UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Probabilistic internal physics models guide judgments about object dynamics

Permalink
https://escholarship.org/uc/item/3ns68302

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Hamrick, Jessica
Battaglia, Peter
Tenenbaum, Josh

Publication Date
2011

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Internal physics models guide probabilistic judgments about object dynamics

Jessica Hamrick (jhamrick@mit.edu), Peter Battaglia (pbatt@mit.edu), Joshua B. Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences, MIT. Cambridge, MA 02139

Abstract

Many human activities require precise judgments about the physical properties and dynamics of multiple objects. Classic work suggests that people's intuitive models of physics are relatively poor and error-prone, based on highly simplified heuristics that apply only in special cases or incorrect general principles (e.g., impetus instead of momentum). These conclusions seem at odds with the breadth and sophistication of naive physical reasoning in real-world situations. Our work measures the boundaries of people's physical reasoning and tests the richness of intuitive physics knowledge in more complex scenes. We asked participants to make quantitative judgments about stability and other physical properties of virtual 3D towers. We found their judgments correlated highly with a model observer that uses simulations based on realistic physical dynamics and sampling-based approximate probabilistic inference to efficiently and accurately estimate these properties. Several alternative heuristic accounts provide substantially worse fits. **Keywords:** intuitive physics, dynamics, perception, model

Introduction

Intuitive physics is a core domain of common-sense reasoning, developing early in infancy and central in adult thought (Baillargeon, 2007). Yet, despite decades of research, there is no consensus on certain basic questions: What kinds of internal models of the physical world do human minds build? How rich and physically accurate are they? How is intuitive physical knowledge represented or used to guide physical judgments?

The kinds of judgments we consider are those necessary to navigate, interact with, and constructively modify real-world physical environments. Consider the towers of blocks shown in Fig. 1. How stable are these configurations, or how likely are they to fall? If they fall, in what direction will the blocks scatter? Where could a block be added or removed from the tower to significantly alter the configuration's stability? People make such judgments with relative ease, yet the literature on intuitive physics has little to say about how they do so.

Classic research focused on the limits of human physical reasoning. One line of work argued that people's understanding of simple object trajectories moving under inertial dynamics was biased away from the true Newtonian dynamics, towards a more "Aristotelian" or "impetus" kinematic theory (Caramazza, McCloskey, & Green, 1981; McCloskey, 1983), yet no precise model of an intuitive impetus theory was developed. Studies of how people judge relative masses in two-body collisions concluded that humans are limited to making physical judgments based on simple heuristics, or become confused in tasks requiring attention to more than one dimension of a dynamic scene (Todd & Jr., 1982; Gilden & Proffitt, 1989a, 1989b, 1994). Neither the impetus accounts nor the simple one-dimensional heuristic accounts attempted to explain how people might reason about complex scenes such as Fig. 1, or gave any basis to think people might reason about them with a high degree of accuracy.

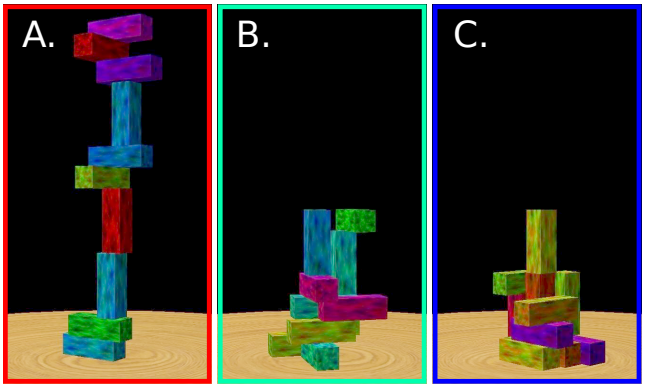

Figure 1: **Three towers of varying height and stability.** Each tower (A, B, C) corresponds to a colored point in Fig. 3. A is clearly unstable, C clearly stable, while B (matched in height to C) is less obvious.

Here we argue for a different view. We hypothesize that humans can make physical judgments using an internal generative model that approximates the principles of Newtonian mechanics applied to three-dimensional solid bodies. They use this model to forward-simulate future outcomes given beliefs about the world state, and make judgments based on the outcomes of these simulations. We believe that only by positing such rich internal models can we explain how people are able to perform complex everyday tasks like constructing and predicting properties of stacks of objects, balancing or stabilizing precariously arranged objects, or intercepting or avoiding multiple moving, interacting objects.

The physical laws of the internal models we propose are essentially deterministic, but people's judgments are probabilistic. Capturing that probabilistic structure is crucial for predicting human judgments precisely and explaining how intuitive physical reasoning successfully guides adaptive behavior, decision-making and planning in the world. We can incorporate uncertainty in several ways. Objects' positions and velocities and their key physical properties (e.g., mass, coefficients of friction) may only be inferred with limited precision from perceptual input. People may also be uncertain about the underlying physical dynamics, or may consider the action of unobserved or unknown exogenous forces on the objects in the scene (e.g., a gust of wind, or someone bumping into the table). We can represent these sources of uncertainty in terms of probability distributions over the values of state variables, parameters or latent forces in the deterministic physical model. By representing these distributions approximately in terms of small sets of samples, uncertainty can be propagated through the model's physical dynamics using only analog mental simulations. Thus a resource-bounded observer

can make appropriate predictions of future outcomes without complex probabilistic calculations. Even though these simulations may approximate reality only roughly, and with large numbers of objects may only be sustainable for brief durations, they can still be sufficient to make useful judgments about complex scenes on short time-scales. Our goal in the present work is to quantitatively compare several such judgments – mainly degree of stability and direction of fall – across human observers, variants of our model, and plausible alternative accounts based on simple, model-free heuristics.

Several recent lines of research suggest approximate Newtonian principles underlie human judgments about dynamics and stability (Zago & Lacquaniti, 2005; Fleming, Barnett-Cowan, & Bülthoff, 2010). Perhaps closest to this study is the work of Sanborn, Mansinghka, and Griffiths (2009), who showed that perception of relative mass from two-body collisions is well-modeled as Bayesian inference in a generative model with Newtonian dynamics. Like us, they frame intuitive physics as a kind of probabilistic Newtonian mechanics in which uncertainty about latent variables gives rise to uncertain predictions of physical outcomes. The main innovation of our work is to capture physical knowledge with a three-dimensional and realistic object-based physics simulation, and to implement probabilistic inference using sample-based approximations; Sanborn et al. used a simpler Bayesian network that was specialized to the case of two point masses colliding in one dimension. Our more general framing allows us to test whether and how a probabilistic-Newtonian framework can scale up to explain intuitive physical reasoning in complex scenes such as Fig. 1.

Model

We frame human physical judgments using a probabilistic model observer (Fig. 2) that combines three components: *perception*, *physical reasoning*, and *decision*. The perception component defines a mapping from input images to internal beliefs about the states of objects in a scene. The physical reasoning component describes how internal physics knowledge is used to predict future object states. The decision component describes how these predicted states are used to produce a desired property judgment. Uncertainty may enter into any or all of these components. For simplicity in this paper we have modeled uncertainty only in the perception component, assuming that observers compute a noisy representation of objects' positions in the three-dimensional scene.[1] When the noise variance σ^2 equals 0, the model's outputs are deterministic and correspond to physical ground-truth judgments. We investigate how the addition of noise, along with several other assumptions about the limitations of realistic observers, might fit human judgments better than the perfect predictions of physical ground-truth.

[1] Similar noise distributions applied to objects' states could also represent other sources of uncertainty, such as unknown latent forces in the world that might perturb the objects' state or uncertainty about specific physical dynamics. Here we do not distinguish these sources of uncertainty but leave this as a question for future work.

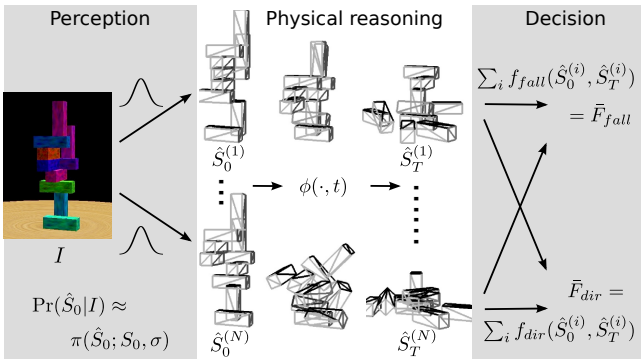

Figure 2: **Model schematic.** Our model has 3 components, perception, physical reasoning, and decision. During perception, an uncertain belief about the tower is inferred from an image. During physical reasoning, tower samples are drawn from this belief distribution, and a physical simulation is applied to each. To make decisions about physical properties, the simulation outcomes are evaluated and averaged.

Our specific experimental focus is on judgments about dynamic events with towers of blocks (Fig. 1), so the relevant object states S_t are the locations and orientations of all blocks in the tower at time t. The effect of Newtonian physics over time on the tower, which includes gravitational forces, elastic collisions, transfer of energy, is represented by the function $\phi(\cdot)$, which inputs S_t and temporal duration T, and outputs the subsequent state $S_{t+T} = \phi(S_t, T)$. Our implementation of physical predictions used the Open Dynamics Engine (ODE, www.ode.org), a standard computer physics engine, which, critically, allows precise simulation of rigid-body dynamics with momentous collisions. The physical properties the observer wishes to predict are represented as predicates over the current and future tower states, $f(S_t, S_{t+T})$. We examine two kinds of judgments about the future state S_T of a tower first observed at $t = 0$:

1. What proportion of the tower will fall, $f_{fall}(S_0, S_T)$?
2. In what direction will the tower fall, $f_{dir}(S_0, S_T)$?

We quantify degree of stability as the proportion of a tower that remains standing following the application of physics for duration T. This definition matches the objective notion that a tower that entirely collapses should be judged less stable than one for which a single block teeters off.

Observer model Predicting a physical tower property means computing $f(S_0, S_T) = f(S_0, \phi(S_0, T))$. In principle, deterministic physics implies that knowledge of S_0 and $\phi(\cdot)$ is sufficient to predict future physical properties perfectly. However, a realistic observer does not have direct access to tower states, S_0, so must rely on uncertain perceptual inferences to draw beliefs about the tower. The observer forms beliefs about S_0 conditioned on an image, I, and represents these beliefs, $\hat{S}_0$, with the distribution, $\Pr(\hat{S}_0|I)$.

Applying physics to the inferred initial state $\hat{S}_0$ induces a future state $\hat{S}_T = \phi(\hat{S}_0, T)$ with distribution $\Pr(\hat{S}_T|I)$. As above, predicting a physical property means computing $f(\hat{S}_0, \hat{S}_T) = f(\hat{S}_0, \phi(\hat{S}_0, T))$. To make decisions about physi-

cal properties, the observer computes the expectation:

$$\mathrm{E}[f(\hat{S}_0,\phi(\hat{S}_0,T)]_I = \int f(\hat{S}_0,\phi(\hat{S}_0,T))\Pr(\hat{S}_0|I)d\hat{S}_0 \qquad (1)$$

which represents the model observer's estimate of the physical property f given I, integrating out the perceptual uncertainty in the initial state $\hat{S}_0$.

Approximating physical inference We model $\Pr(\hat{S}_0|I)$ in a way that reflects perceptual uncertainty and the principle that objects cannot interpenetrate, without committing to particular assumptions about perceptual inference or representations. We approximate $\Pr(\hat{S}_0|I) \approx \pi(\hat{S}_0;S_0,\sigma)$, where $\pi(\cdot)$ is a composition of two terms: a set of independent Gaussian distributions for each block's x and y positions, with variance σ^2 and mean centered on the corresponding block positions in the true world state S_0, followed by a deterministic transform that prevents blocks from interpenetrating. This is clearly a simplified approximation to the observer's perceptual distribution, but it serves as a reasonable starting point, and a place where more sophisticated vision models can be interfaced with our approach in future work.

We approximate Eqn. 1 though a Monte Carlo simulation procedure that draws N "perceptual" samples $\hat{S}_0^{(1,\ldots,N)} \sim \pi(\hat{S}_0;S_0,\sigma)$, simulates physics forward on each sample to time T, and computes the mean value for physical property f across their final states:

$$\bar{F} = \frac{1}{N}\sum_{i=0}^{N} f(\hat{S}_0^{(i)},\phi(\hat{S}_0^{(i)},T)). \qquad (2)$$

Fig. 2 illustrates this computation for one unstable tower.

The simulations depend on 3 parameters: movement threshold, $M = 0.1$m, the distance a block must move before it is considered to have "fallen"; timescale of the simulation, $T = 2000$ms, defined above; and σ, the perceptual uncertainty. When $\sigma = 0$, the $\bar{F}$ predictions deterministically depend on S_0, and represent the ground truth physical property. When a simulation ends, f_{fall} is measured as the number of blocks that have *not* fallen, and f_{dir} is measured as the angle of the mean position of those blocks that have fallen. Pilot analyses determined that our results were insensitive to changes nearby the chosen M and T values, while σ had a substantial effect in comparison with peoples' judgments. We found $\sigma = 0.05$ to provide reasonable fits to all conditions but we explore the effects of varying σ below (Fig. 3).

Heuristics To evaluate alternative explanations of humans' judgments, we test whether several heuristics, i.e. the tower's height, skew, or top-heaviness, may account for their responses. Representing the tower's center of mass in cylindrical coordinates (ρ,θ,z), where the z-axis is vertical, the magnitude of the tower skew is equal to ρ, and the direction of the tower skew is equal to θ. Similarly, the tower's top-heaviness is $\frac{z}{h}$, where h is the height of the tower. We examine the following heuristics:

H_h - height H_θ - skew direction
H_ρ - skew magnitude H_z - top-heaviness

The H_h, H_ρ, and H_z heuristic measures are inversely proportional to physical stability, e.g. tall towers tend to be less stable. For clarity, we negated their values to instead reflect a proportional relationship; this does not affect the correlations beyond changing their sign.

Experiments

General methods Participants were recruited from the MIT BCS human subject pool with informed consent, and were compensated $10/hr. Stimuli were viewed on a standard LCD monitor from an approximate distance of 0.6 meters. All stimuli were rendered in 3D using Panda3D (www.panda3d.org), and physics simulations were computed at 1500Hz using the ODE physics engine.

All trials had 3 phases: *stimulus*, *response*, and *feedback*. All phases depicted a 3D scene that contained a circular 3m radius "ground disk", and a tower of 10 colored blocks (each block $20 \times 20 \times 60$ cm) placed at the disk's center. In the stability experiments, the ground was textured with a wood grain pattern (Fig. 1; in the direction experiment, the ground texture was a visual indicator of the responses (Fig. 5A).

The *stimulus* phase was 3500ms, during which participants passively viewed the tower of blocks from a camera that orbited the tower at 60°/s (total rotation of 180°), so all sides of the tower were made visible to the participant. The camera radius was 7m (stability experiments) or 9m (direction experiment), and the field of view was 40°. No physics simulations were applied, so the only image motion was due to camera rotation. After 3000ms, a cylindrical "occluder" descended vertically over 500ms and rested on the ground plane to obscure the tower from view; this ended the stimulus phase.

The *response* phase then began immediately, and was not limited in time, but ended once the participant depressed a response key.

The *feedback* phase began immediately after the response phase and lasted 2000ms, except in the no-feedback condition (described below), in which it was skipped entirely. During feedback, the occluder ascended vertically out of sight over 500ms, and for the remaining 1500ms the tower was visible. During feedback, physics was turned "on", which meant gravitational acceleration of -9.8m/s was applied to the tower's blocks. Physics caused some towers to collapse, while some remained standing – this indication gave participants feedback about the accuracy of their response. Additionally, in the "stability experiments" (below) the ground pattern was shaded red if the tower fell, and green if the tower remained standing. After 2000ms, the tower was removed from the scene and an inter-trial interval of 500ms was presented, after which the next trial's stimulus phase began.

Before the actual experiment, participants performed an identical unrecorded 20-trial "training" session, with feedback, to be familiarized with the task. All recorded conditions (except *same-height*, see below) were composed of 360 trials of 6 subsessions (60 trials per subsession), with 60 differ-

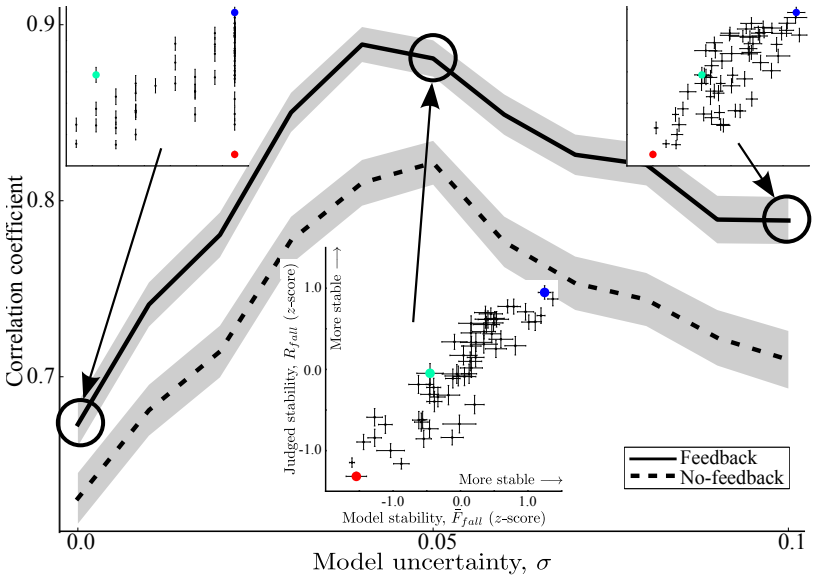

Figure 3: **Stability experiments results.** The lines represent the correlation coefficients (y-axis) between model stability predictions, $\bar{F}_{fall}$, and human judged stabilities, R_{fall}, as a function of the model uncertainty, σ, (x-axis) for the feedback (solid line) and no-feedback (dashed lines) conditions (1 SE bars). The insets depict scatterplots between $\bar{F}_{fall}$ (x) and R_{fall} (y) in the feedback condition for 3 levels of σ (0.0, 0.05, 0.1), where each point is a model/human mean stability score (with 1 SE bars). The colored points show the towers depicted in Fig. 1.

ent towers per subsession randomized in order and repeated across subsessions. Also, each tower's blocks' colors were randomized across different repetitions, as well as the starting angle that the camera faced the tower. Participants were not told that towers were repeated.

Tower stimuli Each tower was constructed by a stochastic process in which 10 blocks were sequentially given positions and orientations that resulted in a stable or unstable stack of blocks. Specifically, each block's placement satisfied two constraints: 1) its center must reside within the 60×60 cm length and width of the tower, and 2) it must be "locally stable", meaning that the block is supported by the blocks beneath it (however adding more blocks on top could cause it to fall). We then scored each tower's "true stability" by simulating physics (i.e. gravity) and measuring whether any blocks in the tower fell within 2000ms – those that had blocks fall were deemed unstable, and those from which no blocks fell deemed stable.

Stability experiments The first set of experiments asked participants to judge whether towers were stable or unstable. The 60 tower stimuli were randomly selected such that 33 were stable and 27 were unstable. Three variations of this experiment were run: *feedback* ($n = 10$ participants), *no-feedback* ($n = 10$), and *same-height* ($n = 9$). The *feedback* condition proceeded exactly as described above. The *no-feedback* condition was identical, except the feedback phase of each trial was omitted. The *same-height* condition was similar to the feedback condition, except that a different set of 108 towers were used as stimuli, where each had the same height of 1.6m. The *same-height* towers were roughly divided into four groups based on simulations from our model: very stable, mildly stable, mildly unstable, very unstable. The *same-height* experimental session was composed of four subsessions of 108 trials (each tower repeated four times, once per subsession).

On each trial participants made graded responses to the question, "Will this tower fall?", by pressing keys on a 1-7 scale to indicate degrees of confidence between "definitely will fall" (1) to "definitely will not fall" (7).

Direction experiment The second experiment asked participants to predict the direction that towers fall in. This experiment used a different set of tower stimuli, in which all were unstable. It further varied from the stability experiments in that the ground disk was divided into four quadrants colored different shades of green, and participants were asked: "Which part of the circle will most of the tower fall on?". They were instructed to depress the number key between 1 and 4 which corresponded to each of the quadrants (labels listed on left of the screen). Fig. 5A shows a screenshot illustrating the setup of this experiment. Feedback was provided in the Direction experiment.

Psychophysical analysis On each trial, we present a tower with state S_0. The human observer responds with the stability property R_{fall} or R_{dir}, depending on condition. We computed each participants' mean R for each tower across the experiment as their physical property judgment. We computed the mean across participant's mean judgments (and SEs, using a bootstrap analysis) to quantify human judgments about each tower's physical property (Figs. 3, 4, 5). We performed correlation analyses using Pearson's correlations, circular correlations, and partial correlations, as noted.

Results

Stability experiments

Feedback To determine whether observers use internal physics knowledge when making stability judgments, we computed the correlation between participants' judgments, R_{fall}, in the *feedback* condition ($n = 10$) and our models' predictions, $\bar{F}_{fall}$ (Fig. 3, insets). In comparison with the ground truth model, which corresponded to zero uncertainty ($\sigma = 0$),

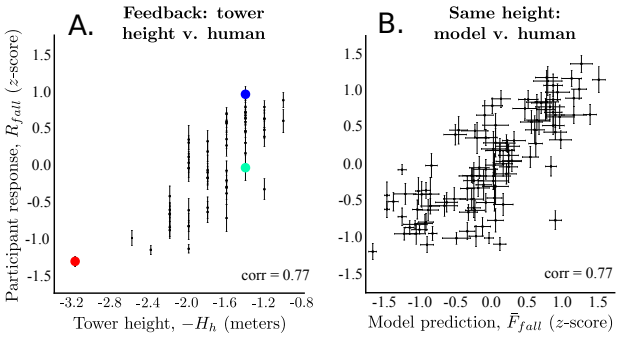

Figure 4: **Effect of height.** Each point is the mean model/heuristic prediction v. human judgment of tower stability for a single tower (SE error bars). Colored points show the towers depicted in Fig. 1. **A.** Feedback condition: height heuristic, $-H_h$ (x) v. human (y). **B.** Same-height condition: model (x) v. human (y).

the correlation coefficient was 0.69 ± 0.088 (standard error, SE). This corresponded to people correctly classifying 66% of towers' stabilities (on the 1-7 scale: 1-3 meant unstable, 5-7 meant stable, response 4 was excluded). The value of σ for which the model best-predicted human responses was 0.05 (which falls within perceptual position discriminability tolerances), and yielded a correlation of 0.89 ± 0.024. Fig. 3 (solid line) shows these correlations for a range of σ values. The high correlation between model and humans supports the hypothesis that people use real physical dynamics knowledge to produce their judgments.

To test the possibility that people use simple heuristics, rather than a richer model, we compared stability judgments predicted by the heuristics introduced in the Model section with human judgments. For $\sigma = 0.05$, the correlation coefficient between humans and: 1) tower height, H_h, was 0.77 ± 0.051, 2) top-heaviness, H_z, was 0.43 ± 0.12, and 3) tower skew magnitude, H_ρ, was 0.22 ± 0.12. Only the correlation for H_h was comparable to the model correlations, the others likely reflect inherent dependencies between the physical properties the heuristics represent and the actual physical stability of the towers. Fig. 4 shows the relationship between people's judgments and height.

Of course, these heuristics are related to the actual physical stability due to the natural structure of the towers, so to decouple the model predictions from the heuristics' effects, we computed the partial correlations between human data and the heuristics, controlling out the model's predictions. The partial correlations between humans and: 1) tower height, H_h, was 0.52 ± 0.092, 2) top-heaviness H_z, was -0.13 ± 0.11, and 3) tower skew magnitude, H_ρ, was -0.12 ± 0.20. Thus, it is clear that height played some role in humans' judgments, but the other heuristics did not.

To evaluate the model without the effect of H_h, we computed the partial correlation between model and participants' responses while controlling out height, which was 0.79 ± 0.044, indicating that people use the physics model independent of height to a significant degree.

Same-height To further examine the effect of height on people's responses, we conducted the *same-height* condition to compare participants' judgments ($n = 9$) to the model's predictions. We computed a correlation coefficient of 0.73 ± 0.039 for $\sigma = 0.0$ and 0.76 ± 0.044 for $\sigma = 0.05$ (Fig. 4B), which is statistically indistinguishable from the height-independent partial correlation computed between humans and model in the *feedback* data (previous section). These significant correlations, coupled with their closeness to the partial correlation, confirms people's judgments are best predicted by the rich, simulation-based physical model.

No-feedback In order to control for possible learning effects in the *feedback* condition we collected participants' judgments in the *no-feedback* condition. We computed the correlation between peoples' judgments ($n = 10$) and the model's predictions, resulting in a coefficient of 0.82 ± 0.030. Fig. 3 (dashed line) shows correlations for a range of σ levels. The model was slightly poorer at predicting their responses, however the correlation between the *feedback* responses and the *no-feedback* responses was 0.95 ± 0.011, suggesting little difference between the two conditions. It may be that the model is better able to predict the *feedback* responses because participants in that condition had opportunity to calibrate their internal physics models. Despite this difference, however, the *no-feedback* correlation is also significant, affirming the hypothesis that physics knowledge plays a large role. Furthermore, the fact that judgments from both conditions are so similar implies people use strategies that are not captured by the model.

Direction experiment

One of the key ideas of the rich simulation-based model is that it is able to easily generalize to many different tasks and situations. In order to assess this flexibility, we compared human judgments ($n = 9$), R_{dir} regarding the direction a tower will fall, with the model's predictions, $\bar{F}_{dir}$, as well as the skew heuristic prediction, H_θ. The circular correlation between R_{dir} and $\bar{F}_{dir}$ was 0.66 ± 0.032 while the correlation between R_{dir} and H_θ was 0.18 ± 0.038. Clearly, the model is far better at explaining humans' direction judgments; Fig. 5B illustrates these results by plotting the differences between R_{dir} and $\bar{F}_{dir}$ for each tower (dots).

The model's predictions about different towers' fall directions vary significantly in confidence, due to the effects perceptual uncertainty on different samples' physical outcomes. Confidence of fall direction judgments can be quantified by circular variance of the model's fall-direction estimates (indicated by dot color in Fig. 5B, or the insets in Fig. 5C) over the N simulations sampled from the same tower. In order to assess model fits on the stimuli for which model predictions are most meaningful, we sorted towers by the circular variance of model predictions and computed model-participant correlations for the k lowest-variance towers, where k was varied from 10 to all 60 towers. Fig. 5C shows the circular correlations between R_{dir} and $\bar{F}_{dir}$, as well as R_{dir} and H_θ, as functions of k. Excluding the 10-20 towers with lowest confidence (i.e., $k = 40, 50$ or less) the correlation between R_{dir}

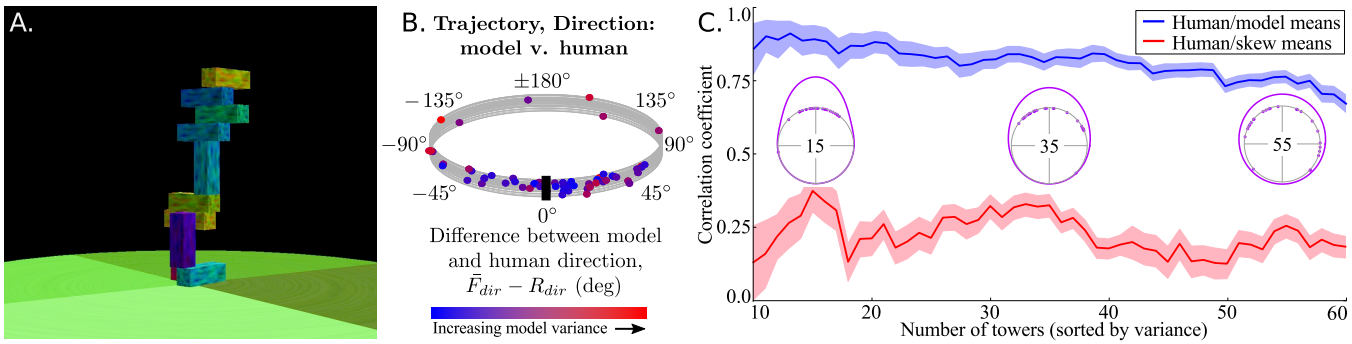

Figure 5: **Direction experiment A.** Experiment stimulus, showing the possible direction choices as quarters of the ground. **B.** Each point is the difference between the mean model/heuristic prediction and the mean human judgment of the tower's fall direction ($\bar{F}_{dir} - R_{dir}$, or $H_\theta - R_{dir}$), and each dot represents a different tower stimulus. **C.** Correlations between humans and model/heuristics. The lines represent the correlations (y-axis) between R_{fall} and $\bar{F}_{dir}$ (blue line, ± 1 SE), and R_{fall} and H_θ (red curve, ± 1 SE), as a function of the number of towers over which the correlation was performed (x-axis, sorted from lowest to high variance towers as predicted by the model). The purple insets show the model's distribution of predicted fall directions (dots), with the best-fit Von Mises distributions (curves), for 3 towers.

and $\bar{F}_{dir}$ is reliably over 0.8. Model predictions for direction of fall thus seem to match human judgments well in those cases where model predictions are meaningfully defined.

Lastly, we evaluated whether the model output's circular variances can predict people's confidence in stability judgments, as measured by variance in responses across participants. There was indeed a correlation of 0.55 ± 0.035 between the circular variances of human judgments and thoses of model predictions across individual towers, which would be expected if humans make judgments by stochastically sampling from possible choices with frequency in proportion to their expected reward.

General Discussion

We find human physical reasoning consistent with a model that uses internal knowledge of physical principles to predict future scenes states, and that internal limitations like uncertainty due to noise can account for deviations from ground truth performance. This consistency may be surprising in light of previous work on intuitive physics with much simpler situations focusing on the ways in which human judgments are biased and error-prone (Todd & Jr., 1982; Gilden & Proffitt, 1989a, 1989b, 1994; Caramazza et al., 1981; McCloskey, 1983). Future work will explore the differences between our tasks and previous intuitive physics studies that might explain this gap, such as differences in the ecological validities of the scenarios, stimuli and tasks (Zago & Lacquaniti, 2005).

While our model is a good predictor of human physical reasoning (Fig. 3), people predict each others' responses even better. This suggests that there is systematic structure to people's judgments that our model does not capture. The model may be limited by its assumption that the brain perfectly models Newtonian dynamics, or its approximations of perceptual inference. One improvement might be to adopt noisy physics simulations, with accuracies diminishing rapidly over time. Another might be to vary perceptual uncertainty for blocks that are visible and occluded, respectively – which can be tested by manipulating participants' viewing conditions.

Though preliminary, this work supports the hypothesis that knowledge of Newtonian principles and probabilistic representations are generally applied for human physical reasoning. Complex tasks like predicting the stability of a tower of blocks are both expressible in our modeling framework and well-matched with human performance. This idea provides rich and flexible foundational groundwork for developing a comprehensive model that naturally scales to a broad class of human physical judgments.

Acknowledgments

We acknowledge the support of our funding sources: Qualcomm, ONR MURI N00014-07-1-0937, ONR grant N00014-09-0124, ARO MURI W911NF-08-1-0242, ONR MURI 1015GNA126.

References

Baillargeon, R. (2007). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development*. Blackwell.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in 'sophisticated' subjects: misconceptions about trajectories of objects. *Cognition*, *9*(2), 117–123.

Fleming, R. W., Barnett-Cowan, M., & Bülthoff, H. H. (2010). Perceived object stability is affected by the internal representation of gravity. *Perception*, *39*, 109.

Gilden, D. L., & Proffitt, D. R. (1989a). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 372–383.

Gilden, D. L., & Proffitt, D. R. (1989b). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 384–393.

Gilden, D. L., & Proffitt, D. R. (1994). Heuristic judgement of mass ration in two-body collisions. *Perception and Psychophysics*, *56*(6), 708–720.

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

Sanborn, A., Mansinghka, V., & Griffiths, T. (2009). A bayesian framework for modeling intuitive dynamics. In *Proceedings of the 31st annual conference of the cognitive science society.*

Todd, J. T., & Jr., W. H. W. (1982). Visual perception of relative mass in dynamic events. *Perception*, *11*(3), 325–335.

Zago, M., & Lacquaniti, F. (2005). Visual perception and interception of falling objects: a review of evidence for an internal model of gravity. *Journal of Neural Engineering*, *2*, S198.