Adaptation to Simulated Hypergravity in a Virtual Reality Throwing Task

MATTI POUKE, University of Oulu, Oulu, Finland
ELMERI UOTILA, University of Oulu, Oulu, Finland
EVAN G. CENTER, University of Oulu, Oulu, Finland
KALLE G. TIMPERI, University of Oulu, Oulu, Finland
ALEXIS P. CHAMBERS, University of Oulu, Oulu, Finland
TIMO OJALA, University of Oulu, Oulu, Finland
STEVEN M. LAVALLE, University of Oulu, Oulu, Finland

According to previous research, humans are generally poor at adapting to earth-discrepant gravity, especially in Virtual Reality (VR), which cannot simulate the effects of gravity on the physical body. Most of the previous VR research on gravity adaptation has used perceptual or interception tasks, although adaptation to these tasks seems to be especially challenging compared to tasks with a more pronounced motor component. This paper describes the results of two between-subjects studies ($n = 60$ and $n = 42$) that investigated adaptation to increased gravity simulated by an interactive VR experience. The experimental procedure was identical in both studies: In the adaptation phase, one group was trained to throw a ball at a target using Valve Index motion controllers in gravity that was simulated at five times of earth's gravity (hypergravity group), whereas another group threw at a longer-distance target under normal gravity (normal gravity group) so that both groups had to exert the same amount of force when throwing (approximated manually in Study 1 and mathematically in Study 2). Then, in the measurement phase, both groups repeatedly threw a virtual ball at targets in normal gravity. In this phase, the trajectory of the ball was hidden at the moment of release so that the participants had to rely on their internal model of gravity to hit the targets rather than on visual feedback. Target distances were placed within the same range for both groups in the measurement phase. According to our preregistered hypotheses, we predicted that the hypergravity group would display worse overall throwing accuracy, and would specifically overshoot the target more often than the normal gravity group. Our experimental data supported both hypotheses in both studies. The findings indicate that training an interactive task in higher simulated gravity led participants in both studies to update their internal gravity models, and therefore, some adaptation to higher gravity did indeed occur. However, our exploratory analysis also indicates that the participants in the hypergravity group began to gradually regain their throwing accuracy throughout the course of the measurement phase.

CCS Concepts: • **Applied computing** → **Psychology**; • **Human-centered computing** → *Laboratory experiments.*

Additional Key Words and Phrases: virtual reality, gravity models, sensory adaptation

1 INTRODUCTION

It appears humans utilize a perceived and/or an internal representation of gravity in a variety of tasks, for example, catching, trajectory estimation, pointing, and body orientation estimation [13]. Furthermore, we appear to be especially well adapted to earth gravity; humans can correctly estimate object trajectories and intercept objects, even when the trajectories are partially occluded, as long as the objects behave according to the familiar

Authors' addresses: Matti Pouke, matti.pouke@oulu.fi, University of Oulu, Oulu, Finland; Elmeri Uotila, University of Oulu, Oulu, Finland; Evan G. Center, University of Oulu, Oulu, Finland; Kalle G. Timperi, University of Oulu, Oulu, Finland; Alexis P. Chambers, University of Oulu, Oulu, Finland; Timo Ojala, University of Oulu, Oulu, Finland; Steven M. LaValle, University of Oulu, Oulu, Finland.

ACM 1544-3558/2024/2-ART
https://doi.org/10.1145/3643849

downwards acceleration of 9.81 m/s^2 (1g) [13, 20, 42]. Acting outside of earth gravity, however, can cause difficulties as various tasks rely on the internal model of gravity calibrated at 1g. For most of us, it is, of course, rather uncommon to experience earth-discrepant gravity in our everyday lives, as such experiences are typically reserved for spacecraft crew or researchers working with simulated gravity environments, such as parabolic flights or rotating rooms.

Virtual Reality (VR) offers everyone the capability to experience something resembling the aforementioned conditions, although its capabilities are mostly limited to audiovisual stimuli, as contemporary VR hardware cannot stimulate internal senses that are affected by true earth-discrepant gravity. The fact that it's impossible to simulate weightlessness or earth-discrepant gravity realistically has often been used as an example of the physical limits of VR technology; no system, no matter how advanced, can manipulate the true effects of gravity for our internal senses (e.g., [36, 37]). However, it has also been argued that the purpose of VR is not to generate perfect artificial stimuli but instead generate *illusions* of alternate realities, such as earth-discrepant gravity [37]. In any case, visual simulation of gravity has been found sufficient in many works of gravity-related research ranging from adaptation studies to astronaut training applications (see below).

Humans' relationship with gravity has attracted a lot of research on various application areas of VR. For example, various training simulations of space-related zero-gravity tasks have been developed for VR (e.g. [5, 22, 26, 34]). Jiang et al. [12] employed VR to study the effects of various colour schemes on cognitive performance and emotion while simulating the effects of altered gravity conditions using tilted bed rest methodology (see e.g. [14]). Aoki et al. [2] studied visual information acquisition and spatial cognition using simulated zero-gravity in VR. VR and artificially manipulated gravity have also been used for rehabilitation training for patients with gait and balance issues [27]. Brubach et al. [4] used altered gravity to manipulate plausibility in a VR experiment. Jörges and López-Moliner [13] argue that in the future, humans are likely to be exposed to earth-discrepant gravity conditions in an increasing manner, either because of actual space travel, inhabiting foreign worlds, or because of various VR and Augmented Reality (AR) systems that visually simulate altered gravity on earth for different purposes. Therefore, they consider it worthwhile to study the human capability to adapt to earth-discrepant gravity, as altered gravity conditions, both visually simulated and actual, significantly affect human performance in various tasks.

1.1 Adaptation to earth-discrepant gravity

The so-called "Prism adaptation" is an example of how the manipulation of the visual field can be used to recalibrate human visuomotor capabilities. In Prism adaptation, the participant's visual field is manipulated so that a forward target appears to be displaced horizontally; when the participant attempts to point or, for example, to throw objects at the target, the participant's aim is similarly displaced. When pointing, an offset from the actual target location is directly observable when the participant's hand is (partially) occluded and one cannot visually guide the hand to the target. After repeated attempts, the participant adapts to the visual displacement and gains the ability to hit the target. However, after the visual displacement is removed and the normal visual field is regained, the participant again temporarily lacks the capability to hit the target because of the adaptation that happened during practice. It appears that age can affect the adaptation effect as well as the time needed to readapt back to normal visual field [1, 32].

Unlike in Prism adaptation, however, humans seem to somewhat resist adapting to altered gravity conditions. Previous studies suggest that humans possess a strong internal model of Earth's gravity and linear acceleration [24]. According to the survey by Jörges and López-Moliner [13], our ability to adapt to different gravity would be limited, especially in VR. Slow adaptation can take place in outer space where both external and internal senses are accommodated equally, however, adaptation would be impossible in VR where only audiovisual stimuli is available. They also argued that in general, tasks with a pronounced motor component, such as point-to-point arm movement, allow for faster adaptation in actual earth-discrepant gravity than interception tasks [8, 13, 19]. It should be noted, though, that several of the VR studies cited in the survey did not utilize immersive hardware similar to contemporary VR systems, but 2D projection walls, for example. In addition, the reported VR studies primarily used interception tasks, whereas studies with tasks having a pronounced motor component were performed in actual earth-discrepant gravity [13]; this suggests VR studies with a pronounced motor component are needed.

McIntyre et al. [23] investigated the capability of astronauts to intercept a vertically moving ball in space. The astronauts performed catching movements earlier in zero gravity than in normal gravity, which indicated using an internal model of gravity in addition to perceptual cues. Adaptation to zero-gravity did occur over the course of two weeks. Gaveau et al. [9] investigated cosmonauts using point-to-point arm movement tasks before and after a spaceflight. They found evidence of the participants adapting their sensorimotor systems during the course of the spaceflight. Adaptation to decreased gravity led to overestimation of earth's gravity upon return, as if their arm kinematics were optimized for increased gravity. They further theorized that abnormal vestibular and proprioceptive information could have caused them to feel abnormally high gravity as the participants could sense an increase in gravity upon return to Earth.

Zago et al. [41] used a projection system to estimate participants' capability to intercept disappearing vertically downward moving targets with randomized laws of motion. They found that participants systematically expected objects to adhere to normal gravity and were surprisingly resistant to adapting to new models of gravity, although by training, the participants were able to adjust the timing of their motor activation. A follow-up study suggested that humans do not adapt to new models of gravity, but merely utilize the default earth gravity model and adjust central processing time when learning to intercept targets moving at constant velocity [42].

Senot et al. [35] utilized immersive stereoscopic VR and tasked participants to use a racket to intercept balls that were moving either in an upwards or downwards motion with either constant, accelerating, or decelerating velocity. Although the participants' performance was best with targets moving in constant velocity, the authors found that participants systematically triggered the racket motion earlier with targets falling from above compared to when they were rising from below, suggesting that participants were anticipating downward moving targets to be influenced by gravity regardless of their true acceleration.

Ye et al. [40] investigated task performance and adaptation in VR using four different tasks in different gravitational settings: striking a ball to hit the target, triggering a ball to hit a target, predicting the landing location of a projectile, and estimating flight duration of a projectile. The participants again showed a tendency to base their physics intuitions according to earth gravity; however, overall, they were able to adapt to the tasks presented in this study in terms of performance and accuracy.

Gravano et al. [10] explored the use of mental imagery in adapting astronauts' sensorimotor capabilities towards 0g conditions by grasping, throwing and catching an imaginary ball; their results suggest that the astronauts' internal gravity models were updated even though the training was performed in normal gravity.

Cano Porras et al. [6] studied simulated gravity and locomotion adaptation, and found that visual-only gravity cues caused by virtual inclines in VR made participants to initially ignore body-based cues and adapt their locomotion as if they were walking on actual inclines. Gradually, the body-based cues took over, however.

Although humans are generally good at predicting object trajectories in motor tasks, it has been argued that humans generally have poor intuitions regarding physics and tend to make bad perceptual judgements based on physics [17, 19]. Ullman et al. [38] argued that humans interpret their physics judgements similarly to contemporary game engines; instead of precise computations, humans utilize shortcuts and "good enough" guesses to predict object trajectories. In VR, it helps if the virtual environment contains rich size cues to help gravity estimation [19]. In fact, it could be that conflicting size cues, such as virtual characters whose size differs from the rest of the virtual scene, can confound gravity judgements [29]. Because of this potential discrepancy between performance in motor tasks and perceptual tasks, La Scaleia [19] investigated whether the internal gravity model only affects the former and not the latter. They used a VR system where a ball was first seen rolling and then falling in different simulated gravity conditions while the ball could be either visible or occluded through the free-fall phase. Participants' perceptual judgements of naturalness, as well as their accuracy in intercepting the ball, were quantified. In the end, they concluded that the internal gravity model was utilized in both perceptual as well as interception tasks.

1.2 The relationship between scale and gravity

Besides simulating earth-discrepant gravity, many VR applications utilize the manipulation of user scale, which can also affect the subjective perception of gravity, provided these applications simulate rigid body dynamics of objects. There are multiple reasons to manipulate users' scale in VR: Applications, such as multiscale Collaborative Virtual Environments (mCVE) allow different users to, for example, investigate architectural and medical visualizations from different scales and perspectives [16, 44]. Scaling of VR users has also been utilized in, for example, locomotion [18], remote collaboration [28], and interior design [43].

However, a realistic simulation of rigid body dynamics produces interesting perceptual effects when the scale of the VR user changes. For example, a human scaled down ten times smaller would observe an object dropped at shoulder-height to hit the ground in roughly 0.17 s. At normal scale dropping an object similarly would take 0.55 s to hit the ground, and at tenfold scale it would take 1.75 s. Therefore, in the eyes of the VR user, object motions appear as if gravity was ten times stronger when the user is ten times smaller, and similarly weaker if the VR user is scaled up, greatly affecting perceived accelerations as well as throwing distances [29, 30]. The human internal model of gravity can thus pose challenges for VR users not only when simulating non-earthlike experiences, but also in mCVEs and other scale-varying applications that simulate gravity. Although in single-user applications, the developer can always adjust simulated gravity to match the user's expectations (provided physical accuracy is not needed), in mCVEs this would not be possible, since simulated gravity would not appear normal for simultaneous users coexisting at different scales. At very small scales, additional physical peculiarities come into play; for example, the work of [25] have utilized VR for teaching nanophysics in robotics operations.

The peculiarities of perceiving rigid body dynamics at various scales have been the focus of our previous studies. In [30] we studied the subjective realism of physics models at various scales and found that when interacting with physically simulated objects, participants tended to consider physics models where gravity was adjusted to match the participant's scale (i.e. gravity was adjusted to be stronger when participants were taller and vice versa) to be more realistic. On the other hand, physics models in which gravity remained unaltered were perceived as unrealistic. This means the participants tended to use their own size as the metric when judging the realism of rigid body dynamics; a model in which objects would behave as if the participants were normal sized and the environment scaled instead was considered as the realistic one. Although it appears humans use their own body as the metric for scale among discrepant size cues in VR, there is some evidence that other virtual characters can disrupt this metric [21, 29].

1.3 Study Overview

Whereas our previous work focused on the subjective perception of realism at 0.1x scale [31], 10x scale [31], as well as 0.2x scale [29], in this work we focus on adapting to gravity conditions similar to those experienced when the participant is at 0.2x scale. We predict that the internal model of gravity could be updated by exposing users to an interaction task that takes place under simulated earth-discrepant gravity in VR. More specifically, we focus on the task of throwing a ball under simulated hypergravity. Somewhat similar to Prism adaptation studies, as well as to the gravity adaptation study of Gaveau et al. [9], we use the aftereffects measured from the results of a pronounced motor task as evidence for adaptation. As for assessing the internal model of gravity, we rely on the human capability to predict partially occluded trajectories [13, 20, 41]. Similarly to earlier studies involving

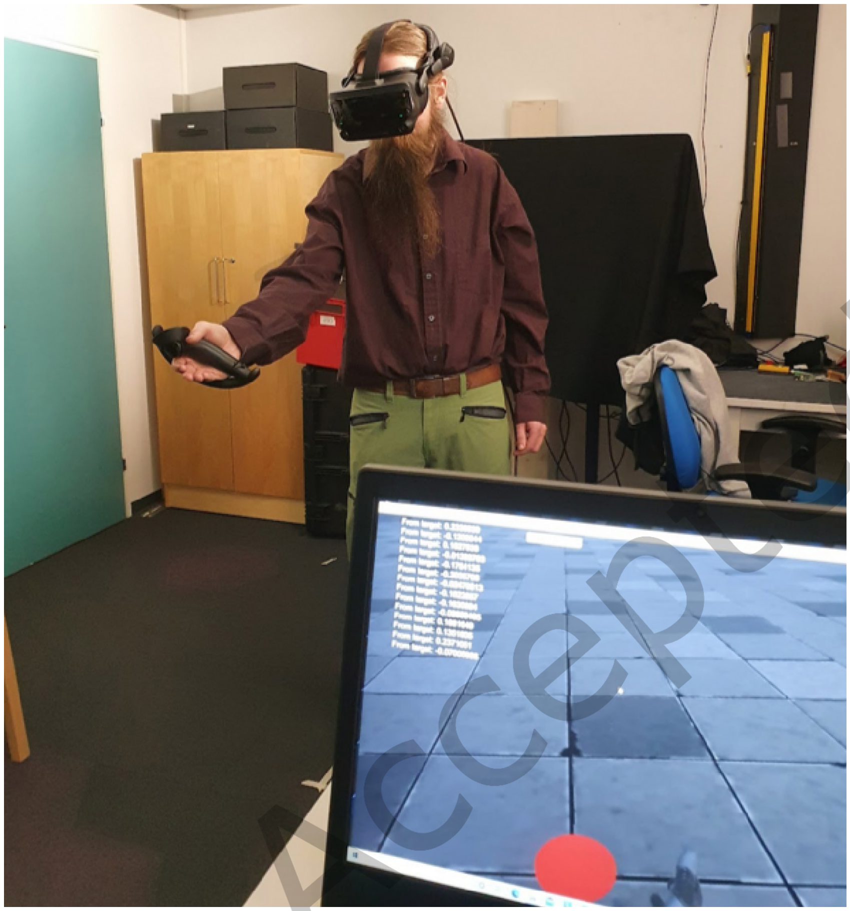

Fig. 1. A participant performing the underhand throw technique that was used throughout the experiment. The experimenter could see the participant's viewpoint overlayed with data and interface information that were visible for the experimenter only.

trajectory estimation (e.g. [41, 42]), our measurement task uses a disappearing ball so that the participants have to rely on their internal model of gravity when performing the throwing task. We choose 5g as our simulated hypergravity, as this roughly corresponds to the subjective experience of gravity at 0.2 scale, which has been utilized in several earlier studies investigating scale (e.g. [29, 39]). We ran two studies mostly identical in content, with the second study serving as a replication of the first one to confirm the robustness of our findings. Our results suggest that our participants' internal models of gravity were at least temporarily updated to some extent. We consider these results as a humble but promising step towards building applications that would allow training and adaptation towards altered gravity conditions using VR-simulated gravity.

2 STUDY 1

2.1 Materials and Methods

2.1.1 Conditions and hypotheses. We designed a between-subjects experiment (n = 60) to examine adaptation to simulated gravity by having participants throw a virtual ball in VR. We used *a priori* power analysis to determine our sample size, which targeted adequate power (80% or greater) to detect large effects (Cohen's d = 0.7 or greater) using either parametric or nonparametric tests. In the adaptation phase, one group of participants was trained to throw a ball under hypergravity (5g), whereas another group instead threw at a longer-distance target under normal gravity (1g). We call these groups *hypergravity* and *normal gravity* groups, respectively. The distances to targets (13.45 m under normal gravity and 2.67 m under hypergravity) were placed so that both groups had to exert roughly the same amount of force to throw the ball at the target; using targets at a similar distance would have caused a confound due to the significantly different amount of strength needed to reach the targets. Although we could have instead opted for similar target distances at the cost of confounding the required amount of strength, we expected that matching the fundamental motor component of the adaptation phase would better equate the experience of both groups and serve as a more robust baseline than opting to match the visual component. The distance of the hypergravity target was approximated manually by a researcher, aiming for a long enough distance to require an actual underhand throw to reach the target, but close enough so that all participants would be able to exert enough strength to do so. The distance to the normal-gravity target was then manually approximated so that a roughly similar amount of strength was needed to reach that target as well.

Then, in the measurement phase, to quantify adaptation, both groups threw at a nearby target at a distance randomly fluctuating between a minimum of 2.54 m and a maximum of 3.54 m between throws. The virtual ball disappeared 0.1 s after throwing so that participants could not adjust their throwing accuracy according to visual feedback but instead had to rely on their internal model of gravity (e.g. [13, 41, 42]). The mean accuracy of the throws was then assessed. We gauged accuracy in terms of both absolute and signed error, as the measures provide two related but distinct channels of information: in the former case, how well a participant performs at the task in general, and in the latter, whether there is a directional bias in terms of consistent under- or over-throwing. The distinction can be illustrated in the hypothetical case of a participant who tends to greatly under- or over-throw the target with equal probability, but never land a direct hit, as averaging this participant's errors would lead to a large absolute error but a signed error near zero. Thus, a difference in absolute errors between groups would signify that increasing the strength of gravity can impair performance, but only a difference in signed errors in the predicted direction would signify that participants truly updated their internal models to be consistent with the new environment.

Our preregistered hypotheses[1] were as follows:

- **H1:** The absolute units from target error of participants conditioned to hypergravity is higher than that of participants conditioned to normal gravity instead, while the ball is hidden.
- **H2:** Participants conditioned to hypergravity will overshoot the target relative to participants conditioned to normal gravity instead, while the ball is hidden.

2.1.2 Experimental prototype. The VR application was made in Unity and used the Valve Index VR Kit. Valve Index HMD and controllers were used as VR hardware. The virtual environment consisted of a single large room with simple graphics, in which the contents could be manipulated by the experimenter in real time. The VR users' hands were visualized with the SteamVR skeletal hand meshes for Unity, which also acted as size cues for the participants. The application was running at 144 Hz to ensure smooth motions even in hypergravity. The application contained multiple scenes that could be switched between practice mode (Figure 3) and the hypergravity and normal gravity targets (Figure 4). The participants could see a yellow circle, which marked

[1] https://osf.io/fr32t/?view_only=5039937bcf33478398725665e5a535f7

the spot where they were instructed to stand. The spot was also visible in the real-world laboratory to ensure all participants would begin the experiment standing in the correct location in VR. In addition, the application allowed the experimenter to toggle simulated gravity between 5g and 1g, as well as toggle the functionality that made the ball disappear automatically 0.1 s after throwing. Simulation of rigid body dynamics at 1g and 5g was handled by Unity's internal physics engine. The virtual ball was grabbed by squeezing the Valve Index controller and thrown by performing an underhand throwing motion and releasing the grip while a strap kept the controller in place; using this controller, the VR user can merely grab the handle of the controller and no actual button presses are needed for either grasping or release gestures. Similarly to default VR throwing implementations in contemporary game engines, the trajectory of the ball is determined by its velocity at the time of release. However, we took steps to improve the default throwing model with the aim of making the throwing more predictable and intuitive. First, we wanted to smooth out inaccuracies in the direction of the throw that often takes place in VR throwing when the velocity of the thrown object is acquired from the time of release or averaged from previous frames. The purpose of our smoothing technique was to mitigate sudden movements at the end of the throw influencing the velocity vector of the ball at the time of release. We defined the virtual ball to fall slightly behind the hand mesh while grabbed, with the distance between the objects increasing with faster motion (this offset was too small to be detectable by the user). This was accomplished by defining the velocity of the held ball at each frame as the vector from ball to hand multiplied by a constant $c1$. The value for $c1$ was acquired empirically by testing different values until we were satisfied with the direction of the ball when thrown. Next, we defined a second constant $c2$ that was used for defining the magnitude of the velocity of the ball when thrown; the velocity vector of the ball was multiplied by constant $c2$ at the time of release. This constant was found empirically, as well. We manually repeated the same throwing motion with a real lightweight object and the virtual ball using the same technique and strength. Repeating this procedure, we incrementally adjusted constant $c2$ until the distances of both objects were approximately equal.

Throw accuracy (error from target) was defined as target distance (distance from the point of release to the target center ignoring depth) subtracted from throw distance (distance from the point of release to the ball landing point ignoring depth, see Fig. 2). Since we wanted to eliminate horizontal errors from our measurements, we did not compute the Euclidean distance from the landing position to the target position. This way, difficulties in throwing in a straight line would not bias the results; our main interest was in throwing strength as that was the only dimension that would be affected by simulated gravity. If the ball did not fly past a visible gap in front of the throwing position, it was not counted as a throw. This was to eliminate accidental drops of the ball from the data.

2.1.3 Experimental process. In Study 1, 60 participants (18 identified as females whereas 42 identified as males, with numbers of males and females balanced in each group) took part in the experiment. Five other participants were discarded due to not completing the study protocol in the intended manner (for example, using an overhanded throwing technique in the measurement phase). The researchers followed a script to instruct the participants sequentially at the beginning of each phase throughout the experimental procedure. The procedure was as follows:

(1) **Instructions phase.** Participants first provided written informed consent per the experimental protocol approved by the local Ethics Review Board (ERB). Participants were then instructed on how to use the Valve Index controllers and the HMD, and the desired underhand throwing technique was demonstrated. They were also instructed to use the same hand for throwing throughout the experiment. Fig. 1 shows an example of a participant performing the intended throwing gesture.
(2) **Practice phase.** Participants were tasked with knocking several balls off of pillars by throwing a ball and hitting them in normal gravity. The purpose of this task was to familiarize the participants with throwing in VR. No data was collected at this phase. We did not limit the number of trials at this point; all participants knocked down all the spheres before proceeding.

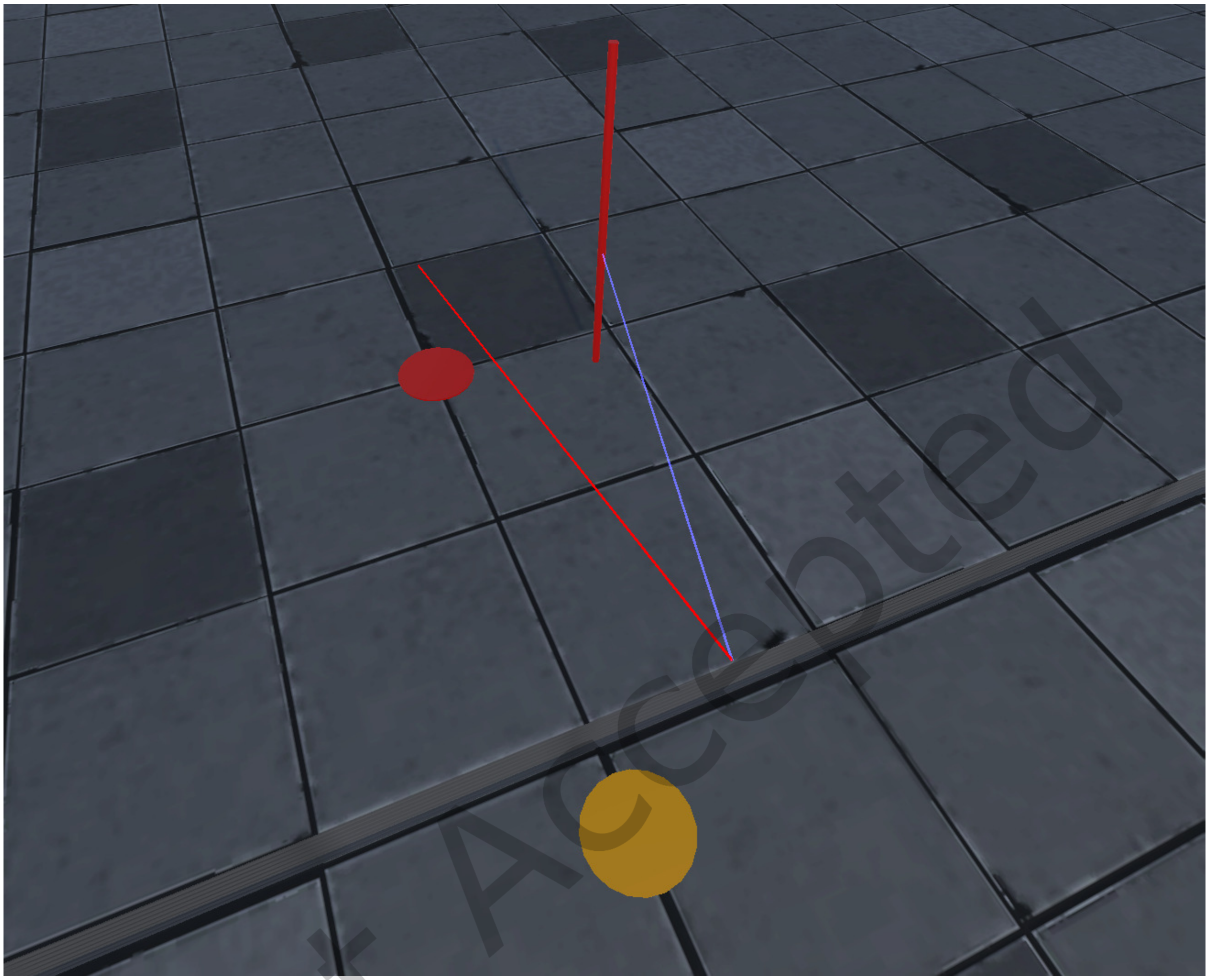

Fig. 2. Throw accuracy computation visualized after throwing the ball purposedly off the target. The blue line denotes the distance from the point of release to the landing position. The red line denotes the distance from the point of release to the target. The final result is the subtraction of the line magnitudes. The visualization was not used in the actual experiments and was therefore not seen either by researchers or participants during data collection.

(3) **Adaptation phase.** Depending on the group, participants either threw the ball at a target close by (2.67 m) under 5g (hypergravity group) or at a target further away (13.45 m) under 1g (normal gravity group), 20 times. As described earlier, the distances were selected so that both groups had to exert approximately similar strength to reach the target as approximated by a researcher. Participants in the hypergravity group were told that there would be an increase in gravity.

(4) **Measurement phase.** Participants threw the ball at a target under 1g while the ball disappeared 0.1 s after release so they had to rely on their internal gravity model when approximating the ball's trajectory (somewhat similar to e.g. [19, 41, 42]). The target would also move slightly between throws so that the distance of the target randomly varied between 2.54 m and 3.54 m. The range of distances was set so that

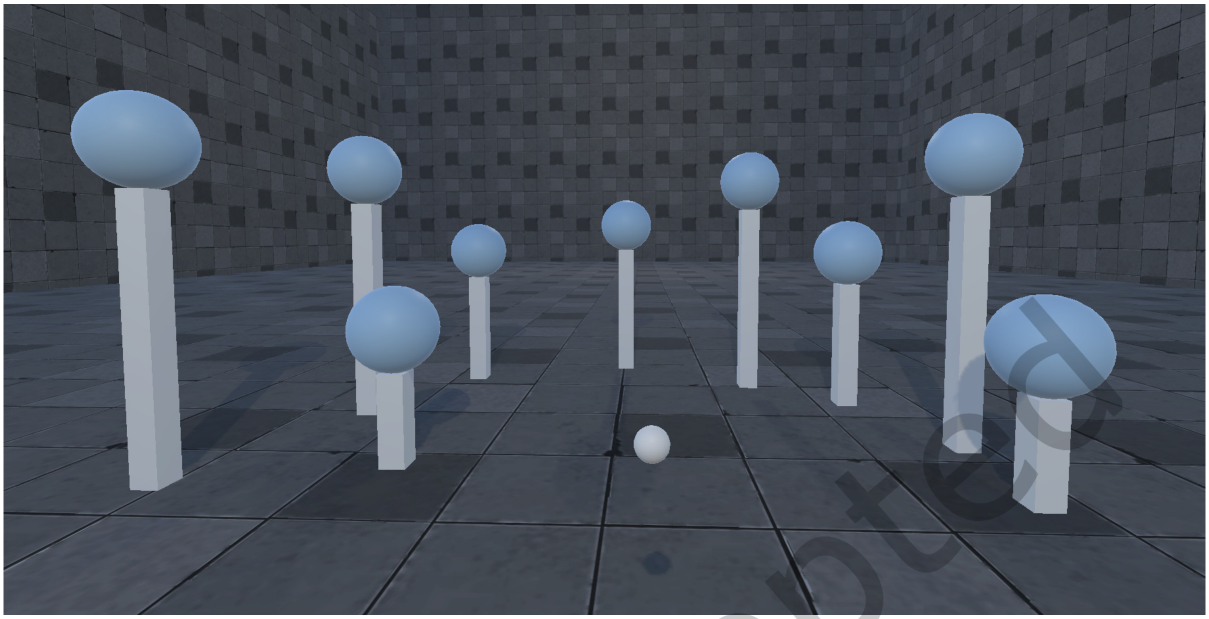

Fig. 3. Practice phase setup including the ball for throwing; screenshot taken from Unity editor (note that the camera is placed somewhat higher than what an average participant was seeing). The furthest column (in the middle) was 6.6 m from the designated throwing location whereas the closest one (third from left) was at 4.0 m. The sphere heights ranged between 1.9 m and 2.9 m (measured from the floor to the sphere center point).

both groups had to exert a significantly different amount of strength compared to the previous phase. Because of this, neither group could rely on muscle memory when throwing. Our scripted instructions ensured that the participants in the hypergravity group were explicitly told that the gravity had now reverted back to normal; once when introducing this phase, and another time as a reminder right before participants started throwing. This phase also involved 20 throws.

(5) **Final phase.** The ball was fully visible again, but otherwise, the conditions remained the same as in the measurement phase. The participants were again tasked with throwing 20 balls at the target. This phase was included for exploratory analysis only.

(6) **Post-experiment phase.** The participants were instructed to fill out a post-experiment questionnaire, debriefed, and compensated with a gift card worth 10€.

The post-experiment questionnaire was administered for the purpose of collecting data for exploratory analysis. This included rating on a 7-point Likert scale their confidence in their accuracy during the measurement phase, how realistic the throwing felt, their level of video game and VR experience, as well as their age. The exact questions for confidence and realism were stated as follows:

"How confident were you in your throwing accuracy when the ball was hidden" where a score of 1 was defined as *"I have no idea where any of my throws landed"* and 7 was defined as *"I feel like most of my throws landed on target"*. We asked the participants to report their confidence twice in the post-experiment questionnaire, using questions *"Before the ball was revealed again"* as well as *"After the ball was revealed again"*.

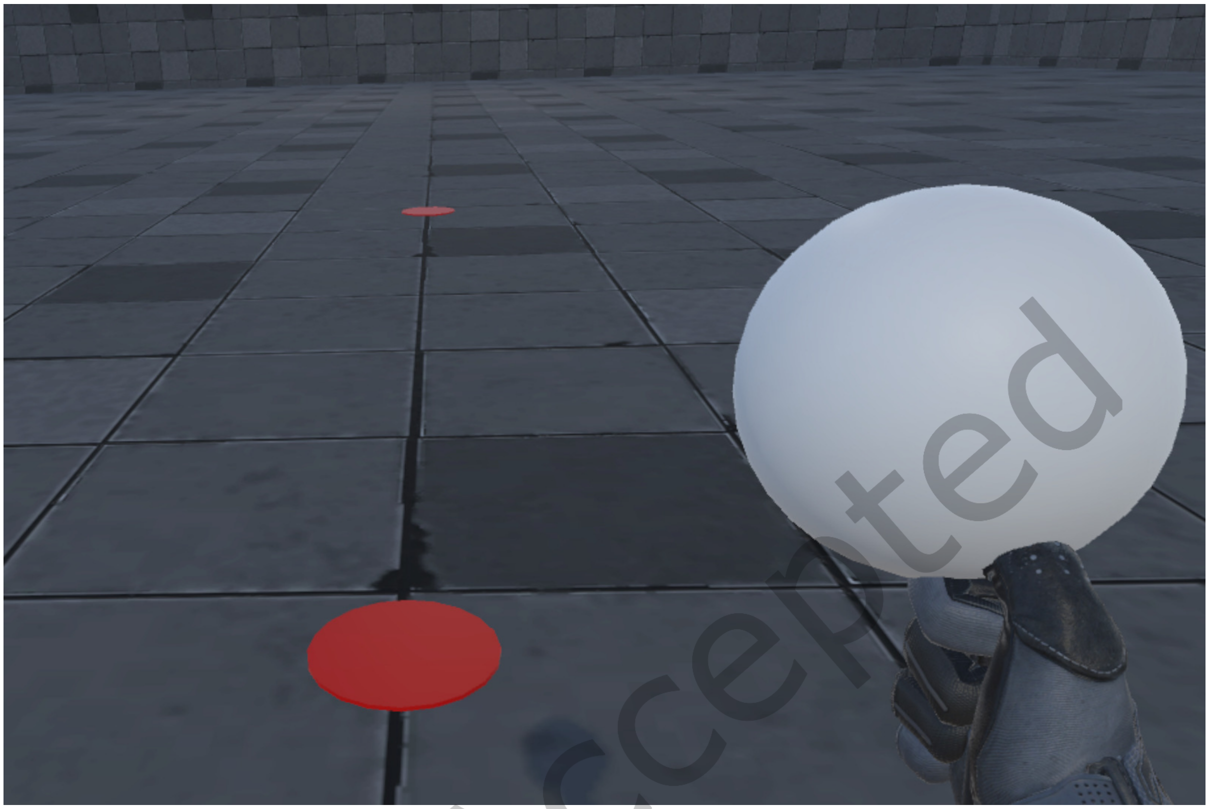

Fig. 4. Hypergravity and normal gravity targets shown at the same time from the point of throwing. Screenshot taken from editor (the targets were not shown to participants simultaneously). Hand and ball mesh models are shown in the foreground.

"How similar did throwing in the experiment feel compared to reality?" where 1 was defined as *"Did not feel real"* and 7 as *"Felt like throwing in real life.*

For Covid-19 precautions, experimenters wore masks at all times. Masks and hand sanitizer were also available for participants. Surfaces were sanitized with alcohol wipes and the HMD were disinfected using a Cleanbox device [2].

2.2 Study 1 Results

Box and whisker plots for each phase of the experiment can be seen in Figures 5, with individual participant means represented as dots. In the signed charts (Figure 5B), positive values indicate overshooting the target, while negative values indicate undershooting. To facilitate visual comparison of hypergravity and normal gravity adaptation phase errors, the errors in the adaptation phase for the normal gravity group were scaled to the target distance of the normal gravity group, as the longer throwing distance in the adaptation phase for the normal gravity group naturally leads to larger errors. Specifically, errors in the adaptation phase for the normal gravity group were divided by the ratio of the target distances in the normal gravity group (13.45 m) to the hypergravity

[2]https://cleanboxtech.com/

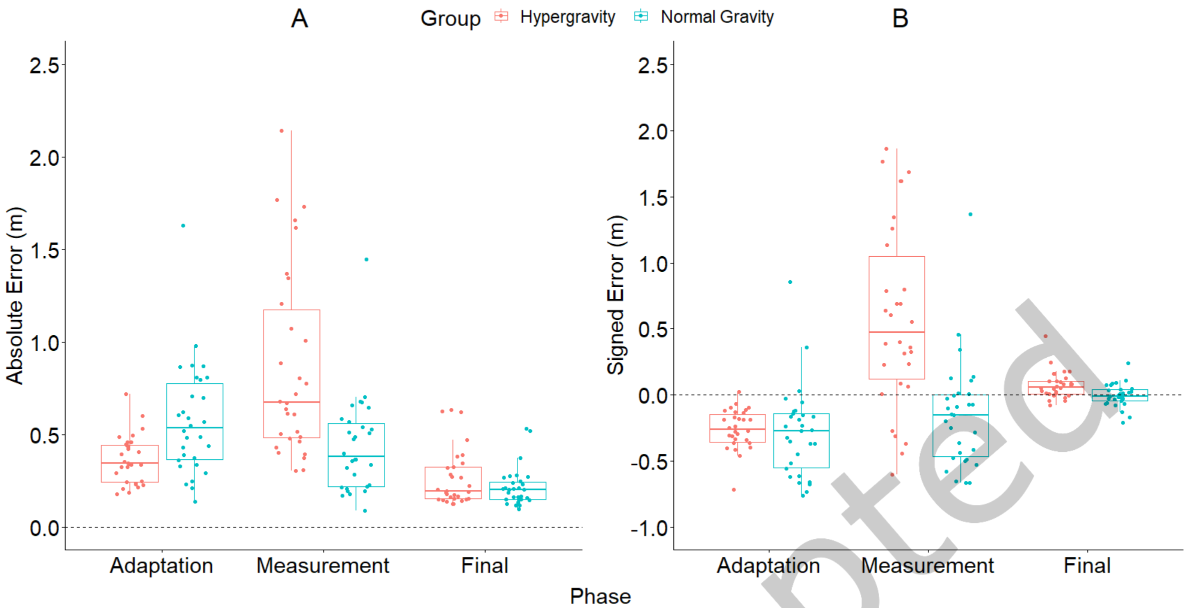

Fig. 5. Boxplots depicting the absolute error (A) and signed error (B) for hypergravity (coral; left pairs) and normal gravity (teal; right pairs) groups, for adaptation, measurement, and final phases of Study 1.

group (2.67 m), resulting in a scaling factor of 5.04. All normal gravity adaptation phase errors were divided by this factor before plotting. No statistical comparisons were performed for the adaptation phase.

2.2.1 Study 1 Confirmatory results. First, we assessed the normality of distributions via graphical inspection and Shapiro-Wilk tests. Q-Q plots revealed deviations from normality for all distributions, and only the signed error distribution from the hypergravity group satisfied the Shapiro-Wilk test ($W = 0.96, p = .23$; all other $ps <= .002$), therefore nonparametric Wilcoxon rank-sum tests [3] were used to compare groups (one-sided, $\alpha = .05$; see preregistration). The difference in absolute errors between hypergravity ($Mdn = 0.67$) and normal gravity ($Mdn = 0.38$) groups in the measurement phase (hypothesis H1) was significant, $Z = 4.10$, $p = 4.15 \times 10^{-5}$, $r = .53$. The difference in signed errors between hypergravity ($Mdn = 0.47$) and normal gravity ($Mdn = -0.15$) in the measurement phase (hypothesis H2) was also significant, $Z = 4.75$, $p = 2.06 \times 10^{-6}$, $r = .61$.

Our results, therefore, support both hypotheses, as gravity adaptation seems to have caused those in the hypergravity group to not only throw less accurately but also to systematically overshoot the target relative to those in the normal gravity group.

2.2.2 Study 1 Exploratory results. Though the differing target distances and associated errors of the groups' adaptation phases preclude meaningful statistical comparison between them, the absolute errors in the final phases did not statistically differ between hypergravity ($Mdn = 0.19$) and normal gravity ($Mdn = 0.20$), $Z = 0.95$, $p = .34$, $r = .12$, indicating that the hypergravity group was able to recalibrate their internal gravity model after receiving visual feedback and that there were no differences in general throwing ability between the two groups.

[3]Referred to as the "Mann-Whitney U" test in the preregistration; also known as the "Mann-Whitney-Wilcoxon," or "Wilcoxon-Mann-Whitney" test.

We also collected questionnaire data regarding sex, age, self-assessed confidence, the realism of the throwing model, and previous experience with video games and VR. The sex ratios of the groups were matched by design, and Wilcoxon-ranked sum tests indicated that the groups also did not happen to differ in age, video game experience, or VR experience (all $ps > .05$). In terms of rating the realism of the throwing model, there were again no differences between the hypergravity ($Mdn = 5$) and normal gravity ($Mdn = 5$), $Z = 0.68$, $p = .49$, $r = .09$, suggesting that both groups "bought in" to the realism of the throwing experience to a similar degree. The two questions concerning confidence were intended to assess participants' confidence in the accuracy of their throws during the measurement phase only, both before their throw trajectories were revealed in the final phase (item 1) and after (item 2). However, given participants' true accuracy and the large increase in scores across all participants for confidence item 2 ($Mdn = 5.5$) relative to confidence item 1 ($Mdn = 3$), $Z = 4.10$, $p = 4.15 \times 10^{-5}$, $r = .53$, it seems possible that item 2 was often misinterpreted as asking for their confidence in their throws while the throw trajectories were visible, and thus participants instead rated their confidence of the final phase.

Confidence before trajectories were revealed was significantly higher for the normal gravity group ($Mdn = 4$) than the hypergravity group ($Mdn = 3$), $Z = 2.11$, $p = .034$, $r = .27$, though for confidence after the reveal, the normal gravity group ($Mdn = 6$) and hypergravity group ($Mdn = 5$) did not differ $Z = 1.08$, $p = .28$, $r = .14$.

2.3 Deficiencies in the experimental prototype

After data collection and analysis, we found limitations within the design of the experimental prototype used in Study 1. Firstly, the throwing distances for the adaptation phase targets for *normal gravity* and *hypergravity* conditions were acquired by a single experimenter gauging the distances manually; this approach is obviously very subjective and limited in accuracy. Secondly, we found a bug in the throwing error computation which subtracted the radius of the target disc from each result. This means that each throw was reported as being 25 cm closer to the center of the target than it should have been, except for the throws that landed inside the target radius, which not only had the wrong distance, but a wrong sign as well (for example, a throw error of 10 cm would become -15 cm). Although we were still somewhat confident of the effect we found, as the issues described affected both groups equally, the deficiencies in the prototype implementation obviously undermined the reliability of the results in Study 1, and a replication study was needed to confirm the existence of the effect we found.

3 STUDY 2

The purpose of Study 2 was to perform a replication of the previous study after fixing the distance calculation bug. Furthermore, we adjusted the distance of the *normal gravity* target according to the formulation below. Finally, we also reworded the questionnaire regarding confidence items 1 and 2.

3.1 Materials and Methods

3.1.1 Conditions and hypotheses. The conditions and hypotheses for Study 2 were the same as for Study 1. We collected data from 42 participants, 21 in each group. According to our power analysis, this sample size gave us adequate power (90% or greater) to detect large effects (Cohen's d = 1.0 or greater) using either parametric or nonparametric tests. The effect size estimate was based on results obtained from the first experiment, although due to the deficiencies reported above, we did not use the exact effect size acquired from Study 1. Based on its results, we did expect a large effect, however. The preregistered hypotheses[4] were similar to those used in Study 1:

- H1: The absolute units from target error of participants conditioned to high gravity is higher than that of participants conditioned to normal gravity instead, while the ball is hidden.

[4]https://osf.io/r7tc6

- H2: Participants conditioned to high gravity will overshoot the target relative to participants conditioned to normal gravity instead, while the ball is hidden.

3.1.2 Determining throwing distances. As explained in Section 2.1.1, the throwing tasks in the adaptation phase were designed to require roughly the same amount of force from each participant, regardless of whether hypergravity or normal gravity was used. The distance for the hypergravity target was experimentally defined as a distance that we assumed all participants would be able to reach by throwing. The throw target of the normal gravity group was therefore set further away compared to the hypergravity group. In order to calibrate precisely the required difference between the target distances, we investigated how the horizontal displacement of a ball with fixed initial altitude and speed is affected by variations of the horizontal angle of its initial velocity under the two different gravity conditions. Note that if the ball is launched from zero height, the initial angle that maximizes the throw distance is always 45 degrees irrespective of the gravity strength. However, since the participants are throwing the ball roughly from their waist level, the optimal initial angle differs from 45 degrees to an extent that depends both on the initial height and the strength of downward acceleration due to gravity.

We ignore air resistance and assume that the trajectory of the ball is completely determined by its initial position $\mathbf{x}_0 = (x_0, y_0)$ in xy-coordinates, its initial velocity $\mathbf{v}_0 := \mathbf{v}(0) = (v_x(0), v_y(0))$, and the angle θ between the horizontal axis and the initial velocity vector. Note that it is not necessary to know the exact force required to throw the ball to the target in the different gravity conditions, but we need to ensure that these forces are approximately equal in both cases. Thus, it is sufficient to determine target distances for which the required initial speeds match under the two gravity conditions. In order to do this, we compute the throw distance as a function of vertical acceleration a, initial position $\mathbf{x}_0$, and initial velocity $\mathbf{v}_0$ (which is defined by the initial throw speed $\|\mathbf{v}_0\|$ and throw angle θ). We then solve for the throw angles that produced the longest throw in each gravity condition, with all the other variables kept fixed. For simplicity, the mass of the ball was assumed to be 1kg, and the initial altitude of the ball was assumed to be 1m. For vertical acceleration, we consider the possibility of regular ($a = g$) or five-fold gravity ($a = 5g$), where $1g = -9.81m/s^2$.

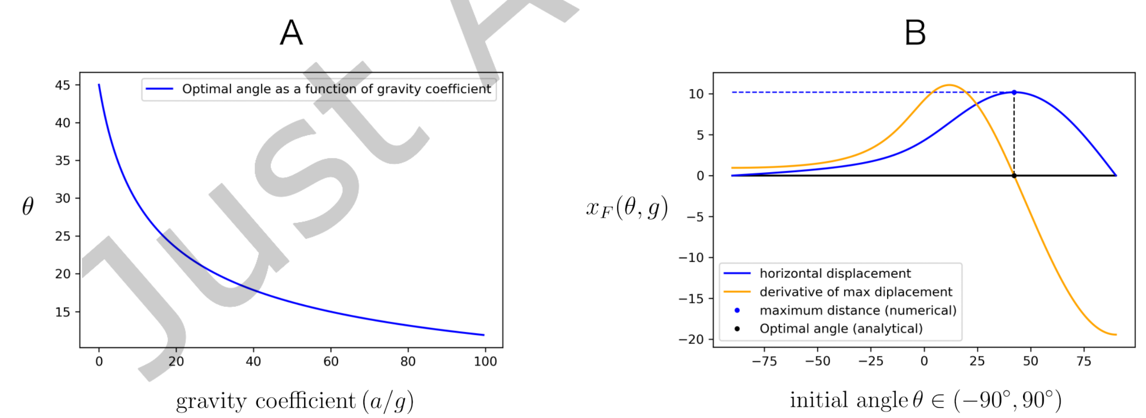

Fig. 6. (A) Dependence of optimal throw angle on gravity coefficient (acceleration divided by gravity). As gravity strength increases, the optimal initial angle approaches zero. Note that if the ball had been thrown from zero initial height, the optimal initial would be 45° regardless of gravity. (B) Throw distances for different initial angles under normal gravity and initial speed 9.524m/s. The optimal angle (blue dot) produces the maximal throw distance 10.197m. The yellow curve shows the derivative of the maximal throw distance as a function of the initial angle.

The time-dependent x- and y-coordinates of the ball are given by

$$x(t) = x_0 + \|\mathbf{v}_0\| t \cos\theta, \qquad y(t) = y_0 + \|\mathbf{v}_0\| t \sin\theta + \frac{1}{2} a t^2. \tag{1}$$

The trajectory of the ball terminates when its altitude reaches zero. The final time t_F thus satisfies $y(t_F) = 0$. Solving this equation, we obtain[5]

$$t_F := t_F(\theta, a) = \frac{-\|\mathbf{v}_0\| \sin\theta - \sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}}{a}. \tag{2}$$

The notation $t_F(\theta, a)$ emphasizes the dependence of t_F on the initial angle θ and the vertical acceleration a. The corresponding final horizontal displacement $x_F := x_F(\theta, a)$ satisfies

$$x_F(\theta, a) = x(t_F(\theta, a)) = x_0 + \|\mathbf{v}_0\| t_F(\theta, a) \cos\theta \tag{3}$$

$$= x_0 + \|\mathbf{v}_0\| \left(\frac{-\|\mathbf{v}_0\| \sin\theta - \sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}}{a} \right) \cos\theta. \tag{4}$$

We next investigate the effect of the initial angle θ on the final horizontal displacement x_F. Formally, we must compute the partial derivative of $x_F(\theta, a)$ with respect to θ. For this, we require the partial derivative of the final time $t_F(\theta, a)$ with respect to θ, which is given by

$$\frac{d t_F(\theta, a)}{d\theta} = \frac{-\|\mathbf{v}_0\|}{a} \left(\frac{\|\mathbf{v}_0\| \cos\theta \sin\theta}{\sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}} + \cos\theta \right). \tag{5}$$

Hence, the partial derivative of the final horizontal displacement with respect to the initial angle is

$$\frac{d x_F(\theta, a)}{d\theta} = \|\mathbf{v}_0\| \left(\frac{d t_F(\theta, a)}{d\theta} \cos\theta - t_F(\theta, a) \sin\theta \right) \tag{6}$$

$$= \frac{-\|\mathbf{v}_0\|^2}{a} \left(\frac{\|\mathbf{v}_0\| \cos\theta \sin\theta}{\sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}} + \cos\theta \right) \cos\theta \tag{7}$$

$$+ \frac{\|\mathbf{v}_0\|^2 \sin\theta + \|\mathbf{v}_0\| \sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}}{a} \sin\theta. \tag{8}$$

The initial angle θ that maximises $d x_F(\theta, a)$ must satisfy $\frac{d x_F(\theta, a)}{d\theta} = 0$, which is equivalent to

$$\frac{-\|\mathbf{v}_0\|^2 \cos^2\theta \sin\theta}{\sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0}} - \|\mathbf{v}_0\| \cos^2\theta = -\|\mathbf{v}_0\| \sin^2\theta - \sqrt{\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0} \sin\theta.$$

The above implies (after simplifying the expressions and squaring both sides)

$$\left(\|\mathbf{v}_0\|^2 (2 \sin^2\theta - 1) - 2 a y_0 \right)^2 \sin^2\theta = \|\mathbf{v}_0\|^2 \left(1 - 2 \sin^2\theta\right)^2 \left(\|\mathbf{v}_0\|^2 \sin^2\theta - 2 a y_0 \right). \tag{9}$$

Finally, a change of variable $x := x(\theta) := \sin^2\theta$, turns (9) into the cubic equation

$$\left(\|\mathbf{v}_0\|^2 (2x - 1) - 2 a y_0 \right)^2 x = \|\mathbf{v}_0\|^2 \left(1 - 2x\right)^2 \left(\|\mathbf{v}_0\|^2 x - 2 a y_0 \right) \tag{10}$$

[5]Note that since we consider the parabolic trajectory in forward time, we need to choose the solution that has the larger x-coordinate. Since the vertical acceleration a has negative sign, we choose the negative sign in the nominator.

in which the second and third order terms cancel out, giving the solution $\sin^2\theta = x(\theta) = \|\mathbf{v}_0\|^2/2(\|\mathbf{v}_0\|^2 - ay_0)$. The partial derivative $\frac{dx_F(\theta,a)}{d\theta}$ thus vanishes at the initial angle

$$\theta = \arcsin\left(\frac{\|\mathbf{v}_0\|}{\sqrt{2(\|\mathbf{v}_0\|^2 - ay_0)}}\right). \tag{11}$$

Note that if one assumes zero initial altitude for the ball ($y_0 = 0$m), this expression simplifies to $\theta = \arcsin 1/\sqrt{2} = \pi/4 = 45^\circ$. In our calculations, we assume $y_0 = 1$m. By inserting into (11) the initial speed $\|\mathbf{v}_0\|= 9.524$m/s (the speed required for the hypergravity participants to reach the target at optimal angle) and initial altitude $y_0 = 1$m for the two gravity conditions $a = g$ and $a = 5g$, we obtain the initial angles $\theta_g = 42.200$ and $\theta_{5g} = 34.727$ which maximize the throw distance for normal gravity and hypergravity, respectively. The corresponding maximal throw distances are $x_F(\theta_g, g) = 10.197m$ and $x_F(\theta_{5g}, 5g) = 2.668m$.

We confirmed the above findings by numerically sampling the lengths of trajectories for different initial angles in the two gravity conditions. Figure 6 (A) shows the dependence of the optimal initial angle on the gravity coefficient, and Figure 6 (B) illustrates the final horizontal displacement (throw distance) of the ball as a function of the initial angle, under normal gravity conditions. Both figures assume that the initial height was 1m and the initial speed was 9.524m/s.

3.1.3 Experimental process. In Study 2, 42 participants (14 identified as females, 27 identified as males, and 1 preferred not to say, with numbers of males and females balanced in each group) took part in the experiment. Four other participants were discarded and replaced; two due to software glitches during the experimental process, one due to an outlier check (as defined in our preregistration), and one for not providing consent for data use. The experiment followed exactly the procedure of Study 1 reported earlier, except that we fixed the distance of the normal gravity target as described above. Depending on the group, participants either threw the ball at a target close by (2.67 m) under 5g (hypergravity group) or at a target further away (10.2 m) under 1g (normal gravity group), 20 times. In addition, we changed the wording of confidence questions 1 and 2 in order to clarify them and the questionnaire now contained an illustration, describing the phases of the experiment from A-D where 'C' denoted the adaptation phase. The participants were compensated with University merchandise worth approximately 10€.

Confidence item 1 was described as follows:
"How confident were you in your throwing accuracy of 'phase C' DURING 'phase C' while the ball was still hidden?
1 = I had no idea where my throws landed
7 = My throws landed where I intended them to land"

Confidence item 2 was described as follows:
"How confident were you in your throwing accuracy of 'phase C' AFTER 'phase C' was complete and you had moved on to 'phase D' where you could again see where your throws were landing?
1 = I had no idea where my throws landed
7 = My throws landed where I intended them to land"

3.2 Study 2 results

Box and whisker plots for each phase of Study 2 can be seen in Figures 7, with individual participant means represented as dots. In the signed charts (Figure 7B), positive values indicate overshooting the target, while negative values indicate undershooting. Errors in the adaptation phase for the normal gravity group were again

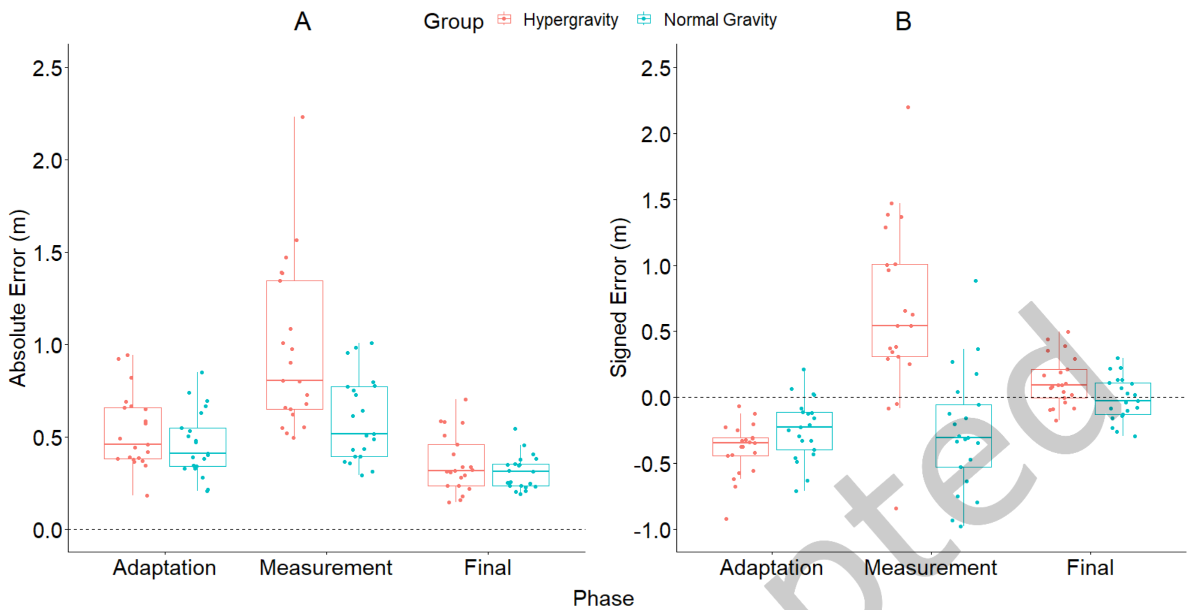

Fig. 7. Boxplots depicting the absolute error (A) and signed error (B) for hypergravity (coral; left pairs) and normal gravity (teal; right pairs) groups, for adaptation, measurement, and final phases of Study 2.

scaled to the target distance of the normal gravity group, where now the errors in the adaptation phase for the normal gravity group were divided by the ratio of the updated target distance in the normal gravity group (10.20 m) relative to the hypergravity group (2.67 m), resulting in a scaling factor of 3.82. All normal gravity adaptation phase errors were divided by this factor before plotting. No statistical comparisons were performed for the adaptation phase, with the exception of the modeling of throwing errors in the exploratory results.

3.2.1 Study 2 confirmatory results. We again assessed the normality of measurement phase distributions via graphical inspection and Shapiro-Wilk tests. Q-Q plots revealed deviations from normality for all distributions, though only the absolute error distribution from the hypergravity group formally failed the Shapiro-Wilk test ($W = 0.88$, $p = .01$; all other $ps > .05$). Nonparametric Wilcoxon rank-sum tests were again used to compare groups (one-sided, $\alpha = .05$; see preregistration). The difference in absolute errors between hypergravity ($Mdn = 0.81$) and normal gravity ($Mdn = 0.52$) groups in the measurement phase (hypothesis H1) was significant, $Z = 3.48$, $p = 4.98 \times 10^{-4}$, $r = .54$. The difference in signed errors between hypergravity ($Mdn = 0.54$) and normal gravity ($Mdn = -0.31$) in the measurement phase (hypothesis H2) was also significant, $Z = 4.91$, $p = 9.13 \times 10^{-7}$, $r = .76$.

After more precisely equating conditions between the two groups, results from Study 2 give further support to our initial hypotheses; those in the hypergravity group not only threw less accurately, but also systematically overshot their targets relative to those in the normal gravity group.

3.2.2 Study 2 exploratory results. The absolute errors in the final phases did not statistically differ between the hypergravity ($Mdn = 0.32$) and normal gravity ($Mdn = 0.31$) groups, $Z = 0.60$, $p = .55$, $r = .09$, again indicating that the hypergravity group was able to recalibrate their internal gravity model after receiving visual feedback and that there were no differences in general throwing ability between the two groups. The sex ratios of the groups were again matched by design, and Wilcoxon-ranked sum tests indicated that the groups also did not happen to differ in age, video game experience, or VR experience (all $ps > .05$). In terms of rating the realism of

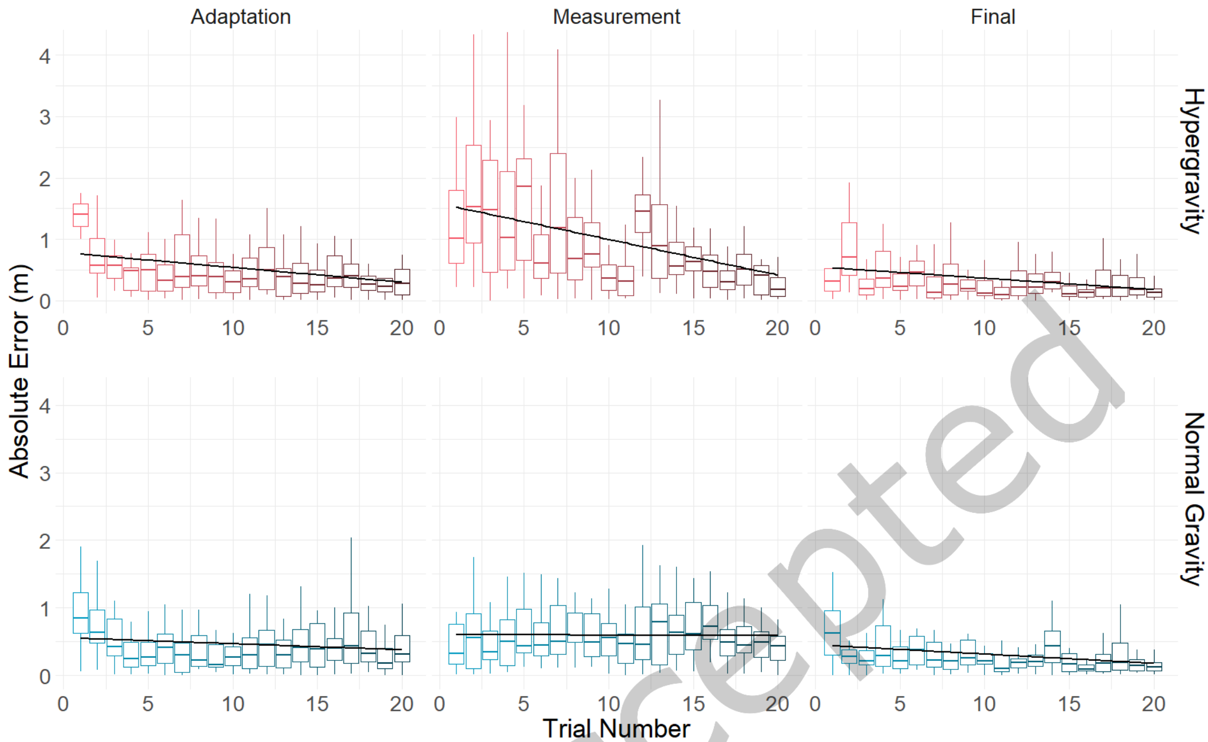

Fig. 8. Boxplots depicting absolute errors across trials (light to dark), phases (left, middle, and center columns), and groups (coral/top row, teal/bottom row). Linear trends are plotted as black solid lines over each cell. Outliers have been omitted from visualizations in order to better highlight central tendencies.

the throwing model, there were again no differences between the hypergravity (Mdn = 5) and normal gravity (Mdn = 6) groups, $Z = 0.61$, $p = .54$, $r = .09$, suggesting that both groups "bought in" to the experience to a similar degree in terms of the perceived realism of the throwing model.

After reformulating our questions concerning confidence before and after throws to remove any possible ambiguities, the same trends remained. We again observed a large increase across all participants from confidence item 1 (Mdn = 3) to confidence item 2 (Mdn = 5), $Z = 4.54$, $p = 5.72 \times 10^{-6}$, $r = .70$. Absolute errors are broken down into individual throws within phases for each group in Figure 8, and for signed errors in 9. Taking a closer look at the trends, participants in the normal gravity group performed decently well in the measurement phase, and although participants in the hypergravity group began with poor accuracy, they trended towards better performance as trials went on until they achieved accuracy levels near their performance level in the final phase. Perhaps then participants correctly understood the confidence items in both Study 1 and 2, and participants in the hypergravity group were merely overweighting their experience towards the end of the measurement phase. Unlike Study 1, confidence before trajectories were revealed was similar for the normal gravity group (Mdn = 3) and the hypergravity group (Mdn = 2), $Z = 0.76$, $p = .45$, $r = .12$, and as well similar after the reveal for the normal gravity group (Mdn = 6) and hypergravity group (Mdn = 5), $Z = 1.41$, $p = .16$, $r = .22$.

Given the trends in Figure 8, we were interested in exploring the effects at the phase and trial levels in finer granularity. The distribution of absolute errors was consistent with a gamma distribution, and so we implemented

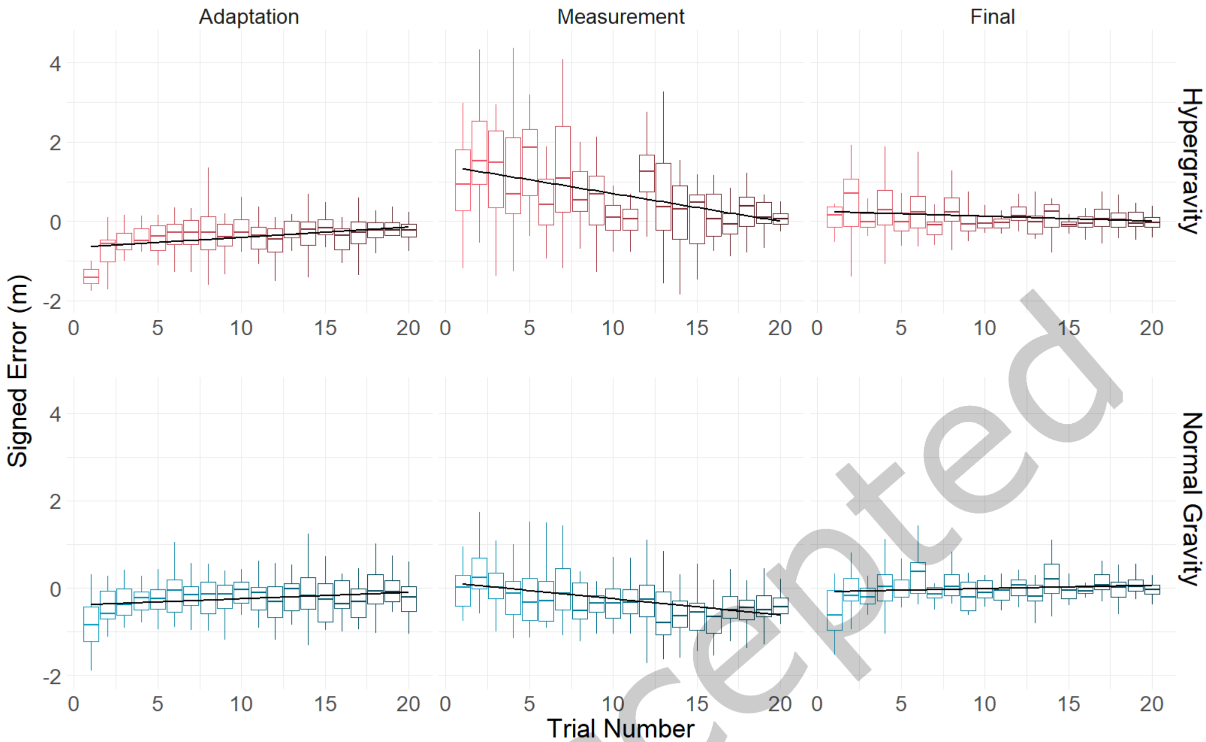

Fig. 9. Boxplots depicting signed errors across trials (light to dark), phases (left, middle, and center columns), and groups (coral/top row, teal/bottom row). Linear trends are plotted as black solid lines over each cell. Outliers have been omitted from visualizations in order to better highlight central tendencies.

a hierarchical gamma model with a log linking function using the "lme4" package in R [3], wherein participants were treated as a clustering variable (level-two) fitted with a random intercept, and group (level-two categorical; levels: hypergravity or normal gravity), phase (level-two categorical; levels: adaptation, measurement, and final), and trial number (level-one integer; numbered 1 through 20 for trials within each phase) were treated as fixed predictors. Some participants did not answer all questionnaire items and therefore needed to be culled from the dataset in order to statistically compare nested models, resulting in the removal of 5 participants in the hypergravity and 3 participants in normal gravity groups. This resulted in a total of 2040 trial outcomes clustered among 34 participants. The trial-level absolute error distribution has the advantage over the (also non-normal) signed error distribution in that the performance dimension only goes in one direction, making the interpretation of model parameter estimates more straightforward (for categorical predictors, a negative β weight corresponds to *smaller* throw error for that level of the variable relative to others, and vice versa; for continuous predictors, a negative β weight indicates that *increases* along the dimension of that variable correspond to *smaller* throw errors, and vice versa.)

We tested the full model family space of group, phase, and trial number predictors and their interactions from the null model all the way to the fully saturated model, and found the fully saturated model (containing the three-way interaction, all two-way interactions, and all main effects) to give the best Akaike information criterion (AIC) score. This outcome is perhaps not too surprising looking at Figure 8; the hypergravity group clearly performs worse, but only during the measurement phase, and in a way such that their errors decreased across trials. This trend suggests that although we were able to alter inner gravity models of the hypergravity group, the change was transitory and their previously held inner gravity model began to reclaim dominance in the absence of visual feedback that would continue to reinforce the new gravity model.

We wanted to explore whether any factors over and above those previously described could have contributed to throwing accuracy. We further compared nested models using chi-squared tests of log-likelihood, iteratively adding new single predictors as main effects to the fully saturated model. None of sex, age, VR experience, videogame experience, confidence before or after the measurement phase, or assessment of realism in the throwing model improved the model fit (all $ps > .05$). With no meaningful contributions from questionnaire items left to consider, we added back the previously removed participants (formerly excluded due to missing questionnaire data) and found that the fully saturated model still gave the best AIC score among the family of models previously tested. Estimates and parameters for the final model containing the full dataset are given in Table 1. Experiment phase and changes across trials accounted for significant variance, yet removing the group interactions and main effect from this model still significantly reduced the model fit, $\chi^2(6, N = 29) = 65.72, p = 3.07 \times 10^{-12}$, further emphasizing the large impact of the gravity manipulation.

4 DISCUSSION

As both initial hypotheses were supported by our results in the original study (Study 1) as well as in the replication study (Study 2), this indicates that adaptation to higher gravity did occur among participants in this experiment. Not only were participants adapting to the throwing task in hypergravity more likely to have worse overall accuracy, but they also appeared to overshoot their throws more often compared to the normal gravity group. This pattern supports the idea that participants' internal model of gravity had temporarily changed during the adaptation phase. We argue that there is evidence for this interpretation specifically since we tested the adaptation *after* exposure to hypergravity so that the participants could not simply use other strategies to compensate for the loss of performance (similarly as in studies by Zago et al. [41, 42], for example). This makes our findings somewhat similar to those made in Prism adaptation studies ([32]), as well as the findings reported by Gaveau et al. [9] in which cosmonauts' internal gravity model was investigated and found to be altered by using a pronounced motor task upon return to normal gravity after exposure to earth-discrepant gravity. In our case, the participants did not appear to compensate after the return to normal gravity, but instead, their performance was congruent with the assumption that simulated gravity remained higher than normal, even though the participants were explicitly informed otherwise before beginning the measurement phase.

According to our exploratory analysis, it appears that returning visual feedback in the final throwing phase recalibrated the internal gravity model back to normal. Furthermore, when inspecting throw accuracies at the granularity of individual trials, it appears that the participants gained accuracy toward the end of the measurement phase, indicating that calibration back to normal gravity began to take place even before the visual feedback was reintroduced. Nevertheless, our findings contradict the argument of Jörges and López-Moliner [13] that adaptation to earth-discrepant gravity would be practically impossible to achieve in VR. However, it should be noted that their survey mostly contained VR studies utilizing interception tasks and not pronounced motor tasks, such as in our study. Moreover, the quantitative change between gravity conditions (1g-5g) in our experiment was rather large compared to previous studies. In addition, Jörges and López-Moliner [13] also argued that adaptation is quicker and easier in pronounced motor tasks, compared to catching and interception tasks. It is very much possible that if we had utilized these tasks in our study, no adaptation would have been observed. On the other hand, our findings are in line with the results of Ye et al. [40] as well as Gravano et al. [10] which suggest that adaptation to tasks taking place in earth-discrepant gravity can be achieved even when training takes place in

Table 1. Model parameters and predictor estimates with 95% confidence intervals for the fully saturated, full dataset model.

	Trial-Level Absolute Throwing Error		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.62	-0.82 – -0.42	**7.802e-10**
group [Hypergravity]	0.39	0.11 – 0.67	**5.931e-03**
phase [Measurement]	0.06	-0.19 – 0.31	6.430e-01
phase [Final]	-0.12	-0.37 – 0.13	3.401e-01
trial n	-0.02	-0.03 – -0.00	**1.655e-02**
group [Hypergravity] × phase [Measurement]	0.75	0.39 – 1.11	**4.394e-05**
group [Hypergravity] × phase [Final]	-0.25	-0.60 – 0.10	1.629e-01
group [Hypergravity] × trial n	-0.03	-0.05 – -0.01	**1.032e-02**
phase [Measurement] × trial n	0.02	-0.00 – 0.04	9.397e-02
phase [Final] × trial n	-0.03	-0.05 – -0.01	**6.035e-03**
(group [Hypergravity] × phase [Measurement]) × trial n	-0.04	-0.07 – -0.01	**6.835e-03**
(group [Hypergravity] × phase [Final]) × trial n	0.02	-0.01 – 0.05	1.166e-01
Random Effects			
σ^2	0.81		
$\tau_{00\ sub}$	0.04		
ICC	0.05		
N_{sub}	42		
Observations	2520		
Marginal R^2 / Conditional R^2	0.187 / 0.228		
AIC	1328.072		

normal gravity. Our findings give further support to the idea that in the future, it might be possible to use VR for gravity adaptation training, at least to some extent.

In our exploratory analysis, we also investigated the contribution of demographics (gender, age, video game experience, and VR experience) and subjective confidence to throwing accuracy. We did not, however, find that these factors would have been able to predict throwing accuracy in both studies. Overall, the gravity in which the adaptation phase took place was the best predictor for throwing accuracy in both studies, giving further support for our hypotheses. Additionally, our exploratory analysis regarding Study 2 did find that progression across trials in the adaptation phase could predict throwing performance for the hypergravity group. This indicates that participants in the hypergravity group gradually regained their throwing accuracy throughout the phase.

4.1 Limitations

Study 1 had substantial limitations that undermined the reliability of its results. Firstly, the throwing distance for the normal gravity group was exaggerated, indicating that participants in the normal gravity group would have had to exert a slightly bigger amount of force compared to the hypergravity group. Secondly, the bug that was related to throwing error computation subtracted 25 cm from the reported errors. The bug therefore artificially increased the reported accuracy of both groups regarding throws that landed outside of the target radius. The results that were less than 25 cm, however, became even more unreliable since it is impossible to know whether these throws had originally landed outside of the target disc radius, or inside the radius, but to the other side of the center. Although we considered the removal of the results of Study 1 altogether, we ultimately decided the paper would be more informative if we retained the result. The reason for this is partly that the mismatch in throwing force could have also introduced bias *against* our hypotheses, not in favor of them: If the normal gravity participants in Study 1 had to use more force in the adaptation phase compared to the hypergravity participants, we assume this would bring the groups' measurement phase results closer together, not further apart. Furthermore, the computation error in Study 1 affected both groups equally. The most significant evidence in favor of our results, however, is the fact that Study 2 was able to replicate the effects of the first study even after fixing both the 1g throw distance and the throw error computation.

A limitation of both studies regarding the generalizability of their results is the fact that the studies were carried out in VR instead of actual hypergravity. Therefore, bodily cues that humans experience during actual abnormal gravity were missing. Furthermore, although we took steps to provide a coherent throwing experience, VR throwing is not the same as real throwing. Our participants did give an acceptable rating of 5 out of 7 for throwing realism in both studies (6 by normal gravity participants in Study 2), however, we cannot expect this simulation to be perfect. The purpose of the custom throwing model was to get throwing in VR to feel somewhat more natural and predictable compared to the Unity default throwing implementation. However, the error between real throwing and VR throwing is difficult to quantify, as many factors that affect throw length are not equated between simulated throwing using a VR controller and actually throwing a physical object. We also did not perform a formal comparison in which participants would have judged our method against the more common method of acquiring projectile velocity directly from a parent object at the time of release. Fortunately, however, the participants did not appear to consider our throwing model to be entirely unrealistic either. Other throwing models could be investigated in future replication studies, perhaps by even calibrating the throwing model for each participant individually. [6]

In both studies, there were a few individual participants who reported that they somehow kept throwing the ball slightly to the left from what they intended during the practice phase. We were not, however, able to reproduce any kind of bug, or glitch that would unintentionally offset the ball's trajectory when throwing. Though it was not formally quantified, we observed large differences in how long it took for participants to finish the initial practice stage which took place before the adaptation phase. This indicates individual differences in either physical throwing skill, or how comfortable participants were with our particular implementation of throwing in VR. We did not observe any statistical difference in accuracy during the final phases of either study. Nevertheless, having a similar task already before adaptation and measurement would have given more evidence regarding equal throwing skills among participant groups. We, however, preferred that neither participant group had experienced the same exact throwing task before the measurement phase. We also acknowledge that participants within the normal gravity group had slightly more "practice" time in throwing due to the practice phase taking place in normal gravity. However, since we found the same pattern of results consisting of a large difference in the measurement phase and no difference in the final phase in both Studies 1 and 2 with different participants, we

[6] An up-to-date link to the application code is maintained at the project repository stored at https://osf.io/6f9ak/

consider it very unlikely that our results could be explained by sampling bias related to differences in participants' throwing abilities.

In both studies, there were some participants who strayed slightly from the experiment procedure in some way or experienced technical or other difficulties. If these issues took place either within the practice phase or adaptation phase, we resolved the issues and carried on with the experimental procedure. However, if these issues took place in the measurement phase, we resolved the issues but did not use data from that participant (we considered these cases as a breach of completing the study protocol as intended). Deviations from protocol included moving outside of the assigned area or failing to stick to the underhand throwing technique. In Study 2, there was one participant who appeared to keep throwing at the same strength after moving from hypergravity adaptation to the measurement phase. This participant's data was most likely removed by our outlier threshold defined in the preregistration. In Study 2, there were two participants who experienced performance glitches; the data from these participants were removed since we had a reason to suspect that these glitches affected throwing performance. After data removal, we collected data from new participants until we reached targets of 60 and 42 participants in Studies 1 and 2, respectively.

A phenomenon that could potentially affect throwing accuracy, especially in the adaptation phase, is the so-called distance compression effect, whereby egocentric distances in VR are generally underestimated [7, 15, 33]. Although newer HMDs appear to be less prone to produce this effect, egocentric distances still appear to be estimated to be only 82% of actual distance [15]. Using a virtual environment that is modeled accurately after a real, familiar place has been shown to mitigate the distance compression effect [11]; therefore we could have possibly placed the experiment, for example, in a familiar campus environment similar to our previous studies to help with distance perception [30]. For this experiment, however, we chose to use a neutral and visually unfamiliar setting to eliminate any visual size cues that could bias participants' perceptions toward either normal gravity or hypergravity (even the floor tiles were sized differently from what one typically sees in real-world architecture). Therefore, the distance compression effect could have potentially affected both of our participant groups in both studies. Replicating Study 2 with different types of environments would be advisable to investigate the potential confound caused by the distance compression effect.

5 CONCLUSION

In this paper, we presented our results concerning adaptation to simulated hypergravity in a VR throwing task. We performed two between-groups experiments ($n = 60$ and $n = 42$). In both studies, two groups of participants were randomly assigned to practice throwing a virtual ball at a target either in hypergravity (5g), or normal gravity (1g). The target distances were placed so that roughly similar throwing strength was required in both groups in order to hit the target. The distances were approximated manually in the first study and mathematically in the second study. We then measured adaptation by having participants of both groups in both stdues throw a virtual ball at a target 20 times under normal gravity so that the virtual ball disappeared shortly after release hence the participants had to rely on their internal model of gravity instead of visual feedback. According to our preregistered hypotheses, we expected participants who practiced under hypergravity to have worse overall accuracy in the measurement phase, as well as more often overshoot the target in comparison to the control group. The results of both studies supported both hypotheses, which indicates that adaptation did occur in the hypergravity groups. Our exploratory analysis further confirmed that gravity manipulation was the most significant predictor of throwing accuracy in both studies. These findings were somewhat surprising since previous literature has argued that adaptation to earth-discrepant gravity would not be possible in VR since bodily gravity cues cannot be simulated [13]. This discrepancy can potentially be attributed to the fact that previous VR studies have mostly implemented catching or interception tasks to investigate adaptation, whereas our study utilized a task with a pronounced motor component, which is known to be less resistant to adaptation.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland projects PIXIE 331822, PERCEPT 322637, SRC of Academy of Finland project COMBAT 293389, Business Finland project HUMOR 3656/31/2019, and the European Research Council project ILLUSIVE 101020977. We wish to thank all our participants.

REFERENCES

[1] Haley Adams, Gayathri Narasimham, John Rieser, Sarah Creem-Regehr, Jeanine Stefanucci, and Bobby Bodenheimer. 2018. Locomotive recalibration and prism adaptation of children and teens in immersive virtual environments. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1408–1417.

[2] Hirofumi Aoki, Takao Yamaguchi, and Ryuzo Ohno. 2001. A study of orientation in a zero gravity environment by means of virtual reality simulation. *AIP Conference Proceedings* 552, 1 (02 2001), 29–34. https://doi.org/10.1063/1.1357901 arXiv:https://pubs.aip.org/aip/acp/article-pdf/552/1/29/11526509/29_1_online.pdf

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[4] Larissa Brübach, Franziska Westermeier, Carolin Wienrich, and Marc Erich Latoschik. 2022. Breaking plausibility without breaking presence-evidence for the multi-layer nature of plausibility. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2267–2276.

[5] Miquel Bosch Bruguera, Valentin Ilk, Simon Ruber, and Reinhold Ewald. 2019. Use of Virtual Reality for astronaut training in future space missions-Spacecraft piloting for the Lunar Orbital Platform-Gateway (LOP-G). In *70th International Astronautical Congress*.

[6] Desiderio Cano Porras, Gabriel Zeilig, Glen M Doniger, Yotam Bahat, Rivka Inzelberg, and Meir Plotnik. 2020. Seeing gravity: gait adaptations to visual and physical inclines–a virtual reality study. *Frontiers in neuroscience* (2020), 1308.

[7] Sarah H Creem-Regehr, Jeanine K Stefanucci, and Bobby Bodenheimer. 2023. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B* 378, 1869 (2023), 20210456.

[8] Frédéric Crevecoeur, Jean-Louis Thonnard, and Philippe Lefevre. 2009. Optimal integration of gravity in trajectory planning of vertical pointing movements. *Journal of neurophysiology* 102, 2 (2009), 786–796.

[9] Jérémie Gaveau, Christos Paizis, Bastien Berret, Thierry Pozzo, and Charalambos Papaxanthis. 2011. Sensorimotor adaptation of point-to-point arm movements after spaceflight: the role of internal representation of gravity force in trajectory planning. *Journal of neurophysiology* 106, 2 (2011), 620–629.

[10] Silvio Gravano, Francesco Lacquaniti, and Myrka Zago. 2021. Mental imagery of object motion in weightlessness. *npj Microgravity* 7, 1 (2021), 50.

[11] Victoria Interrante, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *IEEE virtual reality conference (VR 2006)*. IEEE, 3–10.

[12] Ao Jiang, Yang Gong, Xiang Yao, Bernard Foing, Richard Allen, Stephen Westland, Caroline Hemingray, and Yingen Zhu. 2023. Short-term virtual reality simulation of the effects of space station colour and microgravity and lunar gravity on cognitive task performance and emotion. *Building and Environment* 227 (2023), 109789.

[13] Björn Jörges and Joan López-Moliner. 2017. Gravity as a strong prior: implications for perception and action. *Frontiers in Human Neuroscience* 11 (2017), 203.

[14] Peter D Jost. 2008. Simulating human space physiology with bed rest. *Hippokratia* 12, Suppl 1 (2008), 37.

[15] Jonathan W Kelly. 2022. Distance perception in virtual reality: A meta-analysis of the effect of head-mounted display characteristics. *IEEE transactions on visualization and computer graphics* (2022), 1–13.

[16] Regis Kopper, Tao Ni, Doug A Bowman, and Marcio Pinho. 2006. Design and evaluation of navigation techniques for multiscale virtual environments. In *IEEE Virtual Reality Conference (VR 2006)*. IEEE, 175–182.

[17] Maria Kozhevnikov and Mary Hegarty. 2001. Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review* 8, 3 (2001), 439–453.

[18] Andrey Krekhov, Sebastian Cmentowski, Katharina Emmerich, Maic Masuch, and Jens Krüger. 2018. GulliVR: A walking-oriented technique for navigation in virtual reality games based on virtual body resizing. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 243–256.

[19] Barbara La Scaleia, Francesca Ceccarelli, Francesco Lacquaniti, and Myrka Zago. 2020. Visuomotor interactions and perceptual judgments in virtual reality simulating different levels of gravity. *Frontiers in Bioengineering and Biotechnology* 8 (2020), 76.

[20] Francesco Lacquaniti and Claudio Maioli. 1989. Adaptation to suppression of visual information during catching. *Journal of Neuroscience* 9, 1 (1989), 149–159.

[21] Eike Langbehn, Gerd Bruder, and Frank Steinicke. 2016. Scale matters! Analysis of dominant scale estimation in the presence of conflicting cues in multi-scale collaborative virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 211–220.

[22] Andrey V Maltsev, Evgeny V Strashnov, and Mikhail V Mikhaylyuk. 2021. Methods and technologies of cosmonaut rescue simulation in virtual environment systems. *Scientific Visualization* 13, 4 (2021), 52–65.

[23] Joseph McIntyre, Myrka Zago, Alain Berthoz, and Francesco Lacquaniti. 2001. Does the brain model Newton's laws? *Nature neuroscience* 4, 7 (2001), 693–694.

[24] Daniel M Merfeld, Lionel Zupan, and Robert J Peterka. 1999. Humans use internal models to estimate gravity and linear acceleration. *Nature* 398, 6728 (1999), 615–618.

[25] Guillaume Millet, Anatole Lécuyer, Jean-Marie Burkhardt, D Sinan Haliyo, and Stéphane Régnier. 2008. Improving perception and understanding of nanoscale phenomena using haptics and visual analogy. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 847–856.

[26] Kevin Montgomery, Cynthia Bruyns, and Simon Wildermuth. 2001. A virtual environment for simulated rat dissection: a case study of visualization for astronaut training. In *Proceedings Visualization, 2001. VIS'01.* IEEE, 509–601.

[27] Lars IE Oddsson, Robin Karlsson, Janusz Konrad, Serdar Ince, Steve R Williams, and Erika Zemkova. 2007. A rehabilitation tool for functional balance using altered gravity and virtual reality. *Journal of neuroengineering and rehabilitation* 4, 1 (2007), 1–7.

[28] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[29] Matti Pouke, Evan G. Center, Alexis P. Chambers, Sakaria Pouke, Timo Ojala, and Steven M. Lavalle. 2022. The body scaling effect and its impact on physics plausibility. *Frontiers in Virtual Reality* 3 (2022).

[30] Matti Pouke, Katherine J Mimnaugh, Alexis P Chambers, Timo Ojala, and Steven M LaValle. 2021. The plausibility paradox for resized users in virtual environments. *Frontiers in Virtual Reality* 2 (2021), 48.

[31] Matti Pouke, Katherine J Mimnaugh, Timo Ojala, and Steven M LaValle. 2020. The plausibility paradox for scaled-down users in virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 913–921.

[32] Gordon M Redding, Yves Rossetti, and Benjamin Wallace. 2005. Applications of prism adaptation: a tutorial in theory and method. *Neuroscience & Biobehavioral Reviews* 29, 3 (2005), 431–444.

[33] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. 2013. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 1–40.

[34] Jukka Rönkkö, Jussi Markkanen, Raimo Launonen, Marinella Ferrino, Enrico Gaia, Valter Basso, Harshada Patel, Mirabelle D'Cruz, and Seppo Laukkanen. 2006. Multimodal astronaut virtual training prototype. *International Journal of Human-Computer Studies* 64, 3 (2006), 182–191.

[35] Patrice Senot, Myrka Zago, Francesco Lacquaniti, and Joseph McIntyre. 2005. Anticipating the effects of gravity when intercepting moving objects: differentiating up and down based on nonvisual cues. *Journal of Neurophysiology* 94, 6 (2005), 4471–4480.

[36] Richard Skarbez, Missie Smith, and Mary C Whitton. 2021. Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality* 2 (2021).

[37] Mel Slater. 2014. Grand challenges in virtual environments. *Frontiers in Robotics and AI* 1 (2014).

[38] Tomer D Ullman, Elizabeth Spelke, Peter Battaglia, and Joshua B Tenenbaum. 2017. Mind games: Game engines as an architecture for intuitive physics. *Trends in cognitive sciences* 21, 9 (2017), 649–665.

[39] Björn Van Der Hoort, Arvid Guterstam, and H Henrik Ehrsson. 2011. Being Barbie: the size of one's own body determines the perceived size of the world. *PloS one* 6, 5 (2011), e20195.

[40] Tian Ye, Siyuan Qi, James Kubricht, Yixin Zhu, Hongjing Lu, and Song-Chun Zhu. 2017. The martian: Examining human physical judgments across virtual gravity fields. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1399–1408.

[41] Myrka Zago, Gianfranco Bosco, Vincenzo Maffei, Marco Iosa, Yuri P Ivanenko, and Francesco Lacquaniti. 2004. Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions. *Journal of neurophysiology* 91, 4 (2004), 1620–1634.

[42] Myrka Zago and Francesco Lacquaniti. 2005. Internal model of gravity for hand interception: parametric adaptation to zero-gravity visual targets on Earth. *Journal of neurophysiology* 94, 2 (2005), 1346–1357.

[43] Jingjing Zhang, Thammathip Piumsomboon, Ze Dong, Xiaoliang Bai, Simon Hoermann, and Rob Lindeman. 2020. Exploring spatial scale perception in immersive virtual reality for risk assessment in interior design. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[44] Xiaolong Zhang and George W Furnas. 2005. mCVEs: Using cross-scale collaboration to support user interaction with multiscale structures. *Presence* 14, 1 (2005), 31–46.