Causal Learning and Learning to be Causal

Noah Goodman

Josh Tenenbaum

June, 2006

Special thanks to:

Members of the MIT CoCoSci group,

Rebecca Saxe,

Laura Schulz.

This work is made possible by:

The James S. McDonnell Foundation Causal Learning Collaborative Initiative

ndg@mit.edu

On the origin of causal knowledge

- Hume* observed that there must be some framework knowledge -- a 'causal sense' -- which allows us to infer causal structure from observed conjunctions of events.
- Two questions follow:
  - Specific: how does this causal framework guide the learning of causal structure?
  - General: what is the origin of causal framework knowledge?

*My naive reading of Hume, inspired by Garrett.

ndg@mit.edu

On the origin of causal knowledge

- Hume answered the general question with associational empiricism.
- Recent work suggests a more complex causal framework (cf. Woodward, Pearl, Gopnik, Glymour...).
- This buoys a nativist answer to the general question.
- Whither détente? Can complex causal framework knowledge be learned?

ndg@mit.edu

Objective

- In the following we apply Bayesian techniques to address the in-principle questions of learning framework and specific causal knowledge.
- This will (I hope) give insight into possibilities for the developmental origin of causal knowledge.

ndg@mit.edu

Learnability Analysis

- Bayesian induction is a calculus for rational belief assignment.
- As such it tells us the 'best use' of experience and abstract knowledge for learning, allowing us to answer learnability questions.

ndg@mit.edu

Bayesian Learning

For each hypothesis $h \in \mathcal{H}_F$,

$$P(h|d) \propto P(h)P(d|h)$$

- Framework knowledge, F, determines the hypothesis space, $\mathcal{H}_F$, and prior, $P(h) = P(h|F)$.
- Degree of belief in a hypothesis is then determined by its prior and the probability of the data under that hypothesis.

ndg@mit.edu

The Causal Bayes Nets Framework

- The work of Pearl, Glymour, etc. provides a way to formalize causal framework knowledge:
  - Hypotheses are directed acyclic graphs with an arrow-breaking intervention for each variable -- that is Causal Bayes Nets (CBNs).
  - This way of 'being causal' leaves out some other important, especially domain specific, aspects of causality. (We take the CBNs to be generically parameterized, and use a uniform prior over DAG structures.)

ndg@mit.edu

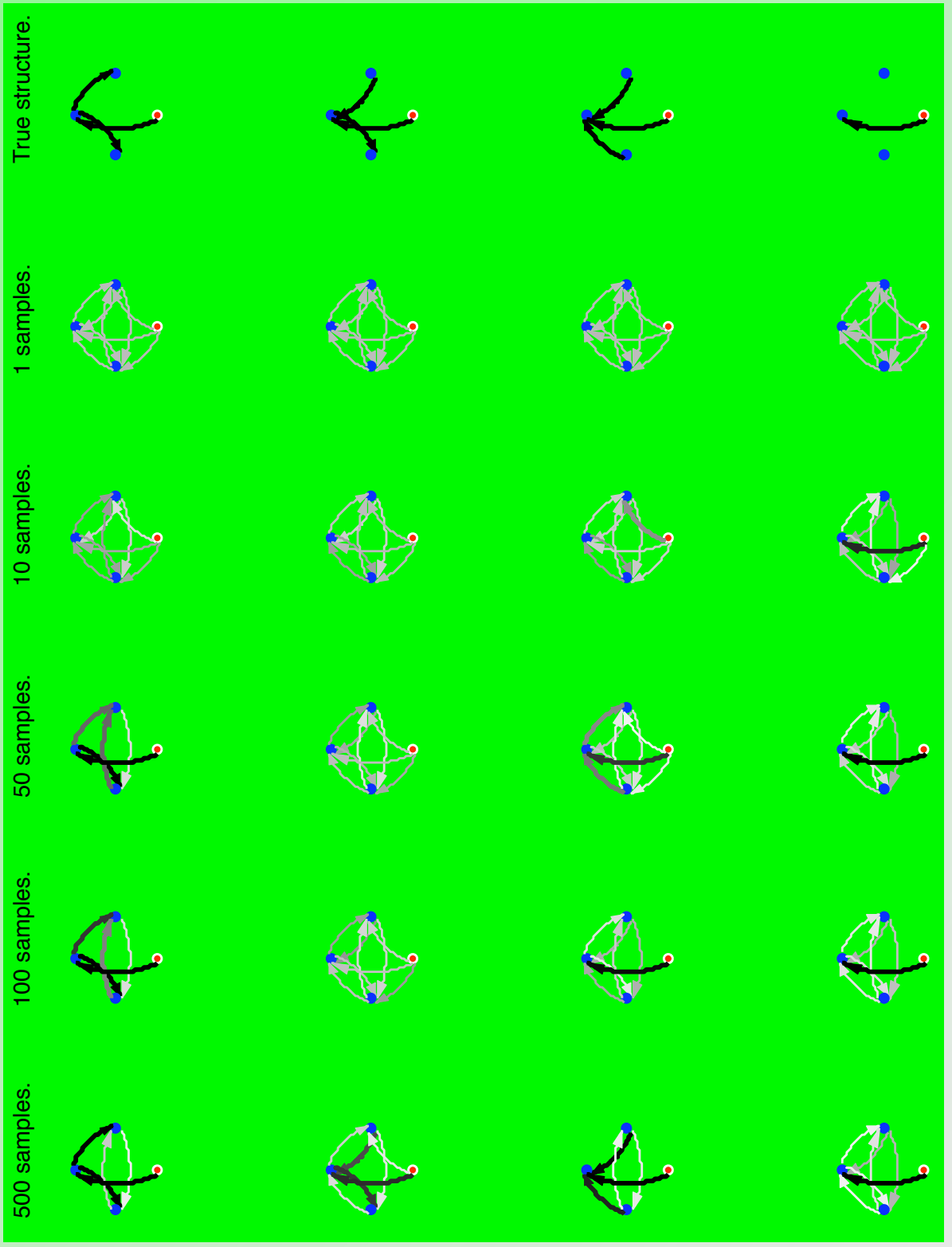
Learning CBNs

- Since each (parameterized) CBN gives a joint distribution, it provides a likelihood function -- allowing Bayesian learning.

ndg@mit.edu

Bayesian learning of specific causal structure, using the CBN framework.

Marginal probability of each edge:

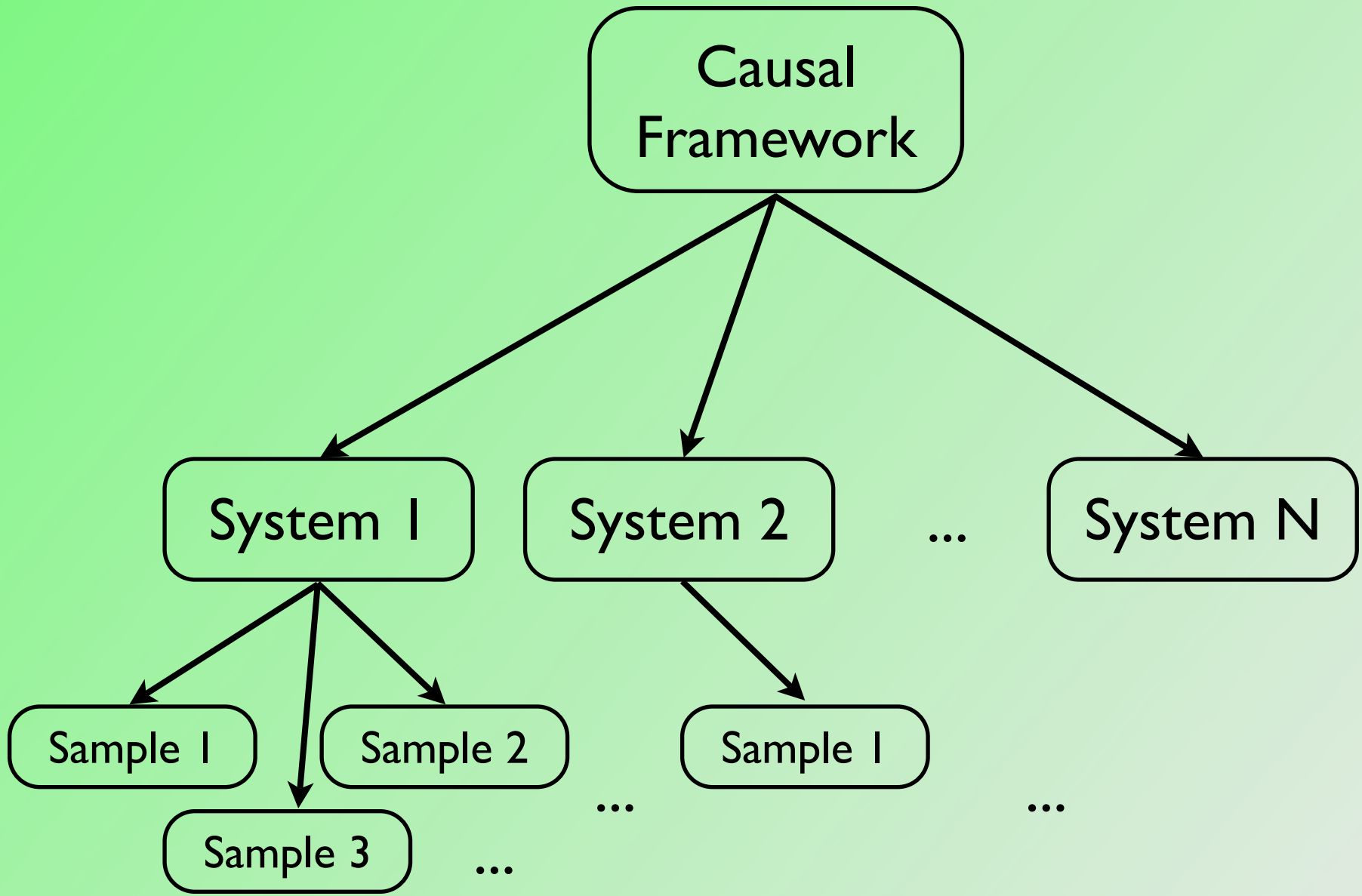


ndg@mit.edu

We've fixed this
CBN Framework
To learn these
System 1
System 2
...
System N
Sample 1
Sample 2
Sample 3
...
Sample 1
...
...
Observed these...

ndg@mit.edu

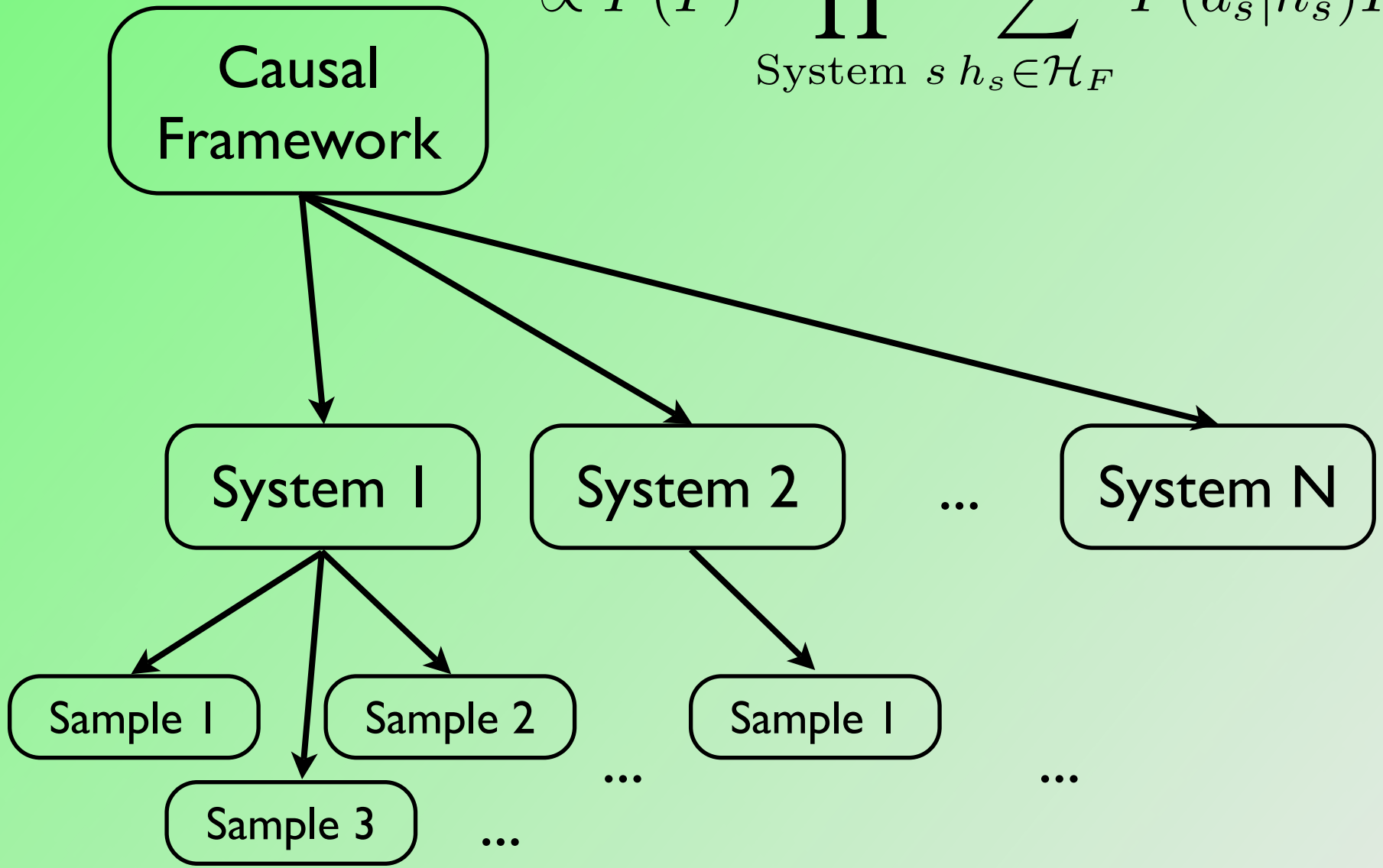
But perhaps we can learn the framework too?



ndg@mit.edu

$$P(F|d) \propto P(F) \prod_{\text{System } s} P(d_s|F)$$

$$\propto P(F) \prod_{\text{System } s} \sum_{h_s \in \mathcal{H}_F} P(d_s|h_s)P(h_s|F)$$



ndg@mit.edu

Learning to be causal

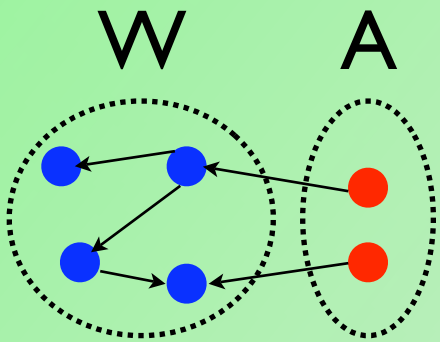
- We need a hypothesis space of alternative causal frameworks that spans from 'not very causal' to 'more causal'.
  - Full Joint: each system is represented as a joint distribution (a list giving the probability for each conjunction of events).
  - DAG: joint distributions represented by local, asymmetric, factorizations (ie. Bayes nets).
  - CBN: DAG plus intervention variables.
- Are these part of a larger, more uniform, hypothesis space?

ndg@mit.edu

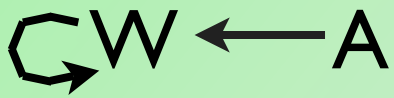
- If the essence of "being causal" is knowing that there are two types of variables, World and Action, and how they relate (eg. Action variables are exogenous to, and arrow-breaking on, World variables).
- Then a hypothesis space of causal frameworks should consist of different ways to understand the distinction between World and Action variables:
  - Is there a distinction?
  - How do the kinds of variables relate?

ndg@mit.edu

Block Structured Causal Frameworks



A Block Description:



"Each W variable may connect to other W variables,
each A variable may connect to W variables,
but the whole network must be acyclic."

ndg@mit.edu

Block Structured Causal Frameworks

We can consider different kinds of block relations:

- ← "may connect"
- ← "must connect"
- ← "may connect once"
- ⇠ "breaks other arrows"

This gives us many frameworks:

Fully Connected: ↻V

DAG: ↻V

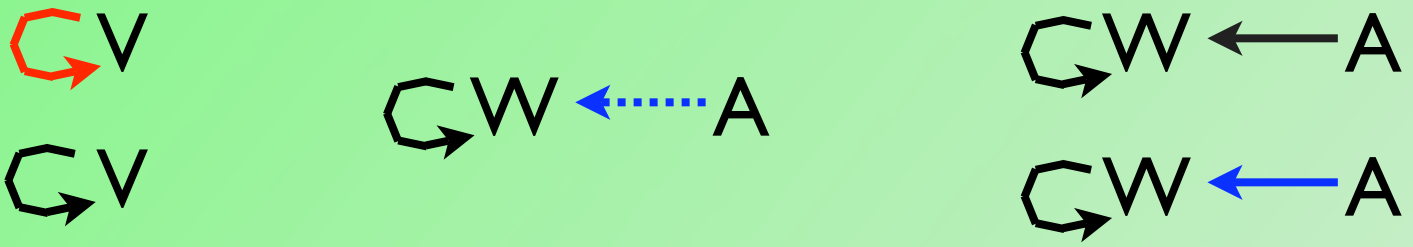
CBN: ↻W ⇠ A

Exogenous Actions: ↻W ← A

Soft Interventions: ↻W ← A

ndg@mit.edu

Block Structured Causal Frameworks

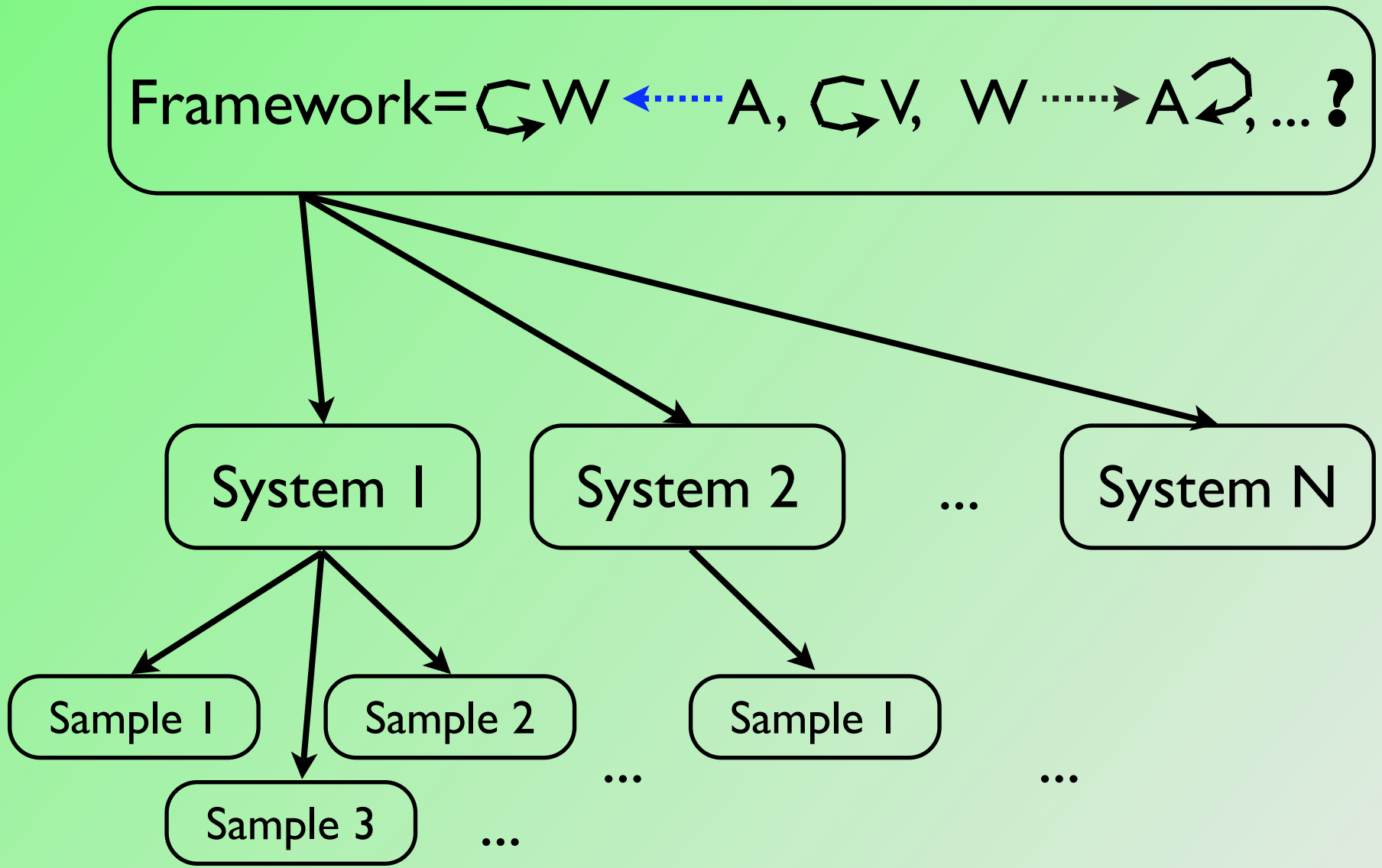


And many other frameworks:

...this simple block language gives 1300+ frameworks.

ndg@mit.edu

So, can we use Bayesian induction to learn the correct framework, using data from a bunch of different systems?



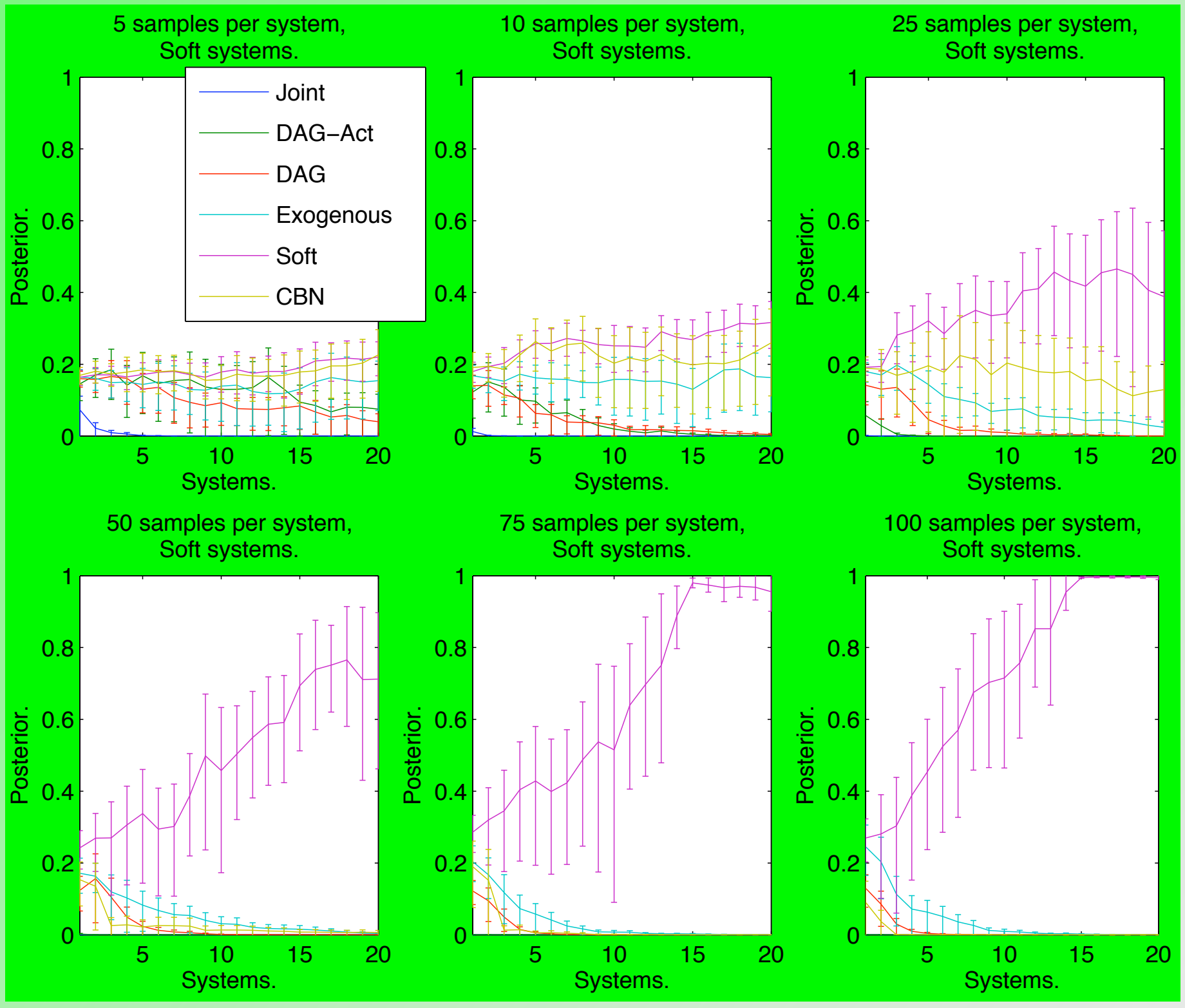
MIT

ndg@mit.edu

Learning to be causal

- We simulated a Bayesian learner in 10 worlds. Each world consisted of 20 systems, with 3 World + 2 Action variables, randomly sampled from the Soft framework. Up to 100 samples were drawn from each system.
- Initially, we consider a restricted hypothesis space of frameworks: Full Joint, DAG ignoring actions, DAG, Exogenous actions, Soft interventions, and CBN.

ndg@mit.edu


5 samples per system,
Soft systems.
10 samples per system,
Soft systems.
25 samples per system,
Soft systems.
50 samples per system,
Soft systems.
75 samples per system,
Soft systems.
100 samples per system,
Soft systems.
Joint
DAG–Act
DAG
Exogenous
Soft
CBN
Posterior.
Systems.
0
0.2
0.4
0.6
0.8
1
5
10
15
20

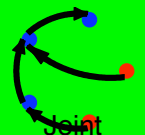
ndg@mit.edu

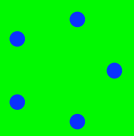
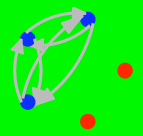
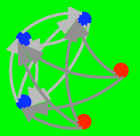
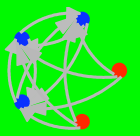
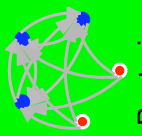
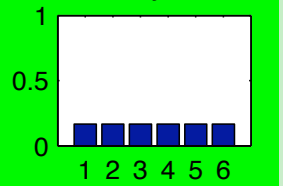
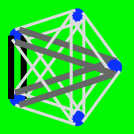
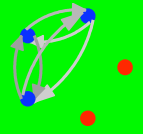
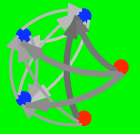
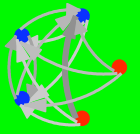
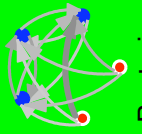
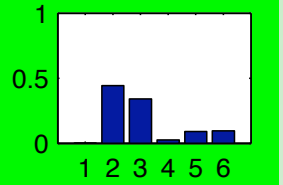
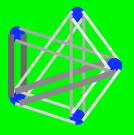
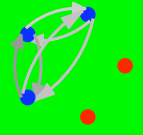
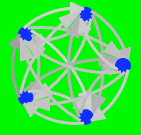
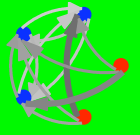
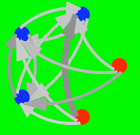
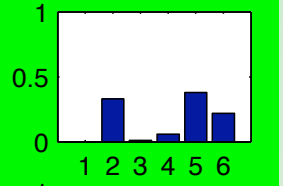
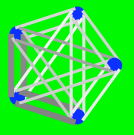
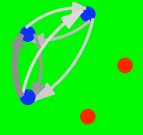
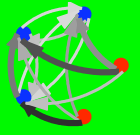
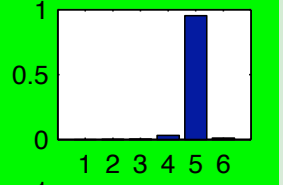
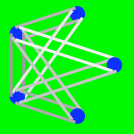
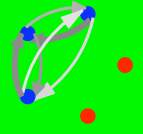
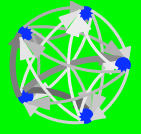
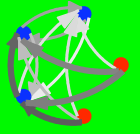
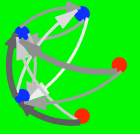
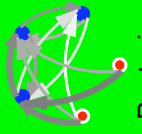
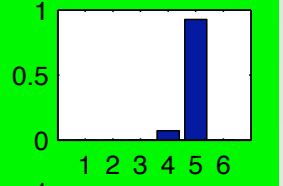
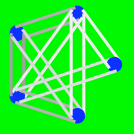
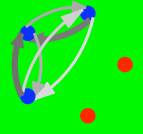
Learning to be causal

- Bayesian learning can also find the correct framework from among the 1300+ frameworks described by simple block structures.

ndg@mit.edu

World 1, System 12,
Soft parametrization.



	Joint framework.	DAG –Act framework.	DAG framework.	Exogenous framework.	Soft framework.	CBN framework.	Framework scores all 20 systems.
1 samples.							
5 samples.							
10 samples.							
25 samples.							
50 samples.							
75 samples.							
100 samples.							

ndg@mit.edu

The Blessing of Abstraction

- In hierarchical Bayesian models learning is often faster at higher levels (the 'blessing of abstraction').
- In our case, the correct causal framework can be learned before the structure of *any* system is learned.

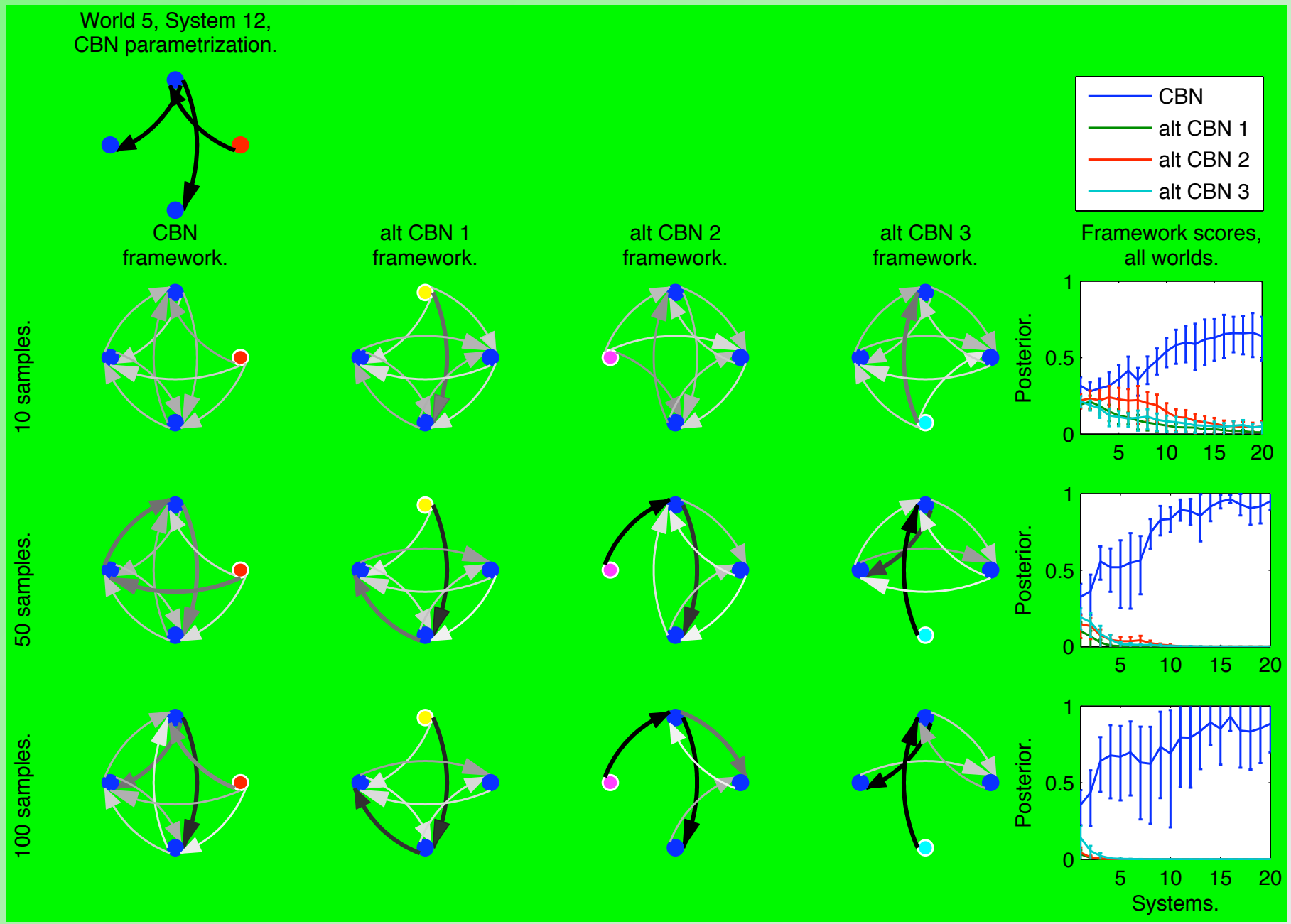
ndg@mit.edu

Action features

- So far I've assumed that there is a special feature marking actions (eg. muscle feedback, 'efficacy').
- However, the same Bayesian methods work when alternative 'distractor' features are available:
  - The inductive problem now is to identify the framework *and* the feature on which the framework is based, concurrently.

ndg@mit.edu

Feature selection: CBN frameworks based on the action feature or one of three distractor features. 10 simulated worlds.



ndg@mit.edu

On Causality and Constructivism

- Causal framework knowledge is learnable, and this knowledge may be acquired faster than any specific causal knowledge.
- Thus the benefits of complex framework knowledge for causal learning are, in principle, available very early in development, *without* requiring that this knowledge be innate.