CIMARRON

Continuous Iterative Modeling and Repair in Response to Novelty

Andy Zane, Kaleigh Clary, Justin Clarke, David Westbrook,
Przemyslaw Grabowicz, David Jensen

Knowledge Discovery Laboratory
University of Massachusetts Amherst

AI Birds Competition
IJCAI 2021

August 25, 2021

UMASS
AMHERST

Agent Overview

The "Big idea"

Novelty detection and repair via causal probabilistic programming

1

Highly expressive languages for causal models (2d physics simulator via SimbaDD)

2

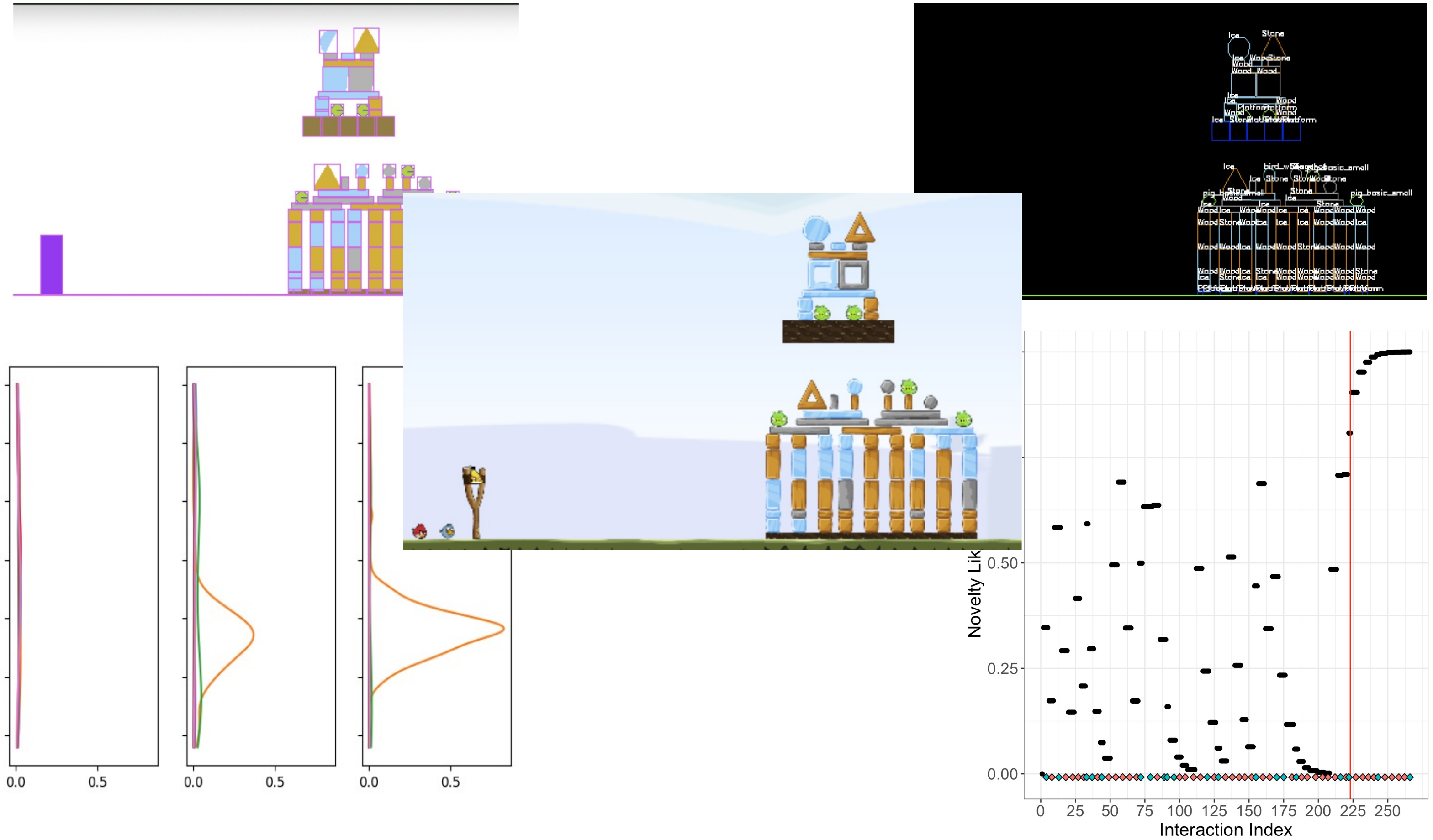
Detection of static novelties and simulation instantiation

3

Reasoning under uncertainty about multiple static and dynamic novelty detections

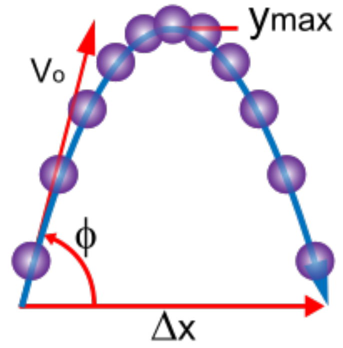
4

Novelty response via program diagnosis and repair (parameter inference for novel ontology components)

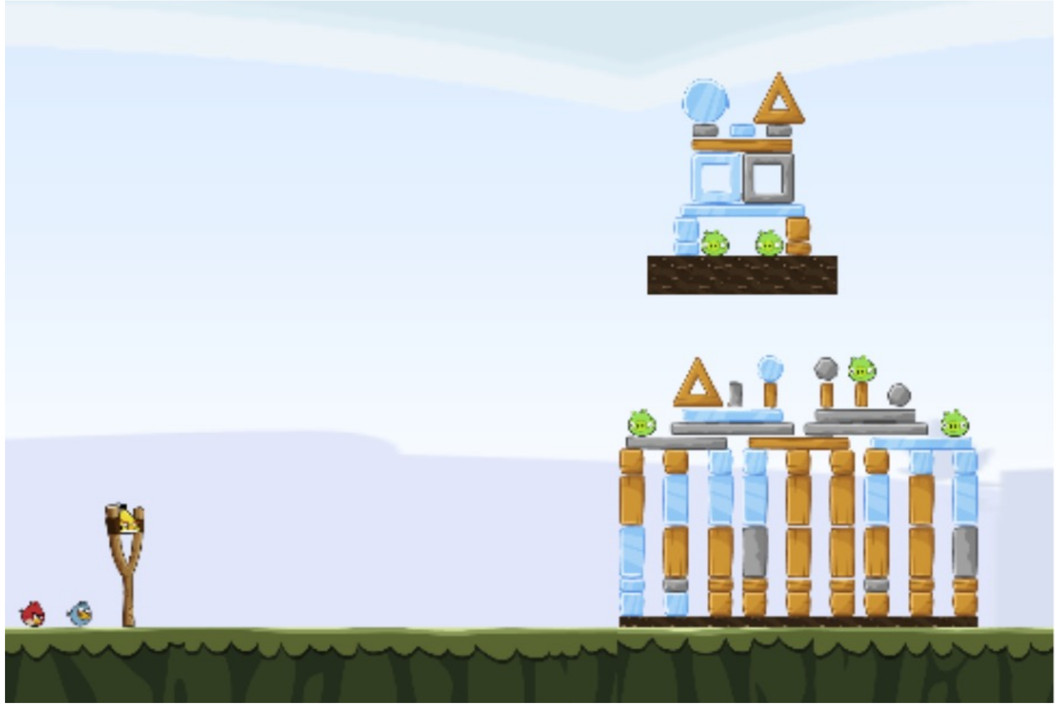


Exploiting classical mechanics (simulator)

- We implemented a set of methods for effective causal reasoning using classical mechanics (AKA "rigid body physics", "the physics you learned in high school", "2D physics engine").
- We use SimbaDD – an Angry Birds simulator developed by the ICCL lab in Dresden
- Classical mechanics provides a pre-existing language for constructing causal models of systems of interacting physical entities. This language has several advantages:



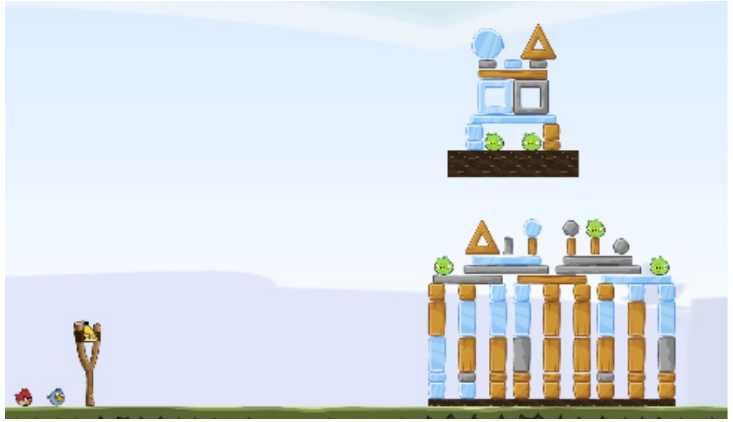
  - **Interpretability** — Classical mechanics provides a large set of known *causal mechanisms* for specifying how entities (e.g., blocks, platforms, birds, pigs) with given attributes (e.g., mass, velocity) affect each other.
  - **Autonomous** — The manner in which entities interact is invariant to how those entities acquired their particular attributes. That is, the causal mechanisms are invariant to intervention on inputs. This is often referred to in the causal inference literature as "autonomy" or "invariance under intervention".
  - **Compositional** — The behavior of large sets of entities can be derived by composing the interactions of smaller numbers of components. This follows directly from autonomy.



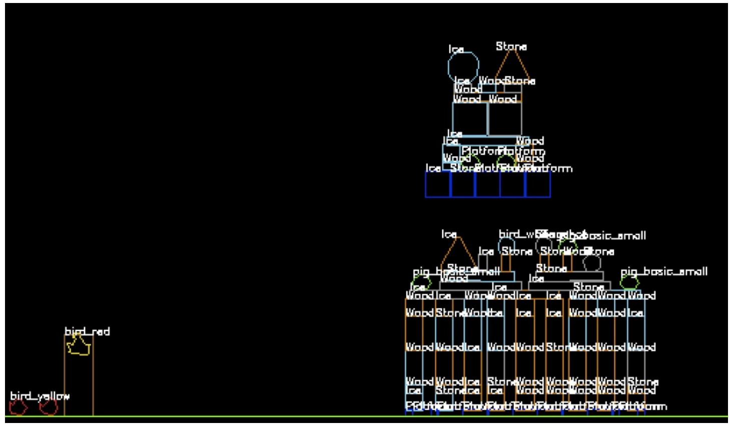
Instantiating complex causal models (simulations)

- We <u>use structured and visual evidence</u> to instantiate complex causal models based on an ontology and known causal mechanisms (in modified SimbaDD simulator).
- This contrasts sharply with nearly all current work in causal models, in which the causal mechanisms are *extremely simple and unknown* (whereas, in our approach, the causal mechanisms are complex and largely known *a priori*).
- We trained and applied a multi-class classification model to infer object type and material in a pre-novelty environment. Object type and material, in turn, imply attribute values (mass, hardness, etc.).

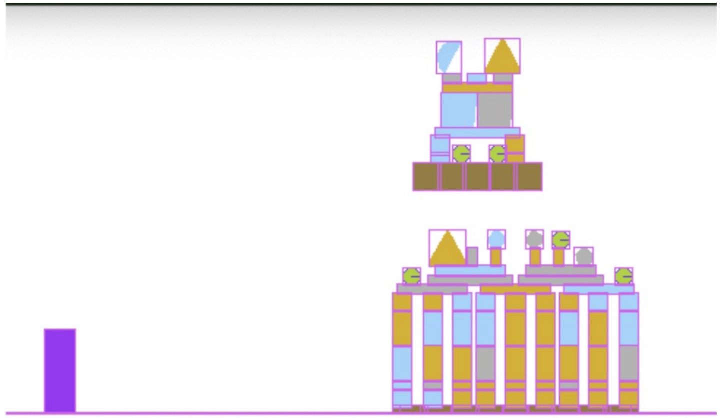
Original Game



Recognized Objects

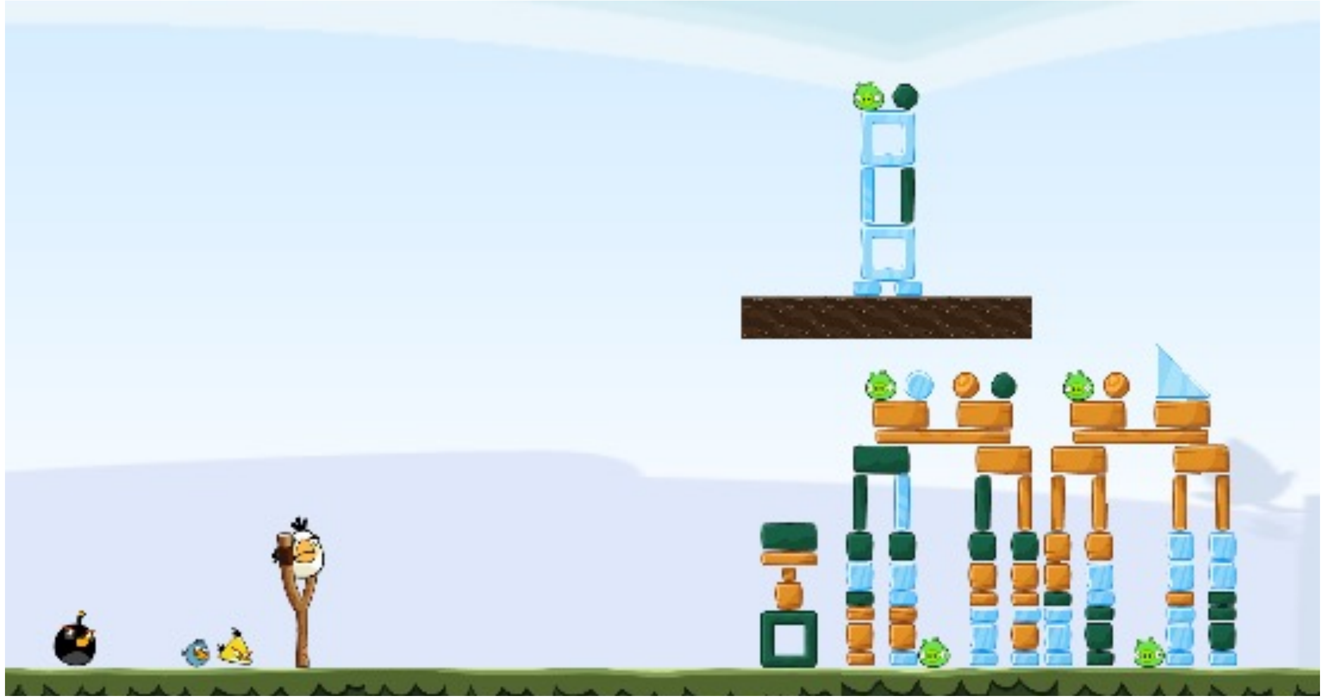


Instantiated Simulation



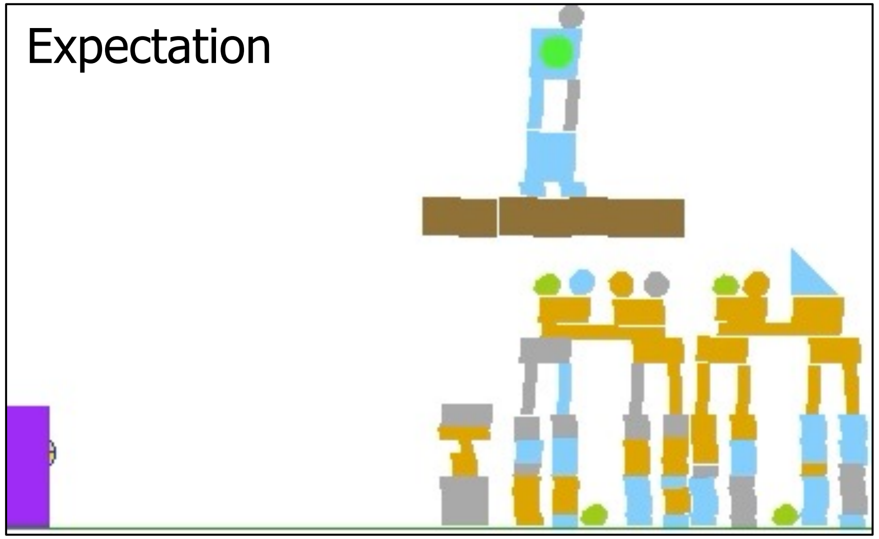
Detecting static visual novelty

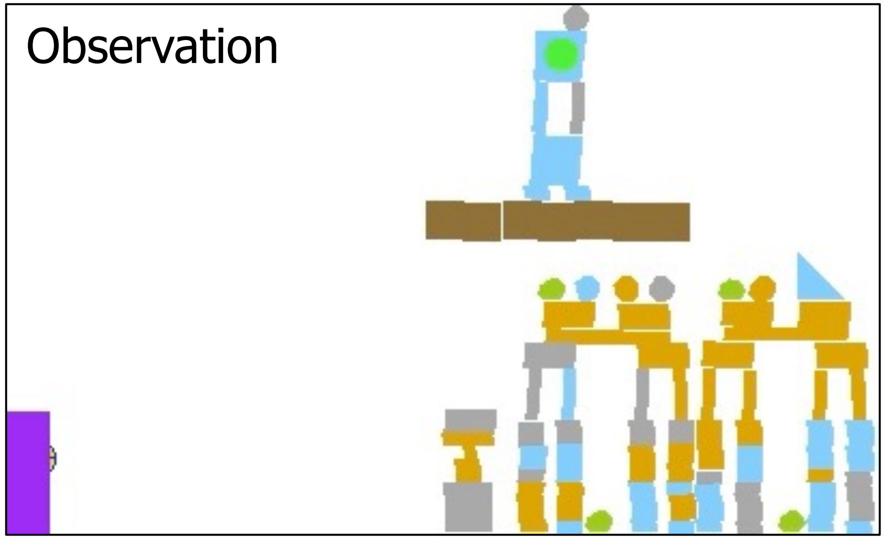
- We detect visual novelty using a multi-class abstaining classifier.
- Nearly all existing classification methods assume a fixed and known ontology of types. We used a multi-class abstaining classifier to identify some objects as novel.
- We replace all detected novel objects with dummy objects (we sample their parameters from some priors) to instantiate and reason about them in simulations



Detecting novelty in dynamics

- We measure the deviation between the behavior of the environment and our causal model in predicting the effects of an action.
- To compute this deviation, we exploit the causal model, which allows what-if simulations. We compare:
    - What would have happened if novelty was not introduced, and an action was taken (expectation)
    - What has happened in the environment given the same action was taken (observation)
    - If the two effects are significantly different, then there is a novelty in the environment

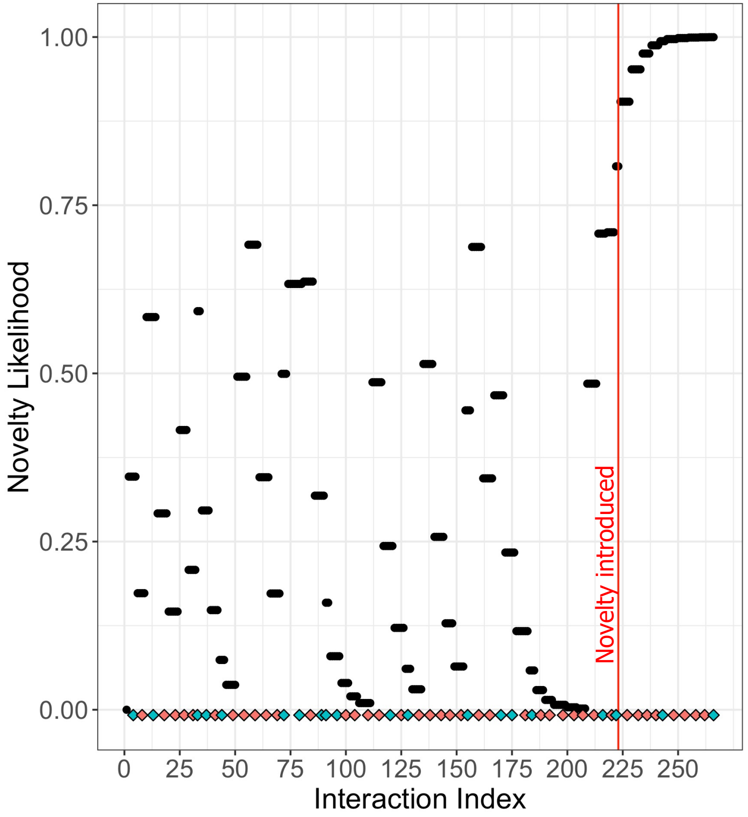

Expectation

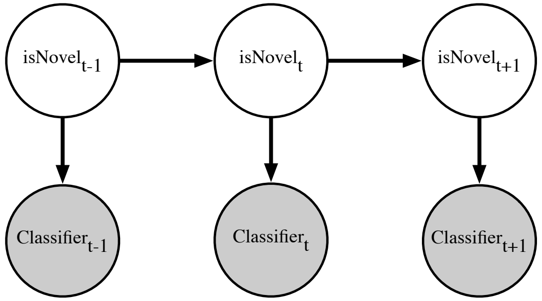

Observation

UMASS AMHERST

Bayesian aggregation of signals from novelty detectors

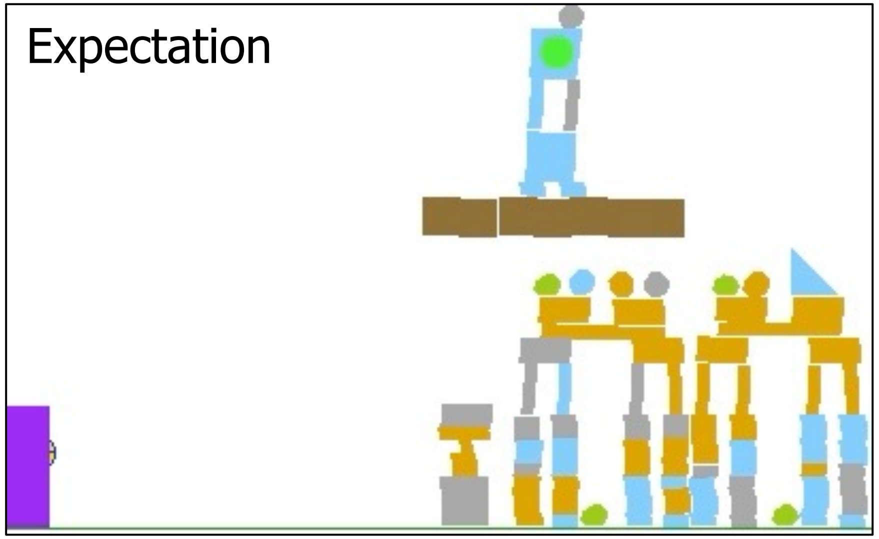
- One of the key challenges of accurate novelty detection is providing accurate inferences about the probability of novelty given a set of imperfect evidence over time.
- Our approach formalizes novelty detection as probabilistic inference in a latent variable model resembling a hidden Markov model
- Object classifiers and level-wise detectors provide probabilistic measurements of whether novelty has occurred within a given level.
- An agent's belief that there is novelty at level *t* informs its inferences at level *t+1*.
- A prior over novelty defines the agent's beliefs over when and if novelty will be introduced in the trial and that, once novelty has been introduced, whether it persists.

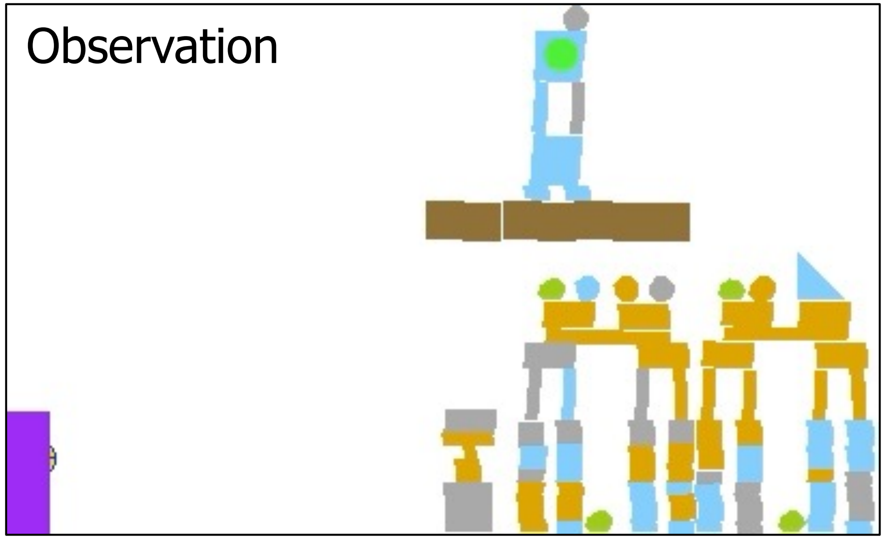




Bayesian inference about novelty

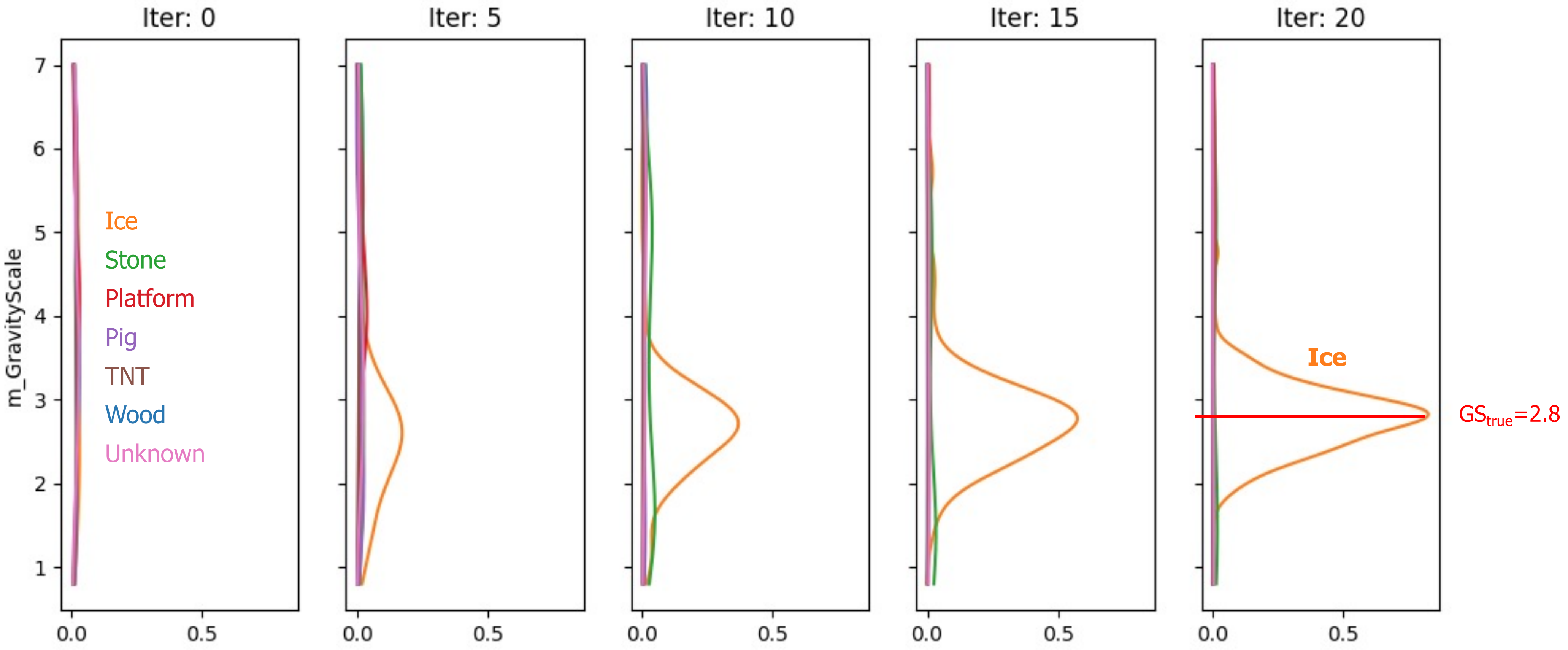
- We prototyped methods for using Bayesian inference to adapt to novelty.
- Given that novelty has been detected, we adapt to that novelty by minimizing the deviation between the behavior of the environment and our simulation (a pseudolikelihood of the data given the model).
- We consider a set of hypotheses about potential novelty. Each hypothesis corresponds to a prior distribution over one or more parameters in the agent's current ontology.
- The manner of novelty detection can influence the prior over hypotheses themselves and over the specific form of the prior on the altered parameters of the ontology.


Expectation


Observation

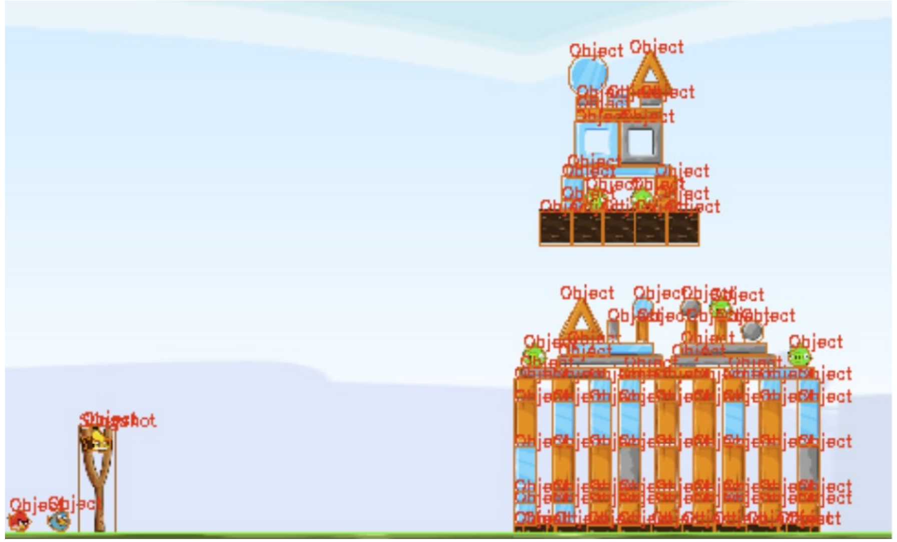
Bayesian inference about novelty (continued)

- In this example, the correct hypothesis (that the gravity scale for ice is novel and equal to 2.8) competes with hypotheses that the gravity scale of each of six other materials is novel.
- As inference iterations progress (left to right), the agent converges on the correct hypothesis.



Insight: New types of causal inference problems

- Work on Angry Birds emphasizes a type of causal inference that is rarely studied, despite its ubiquity in human reasoning: Causal inference that composes known mechanisms into complex and previously unseen configurations that match rich evidence (e.g., object positions, visual appearance, and behavior under intervention).
- This contrasts with the more typical approach that has been widely studied in computer science, statistics, social science, and philosophy: Estimating the parameters of simple but unknown mechanisms in simple configurations that match weak evidence (e.g., conditional independence of variables under observation).
- This new type of causal inference provides opportunities to meet the following goals:
  - *Novelty detection* can be conceptualized as detecting interventions in a nearly correct causal model.
  - *Novelty response* can be conceptualized as augmenting an already substantial ontology of causal mechanisms with one or more novel or altered components.
- These opportunities would not be available under traditional causal models, because nearly everything about the model would need to be learned *de novo*.


Object
Object
Object
Object
Object
Object
Object
Object
Object
Slingshot
Object
Object