Sources of uncertainty in intuitive physics

Kevin A Smith (k2smith@ucsd.edu) and Edward Vul (evul@ucsd.edu)
University of California, San Diego
Department of Psychology, 9500 Gilman Dr.
La Jolla, CA 92093 USA

Abstract

Recent work suggests that people predict how objects interact in a manner consistent with Newtonian physics, but with additional uncertainty. However, the sources of uncertainty have not been examined. Here we measure perceptual noise in initial conditions and stochasticity in the physical model used to make predictions. Participants predicted the trajectory of a moving object through occluded motion and bounces, and we compared their behavior to an ideal observer model. We found that human judgments cannot be captured by simple heuristics, and must incorporate noisy dynamics. Moreover, these judgments are biased in a way consistent with a prior expectation on object destinations, suggesting that people use simple expectations about outcomes to compensate for uncertainty about their physical models.

Keywords: intuitive physics, stochastic simulation, uncertainty, probabilistic inference

Introduction

Predicting how the world will unfold is key to our survival and ability to function on a daily basis. When we throw a ball, cross a busy street, or catch a pen about to fall off of a desk, we must foresee the future physical state of the world to plan our actions. The cognitive mechanisms that help us make these predictions have been termed 'intuitive physics' models.

Although human performance in physical prediction tasks tends to approximate real-world (Newtonian) physics, it does not match exactly: people make systematic prediction errors. While this has been taken as evidence that human models of intuitive physics are non-Newtonian (e.g., McCloskey, 1983), more recently human behavior has been explained by intuitive Newtonian physics models under uncertainty. On this account, human predictions deviate from Newtonian mechanics because of stochastic error – uncertainty about the initial positions or velocity of objects propagates through the non-linear physical model and causes variability and bias in final judgments. For instance, human predictions about the stability of a tower of blocks or the most likely direction for that tower to fall are consistent with a purely Newtonian model of physics with a small amount of uncertainty in the initial positions of the constituent blocks (Hamrick, Battaglia, & Tenenbaum, 2011). Similar models of physics with perceptual noise have been used to explain relative mass judgments in collisions (Sanborn, Mansinghka, & Griffiths, 2009) and infants' expectations for object movement (Téglás et al., 2011).

There are numerous ways in which uncertainty can be introduced into intuitive physical reasoning. We broadly classify these into two categories: perceptual uncertainty and uncertainty about dynamics. Perceptual uncertainty arises because initial measurements of the location and velocity of objects is imperfect; this initial noise propagates through the model. Uncertainty about dynamics reflects noise in the physical model itself. Real object movement and collisions are perfectly deterministic only in an idealized system; in the world, objects can deviate from their ideal path because of multiple, unknowable interactions with the environment (e.g., a ball rolling across gravel will not move in a straight line). Stochastic dynamics could thus reflect such environmental uncertainty.

Our goal is to disentangle the influence of initial noisy percepts and noisy physics on human predictions of object dynamics. We compared human behavior in a simple physical prediction task to a stochastic physics model with parameters reflecting the different types of uncertainty.

Stochastic Physics Model

We designed a model to replicate stochastic physics in a simple environment: a ball bouncing around a two-dimensional box. We based this model on idealized mechanics, but also incorporated the two sources of uncertainty: we added noise to the initial position and velocity to capture perceptual uncertainty, while dynamic uncertainty was captured by jitter in object movement over time, and variability in bounce angles.

Uncertainty Parameters

The model was based on a simple two-dimensional physics engine customized to add our sources of uncertainty. As physical uncertainty goes to zero, this model reduces to laws from idealized mechanics: the ball would continue to move in a straight line at a constant velocity until it hit a wall, at which point it would bounce elastically and with angle of incidence equal to the angle of exit. Uncertainty was captured using four parameters, two for the perceptual error, and two for the stochastic error:

Perceptual Uncertainty At the start of the simulation, the ball's position and velocity were based on where the ball would be in a perfectly deterministic simulation, but with noise added. Position was perturbed by isotropic two-dimensional Gaussian noise parameterized by standard deviation, σ_p. Noise in velocity direction was captured in a von Mises (circular normal) distribution on direction of motion, parameterized by concentration (inverse variance) κ_v. We did not consider uncertainty in the speed of the ball, as this would only affect the timing of the ball's movement but not its destination, which is the prediction we aim to capture.

Dynamic Uncertainty Noise was added during the simulation in two ways. First, at each time step (1000/sec), the direction of the ball was 'jittered' by adjusting its direction using a von Mises distribution with the concentration parameter κ_m. In addition, noise was added during each bounce by assuming that the angle the ball bounced off of the wall was defined by a von Mises distribution centered on the angle of incidence with a concentration parameter κ_b.

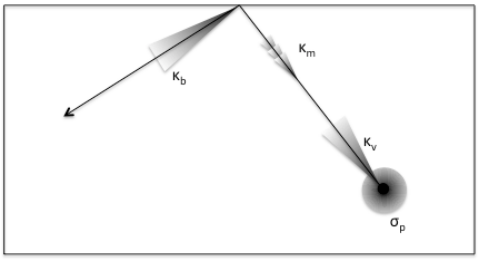

Figure 1: Sources of uncertainty in the stochastic physics model

Experiment

We aimed to test model predictions against human data and to estimate uncertainty parameters in intuitive dynamics. In this experiment, subjects predicted the trajectory of a ball in a two-dimensional environment on a computer screen. This was done in a 'Pong' game where participants tried to catch the ball with a paddle. Crucially, we occluded the latter part of the ball's movement, so that successful prediction of the final position required the mental simulation of the object trajectory. We could estimate the final position predicted by our stochastic physics model with different parameters, and thus compare human behavior to model predictions under varying types and degrees of uncertainty.

In this experiment we parametrically varied both the distance the ball would travel[1] and the number of bounces off of walls while occluded. If intuitive dynamics models are deterministic, then the number of bounces will have no effect on human predictions. The distance manipulation was designed to tease apart the contributions of perceptual uncertainty about velocity and dynamic velocity noise.

Methods

52 UCSD undergraduates (with normal or corrected vision) participated in the experiment for course credit.

Subjects used a computer mouse to control the vertical position of an on-screen 'paddle' to catch a moving ball. The ball moved according to the deterministic physics underlying the stochastic physics model. Both the paddle and the ball were confined to a 1200 by 900 pixel area in the center of the screen. Each trial began with a display of only the paddle, which subjects could move up and down. The paddle was 100 pixels in height and was centered on the vertical position of the mouse before each trial. A mouse click triggered the start of a trial. A ball would then appear on the screen, moving at a constant velocity of 600 pixels/second. After the ball moved 400 pixels, a grey rectangle would occlude the portion of the screen containing the ball (Figure 2). During this period, the ball would continue to move, bouncing perfectly elastically off of the edges of the field, but would not be visible. Once the subjects caught the ball with the paddle, or the ball broke the plane of the paddle, the trial would end and the occluder would be removed, showing whether (and by how far) the subject missed the ball. Upon clicking the mouse, the screen would clear and reset for the next trial. The number of balls caught by the subject was always displayed in the upper right corner as a motivation to perform well.

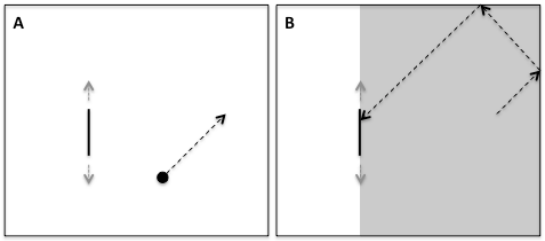

Figure 2: Diagram of a trial. (A) The ball moves unoccluded in a straight line. (B) Once the field is occluded, the ball continues to move until caught or it passes the paddle plane.

Subjects were given 648 trials throughout the experiment. These 648 trials were identical for all subjects, but presented in a randomized order. Each trial had a particular ball trajectory, generated by one of nine conditions. The nine trajectory conditions crossed the distance the ball travelled while occluded (600, 800, or 1000 pixels) with the number of bounces (0, 1, or 2); there were 72 trials of each condition. The specific path for each trial was generated prior to the experiment subject to the constraints of the condition and the constraint that the final position was not in the top 20% or bottom 20% of the enclosed area to avoid bias due to positioning the paddle at the ends of the screen.

Before starting the experiment, subjects were given seven trials without the occluder to demonstrate how the ball would move, then six practice trials with the occluder.

For each trial, we recorded the position of the midpoint of the paddle once the ball was caught or moved past the paddle. From this measure we could calculate, for each trial, (a) the average expected position of the ball, and (b) the variance of predictions around that expectation.

Subject Performance

Accuracy Subjects caught the ball on 43.8% of all trials. Individual subject accuracies varied between 25.6% and 63.7% (chance was 11%). Accuracy also varied by trial condition: subjects were most accurate in the shortest, no

[1] Because the ball always moved at a constant velocity, distance was proportional to duration of occlusion.

bounce condition (69%) and least accurate in the longest, two-bounce condition (32%).

Accuracy improved slightly over time, increasing from 42.7% in the first half of trials to 44.9% on the second half ($\chi^2(1) = 15.9$, $p < 0.001$). However, because this was a small effect, and because in a logistic model predicting accuracy, trial order did not interact with either distance ($\chi^2(2) = 0.72$, $p = 0.70$) or number of bounces ($\chi^2(2) = 4.18$, $p = 0.12$), we do not try to account for this change.

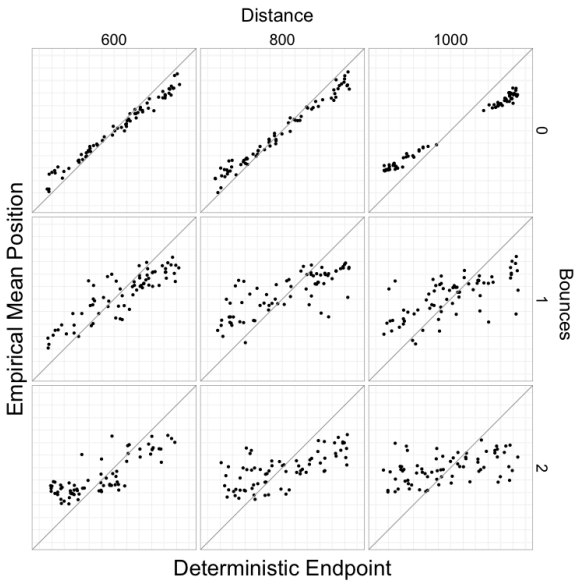

Figure 3: Mean predicted paddle position versus path endpoint using deterministic physics as a function of trial condition. Each point represents a separate trial.

Expected Positions In addition to decreasing accuracy, subjects also showed increasing bias in average predictions as the distance or number of bounces increased. The mean final position of the paddle for each trial shifted towards the center as compared to the final ball position (see Figure 3).[2] The magnitude of this bias toward the center of the screen increased as either distance or number of bounces increased.

Table 1: Percent of distance 'shifted' from actual end ball position towards center by trial condition

		Distance		
		600	800	1000
Bounce	0	24%	44%	53%
	1	23%	60%	70%
	2	41%	63%	84%

[2] There was low lag-one autocorrelation between the position of responses and prior responses (0.11) that did not vary by condition, suggesting that this mean shifting was not driven by subjects leaving the paddle in the same position as their prior trial.

Variance of Responses The variability of subjects' responses around the mean also increased with distance and bounces, but only up to a ceiling - well below the maximum possible spread - once subjects had to take into account even one bounce.

Table 2: Average standard deviation (in pixels) of responses within a trial by condition

		Distance		
		600	800	1000
Bounce	0	65	76	94
	1	111	115	114
	2	115	111	121

Model Application

The coarse results suggest that prediction error and variability increases with distance or number of bounces. But they do not indicate which sources of uncertainty contribute to intuitive physics predictions, nor do they explain why some trials within the same condition produce greater bias and variability than others.

We aimed to tease these factors via our model of stochastic physics. By finding the set of uncertainty parameters that best fits the empirical data, we can compare the relative contribution of the perceptual uncertainty parameters to the dynamic uncertainty parameters. A good model should capture trial-level differences in subjects' performance, and explain trial difficulty based on the interplay of different sources of uncertainty.

Simulation

We replicated the experimental task in the stochastic physics model, simulating the same 648 trials. To mirror this task, each simulation started at the point of occlusion (when subjects could no longer visually track the ball and must predict its path) and ended when the simulated ball crossed the plane of the paddle. On each simulation, we measured the position of the simulated ball along that plane. Because there is no analytic form of the probability distribution over possible trajectories, we simulated each trial 500 times, thus estimating the predictive distribution for each trial via sampling.

No set of uncertainty parameters produced mean estimates of the final position of the ball that were systematically shifted toward the center like the empirical data; as long as Newtonian physics underlies the model, it will in general simulate trajectories that are centered around the actual endpoint, regardless of the uncertainty parameters chosen. Since the magnitude of the center bias scaled with distance and number of bounces, we suspected that subjects were incorporating a prior on final position, producing a center bias proportional to the uncertainty in their physics-

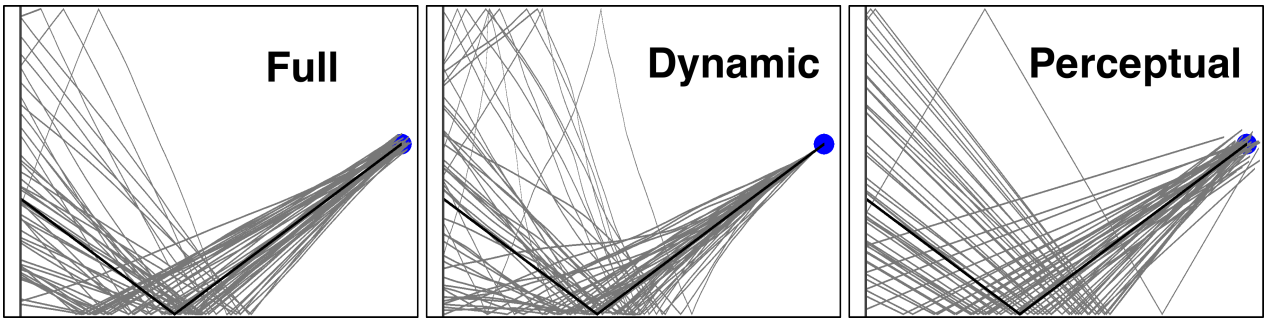

Figure 4: Sample simulation paths for one trial with each model. The grey lines represent individual simulations, the black line represents deterministic simulation. There is no initial uncertainty in the dynamic model, but it builds quickly over time, resulting in wavy paths. The initial position and velocity vary significantly in the perceptual model, but once started, the simulation unfolds deterministically. The full model uses both types of uncertainty and so has more certainty in starting positions than the perceptual model and straighter paths than the dynamic model.

based predictions.[3] People therefore appear to incorporate prior expectations with their intuitive physics models.

We treated this bias as a simple Gaussian prior on the final ball position centered on the middle of the screen, with standard deviation as a free parameter (σ_{prior}). One value of this parameter was used for all trials and conditions.

The final distribution of predictions for each trial was calculated by combining the center-prior with the distribution of predicted positions simulated by the stochastic physics engine. We treated the distribution of predicted positions as a Gaussian and calculated their mean and standard deviation. We could then calculate the mean and standard deviation of the posterior distribution using Bayesian cue combination (e.g., Ernst & Banks, 2002):

$$\sigma_{post}^2 = \left(\frac{1}{\sigma_{prior}^2} + \frac{1}{\sigma_{sim}^2}\right)^{-1}, \mu_{post} = \left(\frac{x_{center}}{\sigma_{prior}^2} + \frac{\mu_{sim}}{\sigma_{sim}^2}\right) * \sigma_{post}^2$$

Using these equations, trials with greater simulation variance will be more affected by the prior, and will shift further towards the screen center. Thus, the model can account for the center-bias in a manner sensitive to prediction uncertainty.

We found the maximum likelihood parameters to fit three quarters of the data (with an equal number of trials from each of the distance by bounce conditions).[4] We also fit two other models: one with only perceptual uncertainty and prior parameters, and one with only dynamic uncertainty and prior parameters. We compared these models based on the likelihood of the 25% of the remaining (cross-validation) data.

Model Results

Model Comparison The stochastic physics model was designed to tease apart how various sources of uncertainty contribute to intuitive physics. Thus we compared the model with both dynamic and perceptual uncertainty to the two nested models with either dynamic or perceptual uncertainty parameters alone to determine which sets of parameters were necessary to best explain the data.

In addition, we tested how well any of the stochastic models captures human behavior by comparing them to a simple regression model with different parameters for each condition. The regression model assumes that people will provide the correct answer, plus some error that varies by condition without regard to individual trial details. We assume that the average reported position will have some variance (estimated independently for each condition), and some bias towards the center (estimated by regressing the average reported position against the deterministic end positions within each condition) – in other words, the regression model is a non-physical error model. This model can capture the gross 'shift' in expected position that was observed in the data in each condition (see Figure 3), but does not treat the shift as an inference done independently on each trial. The spread in responses was assumed to be constant within each condition, and was set at the average empirical standard deviation from that condition. Like the stochastic models, this model was fit on three-quarters of the trials and tested on the remaining data.

Table 3: Model prediction of left-out data

Model	ΔLLH
Full	2,588
Dynamic	2,568
Perceptual	2,197
Regression	2,326
Oracle	3,259

Table 3 shows cross-validation likelihood for the four models. All log-likelihoods are shown as improvement over a baseline assuming that all data came from a single Gaussian. In addition, we included an 'oracle' model that knows the mean and standard deviation of responses for each trial – this serves as the plausible upper limit on how well different models might do. The full stochastic model

[3] The actual endpoint of the ball was uniformly distributed within the space of allowable endpoints. Therefore, this prior is unlikely to have been learned from the experiment.

[4] Numerical optimization techniques can find local minima, so we used multiple starting points and grid search across 1,600 sets of parameters to ensure we were finding the global minimum.

does best, followed closely by a model including only dynamic noise. Both the perceptual noise model and the non-physical model perform worse by many orders of magnitude.

The dynamic model performed nearly as well as the full model for two reasons. First, the parameter representing error in the initial position (σ_p) was set to a small value in the full model and explained very little of the variance in simulations. Second, much of the noise in initial velocity direction can be captured by increasing dynamic velocity noise, and so we cannot say whether any initial velocity noise is required. The model with only perceptual noise did quite poorly because subjects' performance changed with each additional bounce, and thus human performance cannot be captured without dynamic uncertainty.

Trial-Level Simulations Human predictions about individual trials within the same distance-by-bounce condition varied significantly: some had much larger variations in responses or greater shifts toward the center than others. These differences arose from trajectory characteristics other than total distance traveled or number of bounces. For instance, it is harder to predict the end position of a ball that bounces in a corner or balls that approach the paddle at a steep angle. If the stochastic physics model is capturing characteristics of intuitive physics, then it should capture this within-condition variability as well.

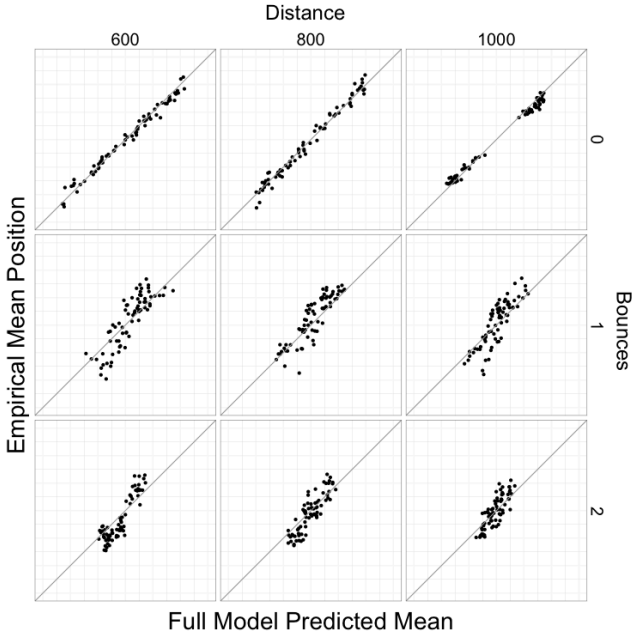

Figure 5: Full model vs. empirical mean position by condition. Each point is a separate trial.

The full stochastic model fit the variation in mean paddle position across trials well (r=0.93), and slightly better than the predictions of the regression model (r=0.90). However, the difference between models is highlighted when considering individual conditions: although both models account for the mean position in the no-bounce conditions, only the full model continues to perform well as bounces and distance are added (see Table 4).

Table 4: Correlation between model and empirical by-trial means within condition

		Full			Regression		
		Distance			Distance		
		600	800	1000	600	800	1000
Bounce	0	.99	.99	.99	.99	.99	.99
	1	.86	.88	.85	.88	.77	.68
	2	.89	.87	.82	.82	.68	.45

The standard deviation of predictions from the full stochastic model was well correlated with the standard deviation of subjects' responses across trials (r=0.79, see Figure 6), albeit with a tendency to overestimate. Moreover, the stochastic physical model also captures the variability across trials within each distance-by-bounce condition (Table 5). Together, these results indicate that human uncertainty about final outcomes accumulates in a manner qualitatively similar to that predicted by a stochastic physical model.

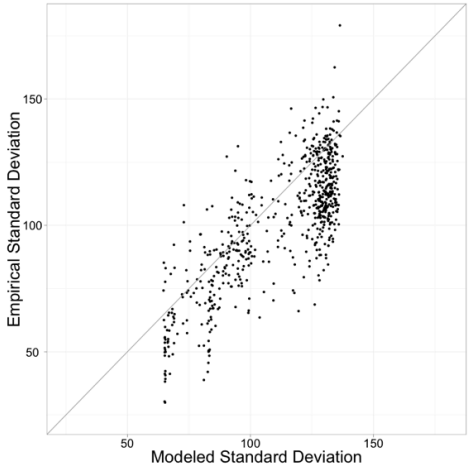

Figure 6: Full model vs. empirical standard deviation by trial

Table 5: Correlation between full model and empirical by-trial standard deviations within condition

		Distance		
		600	800	1000
Bounce	0	.54	.43	.17
	1	.53	.44	.30
	2	.14	.16	.17

In the experimental data, the amount of mean-shifting for each trial is related to the variance of the observations from that trial (Spearman's rho = 0.30), suggesting that people hedge their guesses towards the middle more as the amount

of uncertainty increases. A center-prior captures this behavior by causing more reliance on the prior when there is a wider distribution of model simulations. This has the effect of shifting guesses more towards the center when physical simulations are more uncertain. The stochastic physics model captures this phenomenon by predicting trial-level differences in uncertainty, and is thus better able to describe variation in human responses across trials than a constant mean-shift for each condition (see Figure 5).

General Discussion

We found that human performance on a physical prediction task is captured by a model of stochastic physics with a prior expectation about the final position of objects. Furthermore, we found that bias and variability of human predictions are driven by uncertainty about the dynamics: people use stochastic, rather than deterministic, physics to make predictions. This result supports recent findings that people predict object dynamics using unbiased intuitive physics models (e.g., Hamrick, et al., 2011), and suggests two refinements to this view. First, the internal physics models themselves must be stochastic rather than rely solely on perceptual uncertainty to demonstrate non-determinism. Second, people do not directly use predictions from their physical models, but combine them with simple priors to produce rich behaviors.

Though we found that dynamic uncertainty contributes substantially to predictions in this task, we do not know how people might adjust this uncertainty based on task demands. In this experiment, the ball was easy to see (low perceptual uncertainty) and the background was uniform (suggesting less perturbation during movement). Lower contrast between object and background might cause greater perceptual uncertainty; likewise, backgrounds suggesting a rough surface might cause people to introduce more stochastic movement error into their simulations. An interesting direction for future work is to explore how people adjust the uncertainty within their intuitive physics models to account for different expectations about the world.

We also found that people modulate their physical predictions via prior expectations about the outcomes. Although these expectations could arise in many ways, here we were able to capture human behavior well by using a simple expectation about the final position: despite there being no evidence for it within the experiment, people believed that the ball was more likely to end up in the center of the screen. We made a simplifying assumption that this was a prior expectation about final location; it is possible that this is an approximation of other sorts of priors (e.g., objects tend to travel in a more horizontal direction). More research is required to understand exactly what these prior expectations are, how they develop, and under what conditions they become integrated into models of intuitive physics. Regardless of the prior used, we think that this might reflect a more general strategy that people may adopt to account for their uncertainty in their internal physical model itself: by adjusting model predictions via a simple prior on outcomes, behavior will be more robust to errors in the simulation model. A similar process may suggest a means for combining model-based and model-free predictions (Gläscher, Daw, Dayan, & O'Doherty, 2010): learning simple expectations about the world is a good hedge against model error.

Our models predicted systematically larger variances than those we observed. This may be due to our simplistic choice of the shape of the prior. Gaussian cue combination of the prior and simulated distributions produces dependence between variance and mean-shift: a greater mean-shift arises only from greater variance. Thus to best fit the predicted means, using a Gaussian prior required a biased variance estimate. Further work is required to understand the priors people actually hold (e.g., Stocker & Simoncelli, 2006) to refine the models that people use to simulate the world.

This work supports the hypothesis that intuitive physics models can be built upon a Newtonian framework. Moreover, these models are not deterministic, but incorporate sources of dynamic uncertainty. Furthermore, people do not trust these models entirely, but combine their predictions with simple expectations about the outcome itself. Though just a first step, this provides a framework for disentangling and understanding the various components of intuitive physics models.

Acknowledgments

This work was supported by BIAL Foundation grant to EV

References

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*, 585-595.

Hamrick, J., Battaglia, P., & Tenenbaum, J. (2011). *Internal physics models guide probabilistic judgments about object dynamics*. Paper presented at the Proceedings of the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA.

McCloskey, M. (1983). Intuitive Physics. *Scientific American, 248*(4), 122-130.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2009). *A Bayesian framework for modeling intuitive dynamics*. Paper presented at the Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578-585.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*, 1054-1059.