Intuitive physics via simulation

Peter Battaglia
DeepMind

Intuitive Physics Workshop
What is intuitive physics?
What is intuitive physics?

Part of our physical intelligence:

Our ability to perceive, understand, and act with the everyday environment
“Knowing what is where”
Physical intelligence

Perceive ➔ Understand ➔ Act
Physical intelligence

Perceive → Understand → Act

Inverse rendering
Physical intelligence

Perceive → Understand → Act

Inverse rendering → Inverse dynamics
What is intuitive physics?

Perceive → Understand → Act

Inverse rendering  Intuitive physics  Inverse dynamics
What is intuitive physics?

A core system of knowledge about the physical environment

- Provides internal representations of the world
- Can be queried:
  - “What will happen?”
  - “What has happened?”
  - “Why?”
  - “How?”
  - “What if?”
What makes intuitive physics special?
What makes intuitive physics special?

“Infinite use of finite means”

Combinatorial generalization:
- A few simple rules
- Composable to support a wide range of predictions and inferences
“Infinite use of finite means”
What is the mechanism of human intuitive physics?

“Simulation as an engine of physical scene understanding”
Battaglia, Hamrick, & Tenenbaum (2013) PNAS
What is the mechanism of human intuitive physics?

Will it fall?  In which direction?  Different masses

Complex scenes  Infer the mass  Predict fluids

with: Jess Hamrick, Tom Griffiths, Chris Bates, Josh Tenenbaum
What kind of simulator is good for intuitive physics?

It should be:

- **Flexible**
  - Handle domains beyond what “Bullet” or “PhysX” can handle
  - Even beyond physics: social interaction, other complex systems

- **Learnable**
  - Can improve with experience

- **Generally useful**
  - Support predictions, inferences, planning
Structured models
+ Simulation
+ Deep Learning

Interaction Networks

“Interaction Networks for Learning about Objects, Relations and Physics”
Battaglia, Pascanu, Lai, Rezende, & Kavukcuoglu (2016) NIPS.
Interaction Networks

Relational reasoning
Compute interaction

Object reasoning
Apply object dynamics

Predictions, inferences

$\mathbf{v}_1 \xrightarrow{e} \mathbf{v}_2$
Interaction Networks

- Relational reasoning
  - Objects, relations
  - Compute interaction

- Object reasoning
  - Effects
  - Apply object dynamics

- Relations \((E)\)
- Objects \((V)\)
- Per-object external signals \((U)\)

**Relation model**
- Shared MLP \((f')\) applied per-relation

**Object model**
- Shared MLP \((g)\) applied per-object

Mathematical Formulas:
- \(f(v_1, v_2, e_1) \rightarrow \tilde{e}_1\)
- \(f(v_3, v_2, e_2) \rightarrow \tilde{e}_2\)
- \(0 \rightarrow \tilde{v}_1\)
- \(\tilde{e}_1 + \tilde{e}_2 \rightarrow \tilde{v}_2\)
- \(0 \rightarrow \tilde{v}_3\)

Predictions, inferences
- \(g(v_1, \tilde{v}_1, u_1) \rightarrow v'_1\)
- \(g(v_2, \tilde{v}_2, u_2) \rightarrow v'_2\)
- \(g(v_3, \tilde{v}_3, u_3) \rightarrow v'_3\)
Interaction Networks

- Gravitational forces
- Object features:
  -- Position, velocity, mass

- Rigid collisions
- Object features:
  -- Pos., vel., mass, scale, shape
- Relation features:
  -- Elasticity

- Springs + rigid collisions
- Object features:
  -- Pos., vel., mass, scale, shape
- Relation features:
  -- Spring: spring coefficient, rest length, damping
  -- Rigid: elasticity
- Global features:
  -- Gravitational acceleration
Training and rollouts

Input: Full state(t)
Target: Velocity(t+1)

Rollouts: Predictions back in as inputs, for 1000 time steps

Generalizes:
- To systems of different sizes and structures
- From next-step predictions to 1000+ step rollouts
Details

Datasets: 1000k training, 200k validation, 200k test

Training epochs: 2000 x 10k mini-batches (100 examples per mini-batch)

Architecture:
- Relation model: 4 x 100 units
- Object model: 1 x 100 units

Noise added to inputs
L2 penalty on effects
L2 regularization on parameters
Training: n-body - 6 bodies
Generalization: n-body - 3 bodies

True

Model

Time

DeepMind  @DeepMindAI

Peter Battaglia  @PeterWBattaglia
Generalization: n-body - 12 bodies
Training: Balls - 6 balls, 4 walls
Generalization: Balls - 3 balls, 4 walls

True

Model

Time
Generalization: Balls - 9 balls, 4 walls

True

Model

Time

Time
Training: String - 15 masses, 1 end pinned
Generalization: String - 30 masses, 0 ends pinned
Generalization: String - 30 masses, 2 ends pinned
Dynamics predictions

- **n-body**
  - MSE (log-scale)
  - IN (6 obj): 6
  - IN (3 obj): 3
  - IN (12 obj): 12

- **Balls**
  - MSE (log-scale)
  - IN (6 obj): 6
  - IN (3 obj): 3
  - IN (9 obj): 9

- **String**
  - MSE (log-scale)
  - IN (15 obj, 1 pin): 15
  - IN (5 obj, 1 pin): 5
  - IN (30 obj, 1 pin): 30
  - IN (15 obj, 0 pin): 15
  - IN (15 obj, 2 pin): 15

Legend:
- Black: Constant velocity
- Gray: Baseline MLP
- Red: Dynamics-only IN

DeepMind  @DeepMindAI
Peter Battaglia  @PeterWBattaglia
Global inferences: Potential energy

- Interaction Network far outperforms MLP:
  -- n-body: Interaction Network 1.4 MSE vs. MLP 19 MSE
  -- String: Interaction Network 1.1 MSE vs. MLP 425 MSE
Structured scene understanding

- Interaction networks support scene classification on the basis of relational structure (Raposo, Santoro, Barrett, Rascanu, Lillicrap, Battaglia [Under review] ICLR)

Artificial scene categories, only discriminable by spatial relations

Our experiments showed that interaction networks can:
  - Classify scenes
  - Infer novel scene structures
  - Learn object factorizations from input states or images
  - Support one-shot learning
Imagination-based metacontroller

- Uses interaction network for model-based decision-making
- See Jessica Hamrick’s poster, and talk at 3:30p
Take-homes

● Intuitive physics
  ○ A core system of knowledge about the physical environment
  ○ Can be queried: “What will happen?”, “What has happened?”, …
Take-homes

● Intuitive physics
  ○ A core system of knowledge about the physical environment
  ○ Can be queried: “What will happen?”, “What has happened?”, ...

● Simulation is a mechanism of human intuitive physics
Take-homes

● Intuitive physics
  ○ A core system of knowledge about the physical environment
  ○ Can be queried: “What will happen?”, “What has happened?”, ...

● Simulation is a mechanism of human intuitive physics

● Interaction networks
  ○ First general-purpose learnable physics engine
  ○ Object- and relation-centric reasoning
  ○ Rich generalization
  ○ Also supports:
    ■ Structured scene understanding
    ■ Model-based decision-making
Integrated models of physical intelligence

Perceive ➔ Understand ➔ Act

Inverse rendering ➔ Intuitive physics ➔ Inverse dynamics

eg., PhysNet, Galileo
e.g., Interaction Networks, Neural Physics Engine
e.g., Imagination-Based Metacontroller, Learning Billiards

Act ➔ Perceive ➔ Understand

Relation model ➔ Object model

Inverse rendering

Intuitive physics

Inverse dynamics