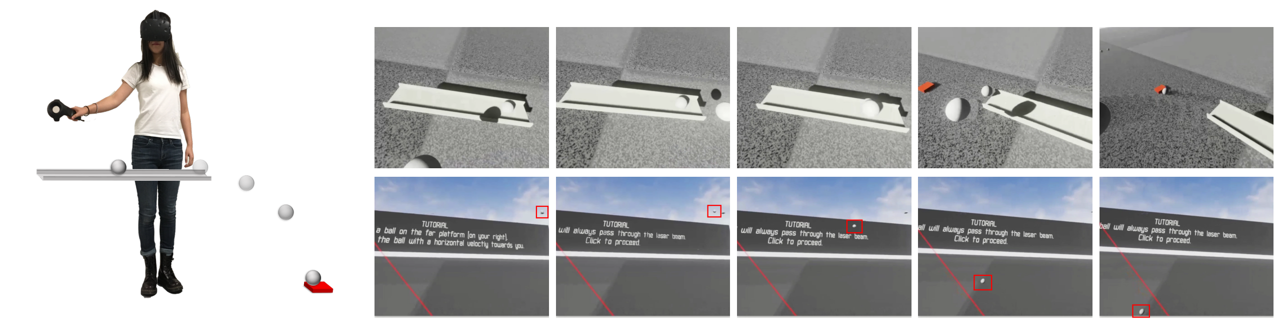
The Martian: Examining Human Physical Judgments Across Virtual Gravity Fields

Tian Ye*, Siyuan Qi*, James Kubricht, Yixin Zhu, Hongjing Lu, and Song-Chun Zhu *Fellow, IEEE*



(a) Illustration of experiment designs (b) Two examples of the experiments: (upper) speed production and (lower) trajectory prediction.

Abstract—This paper examines how humans adapt to novel physical situations with unknown gravitational acceleration in immersive virtual environments. We designed four virtual reality experiments with different tasks for participants to complete: strike a ball to hit a target, trigger a ball to hit a target, predict the landing location of a projectile, and estimate the flight duration of a projectile. The first two experiments compared human behavior in the virtual environment with real-world performance reported in the literature. The last two experiments aimed to test the human ability to adapt to novel gravity fields by measuring their performance in trajectory prediction and time estimation tasks. The experiment results show that: 1) based on brief observation of a projectile's initial trajectory, humans are accurate at predicting the landing location even under novel gravity fields, and 2) humans' time estimation in a familiar earth environment fluctuates around the ground truth flight duration, although the time estimation in unknown gravity fields indicates a bias toward earth's gravity.

Index Terms—Virtual reality, intuitive physics, mental simulation

1 Introduction

Sending a manned spacecraft to Mars would be a fantastic adventure, yet living on another planet could lead to significant challenges to human perceptual and cognitive systems. The change in gravity itself could alter daily activities (e.g., throwing an object toward a desired location or pouring water into a container) that require adjustment of prediction and action in light of changing physical properties on the new planet. Imagine that you are living in an environment with a different gravity field than earth. Would you be able to adapt to it quickly? And how accurate would your predictions about the physical world be compared to when you were on earth?

Consumer-level virtual reality (VR) devices, with rapidly increasing popularity, provide a useful means for researchers to conduct experiments that were traditionally too costly or impossible to carry out in the real world. VR allows users to experience an artificial world in a manner similar to how they experience the real world: i.e., head-mounted displays give the impression of three-dimensional observation, and remote controllers afford interactions with the virtual world from an embodied egocentric perspective. In particular, VR technology allows for both the control of many underlying factors of the virtual world (e.g., time [34] and gravity) and direct measurement of behavioral changes in novel environments.

In this paper, we conducted four experiments to measure human performance in different tasks under novel and familiar gravity fields. In the first two experiments, participants were asked to strike a ball off of a track onto a target location and to trigger a ball to hit a target given a speed rating input. In Experiments 3 and 4, participants were asked to make predictions about the location and flight duration of a projectile given the initial 0.2 seconds of its trajectory. The purpose of the experiments was to examine how humans learn and reason about object motion in novel gravity fields: are humans able to spontaneously habituate to new gravity fields? Do humans implicitly use prior knowledge about earth's gravity to reason about new environments? Are humans implicitly simulating physical motion or predicting the movements using low-level visual features exclusively?

The first pair of experiments in the present work compare human performance in the VR setting with findings in similar real-world situations [21]. The second pair of experiments compare two types of intuitive physical judgments (location predictions and time estimates) under different gravity fields. In summary, this paper made the following contributions: 1) replicated a previous study on speed production and rating in novel virtual environments to demonstrate that VR is a feasible and reliable tool for studying human perception and cognition, 2) carried out a novel experimental design and method which precludes real-world replication, and 3) measured the effect of gravity field on human behavior in tasks varying in their cognitive demands.

The remainder of this paper is structured as follows: Sect. 2 discusses related work in intuitive physics and virtual reality, Sect. 3 describes the experiment methods and results, Sect. 4 provides a comparison of results between experiments, and Sect. 5 concludes the study and outlines proposed directions for subsequent work.

* *Tian Ye and Siyuan Qi contributed equally to this work.*
- *Tian Ye, Siyuan Qi, Yixin Zhu, and Song-Chun Zhu are with UCLA Center for Vision, Cognition, Learning and Autonomy (VCLA). E-mail: etalice@g.ucla.edu, syqi@cs.ucla.edu, yixin.zhu@ucla.edu, sczhu@stat.ucla.edu.*
- *James Kubricht and Hongjing Lu is with UCLA Computational Vision and Learning Lab (CVL). E-mail: kubricht@ucla.edu, hongjing@ucla.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Background and Related Work

2.1 Intuitive Physics

The principles of Newtonian physics accurately describe the behavior of objects in our realizable world, yet people's commonsense beliefs about how objects move are often at odds with ground-truth predictions [8, 18]. For example, when reasoning about the trajectories of moving objects on explicit, pencil-and-paper tasks, people commonly predict that an object will follow a curvilinear path upon exiting a C-shaped tube [28] and that an object dropped from a moving body will follow a linear path downward [27, 29] with a velocity proportional to its weight [37]. Although these misconceptions appear consistent with erroneous physical theories (e.g., medieval impetus or Aristotelian principles), people are not internally consistent in their intuitive physical judgments across related tasks at the explicit level, suggesting adherence to *domain-specific* theories of motion [18, 45]. More recent work proposes that people are susceptible to erroneous theories of motion in the context of *explanation*, but adhere to rational inference on the basis of Newtonian physics in the context of *prediction*: e.g., people are unable to draw the trajectory of an object following release from a pendulum but can successfully place a bucket where they expect the ball to land [40]. Earlier empirical studies also suggest that humans are less susceptible to erroneous theories of motion when physical situations are presented in a familiar context [18] or in an animated format [19, 20]. In addition, although people are inaccurate when reasoning about certain physical situations (e.g., liquid behavior) explicitly, they respond rationally when repeating similar tasks through simulated action [36] or animation-facilitated mental simulation [22].

The present experiments provide two concrete task domains which rely on action and prediction, rather than explanation. Thus, we hypothesized that participants would adhere to Newtonian (rather than erroneous) theories of motion when reasoning about the physical behavior of moving objects at the implicit level. The tasks, however, differ in their cognitive and sensorimotor demands in addition to the spatial information needed to effectively reason about them. For example, Experiments 1 and 2 (replication of previous work; described in Sect. 3) provide participants with a perception-for-action and perception-only task, respectively. Although Krist *et al.*'s original study found no difference between performance in the two tasks [21], there is a breadth of evidence that guided action and perceptual identification rely on two independent neural pathways [14]. More recent fMRI results, however, indicate that object weight (a non-visual, motor-relevant property) can be represented in regions associated with the (ventral) perceptual identification stream from familiar texture cues [11]. Information about the weight of unfamiliar objects, however, is arguably inferred from haptic feedback, a crucial sensory modality absent in virtual environments that has been pursued rigorously over the past three decades [7]. The effect of the absence of haptic feedback on our perception-for-action task is further discussed in the following sections.

The perception-for-action task in Experiment 1 also differs from Krist *et al.*'s original task in that it required the use of a tool (i.e., a controller specifying the position of a virtual ball) to propel a second ball off of a platform. Tool use has been shown to directly affect perceived distance, suggesting that people represent the physical world in terms of their ability to interact with it [44]. Perception-action recalibration has also been reported in studies on predicted walking distance (real and imagined) following adaptation [46], judged hill slant following a loading of weight from a backpack [30] (but see [10]), and illusory reversal of temporal order between actions and sensation [41]. Although the potential effect of tool use on Experiment 1's task is interesting, a more primary aim of the present work is to determine how participants respond and adapt to novel gravitational fields across the four experiments outlined in Sect. 3. Are people biased toward believing that the gravitational acceleration of an unknown environment is equal to that on earth? Does this bias manifest itself across tasks differing in their aforementioned cognitive demands? The potential interplay between prior beliefs about gravity and associated human activity is discussed further in Sect. 2.2.

2.2 Mental Simulation

Previous work has demonstrated that people often employ mental simulation strategies when reasoning about physical situations [16, 17]. For example, expert problem-solvers spontaneously employ mental simulation strategies to anchor assumptions relevant to their explanations when reasoning about a mass-spring system [9] and when people are asked to predict the rotation direction of elements in a pulley system, they intuitively simulate motion in an order corresponding to the machine's causal sequence: i.e., more time is required to reason about motion later in the causal chain [15]. The time-dependent characteristics of people's mental simulations suggest spatial (rather than visual) representation, which quantitatively encodes both latent and observable variables relevant to the physical situation [16, 35].

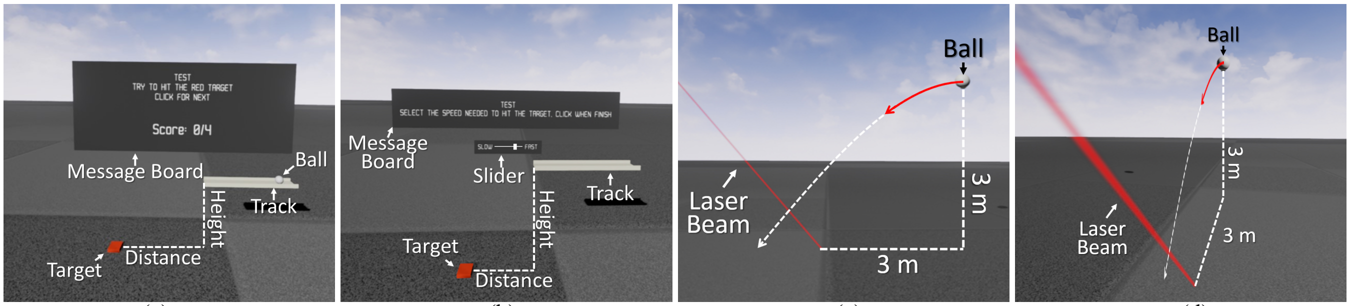
The findings outlined in Sect. 2.1, however, suggest that people's predictions about one-body motion do not always agree with Newtonian physics. Early work on human judgment in two-body interactions (e.g., the collision of two point masses) reported consistent findings [31, 42], although judgment biases were subsequently explained by attention to simplified rules or *heuristics* based on observational cues [13, 32]. Despite the fact that cue-heuristic models explain some human judgment biases qualitatively, several recent models have obtained good quantitative agreement by assuming that people form rational inferences about dynamical systems by combining *noisy* perceptual inputs with Newtonian physical principles, given prior beliefs about spatially represented variables: i.e., the *noisy Newton* hypothesis [33]. Following this hypothesis, quantitative judgments about physical systems can be inferred by simulating initial states (sampled from distributions reflecting noisy perception) forward in time using probabilistic Newtonian physics, querying the output states and aggregating judgments across numerous simulations to form predicted response distributions [5].

Probabilistic simulation models have demonstrated success in predicting human judgments in several domains: physical scene understanding [5], object interactions [23], liquid dynamics [4, 22] and causality in mass-collision displays [12]. Their predictions deviate from ground-truth physics in accordance with biases observed in human experiments. Such findings are generally explained by noisy perceptual inference based on prior beliefs about relevant variables in the physical system. For example, the standard belief that an initially moving object is heavier in a two-body collision is well-explained by a prior belief that objects are more likely to move slow than fast [33]. In the perception-for-action task in Experiment 1, participants must propel a ball to a target location by striking it with their controller. However, they receive no haptic information about the weight of the ball: i.e., the ball *feels* weightless. Thus, we hypothesized that in the absence of haptic feedback, people will underestimate the force needed to propel a ball to a given target location.

In Experiments 3 and 4 of the present work, participants were asked to reason about the trajectory and duration of a projectile moving under different gravity fields (see Sect. 3.3 and Sect. 3.4). One hypothesis is that people predict future physical states by simulating projectile motion forward in time, holding prior beliefs on the underlying physical variables: e.g., velocity and gravitational acceleration. Given robust experience in earth's gravity, we predict that participants' simulations will adhere to a gravitational acceleration biased toward that of earth. Alternatively, people might predict trajectories using more explicit mechanisms based on perceptual identification, such that inferred locations will not be biased toward what would be expected under earth's gravitational field: i.e., gravitational acceleration—and prior beliefs about its magnitude—will not play into the location prediction process.

2.3 Virtual Reality

Virtual reality (VR) technology provides an analog experience in a three-dimensional environment similar to that of the real world. Although the majority of research on VR focuses on the technology itself in order to improve users' experience of hardware and software (e.g., sensing [39, 43], simulation [1–3] and platform integration [24]), recent studies have sought to directly examine human perception and cognition in virtual environments. Previous work has focused on human behavior in situations that are similar between the virtual environment and the real world: e.g., visual perception of egocentric distance in real and virtual environments [25], and human perception of the three components of locomotion (i.e., distance, speed and time) during immersive walkthroughs [6]. Further work has utilized virtual environments to simulate situations that cannot be emulated in the real world, such as the effect of a naturally or unnaturally moving sun on human time judgments [34], the perception and understanding of the exchange of avatars [26], and the visualization of relativity [38].



(a) (b) (c) (d)

Fig. 1: Experiment designs. Notes and measurements overlaid on top of the scenes are for illustration and were not provided to participants. (a) Setting of Experiment 1 in the virtual environment during the testing session: a track at different heights and a red target at different distances were displayed. Participants used a controller to strike the virtual ball on the track onto a red target on the ground. (b) Setting of Experiment 2 during the testing session: similar to Experiment 1, a track and a target were shown to the participants, but no virtual ball was displayed. Participants selected a speed on the slider to indicate at which speed they thought the ball should be projected to hit the target. (c) Setting of Experiments 3 and 4 in the virtual environment: a virtual ball was generated on a platform on the top right side of the display. Once the button on the controller was pressed, the ball was projected horizontally leftward toward the red laser beam. (d) Illustration of Experiments 3 and 4 from the participant's perspective: participants stand next to the laser beam and perceive a virtual ball flying toward them, which is different from inferring the ball speed from a third-person view.

In the present study, two experiments similar to the studies by Krist *et al.* [21] were conducted to compare the behavior of humans in a virtual environment to their corresponding behavior in a real world task (see Sect. 3.1 and Sect. 3.2). Two additional experiments simulated physical situations that preclude real world replication (see Sect. 3.3 and Sect. 3.4).

3 Experiments

Participants and Apparatus A total of 20 participants (8 female and 12 male) participated in the study. Participants were either undergraduate or graduate students at the University of California, Los Angeles. The average age of participants was 22.8 years old with a standard deviation of 2.67. All participants had normal to corrected-to-normal vision.

During the experiments, participants wore an HTC Vive head mounted display (HMD) with two 1080×1200 screens (one per eye), a 90 Hz refresh rate, and a 110° field of view. Participants used a native HTC Vive controller to interact with objects and scenes inside the virtual setting. Responses were automatically tracked by the HTC Vive system and recorded by our programs. Two standard HTC Vive base stations (lighthouses) were mounted on the wall to simultaneously track the pose and dynamics (position, velocity, and orientation) of both the HMD and controllers over time. The virtual environment was designed using the Unreal Engine 4 gaming platform, providing state-of-the-art, physics-based simulation in real time.

Experiment Overview Experiments 1 and 2 were designed to replicate and extend Krist *et al.*'s original study [21] in a virtual environment. The design of our experiments and the previous study by Krist and colleagues was identical except that in the present experiment, participants used an HTC Vive controller to strike a virtual ball (rather than throwing an actual, physical ball) off of a track onto a target location. Participants were free to traverse the virtual environment.

Experiments 3 and 4 differed from typical experiment settings in the literature examining human projectile motion predictions (e.g., in [40]). Rather than presenting pre-recorded videos in 2D displays and collecting responses using a keyboard or mouse, we measured people's performance in an immersive, 3D environment using a laser beam (measurement tool) in VR. The virtual world provided participants with a vivid and realistic environment that enabled several physical interactions between entities and agents. Furthermore, allowing participants to navigate freely inside the virtual environment provided a means to adjust their individual viewing angle so they could view the entire environment.

Experiment Order Participants were asked to complete 3 blocks of experiments in a within-subjects experimental design. In each block, all four experiments shared the same unique gravity field (i.e., gravity field was manipulated between blocks). The gravity field in each environment was selected from 1.5g, 1.0g and 0.5g. Participants were informed that the gravity field for the first block of experiments would be equal to earth's gravity (1.0g). In the subsequent two blocks, half of the participants completed the block under half of earth's gravity field (0.5g) first followed by the block with 1.5 times earth's gravity field (1.5g); the other half of the participants completed the experiments in the counterbalanced order: i.e., 1.5g first and 0.5g second. Participants were told that they would experience unfamiliar gravity fields in the second and third blocks, but information about the specific gravity field was not provided (i.e., whether the gravity field would be greater than or less than earth's gravity).

In each block, Experiments 1 and 2 were conducted prior to Experiments 3 and 4 for all participants. To control for order effects, half of the participants completed Experiment 1 prior to Experiment 2, and the other half completed Experiment 2 first. Similarly, half of the participants completed Experiment 3 prior to Experiment 4, and the other half completed Experiment 4 first. The order stayed the same for all three blacks for the same participant.

3.1 Experiment 1: Direct Action

The first experiment asked human participants to propel a ball toward an indicated target location in a virtual environment under different gravity fields. Participants directly interacted with virtual objects in the VR setting. After providing responses using a Vive controller, participants viewed the full trajectories of the propelled objects. We sought to compare human performance in VR with that in the real world [21] by examining human performance under a gravity field identical to that on earth (i.e., a familiar gravity field). We further examined how well humans perform under gravity fields different from earth's gravity (i.e., unfamiliar gravity fields).

Experiment Setting As illustrated in Fig. 1a, the setting of Experiment 1 consisted of a horizontal track (0.16 meters wide, 0.90 meters long), a red, cubic target (length of each side = 0.12 meters), and a white ball (diameter = 0.08 meters, friction coefficient = 0). In the virtual environment, each controller was represented by a sphere (diameter = 0.10 meters). Participants were instructed to use a Vive controller to hit the white ball on the track onto the red target as indicated in Fig. 6a. The vertical height of the track and the horizontal distance between the track and the target were chosen from a pre-defined discretized set, identical to Krist *et al.*'s previous study [21]. Participants were instructed to hit the ball horizontally so that the ball would exit the track with zero velocity along the vertical axis. They were also informed that only the first collision point of the ball would be counted as a successful hit: i.e., bouncing the ball onto the target or rolling it toward the target would not be counted. The size of the track, the diameter of the ball, and the rendered background environment were held constant across all experiments.

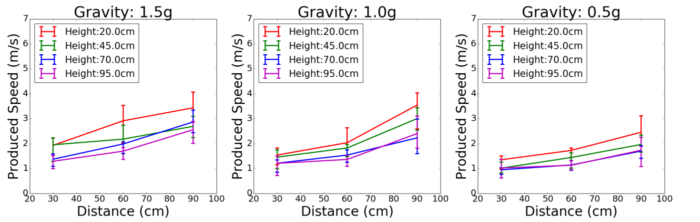


Fig. 2: Mean speeds produced by participants in Experiment 1 under three different gravity fields: 1.5g, 1g, and 0.5g. Error bars indicate standard error of the mean.

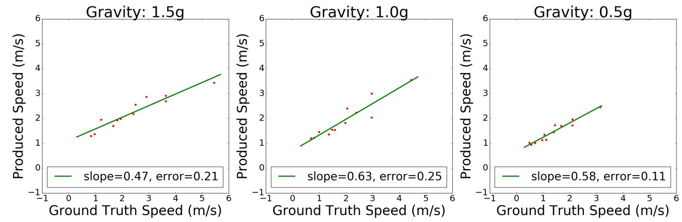


Fig. 3: Mean speeds produced by participants versus ground truth speeds in Experiment 1 under three different gravity fields: 1.5g, 1g, and 0.5g.

Training Session Participants were first given a demonstration and practice trials to familiarize themselves with the virtual environment. Participants were first shown a 3D visual demonstration through the HTC Vive HMD: a ball leaves the track from a height of 1 meter with an initial horizontal speed of 2 meters per second and lands on the ground. Next, they were asked to move the controller to hit the ball on the track, which ensured that participants knew how to use the controller prior to the training and testing trials. Finally, participants were instructed to hit the virtual ball with the same initial configuration three times as far as possible and then three times as close as possible.

After demonstration and practice, participants were given training trials: a target appeared on the floor and the height of the track was adjusted. Participants were instructed to try their best to hit the ball onto the target location. If the ball did not exit the track horizontally in a given training trial, the setting for that trial was presented again at the end of the training session. During the training session, the height of the track was chosen from 0.20, 0.70 and 0.95 meters; the distance between the center of the target and the exit of the track was either 0.30 or 0.90 meters. In total, there were 6 different combinations of distance and height for the training session. During the training session, participants observed the full trajectory of the ball after it left the track (i.e., they received visual feedback in each training trial).

Testing Session After the 6 training trials, participants were presented with 12 testing trials with parameters indicated by the combination of 4 different heights and 3 different distances. Specifically, the height of the track was chosen from 0.20, 0.45, 0.70 and 0.95 meters; the distance between the center of the target and the exit of the track was chosen from 0.30, 0.60, and 0.90 meters. These values were identical to the set of values used in the original study [21]. Similar to the training session, if participants failed to hit the ball so that it exited the track horizontally in a testing trial, that trial was presented again at the end of the testing session. After hitting the ball, participants were provided with full visual feedback: they viewed the full trajectory of the ball and were informed of whether the ball successfully hit the target.

When participants hit the ball using the controller (represented by a virtual sphere; diameter = 0.10 meters), the speed of the controller was measured by the HTC Vive base station and fed into the Unreal Engine. The Unreal Engine then computed the resulting speed of the ball after the collision using an internal physics engine. The speed of the ball was recorded as the produced speed measurement for each testing trial. To examine whether humans behave rationally, a ground-truth model prediction for each trial was calculated analytically based on the physical parameters (i.e., height, distance, and gravity) in each setting.

Results In order to determine whether the speeds produced by participants depended on the magnitude of gravitational acceleration, the height of the track, and the distance between the track and target, we performed a $3 \times 3 \times 4$ (Gravity $\times$ Height $\times$ Distance) analysis of variance (ANOVA) at the $\alpha = 0.05$ significance level. The three-way interaction term was not significant ($F[12,684] = 1.34$, $p = 0.189$). There were significant main effects of gravity ($F[2,684] = 74.05$, $p < .001$), distance ($F[2,684] = 181.79$, $p < .001$), and height ($F[3,684] = 40.31$, $p < .001$). There was also a significant two-way interaction between gravity and distance ($F[4,684] = 4.78$, $p < .001$). However, the interaction between gravity and height ($F[6,684] = 1.10$, $p = .360$) was not significant. Contrary to the ground truth model, the interaction between distance and height was only marginally significant ($F[6,684] = 1.86$, $p = .086$). This result indicates that the relationship between produced speed and distance did not differ significantly across the implemented track heights. Our results demonstrate that participants' produced speeds depended on relevant physical factors: i.e., the height of the track, the distance between the track and target, and importantly gravitational acceleration in the virtual world. In addition, the relationship between produced speed and distance also varied according to specific gravity fields, which agrees with the ground-truth relationships. Fig. 2 depicts the mean speeds produced by participants under three different gravity fields for various height-distance combinations. The three figures qualitatively differ from one another, as evident in the aforementioned Gravity $\times$ Distance interaction.

To analyze the effect of participants randomly selected from the university student population, we performed a random effects ANOVA with two-way interactions. Higher order interaction terms were removed from the analysis due to singular measurements for each height-distance-gravity trial. The analysis showed no significant main effect of participant ($F[19,544] = 1.35$, $p = .245$). There were significant two-way interaction effects between gravity and participant ($F[38,544] = 1.49$, $p = .033$) and between distance and participant ($F[38,544] = 1.54$, $p = .021$), although the two-way interaction between height and participant was not significant ($F[57,544] = 0.74$, $p = .923$). The present results suggest that there are individual differences in perceiving gravity and 3D distance between the target and the track in the VR environment, but people are less variable in their perception of the height of the track in each testing trial. Further comparison between the present results and those of Experiment 2 is provided in Sect. 4.

Regression Analysis To compare participants' produced speeds with the ground-truth model predictions, we performed a linear regression analysis as depicted in Fig. 3. Optimal performance is indicated by a slope of 1.0; a slope less than 1.0 indicates that participants underestimated speed, and a slope greater than 1.0 indicates the opposite. There is a strong linear relationship between speeds produced by human participants and the speeds predicted by the ground-truth model. However, the regression slopes are smaller than 1.0 under each gravity field, indicating that humans move slower than what is optimal given the ground-truth model. Several factors could contribute to the apparent underestimation of speeds in Experiment 1's perception-for-action task. One primary reason could be due to the absence of haptic feedback following interaction with the virtual ball. As the weight of the virtual ball can only be inferred (rather than directly perceived through the sensorimotor system as is typical in real-life situations), participants may have significantly underestimated the weight of the ball. In turn, this may have led participants to underestimate the force needed to propel the virtual ball to hit each target location. Thus, regardless of the gravity field, participants consistently underestimated the speed needed for the task.

A second linear regression analysis was then performed to examine how produced speed varied as a function of distance for each track height. The regression coefficients, standard errors and the r-squared statistics for Experiment 1 are reported in Table 1. The ground-truth slope s_{gt} of the speed-distance relationship was determined by the following expression:

$$s_{gt} = \sqrt{\frac{g'}{2h}}, \tag{1}$$

where g' is the gravitational acceleration and h is the track height.

For each gravity-height condition, the human slope was less than the ground-truth slope, providing converging evidence that speed was underestimated regardless of the environment. For Experiment i, the percent error between the calculated slope s_i and the ground-truth slope s_{gt} for each gravity-height combination was defined by $\frac{|s_i - s_{gt}|}{s_{gt}} \times 100\%$. The mean percent errors (i.e., percent errors averaged across height) for the 1.5g, 1.0g, and 0.5g environments in Experiment 1 were 43.9%, 25.6%, and 34.5%, respectively. The smallest mean percent error was in the 1.0g environment, which agrees with our previous finding where humans performed best in the earth-gravity environment.

Comparison with Original Study The results from Experiment 1 are in general agreement with Krist *et al.*'s previous findings [21]. Both studies (ours in VR and Krist's in the real world), found significant main effects of height and distance, suggesting that humans are sensitive to critical physical variables when interacting with objects, regardless of whether they are virtual or physical. In the virtual environment, however, we found that the interaction effect between height and distance was only marginally significant, whereas the original study (conducted in the real world) revealed a significant interaction between the two factors. The weakened interaction effect was most likely due to participants' underestimation of produced speed, which may have resulted from the absence of haptic feedback from the VR system. Participants may have developed an implicit bias toward believing the virtual ball was weightless, effectively reducing the variance in their responses and the corresponding power of our statistical analyses.

3.2 Experiment 2: Speed Judgment

Although participants were able to move freely in the virtual environment (i.e., all 3D view angles were allowed), they were not able to control or act upon any virtual objects in the present task. The second experiment was designed to examine the human ability to estimate the initial speed of a ball required to hit a target location under different gravity fields. In this task, participants were not asked to perform an action: i.e., striking a ball. Instead, they were asked to give a speed rating.

Experiment Setting Experiment 2 employed the same track, ball, and target as in Experiment 1. Objects in Experiment 2, however, were stationary: i.e., instead of allowing participants to directly interact with the virtual objects, a slider was introduced to gauge human participants' *speed ratings* (see Fig. 1b). For each stimulus, participants estimated the initial speed of the ball required to hit the target on the ground. Participants were asked to report their estimated speed using a slider. The leftmost side of the slider represented the slowest speed (0.1 meters per second), and the rightmost side of the slider represented the fastest speed (5.5 meters per second). In the experiment, participants used a Vive controller to move the slider to indicate their estimated speed. The reading of the slider was converted into speed by the following expression,

$$s = p \times (s_{max} - s_{min}) + s_{min}, \tag{2}$$

where s is the horizontal speed of the ball, p is the reading of the slider (between 0 and 1), and s_{max} and s_{min} are the maximum and minimum speeds, respectively.

Training Session At the beginning of each experiment, three demonstration trials were provided to participants. In each demonstration, a ball traveled along the track at three different speeds: maximum speed (5.5 meters per second), minimum speed (0.1 meters per second), and medium speed (2.8 meters per second). The speed was indicated on the slider at each corresponding position and was visible to the participants.

After observing the three demonstrations, identical settings with the same three pre-defined speeds were shown again but in a random order. This time, the speed of the ball was not explicitly provided to the participants. Instead, the participants were asked to move the slider using their controller to indicate the corresponding speed. If participants did not answer correctly, they were asked to repeat the trial again. If participants failed the second time, they did not proceed to the next part of the experiment. In the real experiment, everyone answered this part correctly after the first or second time. Therefore, every participant proceeded to the next part of the experiment.

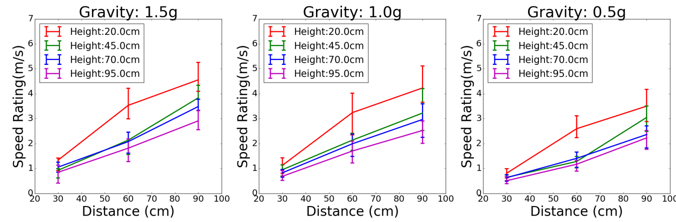


Fig. 4: Mean estimated speed as the function of distance between the track and the target in Experiment 2 under three different gravity fields: 1.5g, 1g, and 0.5g. Error bars indicate standard error of the mean.

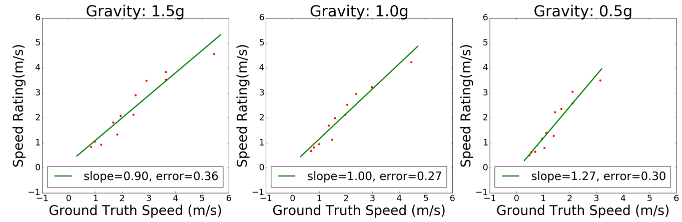


Fig. 5: Correlation between human speed estimates and speeds predicted by the ground truth model in Experiment 2 under three different gravity fields: 1.5g, 1g, and 0.5g.

Similar to Experiment 1, participants completed 6 different training trials, one for each distance-height combination. The task, however, differed in that participants triggered the ball's movement by indicating their speed rating on a virtual slider rather than striking the virtual ball with their controller. In each of the training trials, the trajectory of the ball was displayed after leaving the horizontal track.

Testing Session In the testing session, a track and a target were shown at 12 different height-distance combinations (same combinations as Experiment 1). However, in Experiment 2, after participants selected a speed on the slider, the trajectory of the ball was not displayed after leaving the horizontal track. Hence, no visual feedback was provided to the user by the VR system in Experiment 2. This was done to minimize rapid learning based on previous testing trial outcomes.

Results We carried out the same analysis in Experiment 2 as in Experiment 1. We first performed a $3 \times 3 \times 4$ (Gravity $\times$ Height $\times$ Distance) ANOVA at the $\alpha = 0.05$ significance level to determine whether participants' speed ratings depended on the magnitude of gravitational acceleration, the height of the track, and the distance between the track and target. The three-way interaction term was not significant ($F[12,684] = 0.53$, $p = 0.896$). Results from the ANOVA indicate significant main effects of gravity ($F[2,684] = 69.35$, $p < .001$), height ($F[2,684] = 787.81$, $p < .001$), and distance ($F[3,684] = 105.08$, $p < .001$) which agrees with Experiment 1's results. We found significant two-way interactions between gravity and distance ($F[4,684] = 4.61$, $p = .001$) but gravity did not interact with height ($F[6,684] = 0.60$, $p = .733$). Interestingly, we found a significant Height $\times$ Distance ($F[6,684] = 14.06$, $p < .001$) interaction in Experiment 2, which agrees with the ground-truth model and previous findings in the real-world environment [21]. Fig. 4 depicts the mean estimated speed for different distance-height combinations under three different gravity fields. The influence of speed and distance on human speed estimation appears to qualitatively vary across the four track heights, as evident in the significant Distance $\times$ Height interaction. Furthermore, the relationship also appears to vary across the three gravity fields, as evident in the significant Gravity $\times$ Height interaction.

Running a random effects ANOVA with two-way interactions, we found there was not a significant effect of participant on speed rating ($F[19,544] = 1.59$, $p = .083$). The interaction between gravity and participant ($F[38,544] = 5.57$, $p < .001$), distance and participant ($F[38,544] = 4.34$, $p < .001$), and height and participant ($F[57,544] = 1.91$, $p < .001$) were significant. These findings appear to imply larger individual differences in perceived gravity, distance, and height in the speed judgment task as opposed to the direct action

Table 1: Linear relationships between human estimated speeds and speeds predicted by the ground-truth model in Fig. 2 and Fig. 4, where g denotes ground truth gravity field (1.0 means $1.0g = 9.8m/s^2$), $h(cm)$ is the height of the track, s_{gt} is the ground truth slope, s_1 is the regression coefficient of the data collected in Experiment 1, σ_{s_1} is the standard deviation for the produced speeds in Experiment 1, s_2 is the regression coefficient for the speed ratings in Experiment 2, and σ_{s_2} is the standard deviation of the data collected in Experiment 2.

g	h	s_{gt}	s_1	σ_{s_1}	$r^2_{s_1}$	s_2	σ_{s_2}	$r^2_{s_2}$
1.5	20.0	0.0606	0.0251	0.1097	0.9692	0.0537	0.2762	0.9577
1.5	45.0	0.0404	0.0123	0.0677	0.9522	0.0485	0.1161	0.9905
1.5	70.0	0.0324	0.0249	0.0642	0.9890	0.0405	0.0916	0.9916
1.5	95.0	0.0278	0.0211	0.1069	0.9589	0.0344	0.0306	0.9987
1.0	20.0	0.0495	0.0334	0.2396	0.9212	0.0518	0.2665	0.9577
1.0	45.0	0.0330	0.0258	0.1919	0.9157	0.0379	0.0243	0.9993
1.0	70.0	0.0265	0.0170	0.0865	0.9586	0.0357	0.0452	0.9973
1.0	95.0	0.0227	0.0199	0.2077	0.8468	0.0307	0.0454	0.9964
0.5	20.0	0.0350	0.0185	0.0836	0.9669	0.0449	0.2081	0.9655
0.5	45.0	0.0233	0.0159	0.0222	0.9967	0.0401	0.2668	0.9314
0.5	70.0	0.0187	0.0125	0.0856	0.9280	0.0289	0.0422	0.9965
0.5	95.0	0.0161	0.0119	0.1157	0.8641	0.0287	0.0936	0.9825

task.

Regression Analysis We examined the linear relationship between participants' speed ratings and the ground-truth speed as depicted in Fig. 5. The regression analysis again shows a strong linear relationship between the two speeds under each gravity field. The slope under earth's gravity is 1.0, indicating that participants were highly accurate when triggering a ball to move toward a target location in a familiar environment. The slope in the 1.5g environment is less than 1.0, and the slope in the 0.5g environment is greater than 1.0, indicating that participants under- and overestimated speed in each respective environment when action was not involved. These findings suggest that participants' beliefs about gravitational acceleration in the 1.5g and 0.5g environments were biased toward earth's gravity field when they were asked to mentally estimate the speed rather than physically performing the action. Note that humans underestimated speed under all gravity fields in Experiment 1. However, in Experiment 2, humans' rated speeds under unfamiliar gravity fields showed a strong bias toward earth's gravity and even showed a slope of 1.0 under earth's gravity. The discrepancy between the two experiments is likely due to more implicit reasoning involved in Experiment 1's direct action task and more explicit reasoning based on low-level physical knowledge in Experiment 2's speed judgment task.

Next, we performed a linear regression analysis to quantify how speed rating varies as a function of distance under each gravity field. Calculated regression coefficients and their corresponding standard errors for Experiment 2 are reported in Table 1. The mean percent error between the ground-truth slope and the human slope from the regression analysis was 20.0%, 22.4%, and 58.3% for the 1.5g, 1.0g, and 0.5g environments, respectively. The speed-distance slopes for all track heights in the 0.5g environment were greater than the corresponding ground-truth slope, suggesting a bias toward earth's gravity. In the 1.5g environment, however, the speed-distance slope exceeded the ground-truth slope for three of the four track heights. This appears to indicate a bias away from (rather than toward) earth's gravity field, which disagrees with results from the regression analysis comparing human speed ratings to ground-truth predictions in Experiment 2. However, this discrepancy needs to be interpreted cautiously since humans showed much larger variability in the 1.5g environment. Specifically, participants were increasingly inconsistent when the track was rendered near to the ground in the 1.5g environment, as evident in the large standard errors on the corresponding regression coefficients.

Comparison with Original Paper Results from Experiment 2 are in agreement with Krist *et al.*'s previous findings [21]. The ANOVA results for participants' speed ratings showed the expected results, including significant main effects of height and distance and a significant interaction between the two variables. However, in the explicit reasoning task, we also noticed a strong bias toward earth's gravity field, which suggests the use of low-level, common-sense physical knowledge that over-generalized to novel situations.

3.3 Experiment 3: Contact Location Prediction

Our third experiment was designed to examine the human ability to predict the contact location of a projectile's trajectory under familiar and unfamiliar gravity fields. The ball's trajectory was briefly displayed and then occluded prior to measuring participants' predictions. Thus, participants were required to extrapolate projectile motion according to limited visual input. The aim of the present experiment was to determine the reasoning strategies people employ when predicting future projectile locations: do people propagate spatially represented objects forward in time using a mental simulation engine, or do they rely on more explicit reasoning strategies?

Experiment Setting As illustrated in Fig. 1c, the virtual environment in Experiments 3 and 4 consisted of a tilted laser beam, a launching platform suspended in the air (height = 3 meters), and a white ball (diameter = 0.08 meters, friction coefficient = 0) resting on top of the platform. The angle between the laser beam and the ground was 45°, and the horizontal distance between the bottom of the laser beam and the platform was 3 meters. In the experiment, the white ball moved horizontally with a random initial velocity, and the ball disappeared 0.2 seconds after leaving the platform. Participants were asked to predict the location on the laser beam where they believed the ball would make contact. The trajectory of the ball always intersected with the laser beam. The reason for choosing a specified orientation for the laser beam was to ensure that participants accounted for gravity when estimating the flight duration of the ball in the Experiment 4 (see Sect. 3.4).

Training Session At the beginning of each experiment, participants were shown one full trajectory of a ball moving from the launching platform to the contact location on the laser beam. A second ball was then presented with the same initial speed but disappeared 0.2 seconds after leaving the platform. Participants were then asked to use their controller to indicate where the ball would make contact with the laser beam. The training session was designed to familiarize participants with the controller and the task procedure. Participants were not provided with any feedback on the true contact location nor the accuracy of their decisions.

Testing Session In the testing session, participants were presented with 10 testing trials in a randomized order. In each trial, the ball moved with a different initial speed and disappeared 0.2 seconds after leaving the platform. Participants were then asked to predict the contact location on the laser beam. The location indicated in the virtual environment served as the location prediction measurement for each participant. No feedback was given following each response.

The experiment was conducted under three different gravity fields (1.5g, 1.0g, and 0.5g). The initial speed of the ball, s, for each trial was calculated using the following expression:

$$s = \frac{\tan(\pi/2) \times h}{\sqrt{2h/g'}}, \tag{3}$$

where g' is the gravitational acceleration and h is the height of the contact location on the laser beam. Height was selected from 10 different values: 1.07, 1.17, 1.25, 1.31, 1.37, 1.42, 1.48, 1.54, 1.62, and 1.72 meters. These values were chosen from a Gaussian distribution such that the true contact points were denser in the middle and sparser on both ends of the laser beam. Experiments under different gravity fields shared the same set of heights but in different (randomized) orders.

Results We conducted an ANOVA on the percent error ($\frac{|H_h - H_{gt}|}{H_{gt}} \times 100\%$) between participants' predicted contact locations H_h and the corresponding ground-truth value H_{gt} for each height condition. Results from the ANOVA indicate that the error was not significantly influenced by different gravity fields ($F[2,597] = 0.33$, $p = 0.717$). There was a significant influence of height on the percent error ($F[9,597] = 4.34$, $p < .001$).

Trajectory Models To determine how participants predicted the end point of the trajectory, we compared human performance to four different geometric models. Each model served as a separate hypothesis for predicting the trajectory contact location. Human predictions were compared to each candidate hypothesis.

- **Linear.** The linear model served as a baseline model with the simplest form of contact location prediction.
- **Parabola under current gravity.** The parabola under current gravity model provided the ground-truth contact location for the trajectory in the current environment: 1.5g, 1.0g, or 0.5g.
- **Parabola under earth gravity.** Considering that participants might have had a bias toward earth's gravitational acceleration, we compared each prediction to the contact location for the parabolic trajectory in the earth environment: 1.0g.
- **Circle.** Considering that people have rich experiences with circular motion in daily life, one possible hypothesis is to interpret the observed trajectory components as part of circular object movement. We compared human predictions to the contact location for a circular trajectory.

Trajectory Model Results To test the candidate models, we fit each trajectory to sampled points from the initial 0.2 seconds of the trajectory and computed the mean squared error (MSE) between each model's predicted contact location and human responses. To fit each model, we sampled 10 equally spaced points from the first 0.2 seconds of the observed trajectory. Model parameters were then computed to fit the 10 sampled points such that the MSE was minimized. For the circle model, the least squares method determined a local MSE minimum for the circle's center and radius. The initialization of the parameters influenced the estimated results, so we swept through 20 different circle centers and 30 different radii (600 parameter combinations) for the initialization and picked the parameters that corresponded with minimum MSE.

Fig. 6 depicts the contact locations predicted by each candidate trajectory model and human contact location predictions in each environment. Results indicate that humans are remarkably accurate at predicting contact locations given the initial 0.2 seconds of a projectile's trajectory. Comparing predictions from the four candidate models, the *parabola under current gravity* (ground-truth) model provided the best quantitative fit to human contact location predictions under each gravity field (see Table 2). The average MSE across environments was approximately 10 centimeters, which is fairly accurate given that the cross-section of the Vive controller—which participants used to indicate their contact location predictions—was 11.7 centimeters $\times$ 8.3 centimeters. There was no observed bias toward the *parabola under earth gravity* model in either of the unfamiliar environments (i.e., 1.5g and 0.5g) The present analysis shows that humans can successfully predict future trajectory locations based on limited visual input, and this ability is not hindered in novel physical environments with non-earth gravity fields.

Note that Experiment 3 was different from Experiments 1 and 2 in both the visual inputs provided to participants (e.g., the laser beam, platform, etc.) and its corresponding task demands. Recall that the first two experiment settings were always presented to participants prior to Experiments 3 and 4. Thus, the experimental design made it possible for participants to generalize knowledge about gravity from the first two experimental settings to later experiments since they were informed that all four experiments in each block shared the same gravity field. However, we found that participants showed a global bias toward contact locations predicted by the *linear* trajectory model. If participants employed a prior belief that the gravity field in an environment should correspond with that on earth, one would expect a bias toward the contact locations predicted by the *parabola under earth gravity* model in each of the unfamiliar environments. This result suggests that participants may have employed explicit, perceptual knowledge (e.g., spatial location and velocity) when making their contact location predictions (i.e., prior beliefs about gravitational acceleration did not appear to weigh into participants' contact location predictions).

Table 2: Mean squared error (in meters) of each candidate trajectory model in Experiment 3.

Gravity \ Model	Circle	Line	Parabola	Earth Parabola
1.5g	0.88	1.14	0.07	0.46
1g	0.90	1.26	0.12	0.12
0.5g	0.98	1.51	0.10	1.03

3.4 Experiment 4: Flight Duration Estimation

The fourth experiment was designed to study the human ability to estimate the flight duration of a projectile under familiar and unfamiliar gravity fields. Unlike in Experiment 3, the purpose here was to analyze human time estimation (rather than spatial location prediction) given occluded projectile motion. The aim of the present task was to determine the reasoning strategies people employ when estimating the duration of physical events: do people estimate flight duration using explicit reasoning strategies, as suggested in Experiment 3? Alternatively, do people rely on implicit reasoning strategies (e.g., mental simulation) when performing the temporal task?

Experiment Setting The design of Experiment 4 was similar to that of Experiment 3. However, instead of predicting the contact location of the projectile, participants were instructed to click the trigger on their controller when they believed the occluded ball made contact with the laser beam. The laser beam in Experiment 4 was tilted at a 45° angle, identical to the previous experiment. The laser beam was tilted to ensure participants accounted for gravity when estimating flight duration: i.e., if the laser beam was horizontal, the flight duration would remain constant across trials in a given environment. Moreover, if the laser beam was vertical, flight duration would only depend on the horizontal velocity of the projectile.

Training Session At the beginning of each experiment, participants were presented with one full trajectory of the ball with an unknown initial speed. Participants were instructed to click the controller once the ball made contact with the laser beam. If participants clicked more than 0.05 seconds earlier or later than the true contact time, they were asked to repeat the trial with the same initial speed until they responded within the 0.05 second window. In the second practice trial, participants were presented with another ball with the same initial velocity and were asked to click the trigger when they thought the ball contacted the laser beam. This time, participants were not given feedback, and the ball disappeared 0.2 seconds after leaving the platform. Similar to Experiment 3, the training session was designed to familiarize participants with the controller and flight duration estimation task. The initial speed used in the training session was not observed in the testing session.

Testing Session In the testing session, participants were presented with 10 testing trials with different initial speeds and were asked to click the button on the controller once the ball made contact with the laser beam. The ball disappeared 0.2 seconds after leaving the platform in each testing trial. The heights of the contact locations were chosen from the same set of heights as in Experiment 3. After each prediction, participants were not provided with feedback regarding the ground-truth flight duration nor their accuracy. Flight duration measurements for each participant were determined by subtracting the time the ball left the platform from the response time indicated by participants on their controllers. The experiment was conducted under the three different gravity fields employed in the previous experiments: 1.5g, 1.0g, and 0.5g.

Results In order to examine the effect of gravity on participants' flight duration estimates, we conducted an ANOVA on participants' flight duration errors (i.e., the difference between participants' estimated flight durations and the ground-truth flight durations) in each trial. Results from the analysis revealed a significant main effect of gravity ($F[2,597] = 6.99$, $p = 0.001$). The present results indicate that participants accounted for gravitational acceleration when reasoning about the flight duration of an occluded projectile. Table 3 provides participants' mean flight duration error across height and gravity conditions. As indicated in the rightmost column of the table, participants' flight duration estimates were biased toward flight durations under earth's gravity: i.e., flight duration estimates were over- and underestimated in the 1.5g and 0.5g environments, respectively. Thus, participants' flight duration estimates were biased toward earth-gravity flight durations in the unfamiliar (1.5g and 0.5g) environments.

4 Comparison between Experiments

4.1 Comparison of Experiments 1 and 2

In both Experiments 1 and 2, participants took into consideration all three experimental parameters (i.e., gravity, height, and distance) when

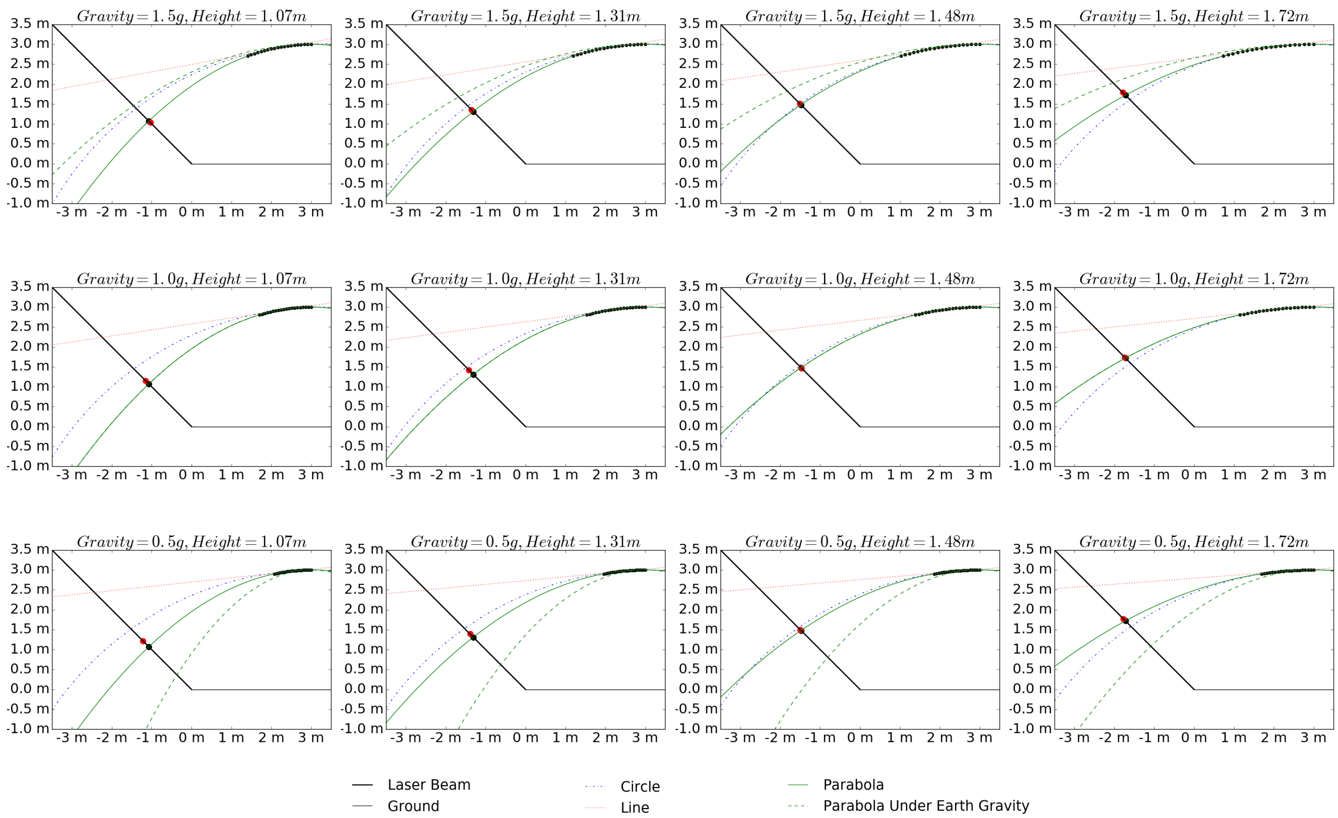


Fig. 6: Contact locations predicted by the four candidate trajectory models in Experiment 3. The black dots on the top right corner indicate the 10 sampled points from the initial 0.2 seconds of each trajectory. The oblique line represents the laser beam, and the horizontal line represents the ground. The red dot on the laser beam indicates participants' mean contact location predictions. Each row depicts results for a different environment: 1.5g, 1.0g, and 0.5g (top to bottom). Each column indicates a different platform height: 1.07, 1.31, 1.48, and 1.72 meters (left to right).

propelling a ball off of a track onto a target location. The ANOVA in both experiments showed interaction effects between gravity and distance but did not show a significant interaction between gravity and height. One interesting difference between the results of Experiments 1 and 2 is that the ANOVA on produced speed in Experiment 1 showed a marginally significant interaction between distance and height, while the ANOVA for Experiment 2 reported a significant result in agreement with previous findings [21]. As mentioned previously, this may have been due to participants significantly underestimating the weight of the projectile ball due to the absence of haptic feedback in Experiment 1. Alternatively, the distance in Experiment 1 may have been underestimated in each environment due to the use of a tool (i.e., a Vive controller). This would be consistent with previous reports that tool use can reduce perceived distance [44].

Comparing the present results with those from the adult group in [21], we found that the distance versus speed rating relationship qualitatively agrees between the VR experiment and previous work in the real-world situation: both experiments revealed a strong linear relation between speed estimates and distance, and people's speed ratings varied across different levels of height. However, we found that participants' produced speeds were slower than they should have been according to the ground-truth physical model in each environment, leading to a somewhat nonlinear trend between produced speed and distance. In Krist *et al.*'s original study, participants physically *pushed* a ball along a track to propel it toward indicated target locations [21]. Thus, participants could adjust their force input—and associated produced speeds—during the testing phase to match their ideal target speed. In the present experiment, participants hit a virtual ball with a second ball (corresponding to the Vive controller's location) in an instant and received no haptic feedback: i.e., they were missing an informative variable in their perceptual-motor representation. This lack of sensorimotor feedback might have made it harder for participants to monitor and adjust the magnitude of their input in real time and perhaps caused them to produce speeds that were biased toward a "moderate" magnitude. This uncertainty in sensory input might have given rise to the nonlinear relationship between produced speed and distance observed in the virtual environment in Experiment 1.

4.2 Comparison of Experiments 3 and 4

To compare participants' contact location predictions and flight duration estimates in the last two studies, we further inferred the height of the contact location of the ball according to the flight duration estimates in Experiment 4 using the following expression:

$$H_{infer} = 3 - 0.5 \times g' \times t^2, \tag{4}$$

where H_{infer} is the inferred height of the contact location, 3 is the initial height (in meters) of the ball, g' is the gravitational acceleration in the environment, and t is the flight duration estimate for each participant. The mean difference, $\delta_{H_{infer}}$, between the inferred height and the ground-truth contact location across trials is presented in Table 4. In each environment, the inferred contact location error from Experiment 4 was at least double the contact location error measured in Experiment 3. If people were using the same reasoning strategy in both tasks, one would expect equivalent magnitudes of error. This suggests that participants did not employ the same reasoning strategy in the the contact location prediction (spatial processing) and flight duration estimation (temporal processing) tasks. Hence, when examining and comparing human behavior across various intuitive physical tasks, it is important to carefully examine the type of processing involved in specific tasks.

5 Conclusion and Future Work

The present work examined human performance in two projectile motion paradigms using state-of-the-art VR technology to produce artifi-

Table 3: Difference between mean flight duration estimates and the ground-truth (earth-gravity) flight duration (in ms) in Experiment 4. For each environment (1.5g, 1.0g, and 0.5g), the first row indicates the difference between participants' mean flight duration estimates and the ground-truth flight duration. The second row indicates the difference between participants' mean flight duration estimates and the flight duration under earth's gravity. Positive and negative values indicate over- and underestimation, respectively. Height is in meters.

Gravity \ Height	1.72	1.62	1.54	1.48	1.42	1.37	1.31	1.25	1.17	1.07	mean (ms)
1.5g	42.88	52.23	22.62	25.99	38.59	-7.65	-25.04	14.95	17.14	32.91	21.46
-	15.42	21.68	-10.44	-9.01	1.65	-46.23	-65.61	-27.62	-28.15	-15.81	-16.41
1.0g	38.73	43.54	10.89	14.71	55.31	-38.06	-16.32	-3.13	-41.42	-33.50	3.07
-	-	-	-	-	-	-	-	-	-	-	-
0.5g	25.83	-10.68	-47.44	-36.42	-12.20	-39.93	-90.63	-98.99	-31.24	-46.69	-38.84
-	123.27	95.56	65.82	82.08	111.56	88.18	42.70	39.54	114.20	107.35	87.03

cial physical scenarios. We assessed human performance under natural and unnatural gravity fields in replicated ([21]; Experiments 1 and 2) and novel (Experiments 3 and 4) settings and systematically examined gravitational effects on human performance. Results in the virtual environment from Experiments 1 and 2 were qualitatively in agreement with previous findings in the real world, although the linear relationship between speed and distance was more pronounced in the speed rating (perception-only) task compared to the direct action (perception-for-action) task. Results demonstrate a strong relationship between produced/rated speeds and target distance, and this linear relationship varied across gravity fields (in accordance with the ground-truth relationship). This indicates that participants consistently attended to gravitational acceleration when producing and rating speeds. Mean speed rating errors were negative in the 1.5g setting and positive in the 0.5g setting, indicating that participants' representations of gravitational acceleration were biased toward earth's gravity in our perception-only task. This finding reinforces the hypothesis that people infer the physical behavior of their environment by combining noisy perceptual inputs with Newtonian principles given prior beliefs about represented variables [33]: e.g., gravity. The present results provide evidence that humans hold a strong prior belief about gravitational acceleration which appears distributed around earth's gravitational acceleration ($\approx 9.8\ \mathrm{m/s^2}$) and leads to apparent biases in their physical intuitions.

The tasks in Experiments 1 and 2 appear drastically different in their cognitive demands, as evident in the qualitative differences between their speed-distance relationships: i.e., the nonlinear trend in Experiment 1 deviates from the ground-truth model and previously reported findings [21]. Furthermore, research in intuitive physics would predict superior performance in the speed production task due to its concrete task domain [18–20, 40]. One key difference in the speed production task, however, is the role of motor input and haptic feedback in producing desired projectile speeds for each experimental trial. In the VR environment, no haptic feedback was provided following propulsion of the ball. Thus, the strength of each hit (manifested in participants' perceptual-motor representations) was inferred rather than directly perceived as is generally the case in daily life. This discrepancy introduces additional uncertainty into the physical environment, which in turn biases the inferred forces toward some prior belief or expectation. Based on these findings, future work in VR environments should work toward providing haptic feedback about experienced forces (perhaps by administering vibrations of variable intensity or other low-level sensory cues; see [7]) in order to create an environment with multisensory input that is more consistent with the real world.

Participants were reasonably accurate, however, when reasoning about future trajectory locations in the subsequent experiments, where they were no longer required to provide motor input to the virtual environment. Furthermore, we found that participants' time judgments were biased toward those that would be expected under earth's gravity, although their location predictions were remarkably accurate. Unlike their flight duration estimates, participants' location predictions did not vary significantly across gravitational fields. Our results imply that participants may infer flight durations by reasoning about each trajectory outside of perception. This strategy was mentioned in subjective reports from some participants that they formed their time estimates by adjusting their gaze according to where they thought the ball *should be*) following occlusion. This strategy agrees with results from physical simulation models, where future physical states are propagated forward in time, given prior beliefs about observable and hidden variables [4, 5, 12], specifically gravity in the present case. Similar to findings from Experiments 1 and 2, responses were once again biased toward what would be expected under earth's gravitational field.

However, participants did not appear to adopt a simulation approach to reason about the *locations* at which the balls contacted the laser beam. Since simulations occur in real time, using a simulation heuristic to predict a future location would be pointless: i.e., future physical states would not be determined until they actually *happened*. Naturally, a baseball player catching a fly-ball does not need much time to move to where he predicts the ball will land. Instead, the location prediction task appears explicit, relying on perceptual cues and prior knowledge that objects tend to move along parabolic trajectories under gravity. Participants in the present study upheld this belief, although their responses were consistently biased (but only slightly) toward the linear trajectory. This agrees with previous findings in the intuitive physics literature, where people can accurately *predict* the end location of a trajectory, although they are less accurate when *explaining* the trajectory on pencil-and-paper tasks [40].

Taken together, the results in the present work demonstrate that humans maintain an impressive ability to habituate to novel physical environments and appear unhindered in predicting future locations of observed trajectories across varying gravitational fields. Our results suggest that humans on a mission to Mars would have minimal difficulty adapting to new gravities, and we conjecture that people's biases toward Earth's gravitational field would diminish over time through learning. We also suspect that this bias would be lessened in the real world since people would feel the weight of their bodies change across environments. Future work should aim to further explore this bias, perhaps by weighing down participants' bodies in gravity fields greater than earth's.

Our experiments provide a set of physical reasoning problems that lend themselves to varying degrees of spatial representation: i.e., participants appeared to represent latent gravitational acceleration in Experiments 1, 2, and 4 but relied on observable position and velocity in Experiment 3. Such representations can be modeled according to noisy human perception, and future work should aim to determine whether associated probabilistic simulation approaches match well to behavioral measurements. In addition, the present work—unlike Krist *et al.*'s previous study [21]—did not explore performance across age groups. It would be interesting for future work to determine how different age groups (e.g., children and the elderly) habituate to novel virtual gravity fields.

Table 4: Mean error of inferred contact location (converted from flight duration estimates in Experiment 4) and predicted contact location (from the results of Experiment 3) in each environment.

Gravity	$\delta_{H_{infer}}$ (m)	$\delta_{H_{predict}}$ (m)
1.5g	0.26	0.07
1.0g	0.24	0.12
0.5g	0.31	0.10

ACKNOWLEDGMENTS

We thank Dr. Chenfanfu Jiang of the UCLA Computer Science and Math Department for useful discussions. The work reported herein was supported by DARPA SIMPLEX grant N66001-15-C-4035, ONR MURI grant N00014-16-1-2007, NSF BCS-1353391and an NSF Graduate Research Fellowship.

REFERENCES

[1] NVidia FleX. https://developer.nvidia.com/flex. Accessed: 2016-12-05.

[2] Unity 3D engine. https://unity3d.com/. Accessed: 2016-12-05.

[3] Unreal engine. https://www.unrealengine.com. Accessed: 2016-12-05.

[4] C. J. Bates, I. Yildirim, J. B. Tenenbaum, and P. W. Battaglia. Humans predict liquid dynamics using probabilistic simulation. In *Proceedings of the 37th annual conference of the cognitive science society*, pp. 172–177, 2015.

[5] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.

[6] G. Bruder and F. Steinicke. Threefolded motion perception during immersive walkthroughs. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pp. 177–185. ACM, 2014.

[7] G. C. Burdea. Haptic feedback for virtual reality. In *Proceedings of the Virtual Reality and Prototyping Workshop*, 1999.

[8] J. Clement. Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50(1):66–71, 1982.

[9] J. Clement. Use of physical intuition and imagistic simulation in expert problem solving. In D. Tirosh, ed., *Implicit and Explicit Knowledge*, pp. 204–244. Erlbaum, Hillsdale, NJ, 1994.

[10] F. H. Durgin, B. Klein, A. Spiegel, C. J. Strawser, and M. Williams. The social psychology of perception experiments: Hills, backpacks, glucose, and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6):1582–1595, 2012.

[11] J. P. Gallivan, J. S. Cant, M. A. Goodale, and J. R. Flanagan. Representation of object weight in human ventral visual cortex. *Current Biology*, 24(16):1866–1873, 2014.

[12] T. Gerstenberg, N. D. Goodman, D. A. Lagnado, and J. B. Tenenbaum. How, whether, why: Causal judgments as counterfactual contrasts. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, pp. 782–787, 2015.

[13] D. L. Gilden and D. R. Proffitt. Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):372, 1989.

[14] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992.

[15] M. Hegarty. Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5):1084, 1992.

[16] M. Hegarty. Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8(6):280–285, 2004.

[17] M. Hegarty and V. K. Sims. Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4):411–430, 1994.

[18] M. K. Kaiser, J. Jonides, and J. Alexander. Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, 14(4):308–312, 1986.

[19] M. K. Kaiser, D. R. Proffitt, and K. Anderson. Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4):795, 1985.

[20] M. K. Kaiser, D. R. Proffitt, S. M. Whelan, and H. Hecht. Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3):669, 1992.

[21] H. Krist, E. L. Fieberg, and F. Wilkening. Intuitive physics in action and judgment: The development of knowledge about projectile motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4):952, 1993.

[22] J. K. Kubricht, C. Jiang, Y. Zhu, S.-C. Zhu, D. Terzopoulos, and H. Lu. Probabilistic simulation predicts human performance on viscous fluid-pouring task. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, pp. 1805–1810, 2016.

[23] W. Liang, Y. Zhao, Y. Zhu, and S.-C. Zhu. Evaluating human cognition of containing relations with physical simulation. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, pp. 782–787, 2015.

[24] J. Lin, X. Guo, J. Shao, C. Jiang, Y. Zhu, and S.-C. Zhu. A virtual reality platform for dynamic human-scene interaction. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, p. 11. ACM, 2016.

[25] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, 11:21–46, 2003.

[26] T. Lopez, R. Bouville, E. Loup-Escande, F. Nouviale, V. Gouranton, and B. Arnaldi. Exchange of avatars: Toward a better perception and understanding. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):644–653, 2014.

[27] M. McCloskey. Intuitive physics. *Scientific American*, 248(4):122–130, 1983.

[28] M. McCloskey, A. Caramazza, and B. Green. Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, 210(4474):1139–1141, 1980.

[29] M. McCloskey, A. Washburn, and L. Felch. Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4):636, 1983.

[30] D. R. Proffitt. Embodied perception and the economy of action. *Perspectives on Psychological Science*, 1(2):110–122, 2006.

[31] S. Runeson. On visual perception of dynamic events. In Almqvist and Wicksell, eds., *Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensia*. Almqvist and Wicksell, Stockholm, Sweden, 1977/1983.

[32] S. Runeson, P. Juslin, and H. Olsson. Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence. *Psychological Review*, 107(3):525, 2000.

[33] A. N. Sanborn, V. K. Mansinghka, and T. L. Griffiths. Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, 120(2):411, 2013.

[34] C. Schatzschneider, G. Bruder, and F. Steinicke. Who turned the clock? Effects of Manipulated Zeitgebers, Cognitive Load and Immersion on Time Estimation. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1387–1395, 2016.

[35] D. L. Schwartz and J. B. Black. Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology*, 30(2):154–219, 1996.

[36] D. L. Schwartz and T. Black. Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1):116, 1999.

[37] B. Shanon. Aristotelianism, Newtonianism and the physics of the layman. *Perception*, 5(2):241–243, 1976.

[38] Z. W. Sherin, R. Cheu, P. Tan, and G. Kortemeyer. Visualizing relativity: The openrelativity project. *American Journal of Physics*, 84(5):369–374, 2016.

[39] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[40] K. Smith, P. Battaglia, and E. Vul. Consistent physics underlying ballistic motion prediction. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 3426–3431, 2013.

[41] C. Stetson, X. Cui, P. R. Montague, and D. M. Eagleman. Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, 51(5):651–659, 2006.

[42] J. Todd and W. Warren. Visual perception of relevant mass in dynamic events. *Perception*, 11(3):325–335, 1982.

[43] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.

[44] J. K. Witt, D. R. Proffitt, and W. Epstein. Tool use affects perceived distance, but only when you intend to use it. *Journal of Experimental Psychology*, 31(5):880–888, 2005.

[45] J. Yates, M. Bessman, M. Dunne, D. Jertson, K. Sly, and B. Wendelboe. Are conceptions of motion based on a naive theory or on prototypes? *Cognition*, 29(3):251–275, 1988.

[46] C. J. Ziemer, M. J. Branson, B. J. Chihak, J. K. Kearney, J. F. Cremer, and J. M. Plumert. Manipulating perception versus action in recalibration tasks. *Attention, Perception, and Psychophysics*, 75(6):1260–1274, 2013.