Toward Domain-General Intuitive Physics: Cognitive Modeling of Deformable Dynamics

Shawn Nordstrom

Yale University

CGSC 2740: Algorithms of the Mind

Professor Ilker Yildirim

December 13, 2025

Abstract

Human observers are able to infer physical properties of soft materials, such as stiffness and damping, from brief visual observations, but it is unclear whether these inferences rely on structured inverse physics or on simpler mappings from motion statistics to "softness." Prior work on cloth perception (Bi et al., 2025) has shown that a probabilistic generative model which infers a cloth simulator can account for human "same material?" judgements better than task-optimized deep networks. In this project I attempt to extend that inverse-physics framework to a minimal, fully controllable soft-body world: a deformable square implemented as a mass-spring system in Taichi with latent material parameters of stiffness and damping. I compare two models that estimate the latent material parameters from simulated impact videos: (1) a generative inverse-physics model that treats the Taichi simulator as a forward model from material parameters to videos and then uses Bayes' rule to infer the most likely parameters given an observed video; and (2) a feature-based baseline that predicts stiffness and damping directly from low-dimensional motion descriptors such as maximum compression and settling time. Using a grid of simulated soft-square impacts with known ground-truth parameters, I evaluate both models in terms of parameter recovery. I discuss how this soft-square world provides a generalizable testbed for modeling soft-object perception and how the comparison between generative and feature-based models can inform future behavioral experiments on material similarity judgements.

Introduction and Prior Work

1.1 Intuitive physics and generative models

Perceiving the physical properties of the world requires more than tracking where objects are and how they move. When we watch a curtain flutter, a pillow compress, or a piece of fruit deform on impact, we also form impressions of *what kind* of material we are seeing, how stiff or soft it is, how quickly it dissipates energy, how it would respond under different forces. These "soft-object" inferences are essential for everyday action and prediction, yet they are difficult to perceive due to the large and complex deformations soft bodies undergo over time.

Many studies have made a general claim about how people might solve problems like this: that perception and cognition rely on internal world models of physics. On this view, observers are not just memorizing how objects tend to move, but implicitly representing hidden physical quantities such as mass, friction, or elasticity, and using an internal "physics engine" to predict how those quantities give rise to observable motion. A large body of behavioral work in intuitive physics suggests that people are sensitive to such latent properties, for example in judging whether stacks of blocks will fall, how objects will collide, or how stable a scene is under gravity. These findings have motivated models in which perception is treated as an inverse problem: given what is seen, infer the hidden physical variables that could have produced it.

Computationally, this idea has been formalized using generative models and probabilistic inference. The mind is hypothesized to maintain a prior over physical variables and a forward model that maps those variables to expected observations, and then to invert this mapping when new data arrive. Lake et al. (2017), for example, argue that many aspects of human learning and thinking are best understood in terms of such structured, causal generative models and internal

simulation, rather than purely task-optimized pattern recognition. In their framework, intuitive physics is one instance of a broader idea of inferring a small set of latent causes that explain the sensory input.

Formally, let θ denote a set of latent physical parameters (for example, stiffness and damping) and V denote an observed video. A generative model specifies a prior over θ and a likelihood over videos given those parameters:

$$\theta \sim p(\theta), \qquad V \sim p(V|\theta)$$

Inference then consists of inverting this mapping with Bayes' rule,

$$p(\theta|V) \propto p(V|\theta)p(\theta),$$

to obtain a posterior distribution over physical parameters given an observed video.

1.2 Soft-object perception and prior work

Soft-object perception provides a particularly stringent test of this "intuitive physics" view. Bi et al. (2025) developed a probabilistic model of a cloth in which a physics engine simulates the motion of cloth with particular mass, stiffness, and external forces, and a Bayesian inference procedure inverts this simulator to estimate those physical properties from observed videos. Given a target cloth animation, their model infers which combinations of physical parameters are most likely to have produced it. To test whether this model is a good candidate for human perception, Bi et al. collected triad judgements in a "same material?" task. On each trial, observers saw a target cloth and two test cloths and chose which test cloth was made of the same material as the target. They then compared several computational accounts, including deep neural networks trained to predict material properties directly from image sequences. The key result was that the simulation-based generative model better matched human trial-by-trial choices than the task-optimized neural networks, suggesting that cloth perception is organized

around a latent space of physical parameters inferred from motion, rather than solely around shallow motion features.

At the same time, it is not obvious that perception always relies on fully structured inverse models. Inverse problems in physics are often underdetermined. Different combinations of latent parameters and nuisance factors (such as initial conditions or unobserved forces) can lead to very similar observable motion. In such cases, relatively simple heuristics or feature-based mappings from motion statistics to an internal "softness" scale might perform quite well on the kinds of stimuli people typically encounter. This prompts the question of, for soft bodies, when we really need an internal physics engine, and when might a sophisticated pattern recognizer be enough?

1.3 Present work

In this project, I extend the inverse-physics framework to a minimal, fully controllable soft-body world and use it to compare a generative model to a simple feature-based baseline. Instead of realistic cloth, I work with a deformable "soft square" implemented as a two-dimensional mass-spring system in Taichi. Its behavior is controlled by a small set of latent material parameters–stiffness and damping. By sampling different parameter settings and simulating impacts with a floor or box, I generate a dataset of videos with known ground-truth material properties.

Then, I compare two models that try to recover those latent parameters from video. The first is an inverse-physics model that treats the Taichi simulator as a forward mapping from material parameters to motion and uses Bayesian inference, given a prior over stiffness and damping, to estimate which values best explain an observed video. The second is a feature-based model that skips explicit simulation and predicts stiffness and damping directly from low-

dimensional motion descriptors such as maximum compression and settling time. Both operate in the same latent space but have different computational assumptions.

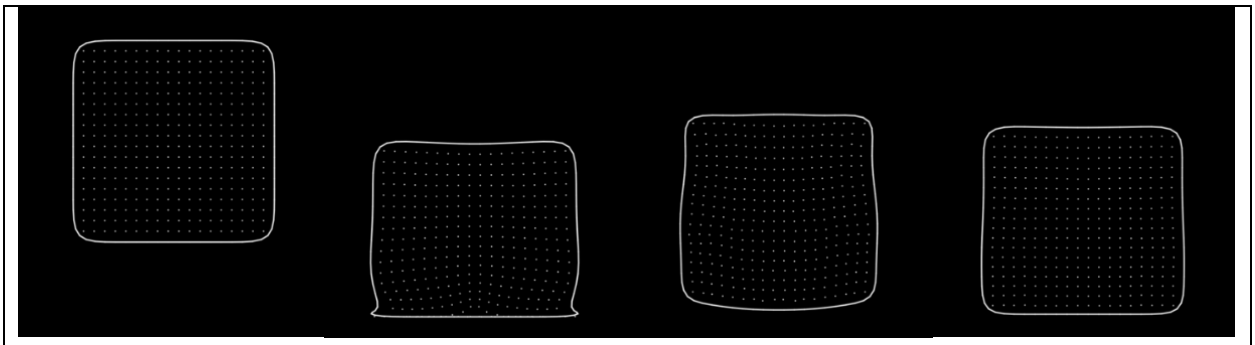

Figure 1. Taichi-based soft-body object simulation visualization, showing the deformable body's motion and shape changes over time under a physics model.

Methods

2.1 Soft-square simulator in Taichi

I modeled a simple soft body as a deformable "soft square" implemented as a two-dimensional mass-spring system in Taichi. The object consists of an $n \times n$ grid of point masses on a regular lattice, connected by springs along horizontal and vertical directions. Each mass i has position $x_i(t) \in R^2$ and velocity $v_i(t)$.

For a spring between masses i and j with rest length L_{ij}, the elastic force is

$$F_{ij}^{spring} = -k(\left|\left|x_i - x_j\right|\right| - L_{ij})\hat{u}_{ij}$$

where k is a global stiffness parameter and $\hat{u}_{ij}$ is the unit vector from x_j to x_i. Each mass also experiences viscous damping

$$F_i^{damp} = -cv_i$$

with global damping parameter c, and a constant gravitational force $F_i^{grav} = mg$. The net force on mass i is the sum of spring, damping, and gravity, and the equations of motion are integrated with a semi-implicit Euler update with time step Δt. The square is dropped under gravity onto a rigid floor; collisions are handled by clamping positions to the boundary and reflecting the normal component of velocity with restitution coefficient r.

The behavior of the soft square is controlled by two latent material parameters,

$$\theta = (k, c),$$

which directly correspond to "stiffness" and "damping" in the rest of the paper. For visualization I render the outer ring of masses as a polygonal outline with three iterations of Laplacian smoothing, but this does not affect the physics at all.

2.2 Dataset

To construct a dataset of impacts, I sampled stiffness and damping from simple ranges

$$k \in [k_{min}, k_{max}], \qquad c \in [c_{min}, c_{max}]$$

using a regular grid of $[N_k]$ stiffness values and $[N_c]$ damping values. For each parameter setting $\theta = (k, c)$, I initialized the square at a fixed height above the floor with a small random horizontal offset and zero initial velocity, then simulated its motion for T time steps at frame rate 60 fps. Each simulated video V is therefore a sequence of 2D positions for all masses over time, together with the ground-truth parameters θ.

In total, the dataset contained 49 videos. I split these into training and test sets in an 8:2 ratio. The training and validation sets were used to fit the feature-based model (Section 2.4); the inverse-physics model (Section 2.3) was evaluated directly on the test set.

2.3 Inverse-physics model

The first model written in the Gen framework the Taichi simulator as a forward generative model from material parameters to videos and performs approximate Bayesian inference over $\theta = (k, c)$ given an observed video V. I use the same structure as in the Introduction:

$$\theta \sim p(\theta), \qquad V \sim p(V|\theta)$$

where $p(\theta)$ is a prior over stiffness and damping defined by the ranges in Section 2.2, and $p(V|\theta)$ is defined implicitly by the simulator plus a simple noise model.

To keep inference tractable, I summarize each video by a low-dimensional vector $\phi(V)$ containing time series statistics, such as the center-of-mass (COM) height over time, maximum compression, and vertical velocity. Given a candidate parameter setting θ, I run the simulator to produce a predicted video$\hat{V}(\theta)$, compute its summary $\phi(\hat{V}(\theta))$, and measure the mismatch between observed and simulated summaries with a squared error:

$$d(V, \theta) = \left|\phi(V) - \phi\left(\hat{V}(\theta)\right)\right|^2.$$

I then define an unnormalized posterior over a discrete grid of candidate parameter values $\{\theta_j\}_{j=1}^{M}$ as $\tilde{p}(\theta_j|V) = \exp(-\lambda d(V, \theta_j))p(\theta_j)$, with scale parameter $\lambda > 0$, and normalize over the grid. The model's estimate for a video is the maximum a posteriori (MAP) parameter:

$$\hat{\theta}_{phys}(V) = \arg\max(\theta_j, \tilde{p}(\theta_j|V)).$$

Therefore, the inverse-physics model explicitly uses the simulator as its likelihood and chooses the stiffness and damping values whose simulated motion best matches the observed video in the summary space.

2.4 Feature-based model

The second model predicts material parameters directly from hand-engineered motion features, without simulating physics. For each video V, I compute a feature vector $f(V) \in R^d$ consisting of low-dimensional descriptors, including maximum compression ratio (minimum vertical extent / initial extent), settling time (time until COM height and velocity remain within a small band), maximum COM speed over the trajectory, bounce count (number of distinct bounces above a threshold).

I then fit a supervised regressor g that maps features to parameter estimates,

$$\hat{\theta}_{feat}(V) = g\big(f(V)\big).$$

In the simplest version, g is a multivariate linear regression model trained on the simulated videos to minimize mean squared error between predicted and ground-truth stiffness and damping. I selected regularization strength and any additional hyperparameters using the validation set, and report results on the held-out test set.

2.5 Evaluation

To assess how well each model recovers the true material parameters, I evaluate their predictions on a held-out test set of simulated trajectories. In the main configuration, stiffness and damping were sampled from a 7×7 grid over $[k_{min}, k_{max}]$ and $[c_{min}, c_{max}]$, resulting in 49 impacts; I randomly split these into an 80/20 train-test grouping. The feature-based model was fit on the training set, while the inverse-physics model was evaluated directly on the test trajectories using the simulator and the grid-based inference described above. For each model $m \in \{phys, feat\}$ and each parameter (stiffness and damping), I compute the R^2 coefficient of determination between true and predicted values.

Results

I evaluated both models on the held-out test set of simulated soft-square impacts described in Section 2. For this first proof-of-concept, I used a very small amount of feature noise, so that the summaries of each trajectory are almost deterministic functions of the underlying parameters. In this report, I used a very small amount of data; I did experiment with more than this, but I think this works well as an initial proof of concept!

R^2 Values	**Inverse-physics model**	**Feature-based model**
Stiffness	1.00	0.69
Damping	0.99	0.98

Figure 2. R^2 values for predicting stiffness and damping using the inverse-physics model versus a feature-based baseline. The inverse-physics model achieves near-perfect fits for both parameters, while the feature-based model performs comparably on damping but substantially worse on stiffness.

As expected in this idealized setting, the inverse-physics model is almost perfectly able to recover both stiffness and damping (Fig. 2). Because the samples are nearly noise-free and the simulator and generator have a matching grid, the posterior scoring over the parameter grid effectively inverts the forward model in this controlled setting. This serves as a sanity check that the implementation behaves as a "correct" inverse for this simple soft-square world.

The feature-based baseline performs surprisingly well given its simplicity. It achieves almost perfect recovery for damping, indicating that the hand-crafted motion summaries make damping essentially linearly readable. Its stiffness predictions are weaker but still reasonable ($R^2 \approx 0.69$), suggesting that, although stiffness is not as directly encoded in these features as damping, there is still a strong systematic relationship.

Fig. 3 shows a higher noise sample with the same configuration. With noisier features, both models remain almost perfectly accurate for damping, but stiffness recovery degrades noticeably. The inverse-physics model's stiffness R^2 drops from 1.00 to 0.88, it seems that the noise causes the stiffness to "slide around" while damping is still correctly identified. The feature model's stiffness R^2 drops from about 0.69 to 0.50. This pattern is consistent with the idea that the chosen motion summaries are highly diagnostic of damping but only weakly informative about stiffness.

R^2 Values	**Inverse-physics model**	**Feature-based model**
Stiffness	0.88	0.50
Damping	0.99	0.98

Figure 3. R^2 values for predicting stiffness and damping with added noise. Both models remain near-perfect on damping, but stiffness degrades—especially for the feature-based model.

Discussion

The current results show that, in this clean and highly controlled setting, the inverse-physics model can almost perfectly recover both stiffness and damping, while a linear feature model already does extremely well on damping and reasonably well on stiffness. When I add more noise to the motion summaries, both models remain almost perfect for damping, but stiffness starts to degrade, especially for the feature model and to a lesser extent for the inverse-physics model. This suggests that, at least for these particular summaries, damping is much more "visible" in the motion than stiffness is.

In these results, damping comes out as relatively easy to read from motion, while stiffness is harder and more entangled with other factors. This seems to suggest that intuitive

physics is more than just a “simulator vs. no simulator” view. Some physical properties may line up naturally with stronger cues in dynamics (for example, how fast things settle or how many times they bounce), and those are properties that both a latent physics model and simple features can recover quite well. Other properties, like stiffness in this setup, seem to be encoded more weakly in the same summaries and can trade off with other variables like damping, making the inverse problem messier even for an idealized model. Taken together, these observations seem to be consistent with the idea that human intuitive physics might involve a mix of strategies, with more simulator-like reasoning for some judgments and more feature-based shortcuts for others.

Overall, I learned a lot from coding a soft-body simulator and treating it as a generative model. Implementing the mass–spring system, defining a latent space, and then trying to invert that simulator was difficult but also fun to do! I also got a better feel for how design choices—like which features you summarize motion with—directly affect identifiability and what a model can or cannot “see.”

There is a lot of room to expand from this report. Technically, I would like to formally present more experiments of the initial configuration, vary scenes that pull apart stiffness and damping more cleanly, and improving on the implementation of the model in Gen. On the cognitive side, I’m extremely interested in finding out more about how these experiments would work. A next step would be to generate soft-square videos, collect human 2AFC or triad similarity judgments, and see where people align more with the inverse-physics model versus a stronger feature-based or neural baseline. That comparison would let us test whether human intuitive physics is better characterized as inverse simulation, feature-based heuristics, or a flexible combination that depends on which physical property or situation is being inferred, both in cases where people are accurate and in cases where they get things wrong.

References:

Bi, W., Shah, A. D., Wong, K. W., Scholl, B. J., & Yildirim, I. (2025). Computational models reveal that intuitive physics underlies visual processing of soft objects. Nature Communications, 16, 6303. https://doi.org/10.1038/s41467-025-61458-x

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253. https://doi.org/10.1017/S0140525X16001837

Goodman, N. D., Tenenbaum, J. B., & The ProbMods Contributors. (2016). Probabilistic Models of Cognition (2nd ed.) [Electronic book]. Retrieved December 10, 2025, from http://probmods.org/

Goodman, N. D., & Tenenbaum, J. B. (n.d.). Generative models (Chapter 2). In Probabilistic Models of Cognition (2nd ed.). Retrieved December 16, 2025, from https://v1.probmods.org/generative-models.html

Cusumano-Towner, M. F., Saad, F. A., Lew, A. K., & Mansinghka, V. K. (2019). Gen: A general-purpose probabilistic programming system with programmable inference. Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI '19), 221–236. https://doi.org/10.1145/3314221.3314642

Kulkarni, T. D., Whitney, W. F., Kohli, P., & Tenenbaum, J. B. (2015). Deep convolutional inverse graphics network. In Advances in Neural Information Processing Systems 28 (NIPS 2015). https://arxiv.org/abs/1503.03167

Erdogan, G., & Jacobs, R. A. (2017). Visual shape perception as Bayesian inference of 3D object-centered shape representations. Psychological Review, 124(6), 740–761. https://doi.org/10.1037/rev0000086

Gallistel, C. R., & King, A. P. (2009). Memory and the computational brain: Why cognitive science will transform neuroscience. Wiley-Blackwell. https://doi.org/10.1002/9781444310498

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. W. H. Freeman.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005