TOWARDS CURIOSITY-DRIVEN LEARNING OF PHYSICAL DYNAMICS

Michael John Lingelbach[1,*], Damian Mrowca[2,*],
Nick Haber[3] , Li Fei-Fei[2] , and Daniel L. K. Yamins[1,2,4]

Department of Neuroscience[1], Computer Science[2], Education[3], and Psychology[4], Stanford, CA
{mjlbach, mrowca}@stanford.edu

ABSTRACT

Throughout our lives, we as humans acquire an intuitive understanding of our physical environments, a capacity that supports our imagination and planning abilities. Driven by our own curiosity, we learn about object motion and properties via self-curated targeted experiments, that teach us what we do not know. Recently, neural network models have been proposed that learn forward object dynamics from observations like humans. Unlike humans, these models do not actively interact with surrounding objects but learn from human-curated datasets as passive observers. In this work-in-progress, we propose a closed-loop system that teaches itself about forward object dynamics without any human intervention. Our model consists of two parts. A forward dynamics model that models the transition between states and a policy model that tries to predict the dynamics model's error conditioned on object interactions as its intrinsic reward. We show that our method is able to train forward dynamics models that generalize to unseen physical scenarios and approaches the upper bound of models trained on human-curated data. The model generates complex behaviors with a preference to novel objects.

1 INTRODUCTION AND RELATED WORK

Cognitive science literature suggests that humans run physics simulations in their mind to plan and imagine the future (Battaglia et al., 2013; Bates et al., 2015; Hamrick et al., 2011; Ullman et al., 2014; Hegarty, 2004; Lake et al., 2017). Experiments show that humans are most likely born with a built-in basic understanding of objects and physics (Spelke & Kinzler, 2007) which gets refined through active experimentation with their environment driven by what is widely known as curiosity (Gopnik et al., 2009; Schmidhuber, 2010). In this work, we build on these insights by proposing a self-learning system that (a) relies on a trainable future predicting world-model with explicitly built-in structure and (b) that is trained through active learning driven by an intrinsic curiosity mechanism.

Recently, several approaches have been proposed to learn about physics from observations (Agrawal et al., 2016; Finn et al., 2016; Byravan & Fox, 2017; Bates et al., 2018; Tacchetti et al., 2018; Sanchez-Gonzalez et al., 2018; Battaglia et al., 2018; Ajay et al., 2019). Given trajectories of physical systems, such as a ball falling on the ground or two cubes colliding, models are trained to predict the future state of the system. Similarly to humans (Spelke & Kinzler, 2007), some of the best forward dynamics models build in object-, part- and relation-centric priors for learning physics (Chang et al., 2016; Battaglia et al., 2016; Mrowca et al., 2018; Li et al., 2018). We seek to combine a model-based reinforcement learning approach with an explicit dynamics model to model human interactive physical learning.

In the following, we explore how artificial agents can teach themselves about physics through object interactions in a box environment. The agent may apply forces on up to three objects in a box as shown in Figure 1. We propose a self-supervised system that trains an accurate forward dynamics model by taking intrinsically-motivated actions. Many curiosity-driven methods have been proposed (Chentanez et al., 2005; Oudeyer et al., 2007; Singh et al., 2010; Schmidhuber, 2010; Frank et al., 2014; Oudeyer & Smith, 2016; Achiam & Sastry, 2017; Pathak et al., 2017; Burda et al., 2018). We

[*] Equal contribution

build on previous work in which curiosity was used to supervise a latent physical dynamics model Haber et al. (2018) to supervise an explicit and interpretable forward dynamics model capable of long range prediction. Using a loss estimation network, we perform a search to find the maximum intrinsic reward for a single-step policy to maximize the model's learning. We show that our interactively trained dynamics model approaches the performance of models trained on manually created expert training datasets when generalizing to unseen scenarios. We demonstrate a preference for hard to learn objects and interactions in the agents behavior.

2 METHODS

Our system consists of a *world-model* (forward predictor) learning online from the agent's interactions with an environment. The agent chooses forces on objects with a *self-model*, which aims to find interesting data for the world-model. In what follows, we describe the environment and hierarchical particle relationship graph representation of objects that our models have access to. We then review the Hierarchical Relation Network Mrowca et al. (2018) world-model, before describing the graph-convolutional self-model architecture.

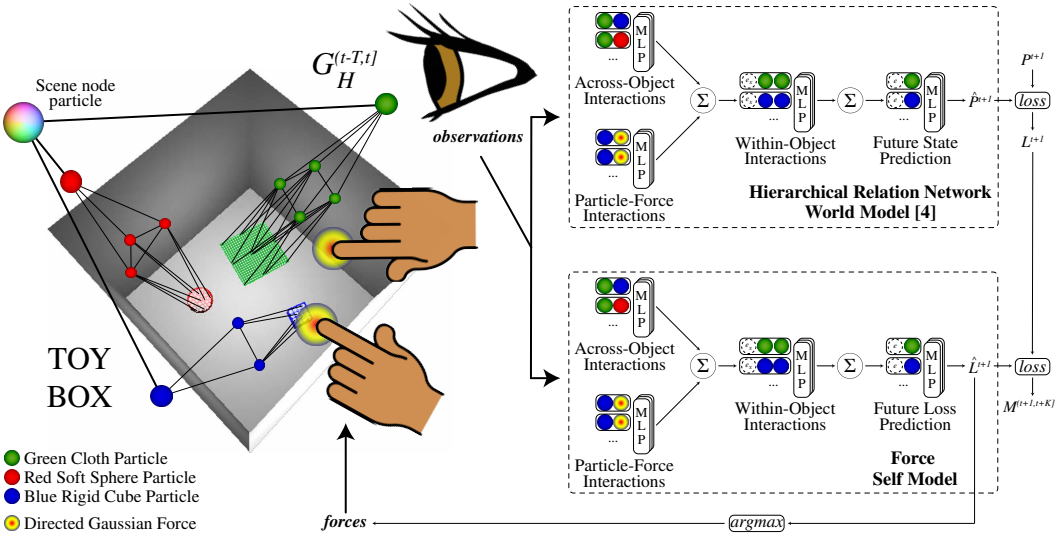

Figure 1: **Environment and Model**. *Left:* Box environment with red ball, green cloth, and blue cube decomposed into a particle relationship graph G_H. The agent can apply forces to the objects to generate physical scenarios of interest. *Right:* Given past observations $G_H^{(t-T,t]}$, the world-model (top right) predicts the next environment state and occurs a loss. The self-model (bottom) predicts the world-model loss given $G_H^{(t-T,t]}$ and randomly sampled actions. The action with the highest predicted loss is greedily executed.

Environment. Our environment (Figure 1 left) consists of a 5 x 5 x 5 unit box in which one to three objects of different shapes (cone, cube, cuboid, cylinder, ellipsoid, prism, pyramid, sphere, stick, torus) and materials (rigid, soft, cloth) are placed. To train its forward predictor, the agent is then tasked to perform short, fixed-length "experiments", with the following interface. (1) At the beginning of the experiment, the objects are randomly placed in the box. (2) The agent can then apply forces anywhere within the box in fixed intervals. The agent chooses the centers, directions, magnitudes, and standard deviations of these forces which are applied to object particles within one standard deviation of the centers, with magnitude proportional to a truncated Gaussian about the center. Intuitively, the agent "pokes" around in the entire box with its fingers and if it hits an object, the object will move. (3) Finally, the scene is reset and the whole process is repeated.

Hierarchical Particle Relationship Scene Graph Representation. In this work, we assume that the agent observes a particle representation of each object in the box. From this particle representation, we construct a hierarchical particle relationship scene graph representation G_H as seen in Figure 1. Graph nodes correspond to either particles or groupings of other nodes and are arranged in a hierarchy,

whereas edges represent constraints between nodes (e.g. materials). For a detailed description of this representation, please see Mrowca et al. (2018).

Forward Predictor World-Model. We use the Hierarchical Relation Network (HRN) as proposed by Mrowca et al. (2018) as the world-model (Figure 1 bottom). The HRN takes a history of hierarchical graphs $G_H^{(t-T,t]}$ as input. The model first computes collision (ϕ_C^W), external force (ϕ_F^W), and past effects (ϕ_H^W) between particles using pairwise graph convolutions. The effects are then propagated through the particle hierarchy using a hierarchical graph convolution module η^W. Finally, the fully-connected module ψ^W computes the next particle states P^{t+1} from the propagated effects and past particle states. The world-model is optimized with an $l2$ loss on next state particle velocities and pairwise distances. The HRN can be unrolled across an arbitrary number of time steps by feeding back its predictions as input.

Action-Proposing Self-Model. We propose a graph convolutional self-model that takes, as input, observation information, and produces a probability distribution over the action space that the agent samples from. The probability distribution is derived as follows: we propose actions for which we explicitly predict the loss of the world-model, and from that we choose that action with the predicted maximum loss. In doing so, the self-model chooses a policy so as to *antagonize* the world model. Intuitively, the world-model loss is the result of interaction of objects, objects and forces, and the coordination of forces applied to objects. Thus, the self-model has three components: an object-object effects component, an object-force effects component, and a force-force effects component. An object- and action-centric representation proved to be crucial. Intuitively, it does not matter if the same force is applied to an object with the left or right hand.

3 EXPERIMENTS

We evaluate the physical understanding and behavior of our self-supervised agent. We first train forward dynamics models on manually constructed action subsets and show that a diverse set of actions is necessary for generalization. We then compare the curiosity-driven agent against the agent trained on hand-designed data and demonstrate the curious agent's superior generalization to unseen scenarios. Finally, we analyze the behavior of our curiosity-driven agent and study its preference towards hard to predict objects and interactions.

Dynamics Model Accuracy and Generalization. To evaluate our forward dynamics model accuracy and generalization ability, we hand design separate training and validation datasets with objects of different shapes (*cube*, *sphere*, *cone*, *octahedron*, *pentagon*, *bowl*, *prism*) and materials (*various levels of softness*) consisting of the following subsets: (1) A **lift subset**, in which objects are repeatedly lifted off the ground and undergo parabolic motion. (2) A **slide subset**, in which objects are repeatedly pushed around on a surface under friction. (3) A **collide subset**, in which objects are repeatedly collided into each other. (4) And a **stack subset**, in which objects are repeatedly placed on top of each other forming a (un)stable stack depending on object geometry. Each subset consists of 90,000 training and 10,000 validation states.

Table 1: Dynamics model particle position mean-squared-error (**MSE**) for 20 time step predictions trained (**T**) and validated (**V**) on different action subsets.

T \ V	**Lift**	**Slide**	**Collide**	**Stack**	**All**
Lift	**0.18**	6.07	5.29	1.93	3.37
Slide	2.32	**0.28**	0.87	2.96	1.61
Collide	1.77	0.51	0.52	18.74	5.39
Stack	5.10	5.56	4.58	0.39	3.91
All	0.21	0.38	0.51	0.36	**0.37**
Shapes $\downarrow$	0.22	0.49	**0.47**	**0.32**	0.38
Materials $\downarrow$	0.22	0.52	0.53	0.35	0.41
# Objects $\downarrow$	0.22	0.66	0.69	0.45	0.51

Given two initial states, the dynamics model is trained to predict the next future state(s) at 100 ms intervals. We train separate models on each train subset and a model on all subsets and evaluate each model on all validation subsets by measuring the mean-square-error (MSE) between predicted and true particle positions. Table 1 summarizes the results for 20 time step predictions. While the dynamics model predicts the action subsets well on which it was trained on, it generalizes poorly to held-out action subsets. Only if trained on all subsets, does the model perform well on all of them and improves on the collide and stack subset benefiting from observing different action types during training. We also explored material, shape, and object number ablations, but found that the dynamics model generalizes well across these axes, due to its particle-centric architecture (see Table 1).

Dynamics Model Performance under Different Policies. As the hand designed data contains only a small subset of all possible object interactions, a curiosity-driven model interacting with objects in a box should be able to generate more varied training data for the dynamics model which thus should generalize better to unseen scenarios. This hypothesis is evaluated in the following on held-out test sets generated using the same procedure as before but with two additional shapes (*pyramid*, *cylinder*).

Table 2: Dynamics model performance on test set (**T**) under different policies (**P**).

P \ T	Lift	Slide	Collide	Stack	#F	#C
RP	0.54	0.69	0.61	0.82	0.12	4.7e-4
ORP	0.33	0.56	0.45	0.80	1.0	0.023
CP	0.46	0.52	0.44	0.59	0.13	4.5e-4
HDP	**0.15**	**0.14**	**0.23**	**0.26**	-	-

We compare dynamics models trained on data following policies that randomly select forces within the box (**RP**), randomly select forces on objects (**ORP**), or follow the hand designed policy (**HDP**) with dynamics models trained on data generated by our curiosity-driven driven policy (**CP**).

The dynamics model trained with RP does poorly. Given the size of the objects and the box, the majority of the box is empty and thus random forces rarely result in object motion. Random forces on objects (ORP) train a better dynamics model that however lacks in accuracy on stacks. Models trained with HDP, are trained on very similar interactions as tested on, providing an upper baseline. CP outperforms RP and ORP but does not reach HDP accuracy. We think this is due to CP not planing multiple steps ahead. CP applies forces which result in object collisions during training but eventually reaches an equilibrium with low force and collision count. Note that this quantitative difference is almost unnoticeable in qualitative rollouts. This is quite remarkable as the agent has no prior knowledge about objects and interactions unlike the expert who hand designed the data. Figure 2 shows how well CP generalizes to the lift and collide subset despite never being explicitly trained on any of those scenarios. A model only trained on the lift subset works well on lift test data but does not generalize to collide data.

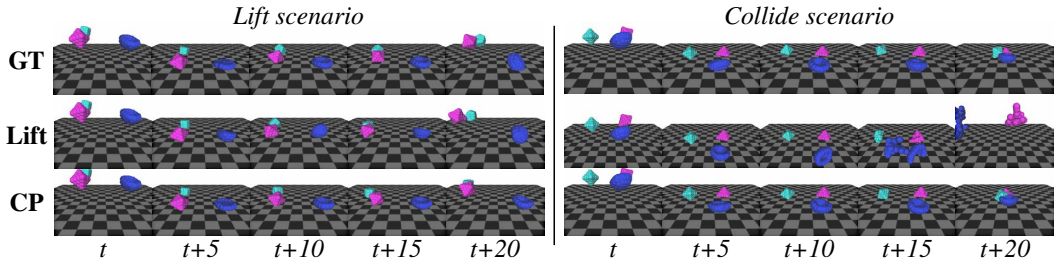

Figure 2: **Lift and Collide test rollouts**. Dynamics model predictions trained on **Lift** and with **CP** are compared against ground truth (**GT**). **CP** works well on both, **Lift** only on the Lift scenario.

Policy Model Behaviors and Preferences. We observe that the curiosity-driven agent learns to apply forces to objects and to collide objects into each other. To examine whether the agent shows a preference towards certain object shapes we systematically juxtapose two objects by placing them in the box and letting the agent choose to apply only one force. We measure how often the agent chooses to apply a force to one shape over the other over multiple trials. Figure 3 summarizes the results. We can see that the agent has a slight preference towards the unseen geometries (*pyramid*, *cylinder*) starred in the figure. Within the seen shapes, the agent develops a hierarchical order of preferences towards certain geometries (*pyramid* over *sphere*, *sphere* over *octahedron*) which indicates that these geometries are particularly hard to learn in comparison to the rejected shape and should be enriched in the training data.

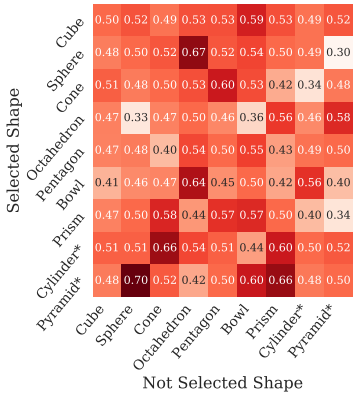

Figure 3: Object preferences.

Conclusion and future work. For future work, we are currently developing a curious tree search method for multi-step planning. Using the world-model, we choose actions at each layer of the tree by sampling according to the expected loss, and sum to the leaves of the tree in order to choose the maximally rewarding trajectory. We find this model shows behavior enriched for long horizon high loss states, such as collisions. Finally, we would like to compare our model's object and action preferences to human preferences with the goal of explaining infant play and human curiosity.

REFERENCES

Joshua Achiam and Shankar Sastry. Surprise-based intrinsic motivation for deep reinforcement learning. *CoRR*, abs/1703.01732, 2017.

Pulkit Agrawal, Ashvin V Nair, Pieter Abbeel, Jitendra Malik, and Sergey Levine. Learning to poke by poking: Experiential learning of intuitive physics. In *Advances in Neural Information Processing Systems*, pp. 5074–5082, 2016.

Anurag Ajay, Maria Bauza, Jiajun Wu, Nima Fazeli, Joshua B Tenenbaum, Alberto Rodriguez, and Leslie P Kaelbling. Combining physical simulators and object-based networks for control. *arXiv preprint arXiv:1904.06580*, 2019.

Christopher Bates, Peter Battaglia, Ilker Yildirim, and Joshua B Tenenbaum. Humans predict liquid dynamics using probabilistic simulation. In *CogSci*, 2015.

Christopher J Bates, Ilker Yildirim, Joshua B Tenenbaum, and Peter Battaglia. Modeling human intuitions about liquid flow with particle-based simulation. *arXiv preprint arXiv:1809.01524*, 2018.

Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. Interaction networks for learning about objects, relations and physics. In *Advances in neural information processing systems*, pp. 4502–4510, 2016.

Peter W Battaglia, Jessica B Hamrick, and Joshua B Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*, 2018.

Arunkumar Byravan and Dieter Fox. Se3-nets: Learning rigid body motion using deep neural networks. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 173–180. IEEE, 2017.

Michael B Chang, Tomer Ullman, Antonio Torralba, and Joshua B Tenenbaum. A compositional object-based approach to learning physical dynamics. *arXiv preprint arXiv:1612.00341*, 2016.

Nuttapong Chentanez, Andrew G Barto, and Satinder P Singh. Intrinsically motivated reinforcement learning. In *NIPS*, pp. 1281–1288, 2005.

Chelsea Finn, Ian Goodfellow, and Sergey Levine. Unsupervised learning for physical interaction through video prediction. In *Advances in neural information processing systems*, pp. 64–72, 2016.

Mikhail Frank, Jürgen Leitner, Marijn F. Stollenga, Alexander Förster, and Jürgen Schmidhuber. Curiosity driven reinforcement learning for motion planning on humanoids. *Front. Neurorobot.*, 2014, 2014.

A. Gopnik, A.N. Meltzoff, and P.K. Kuhl. *The Scientist In The Crib: Minds, Brains, And How Children Learn*. HarperCollins, 2009.

Nick Haber, Damian Mrowca, Li Fei-Fei, and Daniel LK Yamins. Learning to play with intrinsically-motivated self-aware agents. *arXiv preprint arXiv:1802.07442*, 2018.

Jessica Hamrick, Peter Battaglia, and Joshua B Tenenbaum. Internal physics models guide probabilistic judgments about object dynamics. In *Proceedings of the 33rd annual conference of the cognitive science society*, pp. 1545–1550. Cognitive Science Society Austin, TX, 2011.

Mary Hegarty. Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, 8(6): 280–285, 2004.

Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017.

Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B Tenenbaum, and Antonio Torralba. Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. *arXiv preprint arXiv:1810.01566*, 2018.

Damian Mrowca, Chengxu Zhuang, Elias Wang, Nick Haber, Li F Fei-Fei, Josh Tenenbaum, and Daniel L Yamins. Flexible neural representation for physics prediction. In *Advances in Neural Information Processing Systems*, pp. 8813–8824, 2018.

Pierre-Yves Oudeyer and Linda B Smith. How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, 8(2):492–502, 2016.

Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, 11(2):265–286, 2007.

Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *ICML*, volume 70 of *JMLR Workshop and Conference Proceedings*, pp. 2778–2787. JMLR.org, 2017.

Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, and Peter Battaglia. Graph networks as learnable physics engines for inference and control. *arXiv preprint arXiv:1806.01242*, 2018.

J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990 – 2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, Sept 2010.

Satinder P. Singh, Richard L. Lewis, Andrew G. Barto, and Jonathan Sorg. Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Trans. Autonomous Mental Development*, 2(2):70–82, 2010.

Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental science*, 10(1):89–96, 2007.

Andrea Tacchetti, H Francis Song, Pedro AM Mediano, Vinicius Zambaldi, Neil C Rabinowitz, Thore Graepel, Matthew Botvinick, and Peter W Battaglia. Relational forward models for multi-agent learning. *arXiv preprint arXiv:1809.11044*, 2018.

Tomer Ullman, Andreas Stuhlmüller, Noah Goodman, and Joshua B Tenenbaum. Learning physics from dynamical scenes. In *Proceedings of the 36th Annual Conference of the Cognitive Science society*, pp. 1640–1645, 2014.