANS finding no conclusive evidence that ANS training improves symbolic arithmetic [41].

One possible factor behind the lack of generalization of perceptual training in math could reside in the non-ecological quality of the activities used to train this perceptual mechanism. We rarely find ourselves discriminating, for tens of minutes, which set of elements presented on a monitor is the most numerous. These tasks, besides being scarcely ecological, are often monotonous, and are not very suitable for stimulating motivation.

The use of virtual reality (VR) in numerosity perception tasks could provide a controlled and immersive environment for presenting numerical stimuli in ways that are difficult or impossible to achieve with traditional methods. This can be useful for exploring how environmental and contextual factors influence numerosity estimation. This approach allows for a more natural interaction with numerical stimuli, where arrays of objects can be distributed in three-dimensional space rather than being constrained to a two-dimensional screen. This can be particularly useful for simulating real-world environments where participants estimate numerosities in dynamic and spatially complex scenes, such as evaluating the number of objects in a cluttered room or assessing the density of elements in a natural setting. VR-based studies could significantly improve the ecological validity of numerosity perception experiments compared to traditional 2D methods based on monitors.

Moreover, there is recent evidence linking actions to numerosity perception, suggesting the existence of a “sensorimotor numerosity mechanism” integrating sensory numerical information coming from the environment with that internally generated by actions [42–45]. Even if the link between this newly discovered system and math skills is still to be tested, VR stands as a potentially excellent tool for the implementation of setups aimed at promoting the interaction between action and numerical perception. Unlike traditional methods, VR enables direct sensorimotor engagement through embodied interactions such as reaching, grasping, or manipulating numerical stimuli in an immersive space. This could be particularly relevant for studying how proprioceptive and haptic feedback influence numerical estimation [44] and for developing novel training paradigms that leverage multimodal learning.

Furthermore, VR allows for greater experimental control and flexibility in the presentation of numerical stimuli [46]. In the case of numerosity judgments, variables such as object size, spatial arrangement, motion, and environmental lighting conditions can be dynamically isolated or adjusted to investigate their influence on numerosity perception. For example, researchers can test whether numerical estimation differs when elements are presented in peripersonal space versus extrapersonal space or how different levels of visual complexity affect estimation accuracy. The ability to manipulate such parameters in real-time and in an ecologically valid manner makes VR a valuable tool for uncovering underlying cognitive mechanisms that might not be evident with traditional screen-based tasks.

Additionally, VR offers the possibility of incorporating adaptive learning mechanisms, where task difficulty dynamically adjusts based on user performance. This can increase engagement, prevent fatigue, and provide personalized training experiences tailored to individual needs. In general, VR experiences delivered to young populations yield high engagement [47–49], and it is likely that the immersive nature of VR can also enhance motivation by transforming traditionally repetitive numerical estimation tasks into more interactive and gamified experiences, potentially improving long-term learning outcomes.

Taken together, these advantages highlight the potential of VR not only as an alternative to traditional numerosity tasks but as a powerful platform for advancing the understanding of numerical cognition.

At the same time, transposing typical experimental paradigms into a VR system does not come without its challenges. Inconsistent signals from our visual and vestibular systems during space exploration can cause motion sickness [50], which may hamper the participant compliance to long experimental sessions. In addition, the sense of agency—the subjective experience of controlling one’s actions and their outcomes—and presence—the perceptual illusion of being physically immersed in the virtual environment—need to be satisfied, otherwise one risks delivering to the user an immersive but unrealistic environment [51,52]. Further, some VR systems, in particular head mounted displays (HMDs), could provide inconsistent accommodation and vergence signals, which bring about erroneous spatial estimates [53].

All these cognitive and perceptual factors indicate that transferring a research protocol from the typical lab setting to a 3D immersive one may present specific challenges, which may have a crucial impact on the way the systems for numerosity estimation are engaged.

Given these premises the first step that has to be taken before transposing traditional laboratory tests in VR and running lengthy training protocols is to check whether immersive environments yield psychophysical results comparable to those of traditional methods. To the best of the authors’ knowledge, this study is among the first to explore the use of VR in the field of numerosity perception [54]. The main aim of the paper is to replicate previous hallmarks of numerosity tasks performed on traditional 2D screens findings. One is the finding that Weber’s Law for numerosity is present up to a critical numerosity, but then the regime is violated and judgments become more precise [14,35]. The other is the fact that such judgments provide a near constant response time profile at low numerosities and then speed up at higher numerosities [34].

These findings not only have been replicated multiple times in the literature and may provide a useful benchmark for the VR setup, but they also have theoretical importance as this dual behavior is likely reflecting the presence of two separate systems that handle numerosity judgments, one more invariant to spatial position, contrast, and low level features for moderate numerosities, and one more capable of deriving global low level statistics operating with sufficiently dense display [14].

Replicating these features has the twin scope of both demonstrating that VR environments carry over results obtained in two-dimensional displays, but also to demonstrate that 3D cues do not have a major impact in core features of the system for determining the numerosity of items surrounding us.

## 2. Materials and Methods

### 2.1. Experimental Devices

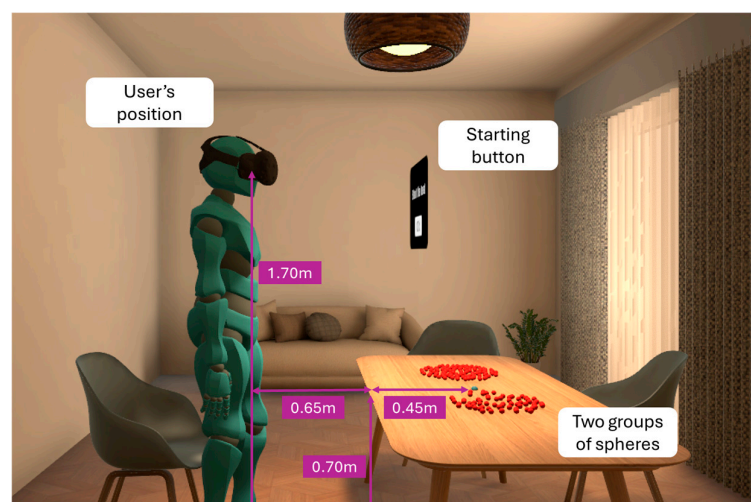
In this study, a head-mounted display, specifically the VIVE Focus 3, was selected to provide participants with a virtual reality environment, aligning with the study's objective of delivering an immersive experience with higher ecological validity, resembling real-world conditions rather than a conventional laboratory setting. Compared to other VR technologies, such as Cave Automatic Virtual Environment (CAVE) systems, which require specialized infrastructure and substantial space [55], the selected solution offers a balance of high immersion and practicality.

The VIVE Focus 3 was particularly chosen due to its ability to track users' body, hand, and head movements, which was essential for capturing natural interactions in the numerosity task assessment. Additionally, the device's high-resolution panels and 90 Hz refresh rate help reduce motion sickness symptoms (e.g., dizziness or headaches), ensuring a comfortable experience for participants [56]. Observers used the device's two handheld controllers to interact with elements in the virtual scene and perform selections for the numerosity task assessment.

To support real-time rendering of complex virtual environments without compromising performance, the system was connected to an MSI GE66 Raider laptop (Intel Core i7-10870 x64, 32 GB RAM, NVIDIA GeForce RTX 3060, Intel, Santa Clara, CA, USA) via the VIVE Business Streaming application. This setup enabled low-latency interaction and ensured that the VR experience remained smooth and responsive, which is critical for the accuracy and validity of the experimental tasks [57].

### 2.2. Immersive Environment

The virtual environment aims to be a first step in moving from traditional 2D numerosity assessment tasks to 3D settings. For this reason, the environment replicates a living room with a table, chairs, lights, and furnishing elements (Figure 1).



**Figure 1.** Immersive VR environment developed for the study. The measures indicate the position of a participant with respect to the fixation point in the middle of the table and a pair of sphere groups.

The application has been developed using the software Unity 6, combined with the SteamVR plugin to include the VR settings. The level of detail used in the environment was the balance point between photorealism and performance. Lights and shadows have been precomputed and baked using an adaptive probe volume. The method consists of sampling the lighting at strategic points in the room, denoted by the positions of the probes. The lighting at any point is approximated by interpolating between the samples obtained by the nearest probes. The interpolation is fast enough to be used during the tests without notably affecting the refresh rate of the HMD.

The interaction mechanisms within the virtual environment have been kept as simple as possible to minimize potential difficulties for participants unfamiliar with VR technology. To confirm selections and start the experience, only the trigger button on the handheld controller is used.

### 2.3. Procedure

Six participants (five males and one female) took part in the study. The average age of the participants was 34.67 years (range: 28–49 years). All participants had normal or corrected-to-normal vision. The study was approved by the local health service ethics committee (“Comitato Bioetico dell’Università di Pisa”, 24 September 2021, n. 31.) and was conducted in accordance with the Declaration of Helsinki.

Prior to the experiment, a calibration phase was conducted to adjust the HMD for each participant, specifically by setting the interpupillary distance to match their individual needs. This calibration ensured accurate visual alignment and optimal comfort. Participants were then given time to familiarize themselves with the virtual environment, the VR controllers, and the general mechanics of the study.

The task involved judging two groups of small spheres placed on a virtual table and determining which group contained more spheres. The participant’s position, while interacting with the virtual environment, is visually represented by the avatar in Figure 1. It is important to note that this avatar was added only for illustrative purposes in the figure and was not present in the actual virtual environment during the study.

Participants began the task by standing in front of a starting panel positioned near the virtual table. To start the evaluation, they confirmed their readiness by pressing a virtual button on the panel using the VR controller. Following this confirmation, a series of pairs of red sphere groups appeared on the table. The virtual table measured  $0.9 \times 1.9 \text{ m}^2$  in area and was set at a height of 0.7 m. At the center of the table, a light blue semi-sphere served as a fixation point to help participants maintain their gaze at a consistent location.

Participants were instructed to stand 0.65 m away from the table’s edge (1.1 m from the blue fixation point). Their standing height ranged approximately from 1.70 to 1.85 m, resulting in eye heights between 1.58 m and 1.70 m above the floor. A chair positioned directly behind the participants provided a fixed reference point to standardize positioning during the task.

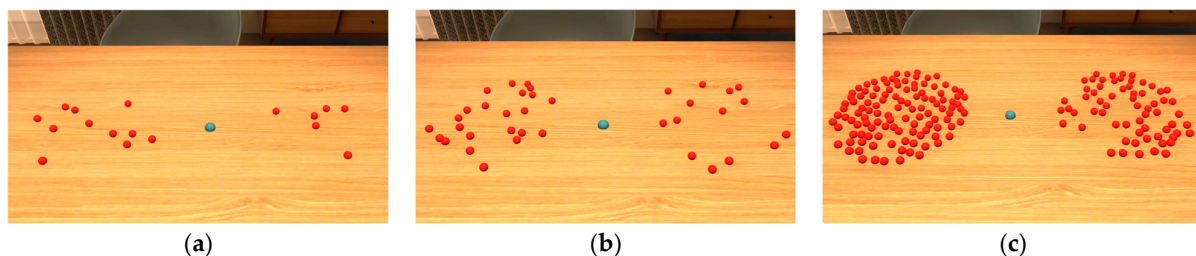
The red spheres occupied a circular area with a radius of 18 cm, and the center of this circular area was located 30 cm from the blue fixation point. Each sphere had a diameter of 2 cm. Depending on the experimental condition, the number of spheres within each group varied from 5 to 130. The spatial positions of the spheres were precomputed using MATLAB R2024a (a high-level programming language designed for numerical computation, data analysis, and visualization, widely used in engineering, scientific research, and mathematics), based on a validated algorithm previously employed in similar studies by the research group [35]. These coordinates were then imported into Unity prior to the experiment to ensure precise and reproducible spatial arrangements.

The VR system dynamically calculated the participant's viewpoint in real-time, rendering the virtual scene based on their physical position. This process ensured that the retinal projection of the spheres remained accurate for each individual observer. As a result, a single sphere subtended a degree of visual angle (dva) of 0.8 degrees, with an average eccentricity of 12.2 dva. The overall spatial extent of the sphere groups spanned approximately 14.5 dva horizontally and 10.6 dva vertically. The dva is a measure of the angle subtended by an object on the retina, determined by its size and distance from the observer.

During the task, participants were required to determine which group of spheres was more numerous by pressing the trigger button on the corresponding side of the controller. Each pair of sphere groups was presented for 240 ms, but participants were allowed to respond at any time after the initial presentation.

#### 2.4. Experimental Conditions

The experiment involved testing three distinct numerosity conditions, each subdivided into three different blocks of 40 trials. In the low numerosity condition, the numerosity of the groups could vary between 5 and 11 (average 8), in the intermediate numerosity condition between 10 and 26 (average 18), and in the high numerosity condition between 50 and 130 (average 90). To ensure consistency and control over the experimental variables, the numerosity of the two groups—left and right—was determined in advance, together with the coordinates of the positions of all the spheres. Importantly, the numerosity and positioning were chosen independently for the left and right sets, ensuring no bias between the two groups. An example of the experimental setup as viewed through the participant's HMD is shown in Figure 2, where the test is depicted from the perspective of the observer in all three numerosity conditions that were evaluated. This illustration provides a visual representation of how the varying quantities of items appeared in the three conditions, highlighting the differences in numerosity and their potential impact on the observer's perception.



**Figure 2.** Sample configurations for the three numerical conditions. Participants compared the numerosity of two groups of spheres positioned to the left and right of the fixation point (blue dot). The numerosity levels were categorized as low (a), intermediate (b), or high (c).

#### 2.5. Data Analysis

Each participant estimated numerosity by choosing either “right more numerous” or “left more numerous” across 120 trials for each of the three evaluated conditions. Data recorded included the numerosity values of each pair of sphere groups, the participants' binary choices (1: right more numerous, 0: left more numerous), and the response time for each decision.

The responses indicating “right more numerous” were plotted as a function of the numerosity difference between the right and left groups of spheres, normalized by the average numerosity of the range. These response curves were then fitted with a cumulative Gaussian function, which served to model the relationship between the numerosity difference and the likelihood of perceiving the correct group as more numerous. The median of the Gaussian function represents the point of subjective equality (PSE), which is the

numerosity difference at which participants are equally likely to judge either the left or right group as more numerous. The steepness of the Gaussian curve is indicative of the precision or sensitivity of the participant's judgment, with steeper curves reflecting more precise judgments and broader curves suggesting greater uncertainty or noise in the response.

The Weber's fraction, which measures the smallest noticeable difference between two stimuli relative to the magnitude of the original stimulus, can be used to quantify how precisely individuals distinguish between different quantities. To calculate the Weber's fraction, the just noticeable difference (JND) was extracted from the fitted Gaussian curve. The JND represents the smallest change in numerosity required for the participant to reliably distinguish between the two groups, reaching a 75% correct response rate. The JND was then divided by the average numerosity of the respective condition (i.e., 8 for the very low numerosity condition, 18 for the low numerosity condition, and 90 for the high numerosity condition) to compute the Weber's fraction. This fraction provides a measure of the relative sensitivity to numerosity differences, with larger values indicating poorer discrimination ability.

To estimate the variability and ensure robust statistical analysis, error bars were calculated using bootstrapping, which involved resampling the original data 1000 times. This technique enabled the estimation of confidence intervals for each measure, providing a more reliable representation of the underlying variability in participants' responses [58].

To statistically assess the effects on precision and response times, Weber fractions and response times were analyzed using a one-way repeated measures ANOVA to determine whether statistically significant differences exist among the groups. This was followed by Bonferroni-corrected post hoc *t*-tests (two-tailed) to compare group pairs.

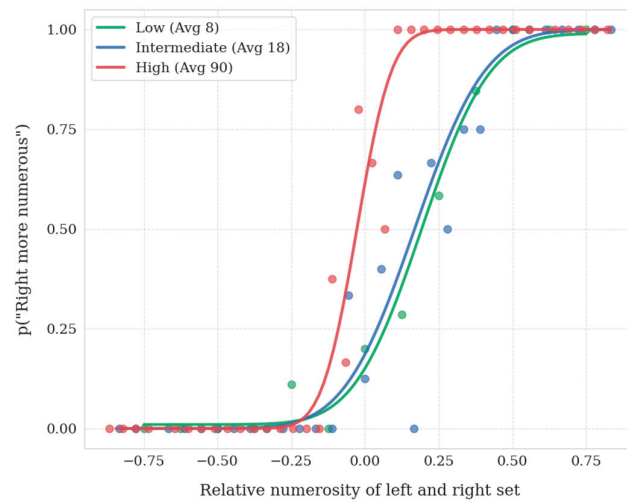
Frequentist *t*-test were complemented with the Bayesian statistics, repeated measures ANOVA and post hoc *t*-test, with estimation of Bayes Factors (BF) [59], which quantify the evidence for or against the null hypothesis by comparing the likelihoods of the alternative hypothesis ( $H_1$ ) and null hypotheses ( $H_0$ ). By convention, Bayes Factors greater than 3 indicate moderate evidence in favor of the alternative hypothesis (greater than 10 for strong evidence), whereas values below 1/3 suggest moderate evidence against it (below 1/10 for strong evidence).

To estimate the required sample size ( $N$ ) for this preliminary evaluation, a large effect size ( $\geq 2$ ) was considered, along with a high statistical power ( $1 - \beta = 0.95$ ) and a significance level ( $\alpha = 0.05$ ). Given the context of this study and the use of a paired *t*-test, the required sample size was determined to be six participants [60].

### 3. Results

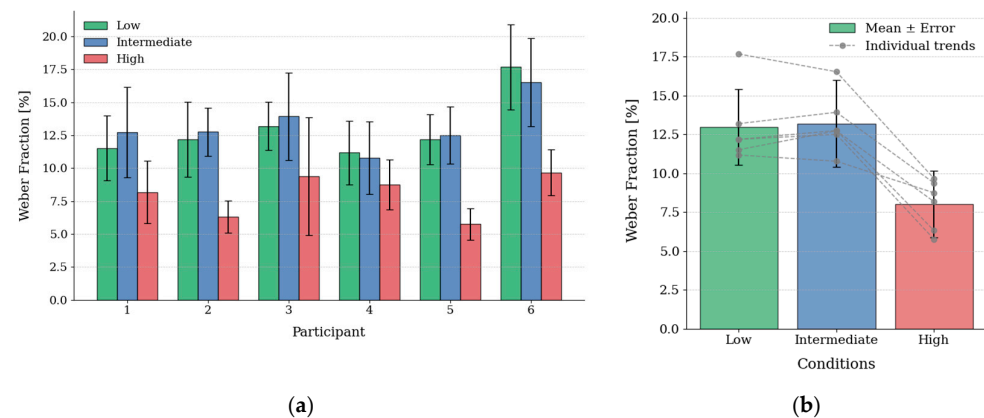
Six participants took part in a replication of the Anobile et al. study [35], which tested numerosity precision (i.e., Weber fractions) over a range of numerosities, and investigated whether humans perceive numerosity directly or infer it indirectly from texture density. Participants were required to choose which of two sets of red spheres was more numerous and their preference for the right set was plotted against the numerosity difference between the numerosity of the right and left sets normalized by the average numerosity of the range (Figure 3). These responses were fitted by a cumulative Gaussian function whose steepness indicates the precision in performing the judgment. The range necessary to move from 50% to 75% "right more numerous" responses indicate the extent of the zone of uncertainty reflecting the precision of the judgment. As visible from the representative participant shown in Figure 3, the three ranges considered in this study—low (avg 8), intermediate (avg 18), and high (avg 90)—produced different results. Whereas the initial two ranges produce similar curves (green and blue), the curve obtained for the high numerosity range

(red) was steeper, indicating that judgments involving higher numerosities yield more precise decisions.



**Figure 3.** Sample psychometric curves for a representative participant for the three numerical conditions. Raw data for the three numerosity ranges are plotted with different colors: green (low), blue (intermediate), and red (high). Raw judgments are fitted with a cumulative Gaussian curve (continuous lines).

To quantify this effect, for each participant, the Weber fraction (i.e., the JND/average numerosity) has been computed and plotted separately for the three numerosity ranges (Figure 4). As visible in the plots for low and intermediate numerosities, Weber fractions are around 13%, whereas for high numerosity they decrease to approximately 8%.



**Figure 4.** Weber fractions of participants for the three numerosity conditions. Weber fraction values and error bars (a), and combined using average values and trend lines (b).

The frequentist statistical analysis with one-way repeated measures ANOVA resulted in the presence of statistically significant differences among the three numerosity conditions for Weber fraction values ( $F(10,2) = 35, p < 0.001$ ). It is also confirmed by the Bayesian statistics for the model, showing extremely strong evidence in favor of the alternative hypothesis ( $BF_M = 1665$ ).

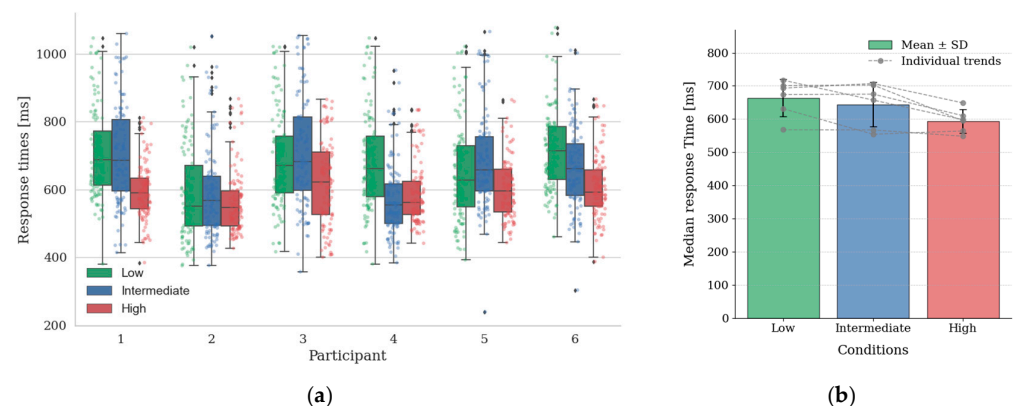
Bonferroni corrected post hoc *t*-test for the various conditions indicate that the difference between low and intermediate range is not statistically significant ( $t(5) = 0.31, p > 0.5$  two-tails) and actually there is moderate evidence for the null hypothesis ( $BF_{10} = 0.26$ ); instead, the difference between the lower conditions and the highest numerosity range is significant (low vs. high  $t(5) = 7.14, p < 0.001$ , intermediate vs. high

$t(5) = 7.45, p < 0.001$ ) and provide a strong support against the null hypothesis ( $BF_{10} = 12.6$  and  $BF_{10} = 22.5$ , respectively).

The Cohen's  $d$  values were calculated to estimate the pairwise effect sizes between group means. The results confirm the frequentist and Bayesian analyses. They indicate a very large effect for the comparisons between low vs. high ( $d = 2.44$ ) and intermediate vs. high ( $d = 2.93$ ), suggesting substantial differences in means. In contrast, the comparison between low and intermediate groups yielded a very small effect size ( $d = 0.09$ ), implying minimal difference between these conditions.

In summary, there is no significant difference between low and intermediate numerosities, while high numerosities are perceived significantly more accurately than low and intermediate conditions.

Recent evidence suggests that another critical index of psychophysical performance is response times [45]. Previously, our group has demonstrated that response times for numerosity verbal estimation follows a descending pattern. An initial plateau at low and moderate numerosities is followed by a second phase where response times are faster [34]. In this study, a similar approach was followed, analyzing the response times for numerosity discriminations in VR. Figure 5 illustrates the response times recorded for each participant with a box plot, and the median values in the three conditions. The box plot visually represents data distribution by displaying the median (Q2) and the interquartile range (IQR), which is the difference between the third quartile (Q3) and the first quartile (Q1). The whiskers extend to the maximum and minimum values within 1.5 times the IQR from Q3 and Q1, while outliers beyond this range are shown as individual black dots.



**Figure 5.** Response times of participants for the three numerosity conditions. Box plot of the response times plotted individually with raw values overlapped (a). Bar chart showing median values averaged with individual trend lines (b).

The frequentist statistical analysis with one-way repeated measures ANOVA resulted in the presence of statistically significant differences among the three numerosity conditions ( $F(10,2) = 10.4, p = 0.004$ ). It is also confirmed by the Bayesian statistics, showing strong evidence in favor of the alternative hypothesis ( $BF_M = 11.17$ ).

Bonferroni corrected post hoc  $t$ -test for the three conditions founds again a slight descending pattern, with the two lower numerosities presenting a similar response time (663 and 643 ms  $t(5) = 1.27, p = 0.69, BF_{10} = 0.40$ ) and the higher numerosity leading to faster responses by about 60 ms (average 593 ms), which is different from the two lower numerosities (low vs. high  $t(5) = 4.42, p = 0.004, BF_{10} = 6.1$ ; intermediate vs. high  $t(5) = 3.1, p = 0.031, BF_{10} = 1.8$ ). Also with the response times, there is no significant difference between low and intermediate conditions, while high is significantly different from both intermediate (limited evidence) and low conditions (strong evidence).

This is also confirmed by the Cohen's  $d$  values for the pairwise effect sizes. The results indicate a huge effect for the comparisons between low vs. high ( $d = 1.50$ ) and a very large effect for intermediate vs. high ( $d = 0.93$ ), suggesting relevant differences in means. In contrast, the comparison between low and intermediate conditions yielded a small effect size ( $d = 0.33$ ).

#### 4. Discussion and Conclusions

The aim of the current study was to assess the viability of conducting numerosity decisions in a VR setup and to determine whether the key characteristics of numerosity judgments observed in typical laboratory settings can be replicated.

The main finding is that the VR platform, powered by Unity and the VIVE Focus 3 headset, proved flexible enough to incorporate the essential features of a psychophysical experiment. In this paradigm, the classical method of adjustment was employed, where the experimenter determines the test trials and their randomization before the experiment begins. This method functioned flawlessly.

However, implementing a modern psychophysical experiment in VR presents certain challenges. Numerosity judgments typically require that dots forming a cloud are randomly distributed over a region of interest while adhering to specific constraints (e.g., ensuring objects do not overlap). This real-time computational demand was addressed by pre-calculating a list of coordinates in MATLAB, which was then integrated into the VR pipeline. Although this approach may seem labor-intensive, it is entirely manageable. The Unity application successfully read and interpreted the MATLAB output without issues.

At the same time, it is worth noting that over the years several efficient methods have been developed to expedite data collection. To incorporate these, it will certainly be needed to change the work pipeline. Many of these are adaptive methods that analyze stimulus-response histories to suggest efficient stimulus intensities for subsequent data collection. While such algorithms are widely available in various MATLAB toolboxes, they lack proper integration with systems like Unity, which was originally designed for application and game development. In addition, the process of generating real-time stimuli that obey multiple constraints may be challenging and this becomes more complex when the algorithms reside externally. Nonetheless, since these algorithms are based on mathematical and statistical principles, translating them into code compatible with Unity is feasible. This will require additional development and testing to ensure optimal system performance while maintaining a seamless VR experience.

From a psychophysical perspective, the experiment was successful. Two distinctive patterns commonly observed in typical 2D displays were successfully replicated, suggesting that the core features of the numerosity estimation system, previously studied only in 2D environments, are preserved in more realistic VR setups, which are inherently three-dimensional [34,35].

The previous Anobile et al. study [35] provided evidence that numerosity and density judgments are governed by separate perceptual mechanisms with different psychophysical characteristics. In particular, for densities up to  $0.25$  dots/deg<sup>2</sup>, Weber fractions remained constant, supporting a direct perception of numerosity; while beyond  $0.25$  dots/deg<sup>2</sup>, Weber fractions decreased, indicating a transition to texture-density mechanisms. The drop in Weber fractions reported here (about a factor of 2) is entirely consistent with these previous findings, which show that, after a breakpoint judgments, the Weber fractions values follow a regime change and obey a square-root law [14,35]. Interestingly, the quantitative similarity between the current dataset and previous datasets resides not only in the drop of Weber Fractions at higher numerosities but also in the average values.

Previous studies have reported Weber fractions ranging from approximately 15% for low and intermediate conditions to approximately 8% in high numerosities conditions [61].

Also, response times exhibited a trend similar to previously published data from the group. In [34], the authors investigated how response times vary across different numerosity perception regimes: subitizing, estimation, and texture. For numerosities greater than four (estimation range) response times are higher due to a higher cognitive demand for estimation tasks. At very high numerosities ( $\geq 50$ ) with dense item packing, response times decreased, implying a transition to texture perception mechanisms. Importantly, in that dataset observers were required to verbally estimate dot numerosity, which led to higher response times, typically of more than one second. Even if a direct comparison is not possible, the patterns emerging from the two paradigms are very similar indicating that one is probing similar sensory systems.

This study focused on numerosity discrimination tasks, which are fundamental for neuroscientists to infer the quality of underlying sensory representations and are valuable in basic research. In principle, any property of the numerosity system can be investigated just as effectively in a VR environment.

Of the two key results reported—an increment in precision at high numerosities and a decrease in response times—the replication of the Weber fraction pattern as a function of numerosity is particularly significant. Many training protocols aim to enhance cognitive and perceptual abilities related to numerosity, which directly depends on the ability to measure and manipulate sensory precision.

Although based on a limited sample, our results contribute to expanding current knowledge. Notably, the observed decrease in response times with increasing numerosity was tested using a two-alternative forced-choice (2AFC) paradigm, which does not require verbal responses. This is intriguing because previous research observed similar patterns with verbal estimates. In those cases, faster responses at higher numerosities could be attributed to coarse estimation strategies (e.g., estimating in tens), leading to quicker judgments. The fact that this pattern also emerges in a 2AFC setting suggests that the effect originates at a perceptual level rather than from the process of selecting a verbal estimate.

The findings of this study addressed the dual research questions: first, demonstrating that VR environments can replicate results obtained using the traditional approach based on two-dimensional displays; and second, suggesting that the core features of numerosity perception remain consistent across the two different display methods. However, VR technology extends the scope of research by enabling the study of additional variables that cannot be tested in traditional 2D settings. In particular, VR is highly beneficial for simulating real-world environments where participants estimate numerosities in dynamic and spatially complex scenes, thereby significantly enhancing the ecological validity of numerosity perception experiments.

Moreover, VR solutions provide greater experimental control and flexibility in presenting numerical stimuli. In numerosity judgment studies, key variables such as object size, spatial arrangement, motion, and environmental lighting conditions can be dynamically modified to assess their influence on numerosity perception.

Future research will aim to directly compare the VR-based paradigm with traditional 2D methods to verify whether there is a correlation of the Weber fractions between the two approaches, thereby assessing the consistency and reliability of VR-based psychophysical experiments. Additionally, it will be important to evaluate the comfort of the VR experience relative to classical methods, providing insights into user experience and identifying any potential ergonomic or cognitive load challenges associated with immersive environments and comparing them with the ones of the traditional setup. Finally, extending the analysis to a more diverse participant pool will enhance the generalizability of

the findings, and a deeper investigation into response times will help uncover the cognitive and perceptual mechanisms underlying numerosity judgments in an immersive VR environment.

**Author Contributions:** Conceptualization, B.A., G.M.C. and G.A.; methodology, B.A., G.M.C. and G.A.; software, B.A. and G.M.C.; formal analysis, G.M.C.; visualization, B.A.; writing—original draft preparation, B.A., G.M.C. and G.A.; writing—review and editing, B.A., G.M.C., G.A., A.V.R. and M.B.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of “Comitato Bioetico dell’Università di Pisa”, 24 September 2021, n. 31.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. OECD. *The High Cost of Low Educational Performance. The Long-Run Economic Impact of Improving PISA Outcomes*; OECD: Paris, France, 2010. [CrossRef]
2. Franklin, J. *Counting on the Recovery. The Role for Numeracy Skills in ‘Levelling up’ the UK*; Pro Bono Economics: London, UK, 2021.
3. Castaldi, E.; Piazza, M.; Iuculano, T. Learning disabilities: Developmental dyscalculia. *Handb. Clin. Neurol.* **2020**, *174*, 61–75. [CrossRef] [PubMed]
4. Ashcraft, M.H. Math Anxiety: Personal, Educational, and Cognitive Consequences. *Curr. Dir. Psychol. Sci.* **2002**, *11*, 181–185. [CrossRef]
5. Schleicher, A. Why Is Numeracy Important? Available online: <https://www.nationalnumeracy.org.uk/what-numeracy/why-numeracy-important> (accessed on 1 April 2025).
6. Butterworth, B. *Dyscalculia: From Science to Education*; Routledge: London, UK, 2018.
7. Piazza, M. Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* **2010**, *14*, 542–551. [CrossRef] [PubMed]
8. Dehaene, S. *The Number Sense: How the Mind Creates Mathematics*, Revised and Updated ed.; Oxford University Press: Oxford, UK, 2011.
9. Feigenson, L.; Dehaene, S.; Spelke, E. Core systems of number. *Trends Cogn. Sci.* **2004**, *8*, 307–314. [CrossRef]
10. Burr, D.; Ross, J. A visual sense of number. *Curr. Biol.* **2008**, *18*, 425–428. [CrossRef]
11. Piazza, M.; Izard, V.; Pinel, P.; Le Bihan, D.; Dehaene, S. Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron* **2004**, *44*, 547–555. [CrossRef]
12. Piazza, M.; Eger, E. Neural foundations and functional specificity of number representations. *Neuropsychologia* **2016**, *83*, 257–273. [CrossRef]
13. Cicchini, G.M.; Anobile, G.; Burr, D.C. Spontaneous perception of numerosity in humans. *Nat. Commun.* **2016**, *7*, 12536. [CrossRef]
14. Anobile, G.; Cicchini, G.M.; Burr, D.C. Number as a Primary Perceptual Attribute: A Review. *Perception* **2016**, *45*, 5–31. [CrossRef]
15. Stoianov, I.; Zorzi, M. Emergence of a ‘visual number sense’ in hierarchical generative models. *Nat. Neurosci.* **2012**, *15*, 194–196. [CrossRef]
16. Ferrigno, S.; Jara-Ettinger, J.; Piantadosi, S.T.; Cantlon, J.F. Universal and uniquely human factors in spontaneous number perception. *Nat. Commun.* **2017**, *8*, 13968. [CrossRef] [PubMed]
17. Pica, P.; Lemer, C.; Izard, V.; Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **2004**, *306*, 499–503. [CrossRef] [PubMed]
18. Nieder, A. The neuronal code for number. *Nat. Rev. Neurosci.* **2016**, *17*, 366–382. [CrossRef] [PubMed]
19. Liao, D.A.; Brecht, K.F.; Veit, L.; Nieder, A. Crows “count” the number of self-generated vocalizations. *Science* **2024**, *384*, 874–877. [CrossRef]
20. Rugani, R.; Vallortigara, G.; Priftis, K.; Regolin, L. Number-space mapping in the newborn chick resembles humans’ mental number line. *Science* **2015**, *347*, 534–536. [CrossRef]
21. Potrich, D.; Zanon, M.; Vallortigara, G. Archerfish number discrimination. *eLife* **2022**, *11*, e74057. [CrossRef]

22. Howard, S.R.; Avarguès-Weber, A.; Garcia, J.E.; Greentree, A.D.; Dyer, A.G. Numerical ordering of zero in honey bees. *Science* **2018**, *360*, 1124–1126. [[CrossRef](#)]
23. Bortot, M.; Agrillo, C.; Avarguès-Weber, A.; Bisazza, A.; Miletto Petrazzini, M.E.; Giurfa, M. Honeybees use absolute rather than relative numerosity in number discrimination. *Biol. Lett.* **2019**, *15*, 20190138. [[CrossRef](#)]
24. Halberda, J.; Mazocco, M.M.M.; Feigenson, L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* **2008**, *455*, 665–668. [[CrossRef](#)]
25. Mazocco, M.M.; Feigenson, L.; Halberda, J. Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* **2011**, *82*, 1224–1237. [[CrossRef](#)]
26. Piazza, M.; Facoetti, A.; Trussardi, A.N.; Berteletti, I.; Conte, S.; Lucangeli, D.; Dehaene, S.; Zorzi, M. Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* **2010**, *116*, 33–41. [[CrossRef](#)] [[PubMed](#)]
27. Anobile, G.; Cicchini, G.M.; Gasperini, F.; Burr, D.C. Typical numerosity adaptation despite selectively impaired number acuity in dyscalculia. *Neuropsychologia* **2018**, *120*, 43–49. [[CrossRef](#)] [[PubMed](#)]
28. Decarli, G.; Sella, F.; Lanfranchi, S.; Gerotto, G.; Gerola, S.; Cossu, G.; Zorzi, M. Severe Developmental Dyscalculia Is Characterized by Core Deficits in Both Symbolic and Nonsymbolic Number Sense. *Psychol. Sci.* **2023**, *34*, 8–21. [[CrossRef](#)] [[PubMed](#)]
29. Piazza, M.; Pinel, P.; Le Bihan, D.; Dehaene, S. A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* **2007**, *53*, 293–305. [[CrossRef](#)]
30. Cochran, A.; Cui, L.; Hubbard, E.M.; Green, C.S. “Approximate number system” training: A perceptual learning approach. *Atten. Percept. Psychophys.* **2019**, *81*, 621–636. [[CrossRef](#)]
31. Park, J.; Brannon, E.M. Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition* **2014**, *133*, 188–200. [[CrossRef](#)]
32. Togoli, I.; Fedele, M.; Fornaciai, M.; Buetti, D. Serial dependence in time and numerosity perception is dimension-specific. *J. Vis.* **2021**, *21*, 6. [[CrossRef](#)]
33. Caponi, C.; Castaldi, E.; Grasso, P.A.; Arrighi, R. Feature-selective adaptation of numerosity perception. *Proc. R. Soc. B Biol. Sci.* **2025**, *292*, 20241841. [[CrossRef](#)]
34. Pomè, A.; Anobile, G.; Cicchini, G.M.; Burr, D.C. Different reaction-times for subitizing, estimation, and texture. *J. Vis.* **2019**, *19*, 14. [[CrossRef](#)]
35. Anobile, G.; Cicchini, G.M.; Burr, D.C. Separate mechanisms for perception of numerosity and density. *Psychol. Sci.* **2014**, *25*, 265–270. [[CrossRef](#)]
36. Libertus, M.E.; Odic, D.; Feigenson, L.; Halberda, J. Effects of Visual Training of Approximate Number Sense on Auditory Number Sense and School Math Ability. *Front. Psychol.* **2020**, *11*, 2085. [[CrossRef](#)]
37. Park, J.; Brannon, E.M. Training the approximate number system improves math proficiency. *Psychol. Sci.* **2013**, *24*, 2013–2019. [[CrossRef](#)] [[PubMed](#)]
38. Szkudlarek, E.; Park, J.; Brannon, E.M. Failure to replicate the benefit of approximate arithmetic training for symbolic arithmetic fluency in adults. *Cognition* **2021**, *207*, 104521. [[CrossRef](#)] [[PubMed](#)]
39. Lindskog, M.; Winman, A. No evidence of learning in non-symbolic numerical tasks—A comment on Park and Brannon (2014). *Cognition* **2016**, *150*, 243–247. [[CrossRef](#)]
40. Merkley, R.; Matejko, A.A.; Ansari, D. Strong causal claims require strong evidence: A commentary on Wang and colleagues. *J. Exp. Child Psychol.* **2017**, *153*, 163–167. [[CrossRef](#)]
41. Szűcs, D.; Myers, T. A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends Neurosci. Educ.* **2017**, *6*, 187–203. [[CrossRef](#)]
42. Benedetto, A.; Chelli, E.; Petrizzo, I.; Arrighi, R.; Anobile, G. The role of motor effort on the sensorimotor number system. *Psychol. Res.* **2024**, *88*, 2432–2443. [[CrossRef](#)]
43. Yang, H.; Jia, L.; Zhu, J.; Zhang, J.; Li, M.; Li, C.; Pan, Y. The interplay of motor adaptation and groupitizing in numerosity perception: Insights from visual motion adaptation and proprioceptive motor adaptation. *PeerJ* **2024**, *12*, e16887. [[CrossRef](#)]
44. Anobile, G.; Arrighi, R.; Castaldi, E.; Burr, D.C. A Sensorimotor Numerosity System. *Trends Cogn. Sci.* **2021**, *25*, 24–36. [[CrossRef](#)]
45. Maldonado Moscoso, P.A.; Cicchini, G.M.; Arrighi, R.; Burr, D.C. Adaptation to hand-tapping affects sensory processing of numerosity directly: Evidence from reaction times and confidence. *Proc. Biol. Sci.* **2020**, *287*, 20200801. [[CrossRef](#)]
46. de la Rosa, S.; Breidt, M. Virtual reality: A new track in psychological research. *Br. J. Psychol.* **2018**, *109*, 427–430. [[CrossRef](#)] [[PubMed](#)]
47. Akman, E.; Cakir, R. The effect of educational virtual reality game on primary school students’ achievement and engagement in mathematics. *Interact. Learn. Environ.* **2023**, *31*, 1467–1484. [[CrossRef](#)]
48. de Castro, M.V.; Bissaco, M.A.S.; Panccioni, B.M.; Rodrigues, S.C.M.; Domingues, A.M. Effect of a Virtual Environment on the Development of Mathematical Skills in Children with Dyscalculia. *PLoS ONE* **2014**, *9*, e103354. [[CrossRef](#)] [[PubMed](#)]
49. Liu, R.X.; Wang, L.; Koszalka, T.A.; Wan, K. Effects of immersive virtual reality classrooms on students’ academic achievement, motivation and cognitive load in science lessons. *J. Comput. Assist. Learn.* **2022**, *38*, 1422–1433. [[CrossRef](#)]

50. Lackner, J.R. Motion sickness: More than nausea and vomiting. *Exp. Brain Res.* **2014**, *232*, 2493–2510. [[CrossRef](#)]
51. Monti, A.; Aglioti, S.M. Flesh and bone digital sociality: On how humans may go virtual. *Br. J. Psychol.* **2018**, *109*, 418–420. [[CrossRef](#)]
52. Pan, X.N.; Hamilton, A.F.D. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *Br. J. Psychol.* **2018**, *109*, 395–417. [[CrossRef](#)]
53. Renner, R.S.; Velichkovsky, B.M.; Helmert, J.R. The Perception of Egocentric Distances in Virtual Environments—A Review. *ACM Comput. Surv.* **2013**, *46*, 1–40. [[CrossRef](#)]
54. Albert, L.; Potheegadoo, J.; Herbelin, B.; Bernasconi, F.; Blanke, O. Numerosity estimation of virtual humans as a digital-robotic marker for hallucinations in Parkinson’s disease. *Nat. Commun.* **2024**, *15*, 1905. [[CrossRef](#)]
55. Cruzneira, C.; Sandin, D.J.; Defanti, T.A.; Kenyon, R.V.; Hart, J.C. The Cave—Audio-Visual Experience Automatic Virtual Environment. *Commun. ACM* **1992**, *35*, 64–72. [[CrossRef](#)]
56. Jensen, L.; Konradsen, F. A review of the use of virtual reality head-mounted displays in education and training. *Educ. Inf. Technol.* **2018**, *23*, 1515–1529. [[CrossRef](#)]
57. Aruanno, B.; Tamburrino, F.; Neri, P.; Barone, S. VR Lab: An Engaging Way for Learning Engineering and Material Science. In Proceedings of the International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering, Florence, Italy, 6–8 September 2023; pp. 389–396. [[CrossRef](#)]
58. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
59. Rouder, J.N.; Speckman, P.L.; Sun, D.; Morey, R.D.; Iverson, G. Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **2009**, *16*, 225–237. [[CrossRef](#)] [[PubMed](#)]
60. 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. *Behav. Res. Methods* **2007**, *39*, 175–191. [[CrossRef](#)] [[PubMed](#)]
61. Anobile, G.; Turi, M.; Cicchini, G.M.; Burr, D.C. Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *J. Vis.* **2015**, *15*, 4. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.