On the Learning Mechanisms in Physical Reasoning

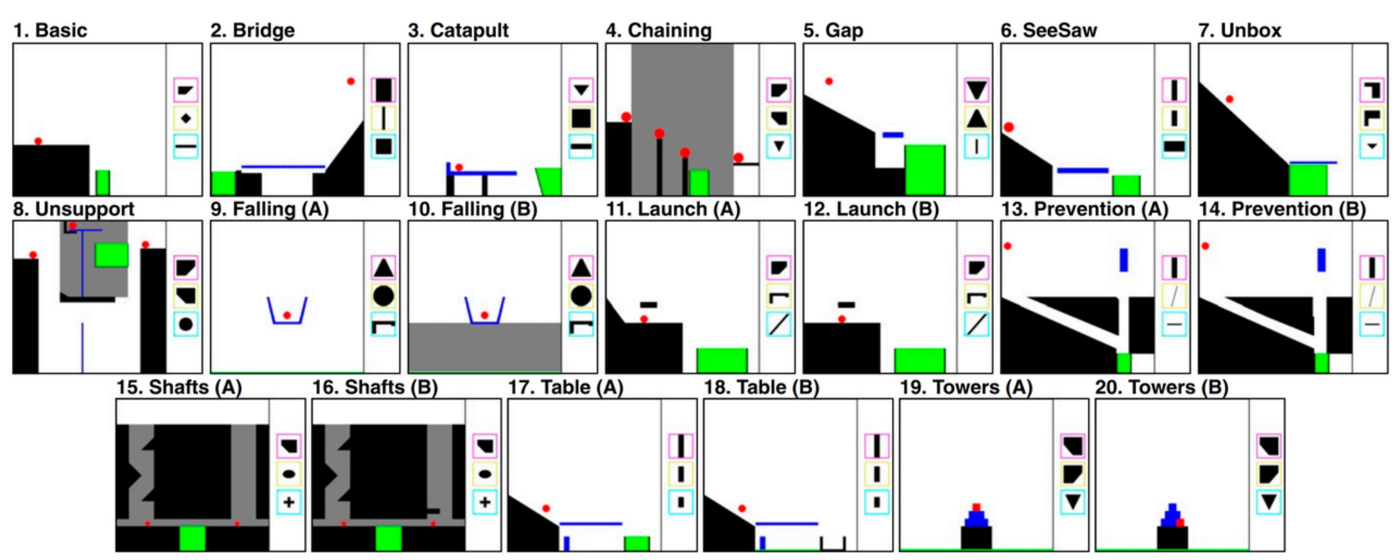
Shiqian Li[*,1,2,5], Kewen Wu[*,3,5], Chi Zhang[4,5] ✉, Yixin Zhu[1,2] ✉

[1]School of Intelligence Science and Technology, Peking University
[2]Institute for Artificial Intelligence, Peking University
[3]Department of Automation, Tsinghua University
[4]Department of Computer Science, University of California, Los Angeles
[5]Beijing Institute for General Artificial Intelligence (BIGAI)

NEURAL INFORMATION PROCESSING SYSTEMS

There are two reasoning processes in solving physical puzzles

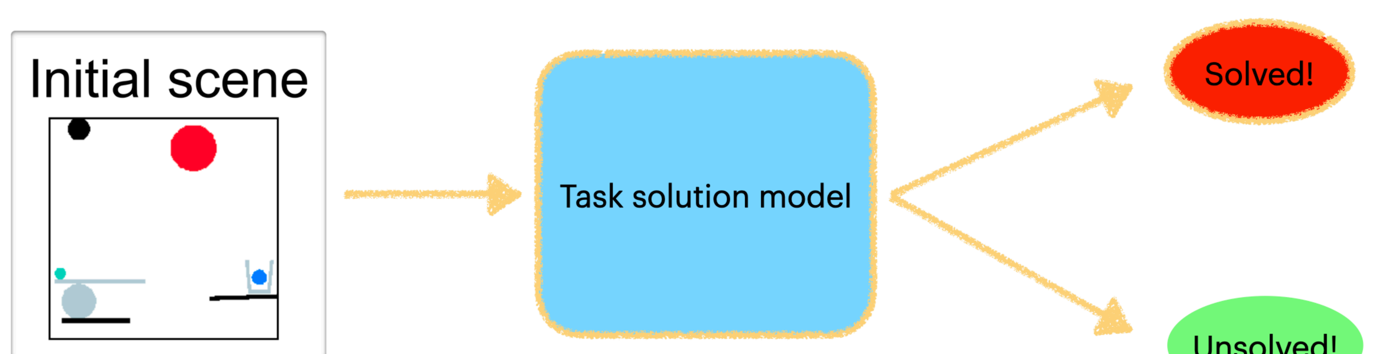
Physical intuition at a glance without much thinking.



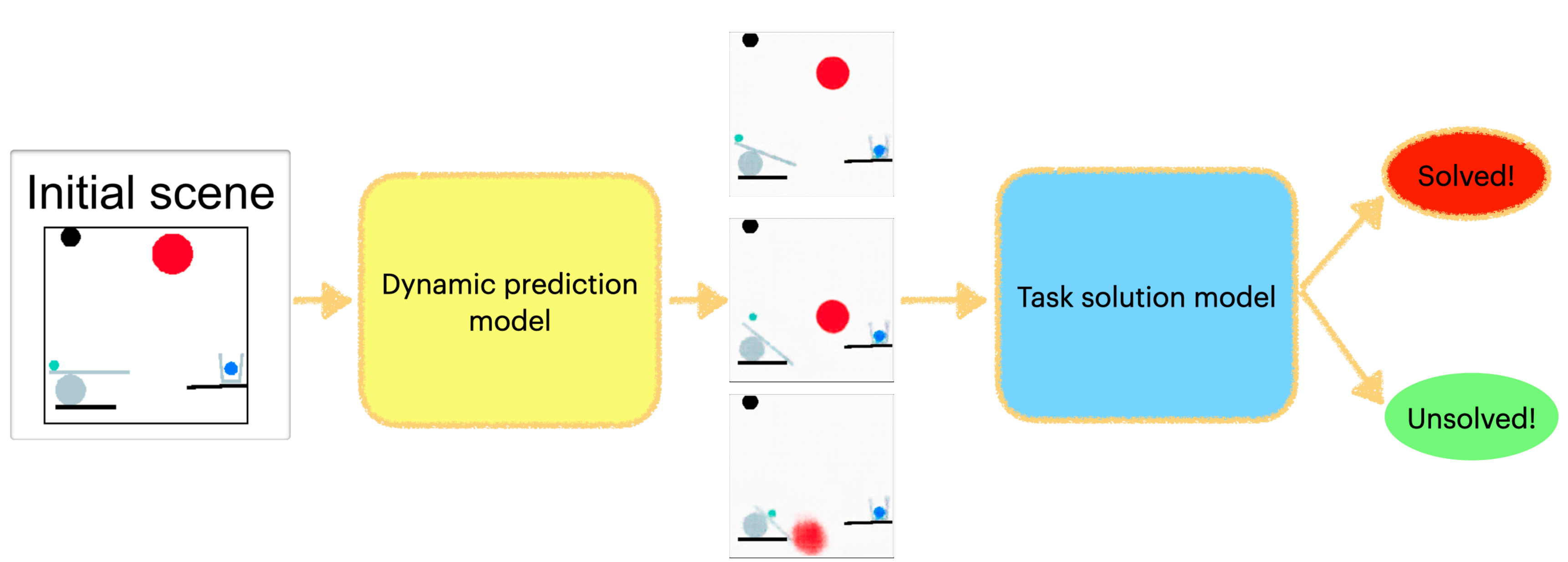
Unfolding of states under the assumed physical dynamics

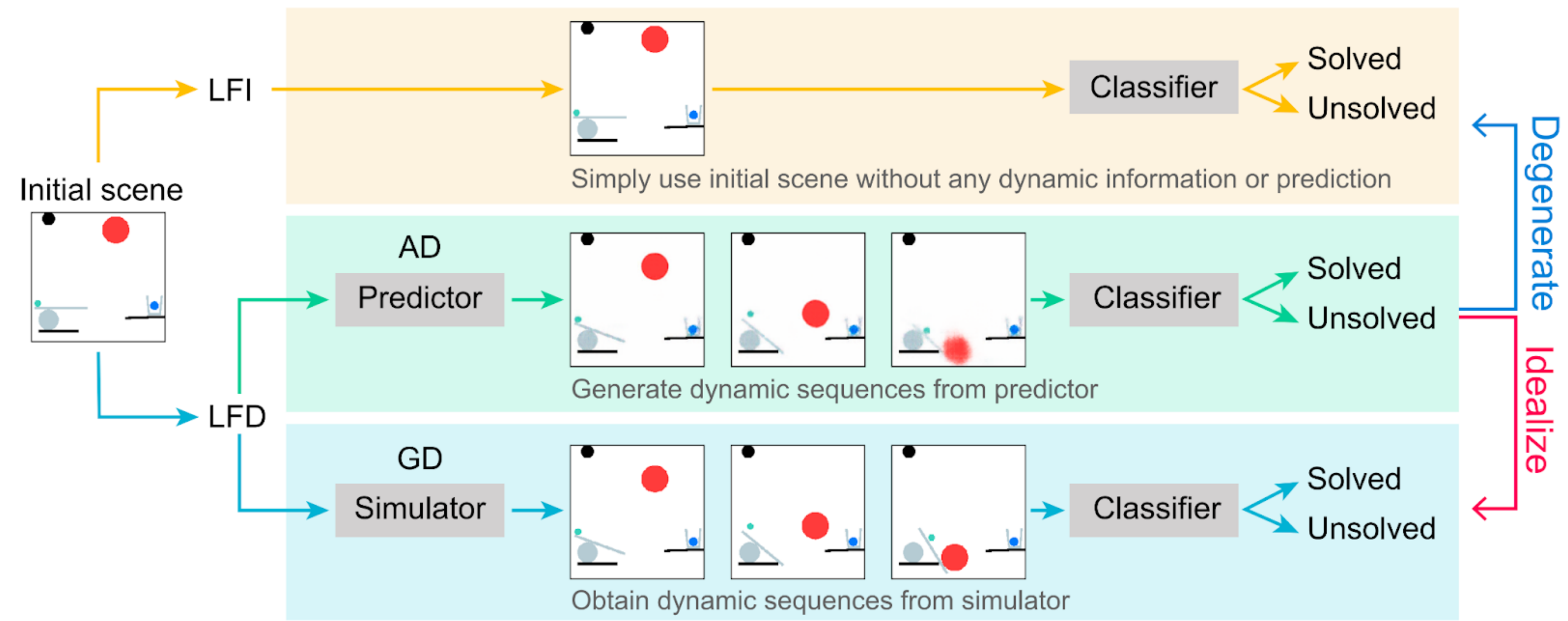
Definition

- Learning from Intuition (**LfI**)



- Learning from Dynamics (**LfD**)





Experimental Design

Exp1: LfD vs LfI

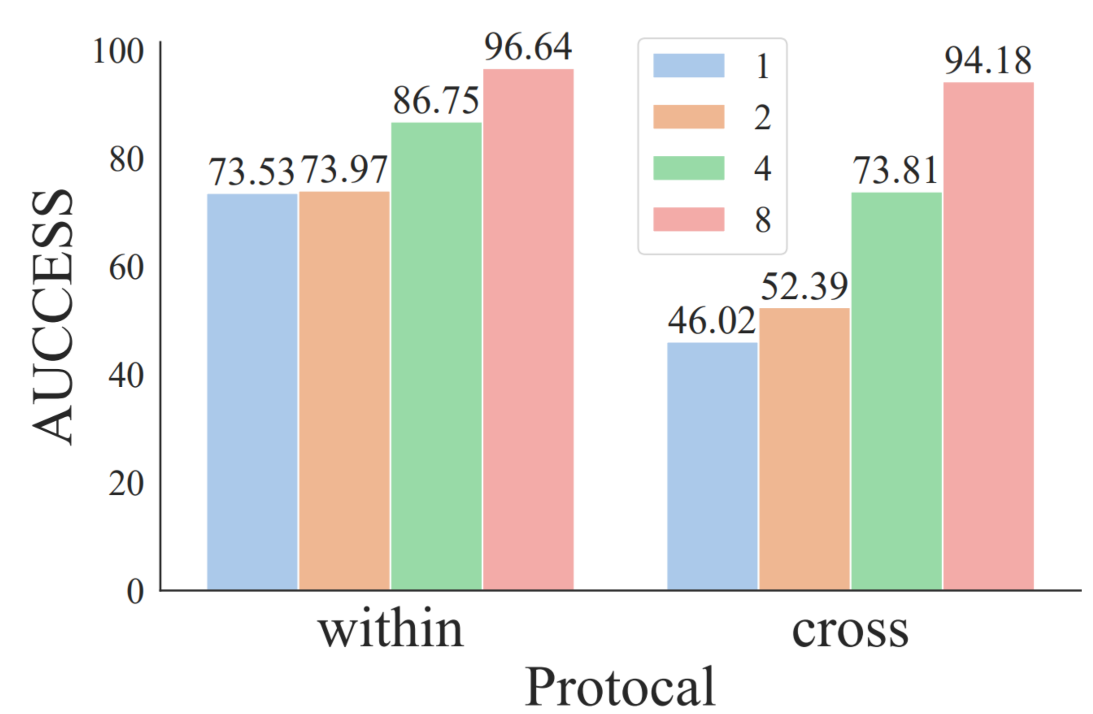
Challenge the previous thought

Model	Mechanism	Input	Supervision	Within	Cross
RPIN	Learning from Dynamics (LfD)	Initial scenes, bboxs	Bboxs, masks, outcomes	85.49	50.86
ViT	Learning from Intuition (LfI)	Initial scenes	Outcomes	84.16±0.30	56.31±1.95

A simple binary-classification Vision Transformer, which represents LfI, reaches or outperforms SOTA dynamic-based RPIN, which represents LfD.

Exp2: LfD under GD

Do ground-truth dynamics help make better decisions than intuition?



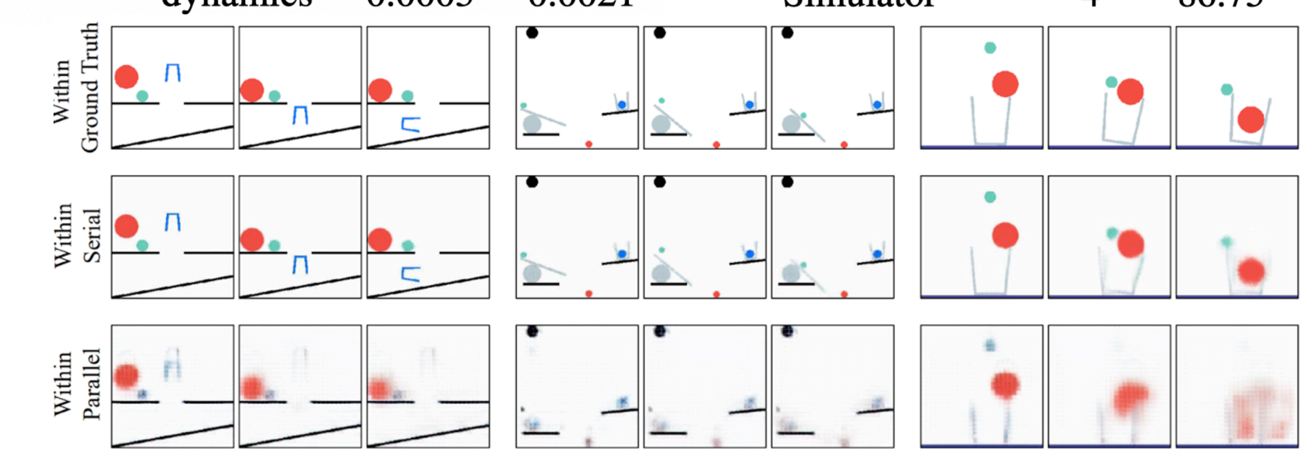
The second experiment serves as a diagnostic test for the efficacy of dynamics. We assume an ideal dynamics prediction model that accurately predicts the future. Specifically, we supply the model with ground-truth dynamics. The performance is significantly boosted with four or more input frames. Therefore, we conclude that accurate dynamics do help problem-solve in physical reasoning.

Exp3: LfD under AD

How do approximate dynamics perform?

Opt	Loss	Within	Cross
Parallel	entropy	0.0638	0.5726
	dynamics	0.0039	0.0049
Serial	entropy	0.1285	0.6554
	dynamics	0.0003	0.0021

Prediction	NF	Within	Cross
PredRNN (parallel)	4	75.22	46.42
PredRNN (serial)	4	64.90	44.33
/	1	73.53	46.02
Simulator	4	86.75	73.81



In the third experiment, we train the LfD pipeline using two optimization schedules, parallel and serial. The results show that independent of the optimization schedule used, LfD using approximate dynamics falls far behind LfD using ground-truth dynamics and performs equally or even worse than LfI, indicating that approximate dynamics do little help for the task-solution model in making better judgments.

Exp4: More on LfI

How does LfI perform?

Model	Mechanism	Object Info	Supervision	Within	Cross
ViT	LfI	False	Outcome	84.16±0.30	**56.31±1.95**
Swin	LfI	False	Outcome	84.71±0.33	54.92±2.30
BEiT	LfI	False	Outcome	83.59±0.09	54.07±1.88
Dec [Joint]	LfD under Approximate Dynamics (AD)	False	Dynamics & Outcome	79.73	52.64
RPIN	LfD under AD	True	Dynamics & Outcome	**85.49**	50.86

In the fourth experiment, we consider testing additional visual classification models to verify the effectiveness of LfI. The results show that LfI models are competitive with the SOTA LfD model and even outperform SOTA in unseen tasks. Besides the promising performance, LfI models also demonstrate merits: it is design-efficient, requires no extra task-specific prior knowledge, and can be easily pre-trained. Thus, we view LfI as a simpler and more effective paradigm for physical reasoning.

References:

Allen, Kelsey R., Kevin A. Smith, and Joshua B. Tenenbaum. "Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning." PNAS (2020)

Battaglia, Peter W., Jessica B. Hamrick, and Joshua B. Tenenbaum. "Simulation as an engine of physical scene understanding." PNAS (2013)

Bakhtin, Anton, et al. "Phyre: A new benchmark for physical reasoning." NeurIPS (2019)

Qi, Haozhi et al. "Learning Long-term Visual Dynamics with Region Proposal Interaction Networks", *ICLR* (2021)

Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, Joshua B. Tenenbaum, Song-Chun Zhu. "Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Humanlike Common Sense. " *Engineering* (2020)

Shiqian Li, Kenwen Wu, Chi Zhang, Yixin Zhu. "On the Learning Mechanisms in Physical Reasoning." NeurIPS (2022)