

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Investigating the Manifestation of naïve Physics at the Embodied Level Through Motor and Judgment Tasks in Virtual Reality

Permalink

<https://escholarship.org/uc/item/5v6214kh>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Wang, Fan

Chatain, Julia

Varga, Rudolf

et al.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Investigating the Manifestation of Naïve Physics at the Embodied Level Through Motor and Judgment Tasks in Virtual Reality

Fan Wang (fan.wang@gess.ethz.ch)¹, Julia Chatain², Rudolf Varga¹, Bruno Herbelin³, Olaf Blanke³ & Manu Kapur¹

¹Department of Humanities, Social and Political Sciences, ETH Zurich

²Faculty of Mathematics and Computer Science, UniDistance Suisse

³Laboratory of cognitive neuroscience, EPFL

Abstract

People often hold the intuition that heavier objects fall faster, yet can still accurately catch objects of different masses. Such coexistence of naïve physics and the internal model of Earth's gravity is well documented, but how these are weighted when they conflict remains unclear. We designed a Virtual Reality ball-catching experiment to examine how acceleration manipulations and the presence of a second, visible ball that enabled visual comparison modulate motor catching and perceptual judgment. We found that, when participants attempted to catch a single occluded ball, both motor catching and naturalness judgments showed a partial reliance on the internal model and were especially successful in distinguishing slow accelerations. Introducing a second ball with manipulated acceleration shifted judgments towards naïve physics expectations, while motor catching accuracy remained relatively robust. Our findings highlight how naïve physics can manifest at the perceptual level and inform motor-level embodied learning designs to initiate conceptual changes.

Keywords: internal model; naïve physics; embodied cognition

Introduction

It is not uncommon for people to hold the intuition that “heavier objects always fall faster”, yet being able to catch a falling cup or a flying tennis ball despite their mass differences. Humans form multi-modal knowledge about the physical world (Zago & Lacquaniti, 2005). Some knowledge take the form of abstract beliefs, while others are more implicit and embodied. In this study, our aim is to investigate human perceptual and motor behavior when different forms of knowledge lead to contradictory predictions.

Humans often demonstrate heuristic-based naïve physics judgments that are fragmented, context-dependent, and prone to misapplication (diSessa, 1993). Such intuitions also manifest in perceptual level. For example, in a lab study where polystyrene or wooden spheres fall on a projected screen, the participants exhibited a heavy-fast, light-slow heuristic when indicating the landing time (Vicovaro et al., 2019).

In contrast, when acting, humans also demonstrate internal models consistent with Newtonian physics (Neupärtl et al., 2020). Earth's gravity is one of such strong priors (Jörges & López-Moliner, 2017). Studies have shown that people tend to expect downward-moving objects to accelerate at 1g even when the velocity is actually constant (Zago et al., 2004). Specifically, people can successfully hit the ball falling from an incline when the trajectory is occluded (La Scaleia et al.,

2015) and accurately judge whether the ball motion is natural in a Virtual Reality (VR) simulation (La Scaleia et al., 2020). Compared to naïve heuristics, this embodied knowledge is typically more robust, and readily expressed at the motor level (Vicovaro, 2023).

A central question in intuitive physics research is whether and how human behavior integrates reactive, heuristic processes, and internalized, model-based control (Belousov et al., 2016; Vicovaro, 2021; Zhao & Warren, 2015). In the probabilistic view, humans learn by minimizing prediction errors through interaction with the environment, gaining more accurate internal models (Friston, 2010; Smith et al., 2024). What remains unknown is how the different forms of knowledge are weighed under specific task conditions, especially when they might generate conflicting predictions.

We designed a two-phase VR experiment that manipulated free-fall accelerations and object material. The first phase was to design and test conditions in which the embodied internal model of Earth's gravity can be activated when catching a single falling ball with partially occluded trajectory. As a design rationale, we emphasize the embodied nature of the task by affording direct interaction between participants and falling balls. Participants were immersed in VR and stood directly below the falling ball, acquiring a first-person perspective that made the action more immediate and grounded in prior experience.

The second phase was designed to elicit conflicting predictions from perceptual-level naïve physics and motor-level internal model of Earth's gravity. Building on Phase 1, Phase 2 introduced an additional visible comparison ball of the different material that always fell under the acceleration of 1g. In each trial, two tasks were included. Task1 was to try to catch the occluded ball (motor catching). Task2 was to judge the naturalness of the motion of this ball (naturalness judgment). Participants were unaware of the balls' acceleration amplitudes and were explicitly instructed to give the judgment only on the occluded ball. The motor catching task required spatial-temporal coordination under limited visual information, and thus preferred the internal model of gravity. In contrast, with another visible ball providing contextual visual reference, the naturalness judgment task encouraged comparative evaluation. This increased potential application of heuristics such as “heavier things fall faster”.

Informed by the literature that motor catching is more robust across contexts to follow the internal model of Earth's

gravity, we formulate our hypotheses based on the assumption that participants mainly rely on this internal model in motor catching tasks. In addition, the naturalness judgment tasks in Phase 1 were also hypothesized to be mainly based on this internal model aligned with Earth's gravity, as there was no distracting or confusing contextual cue to intentionally elicit the naïve physics knowledge.

RQ1(Phase 1): In phase 1 (single occluded ball), how do participants' motor catchings and naturalness judgments differ across simulated free-fall acceleration (0.5 g, 1 g, 1.5 g) and ball material (heavy, light)?

H1.1(accuracy): The motor catching will be the most successful in 1 g condition. Ball material will not have a main effect on accuracy. **H1.2(direction):** The motor catching will be earlier than the successful catching in 0.5 g condition and later than the successful catching in 1.5 g condition. Ball material will not have a main effect on motor catching direction. **H1.3(judgment):** Participants will judge the 1 g condition as the most natural motion. Ball material will not have a main effect on naturalness judgment.

RQ2(Phase 2): In phase 2 (two ball comparison), how do participants' motor catchings and naturalness judgments differ across the occluded ball's simulated free-fall acceleration (0.5 g, 0.8 g, 1 g) and ball material (heavy, light)?

H2.1(accuracy): The motor catching will be the most successful in 1 g condition. Ball material will not have a main effect on accuracy. **H2.2(direction):** The motor catching will be earlier than the successful catching in both 0.5 g and 0.8 g conditions. Ball material will not have a main effect on motor catching direction. **H2.3(judgment):** Because of the coexistence of the internal model of Earth gravity and the inaccurate conceptions that "heavier objects fall faster", we leave this hypothesis non-directional. Judgment patterns will be exploratorily examined to assess how conflicting cues are resolved when two objects fall simultaneously.

Methods

Participants

We recruited 25 participants (aged 19–37 years, $M = 24.24$; 18 female) from a university participant pool. The sample size was determined based on similar previous studies (La Scaleia et al., 2020; Vicovaro et al., 2019) which revealed that motor catching relied on an internal model of Earth's gravity, and that perceptual judgments showed a heavy-fast, light-slow heuristic. None of the participants were frequent VR users. This study has been approved by the ETH Zurich Ethics Commission (proposal 25 ETHICS-339).

VR Setup

We implemented the activity using Unity (version 2022.3.62f2) with the Meta SDK for VR development. In VR, a table was adjusted to participants' height so that the tabletop was 55 cm below their headset level. A release scaffold was positioned 2.3 m above the table. An occlusion box around 1.4 m tall was placed on the table directly beneath the

falling ball. Figure 1 demonstrates a visual representation of the setup that we detail below.

In all trials in both phases, participants were required to stand close to the table and look up to trigger the balls' fall. The initial fall of one or two balls (for Phase 1 and 2 respectively) was visible, the remainder occluded by a tall box (Figure 1). Participants were instructed to press the trigger button on the controller with the middle finger of their dominant hand, represented in VR as a ring, at the moment they believed that the center of the ball with the occluded trajectory would pass the ring, as if they were attempting to catch it. Specifically, participants were asked to place their virtual hand inside the occlusion box, which prevented them from seeing their hand when catching and knowing the outcome of their attempts. This constituted one motor catching trial. After a delay, participants grasped the ball as it came out of the occlusion box and placed it into a corresponding box to indicate whether they considered the motion of the ball natural to them. This constituted one naturalness judgment trial.



Figure 1: VR setup of **Phase 2 (two-ball comparison)**. (a) A tall occlusion box is placed on a table directly beneath the falling ball. Participants hold the controller with their dominant hand. The controller is virtually represented as a ring in the VR environment. Participants need to stand closely in front of the table, putting their virtual hand, i.e., the ring, on top of the table and inside the box, and look up to trigger the balls' falling. When they believe that the center of the occluded ball passes the ring, they press the trigger button. They do not see the ring during this catching process. (b) After a delay, the ball comes out from the side of the box, and participants grab the ball to put it into corresponding box that reflects their naturalness judgment.

Procedure

The experiment used a within-subjects design. The participants first went through a VR tutorial that familiarized them with the environment and the controller. Before the main experiment, participants completed a baseline reaction time task using a generic white ball without texture and with a fully

visible trajectory. We used a yellow-green tennis ball and a dark-blue metal ball as visual representations of the light and heavy balls, respectively. Before the start of Phase 1, participants were asked to carefully observe the appearance of the balls and describe their impressions. All participants reported that the dark-blue ball appeared heavier than the yellow-green ball. Then they were prompted to consider the typical heaviness of a tennis ball and a metal ball corresponding to these visual representations. We also informed the participants that one or two of these balls would fall under different simulated motions in the following trials without specifying the speeds.

Phase 1 of the experiment involved a single ball falling into an occlusion box. Ball materials (heavy, light) and simulated accelerations (0.5 g, 1 g, 1.5 g) were varied in randomized order. Participants' motor catchings and naturalness judgments were collected as dependent variable measures.

Phase 2 of the experiment involved two balls of different materials falling from the same height at the same time. One ball was occluded and its acceleration was manipulated, while the other remained visible and always fell under 1 g. The material of the occluded ball (heavy, light) and its simulated acceleration (0.5 g, 0.8 g, 1 g) were varied in randomized order as independent variables. The measures were the same as Phase 1. Specifically, here, we asked participants to judge the naturalness of the motion of the occluded ball. They were unaware that the visible ball always fell under 1 g.

For each condition in both Phase 1 and Phase 2, participants completed a block of six repeated trials. Within these blocks, the occlusion box appeared three times on the left side and three times on the right side of the table. In total, each phase consisted of 36 trials. The order of the blocks was randomized and the positions of the occlusion box were balanced between left and right within blocks.

At the end of the VR session, participants judged the naturalness of experimental conditions involving two visible balls. Following the VR session, we interviewed the participants to gain their reflections about their experience and information on their prior physics knowledge. In this paper, we focus on the quantitative results from the two VR phases.

Data Analysis

We quantified participants' motor catching using time error (TE), defined as the difference between the time at which the participants attempted to catch the ball and the expected arrival time of the ball at the catching height under the corresponding acceleration. Specifically, we separately looked at the absolute value of TE, which represented the success of the catch effort, and the sign (direction) of TE, which represented whether the catch was early or late. For the judgment of naturalness, we recorded the answers as "natural" or "unnatural".

We analyzed the data using Bayesian multilevel generalized linear models implemented in *brms* package for R (Bürkner, 2017). The Bayesian framework was chosen to allow for direct probabilistic inference of planned contrasts while accounting for participant-level variability through hierarchical random

effects. All models were estimated using Hamiltonian Monte Carlo sampling with the Stan platform. We ran four independent Markov chains with 4,000 iterations each (including 2,000 warm-up samples), totaling 8,000 post-warm-up draws. Convergence was monitored via the $\hat{R}$ diagnostic, where all values were 1.00.

The fixed effects included simulated acceleration, ball material, and their interaction. To capture the learning effect and the baseline reaction time, we included a standardized global trial index, a standardized within-block repetition index, and a standardized baseline reaction time measured in the experiment as covariates. To account for individual differences in sensitivity to experimental manipulations, participant-specific random intercepts and random slopes for acceleration, ball material, and their interaction were included. For the absolute TE, because they were strictly non-negative and right-skewed, a Hurdle-Gamma distribution with a log link was used. For the direction of TE and the naturalness judgments, due to the binary nature, we used a Bernoulli distribution with a logit link to model them.

For the regression coefficients of the main model, we report the posterior medians as point estimates (β) alongside 95% Credible Intervals (CI). For hypothesis testing, We computed Estimated Marginal Means (EMMs) from the main Bayesian model, marginalizing over the corresponding factor (acceleration or ball material). The planned contrasts between the hypothesized levels were evaluated using posterior samples. We report posterior medians as contrast estimates using point estimates (β) with 95% Highest Posterior Density (HPD) intervals, which provide the most probable range of the posterior distribution for these effects. We always set 1 g, light ball, and "unnatural" judgment, as the reference for the models.

Results

We removed 1.3% of trials (14 in Phase 1, 10 in Phase 2) because they were catching attempts after the judgment phase started (1 second after the ball landed). The descriptive statistics for Phase 1 and Phase 2 are shown in Table 1.

Absolute Timing Error

Phase 1 (Single Occluded Ball, Figure 2A) Across conditions, the absolute TE of the motor catchings was $M = 139.7$ ms, $SD = 107.6$ ms, suggesting a substantial deviation from successful catching. The Bayesian model for absolute TE showed that several control variables contributed to accuracy variability. The posterior distribution for the baseline reaction time was shifted towards positive values ($\beta = 0.31$, 95% CI [0.06, 0.57]), indicating that participants with slower baseline responses exhibited higher absolute TE. Furthermore, the posterior estimate for the effect of the global trial was slightly below zero ($\beta = -0.06$, 95% CI [-0.12, -0.00]), suggesting an improvement in timing accuracy as the session progressed. Individual differences in catching accuracy were moderate, with variability in the reference level

Table 1: Descriptive statistics for Phase 1 and Phase 2. Values represent the Mean (*SD*) of participant-level averages. TE(Abs) is the absolute timing error (measured in milliseconds). TE(Dir) is the direction of timing error (percentage of late response). Judgment is the naturalness judgment (percentage of “natural” responses).

Condition	TE(Abs)/ms	TE(Dir)/%	Judgment/%
Phase 1			
0.5 g, Light	140.7 (100.3)	77.3 (26.3)	38.7 (28.8)
0.5 g, Heavy	151.7 (96.9)	81.3 (27.4)	35.3 (38.6)
1 g, Light	148.3 (106.7)	85.2 (27.4)	73.9 (32.3)
1 g, Heavy	144.9 (141.6)	76.7 (30.7)	67.1 (31.3)
1.5 g, Light	135.5 (111.5)	85.2 (25.2)	59.7 (36.6)
1.5 g, Heavy	117.4 (89.9)	80.9 (27.9)	65.2 (36.2)
Phase 2			
0.5 g, Light	157.3 (119)	88.5 (19.7)	38.7 (42.7)
0.5 g, Heavy	184.7 (139.4)	85.3 (21.1)	15 (26.8)
0.8 g, Light	135.6 (94.7)	82 (29.2)	67.1 (35.6)
0.8 g, Heavy	141.6 (95.2)	78.7 (27.4)	38.7 (35.9)
1 g, Light	131.6 (98.9)	81.3 (30.6)	70 (33.3)
1 g, Heavy	134.8 (96.4)	73.3 (34.7)	58.7 (38.2)

($\sigma = 0.71$, 95% CI [0.52, 0.99]) exceeding that of condition-specific slopes (all $\sigma \leq 0.37$).

Regarding acceleration factor, the planned contrast between the acceleration conditions 0.5 g and 1 g yielded a posterior estimate centered close to zero ($\beta = 0.06$, 95% HPD [-0.10, 0.21]), indicating substantial overlap with no effect. In contrast, the posterior distribution for the comparison between 1.5 g and 1 g was shifted towards smaller absolute TE values ($\beta = -0.15$, 95% HPD [-0.30, -0.01]), suggesting an increased timing accuracy in this higher gravity. These results did not support **H1.1** about simulated acceleration.

Regarding ball material factor, the planned contrast was centered near zero ($\beta = -0.01$, 95% HPD [-0.17, 0.14]). This supported the second part of **H1.1** in that ball material did not substantially bias motor catching accuracy in this phase.

Phase 2 (Two-Ball Comparison, Figure 3A) Across conditions, the absolute TE was $M = 147.6$ ms, $SD = 108.3$ ms, again deviated from successful catchings. The Bayesian model identified baseline reaction time as the only strong contributor to absolute TE, with the posterior shifted towards positive values ($\beta = 0.32$, 95% CI [0.09, 0.55]). Individual differences in catching accuracy were moderate, with variability in the reference level ($\sigma = 0.62$, 95% CI [0.44, 0.85]) exceeding that of condition-specific slopes (all $\sigma \leq 0.39$).

Regarding acceleration factor, planned contrast between 0.5 g and 1 g was shifted towards higher absolute TE values ($\beta = 0.24$, 95% HPD [0.08, 0.38]), whereas the contrast

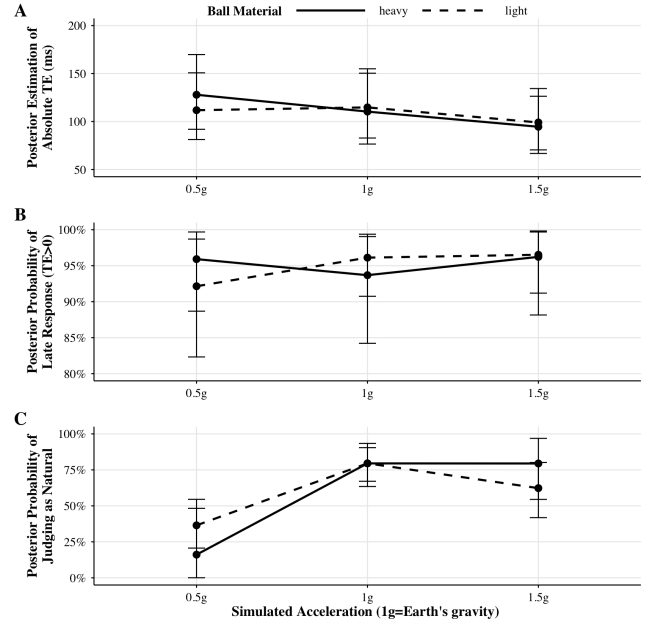


Figure 2: **Phase 1 (single occluded ball)** Bayesian posterior estimation of: A) absolute timing error; B) probability of late response; C) probability of judging the motion of the occluded ball as natural.

between 0.8 g and 1 g overlapped zero ($\beta = 0.08$, 95% HPD [-0.07, 0.23]). This might have partially supported **H2.1**, indicating that participants were able to perform accurately near earth gravity with another ball’s presence.

Regarding ball material factor, the planned contrast centered near zero ($\beta = -0.08$, 95% HPD [-0.21, 0.05]). This supported the second part of **H2.1** in that ball material did not substantially bias motor catching accuracy in this phase.

Timing Error Direction

Phase 1 (Single Occluded Ball, Figure 2B) Across conditions, there were $M = 81.1\%$, $SD = 27.2\%$ positive TE records, suggesting that most of the attempts were late. The Bayesian model of the TE direction showed posterior evidence for an interaction between the 0.5 g condition and the heavy-ball condition, calling for further analysis. Regarding control variables, the baseline reaction time again showed a strong association with the outcome ($\beta = 2.14$, 95% CI [0.89, 3.67]), indicating that individual response latencies strongly influenced how many late responses they performed. The variability between participants in the reference level ($\sigma = 2.00$, 95% CI [1.17, 3.15]) indicated substantial individual differences in the number of late attempts.

Planned contrasts on TE direction showed substantial overlap with zero for comparisons between 1 g and 0.5 g ($\beta = 0.15$, 95% HPD [-0.62, 0.81]), between 1.5 g and 1 g ($\beta = 0.33$, 95% HPD [-0.47, 1.30]), and between light and heavy ($\beta = -0.01$, 95% HPD [-0.91, 0.77]). Post-hoc simple effect con-

trasts showed a broad overlap with zero at all acceleration levels and ball materials (all 95% HPD included zero), suggesting that the interaction between the 0.5 g condition and the heavy-ball condition was driven by relative differences across conditions rather than effect at a single level. For light balls, reducing the acceleration from 1 g to 0.5 g shifted the posterior towards earlier responses ($\beta = -0.74$, 95% HPD $[-1.58, 0.13]$, Odds Ratio Upper HPD = 1.00), while for heavy balls, the same acceleration change did not have such an effect ($\beta = 0.45$, 95% HPD $[-0.54, 1.57]$). Specifically, at 1 g, the contrast between heavy and light balls overlapped with zero ($\beta = -0.53$, 95% HPD $[-1.58, 0.47]$). Together, the results partially supported **H1.2**, indicating that ball material did not substantially bias motor catching latency under 1 g and 1.5 g.

Phase 2 (Two-Ball Comparison, Figure 3B) There were $M = 81.5\%$, $SD = 27.6\%$ positive TE records, again indicated late attempts. Bayesian model indicated both baseline reaction time ($\beta = 1.4$, 95% CI $[0.36, 2.56]$) and within-block repetition ($\beta = 0.49$, 95% CI $[0.23, 0.74]$) contributed to TE direction. The latter suggested that within a block of trials, participants tended to give more late responses. The variability between-participants in the reference level ($\sigma = 2.42$, 95% CI $[1.55, 3.64]$) indicated substantial individual differences in the number of late attempts.

The planned contrasts overlapped zero for all the hypothesized comparisons. Post-hoc simple-effect contrasts at 1 g showed a posterior shifted slightly towards positive values ($\beta = 0.86$, 95% HPD $[0.00035, 1.77]$), consistent with a tendency towards more late responses for light balls than for heavy balls at Earth's gravity, though the evidence was modest. Together, the results partially supported **H2.2**, indicating that ball material did not substantially bias motor catching latency under 0.5 g and 0.8 g.

Naturalness Judgment

Phase 1 (Single Occluded Ball, Figure 2C) The Bayesian model for the naturalness judgment showed that the posterior distribution for the global trial number shifted towards negative values ($\beta = -0.64$, 95% CI $[-0.97, -0.33]$), suggesting that as the experiment progressed, participants generally became more critical in their naturalness ratings. Participants differed strongly in judgments, especially for heavy ball under 0.5 g ($\sigma = 4.75$, 95% CI $[2.82, 7.72]$, $\sigma = 1.35$ in the reference level), indicating a high heterogeneity in how people perceived the naturalness of ball motions.

Regarding acceleration factor, the planned contrasts showed a clear shift favoring higher naturalness ratings for 1 g relative to 0.5 g ($\beta = 2.50$, 95% HPD $[1.20, 4.04]$). In comparison, the difference between 1 g and 1.5 g was much smaller and its uncertainty interval overlapped zero ($\beta = 0.43$, 95% HPD $[-0.39, 1.25]$). These results partially supported **H1.3**, indicating that participants were able to differentiate substantially lower simulated acceleration and Earth gravity acceleration.

Regarding ball material factor, the pairwise comparison between light and heavy balls was centered near zero ($\beta = 0.10$,

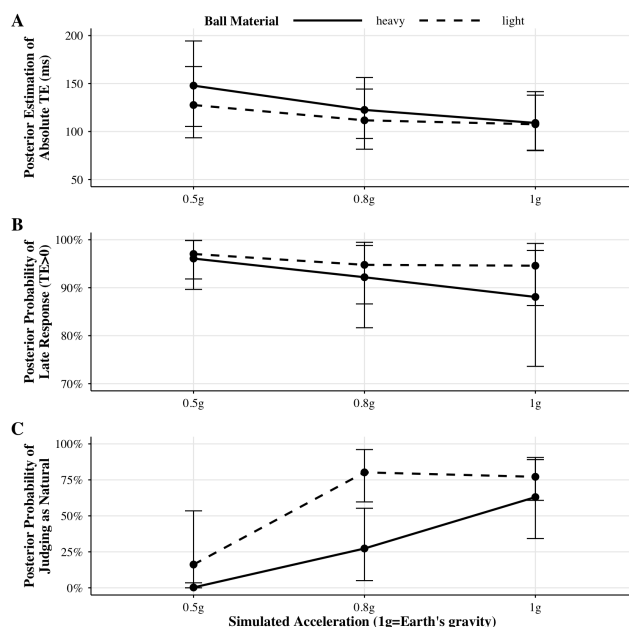


Figure 3: **Phase 2 (two-ball comparison)** Bayesian posterior estimation of: A) absolute timing error; B) probability of late response; C) probability of judging the motion of the occluded ball as natural.

95% HPD $[-0.88, 1.16]$, suggesting that the participants considered the motion of the heavy and light balls to be similarly natural. This supported the second component of **H1.3** in that ball material did not systematically bias naturalness judgment in this phase.

Phase 2 (Two-Ball Comparison, Figure 3C) Bayesian model again showed a global trial effect shifted towards negative values ($\beta = -0.37$, 95% CI $[-0.72, -0.02]$), indicating that participants became generally more critical over time. Interestingly, the main model revealed strong interaction effects, motivating follow-up simple-effects comparisons. Between-participants variability was extremely large across conditions (σ ranging from 1.5 to 5.81), indicating that participants differed greatly in their perceptual judgments.

For acceleration factor, occluded balls falling under 1 g were judged to be more natural than under 0.5 g ($\beta = 4.73$, 95% HPD $[2.32, 7.77]$), but showed little difference compared to 0.8 g ($\beta = 0.68$, 95% HPD $[-0.33, 1.67]$).

For ball material factor, conditions where light balls were occluded (with heavy balls always visible and falling at 1 g) were systematically judged to be more natural than those with heavy balls occluded ($\beta = 2.49$, 95% HPD $[0.93, 4.51]$). This aligned with the persistence of the inaccurate conception that heavier things fall faster, and thus the scenario of the light ball falling under 0.8 g was judged to be the most natural. In response to the open hypothesis **H2.3**, despite being explicitly instructed to base judgments only on the occluded ball, the inaccurate conception prevailed.

Discussion

Overall, the results indicate that motor-level knowledge of Earth's gravity is more robust than perceptual judgment. We observed that, in both phases, motor catching showed similar accuracy for heavy and light balls. However, naturalness judgments, while being able to distinguish slow accelerations correctly in Phase 1, shifted towards heuristic of "heavier objects fall faster" in Phase 2. Together, these findings suggest that motor-based representations are more stable and aligned with canonical physics than perceptual ones.

Phase 1 - Single Occluded Ball

The Phase 1 results regarding acceleration conditions were partially aligned with canonical physics. Absolute TE analysis suggested that catch were most successful in 1.5 g, which might be due to the greater effort in reaction to faster dynamics. In contrast, TE direction analysis showed an interaction effect between 0.5 g and the heavy-ball condition. This might be due to participants' over-compensating at 0.5 g for light balls, which gave early catches, while heavy balls were more robust across accelerations. Naturalness judgments showed clearer alignment with canonical physics. The 1 g condition was judged as most natural and 0.5 g unnatural. People also tended to judge the motion at 1.5 g as natural, which might be due to capacity limit in perceiving fast movements.

Regarding ball material factor, both motor catchings and naturalness judgments aligned with canonical physics. This contrasted with Vicovaro et al. (2019), who demonstrated a "heavy-fast, light-slow" heuristic. Two key methodological differences may account for this discrepancy. First, the relative mass difference in our study was smaller (tennis balls as the light condition rather than very light polystyrene spheres). Second, our design created a direct interaction with the ball by having participants immersed in VR and stand close under the falling ball(s). This made ball-catching more action-oriented and embodied (Zago & Lacquaniti, 2005), instead of distant and perceptual. Taken together, these findings provided a useful baseline for interpreting the Phase 2 results.

Phase 2 - Two-Ball Comparison

Introducing an additional contextual cue, i.e., a visible ball of different material falling at 1 g, could be both helpful and distracting. The results of absolute TE showed that people were more successful in catching balls falling under 1 g and 0.8 g regardless of the ball materials. Most, except one of the participants, reported in the interview that they primarily focused on the occluded ball, which suggested that active predictions were still involved. Together, the results of absolute TE may reflect ongoing engagement of the internal gravity prior at the motor level near Earth gravity.

However, this conclusion should be treated with caution. The post-hoc analysis of the TE direction revealed more late responses for light balls compared to heavy balls under 1 g. This suggested that naïve physics may partially influence motor catching under this condition.

At the perceptual level, Phase 2 revealed a clearer systematic tendency consistent with naïve physics heuristics. Participants judged scenarios in which light balls fell at 0.8 g as most natural, reflecting strong mass-based expectations. One possible explanation might be that people have limited capacity to process multiple visual or imagined objects (Balaban & Ullman, 2025) and thus resort to heuristics in perceptual tasks. At 0.5 g, all conditions were considered unnatural. This indicated that the internal model of Earth's gravity remained the dominant factor when the deviations were large.

Learning Effect and Baseline Reaction Time

Across motor catching results, the baseline reaction time was consistently identified as a strong predictor of catching performance. This suggested substantial individual differences in temporal processing. Regarding judgments, participants became progressively more strict or critical over time in both phases, which might explain the general low ratio of judging ball motions as natural in both phases. Because we did not provide feedback for participants' answers, this tendency might come from their own meta-cognitive reflections on the experimental designs or simply fatigue over repeated trials.

Interestingly, although participants did not receive explicit feedback on catching success, a modest practice effect was observed in Phase 1 absolute TE ($\beta = -0.06$, 95% CI $[-0.12, -0.00]$). This indicated that participants were able to improve catching accuracy based solely on visual information from the visible part of the trajectory combined with their internal physics model, reflecting implicit learning and adaptation to the designed tasks.

Limitations and Future Direction

Several limitations of the current study should be noted. First, the experimental settings, including ball materials and acceleration levels, were selected based on prior literature and standard classroom teaching practices, to maintain practical relevance to naïve physics. Future work could benefit from exploring a wider and more continuous range of conditions to better characterize general patterns. Second, we cannot be certain that all participants employed the hypothesized motor catching strategy. Triangulating these findings with qualitative data and computational modeling, or employing eye-tracking for future studies, will be valuable for deeper insight. Moreover, future studies can look at computational modelings to more directly characterize the behavior we found here, for example, by comparing linear models held by naïve physics with quadratic models derived from Newtonian physics.

Our findings also suggest promising directions for learning interventions. Naïve physics remains a significant challenge in formal education, where the persistence of embodied knowledge might be the key reason as well as the solution. Through carefully designing learning contexts informed by embodied cognition research, learners might selectively rely more on the accurate internal model of Earth's gravity in the motor level to achieve effective learning outcomes.

Acknowledgments

We are grateful to the Decision Science Laboratory in ETH Zurich for their assistance in recruiting participants for this study. We also thank all participants for their time and for providing valuable reflections and feedback. Maria Ioanna Magkouta, Xiaoxuan Li, Laura Bock, and Ipek Paksoy contributed as pilot users and offered helpful input on the study design and protocol, which we greatly appreciate. Special thanks go to Neals Chatain for insightful suggestions that substantially improved the usability and overall design.

We used AI during manuscript preparation for limited support tasks, including brainstorming of theoretical alternatives, Bayesian model debugging, refinement of Bayesian language reporting, grammar and spelling correction, and enhancement of the resolution of Figure 1. We take full responsibility for the conception, methodology, implementation, analysis, interpretation, and final content of this study.

All relevant data reported in this paper can be found in OSF project: Investigating the Manifestation of Naïve Physics at the Embodied Level Through Motor and Judgment Tasks in Virtual Reality.

References

- Balaban, H., & Ullman, T. D. (2025). The capacity limits of moving objects in the imagination. *Nature Communications*, 16(1), 5899. <https://doi.org/10.1038/s41467-025-61021-8>
- Belousov, B., Neumann, G., Rothkopf, C. A., & Peters, J. R. (2016). Catching heuristics are optimal control policies. *Advances in Neural Information Processing Systems*, 29.
- Bürkner, P.-C. (2017). **Brms** : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2-3), 105–225. <https://doi.org/10.1080/07370008.1985.9649008>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Jörges, B., & López-Moliner, J. (2017). Gravity as a Strong Prior: Implications for Perception and Action. *Frontiers in Human Neuroscience*, 11, 203. <https://doi.org/10.3389/fnhum.2017.00203>
- La Scaleia, B., Ceccarelli, F., Lacquaniti, F., & Zago, M. (2020). Visuomotor Interactions and Perceptual Judgments in Virtual Reality Simulating Different Levels of Gravity. *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.00076>
- La Scaleia, B., Zago, M., & Lacquaniti, F. (2015). Hand interception of occluded motion in humans: A test of model-based vs. on-line control. *Journal of Neurophysiology*, 114(3), 1577–1592. <https://doi.org/10.1152/jn.00475.2015>
- Neupärtl, N., Tatai, F., & Rothkopf, C. A. (2020). Intuitive physical reasoning about objects' masses transfers to a visuomotor decision task consistent with Newtonian physics. *PLOS Computational Biology*, 16(10), e1007730. <https://doi.org/10.1371/journal.pcbi.1007730>
- Smith, K. A., Hamrick, J. B., Sanborn, A. N., Battaglia, P. W., Gerstenberg, T., Ullman, T. D., & Tenenbaum, J. B. (2024, November). Intuitive physics as probabilistic inference. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian models of cognition : Reverse engineering the mind* (pp. 369–393). MIT Press.
- Vicovaro, M. (2021). Intuitive physics and cognitive algebra: A review. *European Review of Applied Psychology*, 71(5), 100610. <https://doi.org/10.1016/j.erap.2020.100610>
- Vicovaro, M. (2023). Grounding Intuitive Physics in Perceptual Experience. *Journal of Intelligence*, 11(10), 187. <https://doi.org/10.3390/jintelligence11100187>
- Vicovaro, M., Noventa, S., & Battaglini, L. (2019). Intuitive physics of gravitational motion as shown by perceptual judgment and prediction-motion tasks. *Acta Psychologica*, 194, 51–62. <https://doi.org/10.1016/j.actpsy.2019.02.001>
- Zago, M., Bosco, G., Maffei, V., Iosa, M., Ivanenko, Y. P., & Lacquaniti, F. (2004). Internal Models of Target Motion: Expected Dynamics Overrides Measured Kinematics in Timing Manual Interceptions. *Journal of Neurophysiology*, 91(4), 1620–1634. <https://doi.org/10.1152/jn.00862.2003>
- Zago, M., & Lacquaniti, F. (2005). Cognitive, perceptual and action-oriented representations of falling objects. *Neuropsychologia*, 43(2), 178–188. <https://doi.org/10.1016/j.neuropsychologia.2004.11.005>
- Zhao, H., & Warren, W. H. (2015). On-line and model-based approaches to the visual control of action. *Vision Research*, 110, 190–202. <https://doi.org/10.1016/j.visres.2014.10.008>