Graduate Students: Project Reminder

Midpoint report due on **Monday**! Slides for presentations due **Sunday**.

Schedule time to chat if you are stuck.

Book now!

- **today** 2-3:30, 4:30-5
- **tomorrow** 9-11, 12-2:30
- **Friday** 1-2, 3:30-5

Everyone: include questions about vocab & concepts in your reviews for Friday.

CS 295B/CS 395B Systems for Knowledge Discovery

From Causal Inference to Experimental Design

The University *of* Vermont

High level idea for today

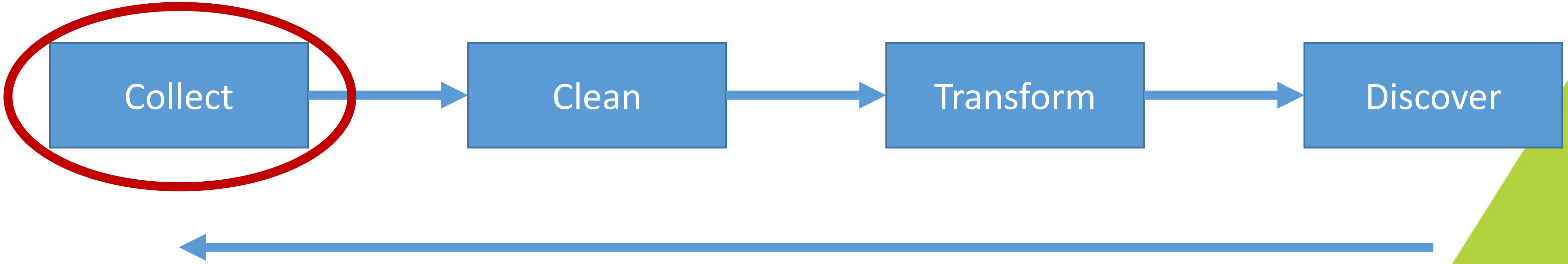
Context… but NOT about specific papers

Things that were surprising to me

NOT about vocab

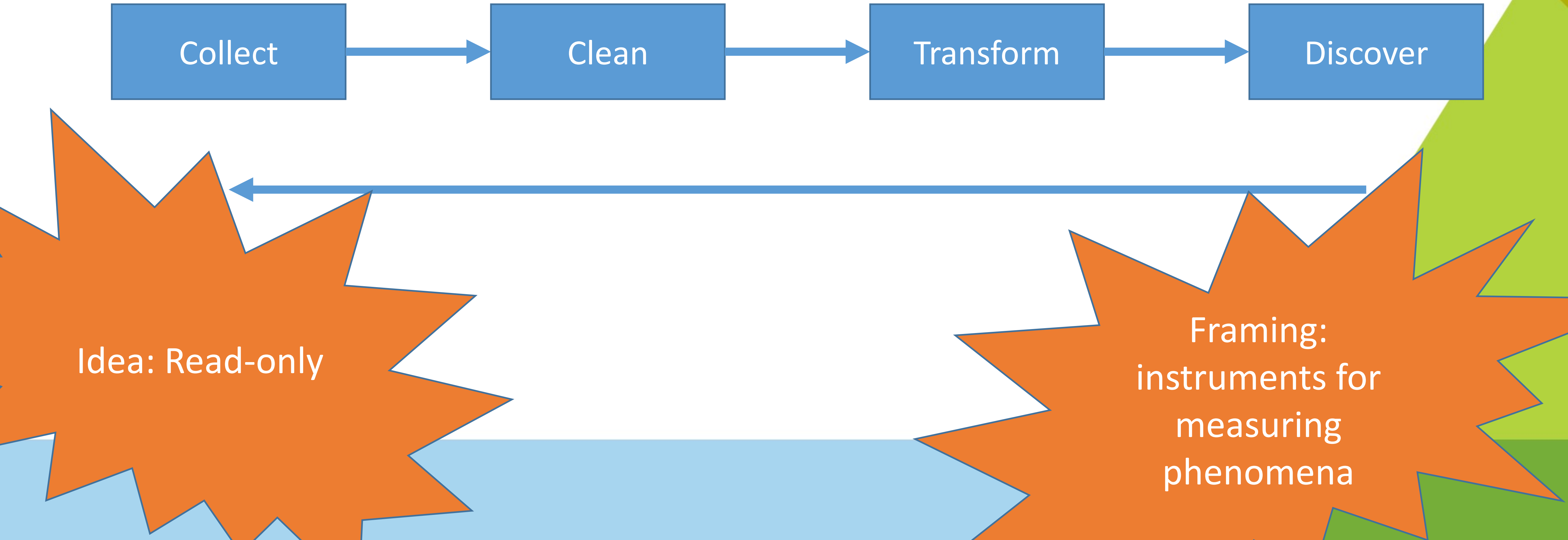
More about framework for reading the papers

Recall…



Recall…

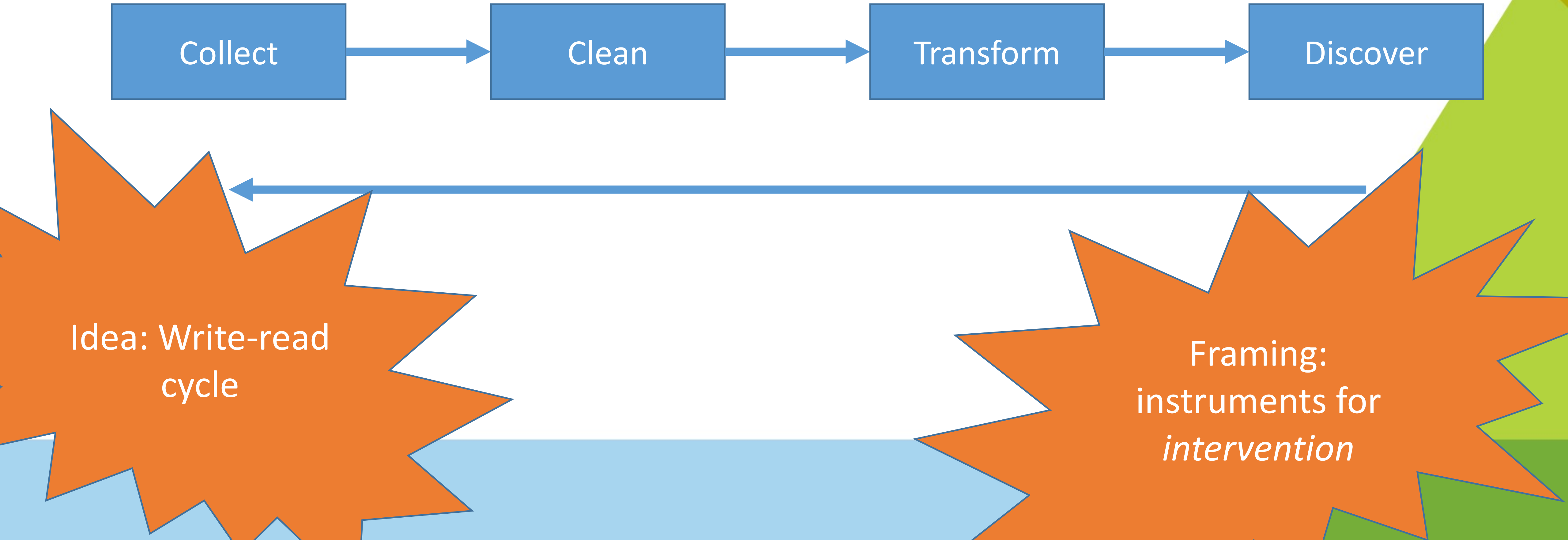
Snorkle AutoMan
DeepDive SurveyMan



Now: Experiments

PlanOut

PlanAlyzer



Experimentation ties together

data collection & decision-making

much more closely than we've seen before

Recall: Research Questions

Descriptive

- What are the characteristics of <phenomena>?

Associative

- What is the relationship between…?
- Under what conditions…

Causal

- Can we <verb> <noun> such that <dependent clause>?
- Does X cause Y when Z…?

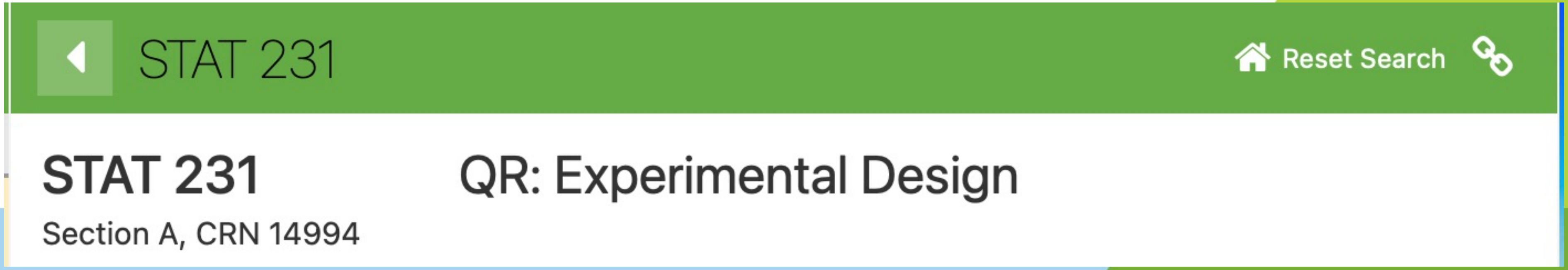
Previously: **implicitly** causal questions

Now: **explicitly** causal questions

Today's topics

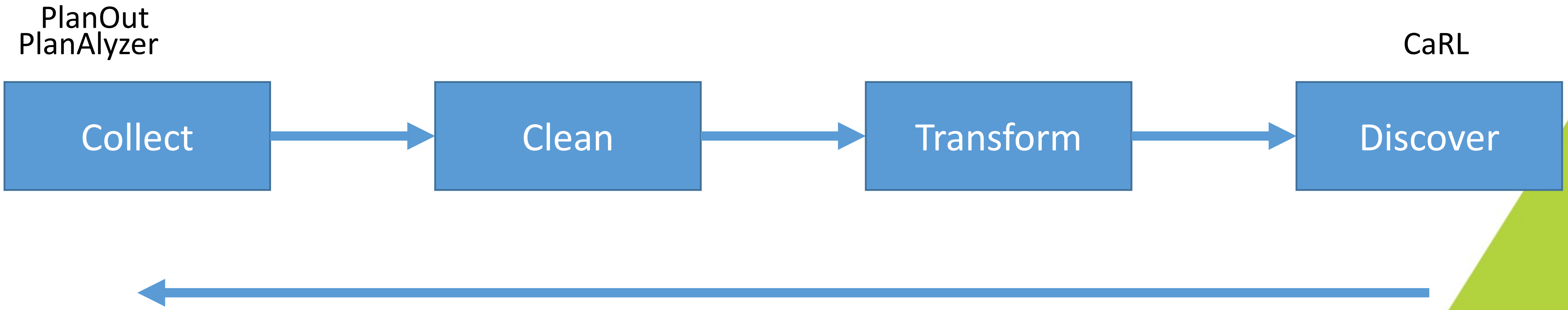
Recap: Causal inference

Thinking like an experimentalist



Causal Inference → Experimentation

Now: Experiments



Recall: **CaRL** == declarative language for **Ca**usal **R**elational **L**earning

Recall: Causal Models

- Assuming model **structure** and **parameters** are known (completely known mechanism)…
  - Experiments are just the *do-calculus*

- Fundamental challenge: structure and parameters are unknown
  - Causal Inference: problem of inferring model structure and parameters from ***purely observational data***

Example do-calculus on board

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data...up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data...up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

Pros

- Complete model
  - Can answer any Q!
- No need to fetch more data
  - Experiments purely simulated!
- Mechanism is interpretable
  - Can explain why!

Cons

- Rarely known *a priori*
  - Can use domain knowledge, but what if you are wrong?
  - Can be learned from data…up to a certain point
  - May have incomplete data
- Complete model may be complex
  - Can build a simpler model, but higher variability

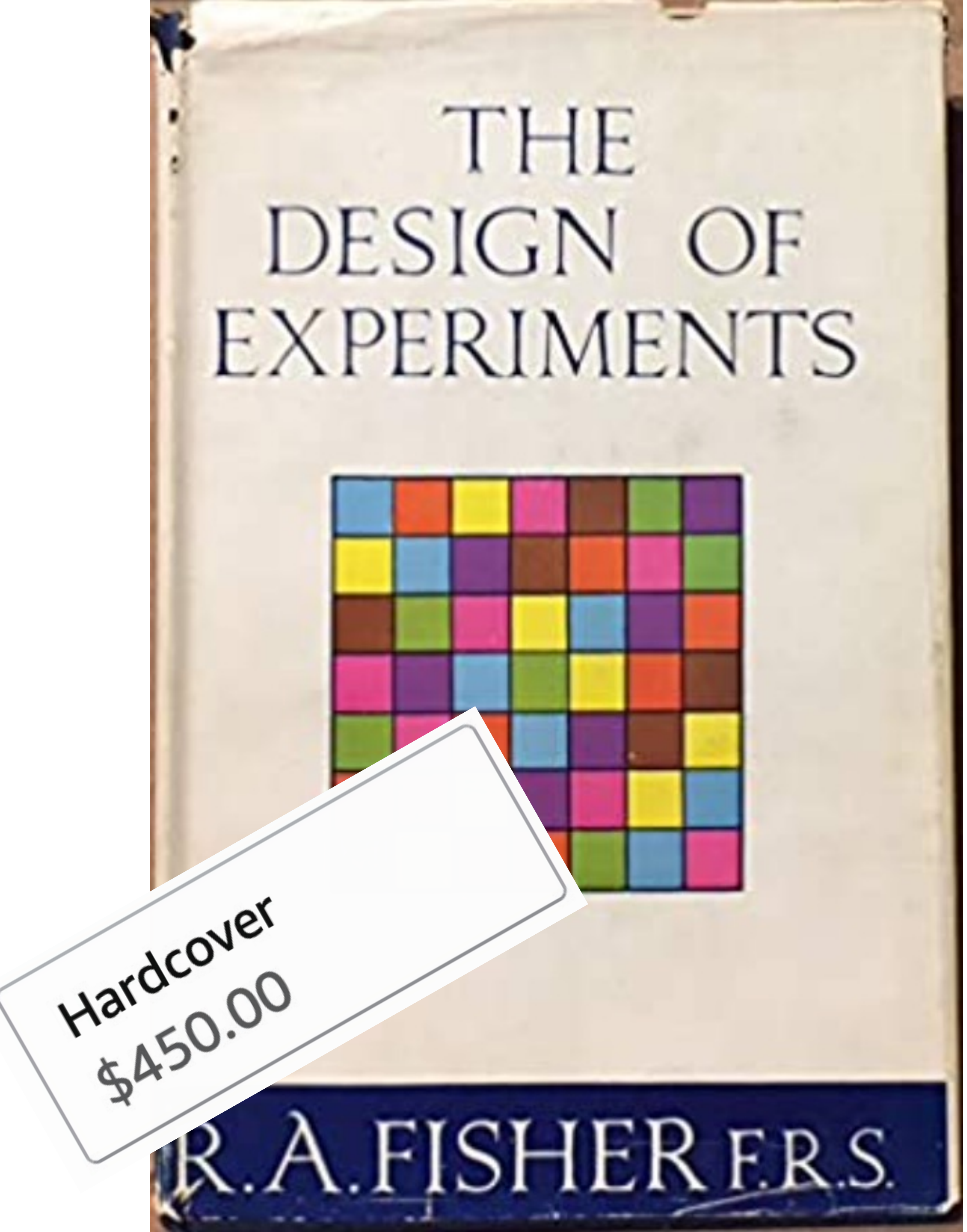
Bivariate case

If all you want is to measure the effect of a

treatment variable on an **outcome variable**...

…and can set the treatment variable…

Then randomized field experiments are your friend.

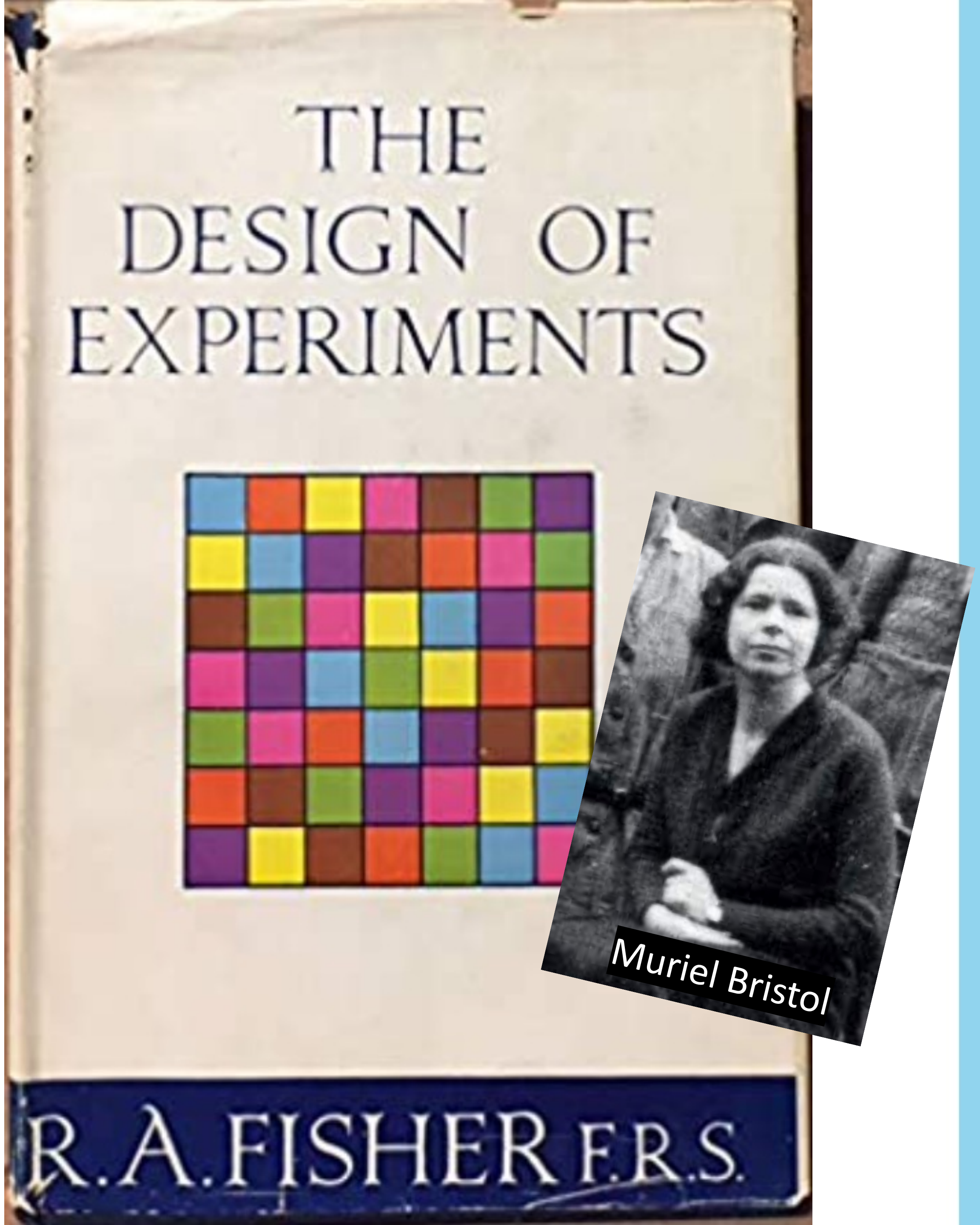


Two big ideas

“Factorial Design”

(vs. OFAT)

Randomized Assignment



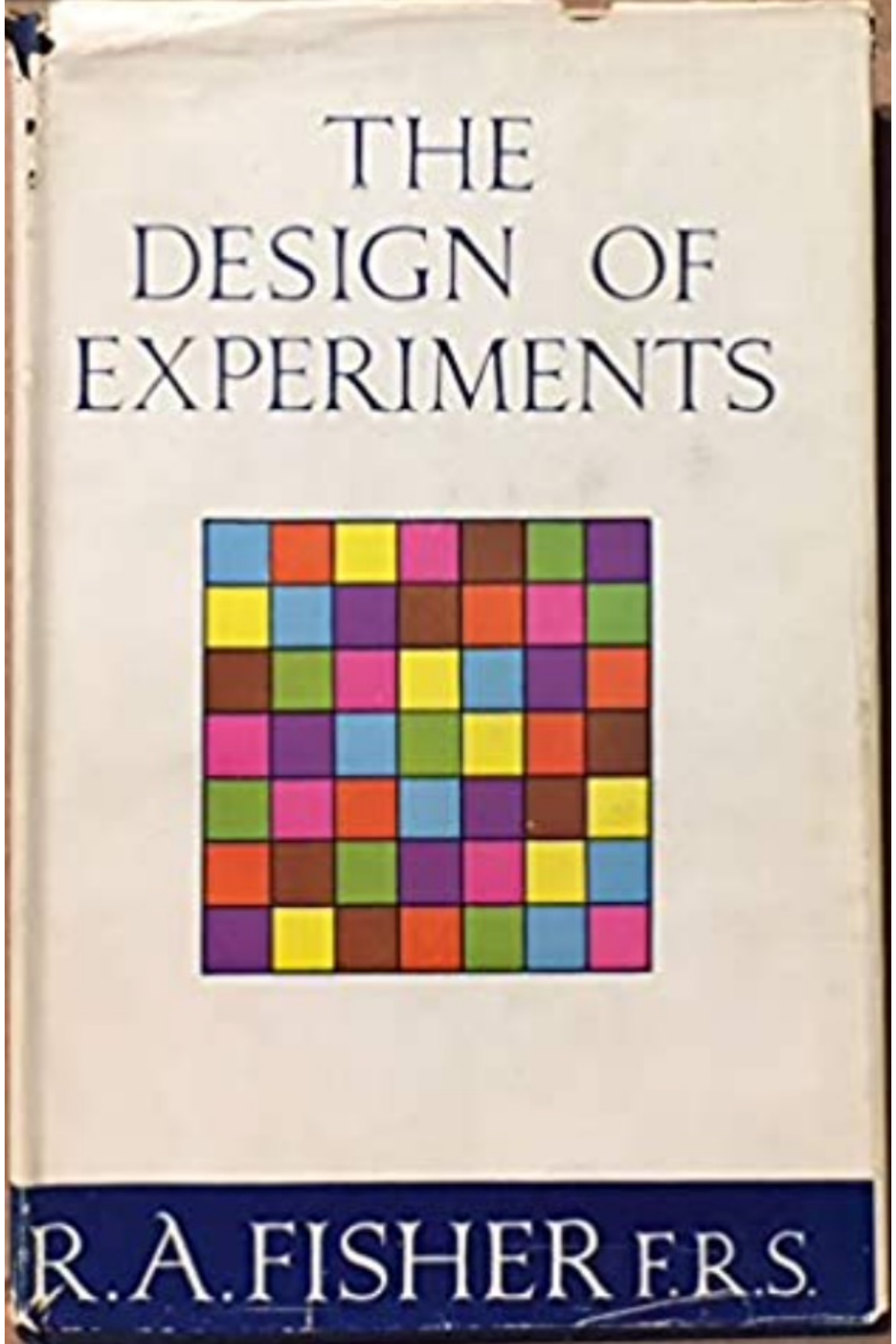
Two big ideas

"Factorial Design"

(vs. OFAT)

Randomized Assignment

(Original was one factor, but let's be creative)

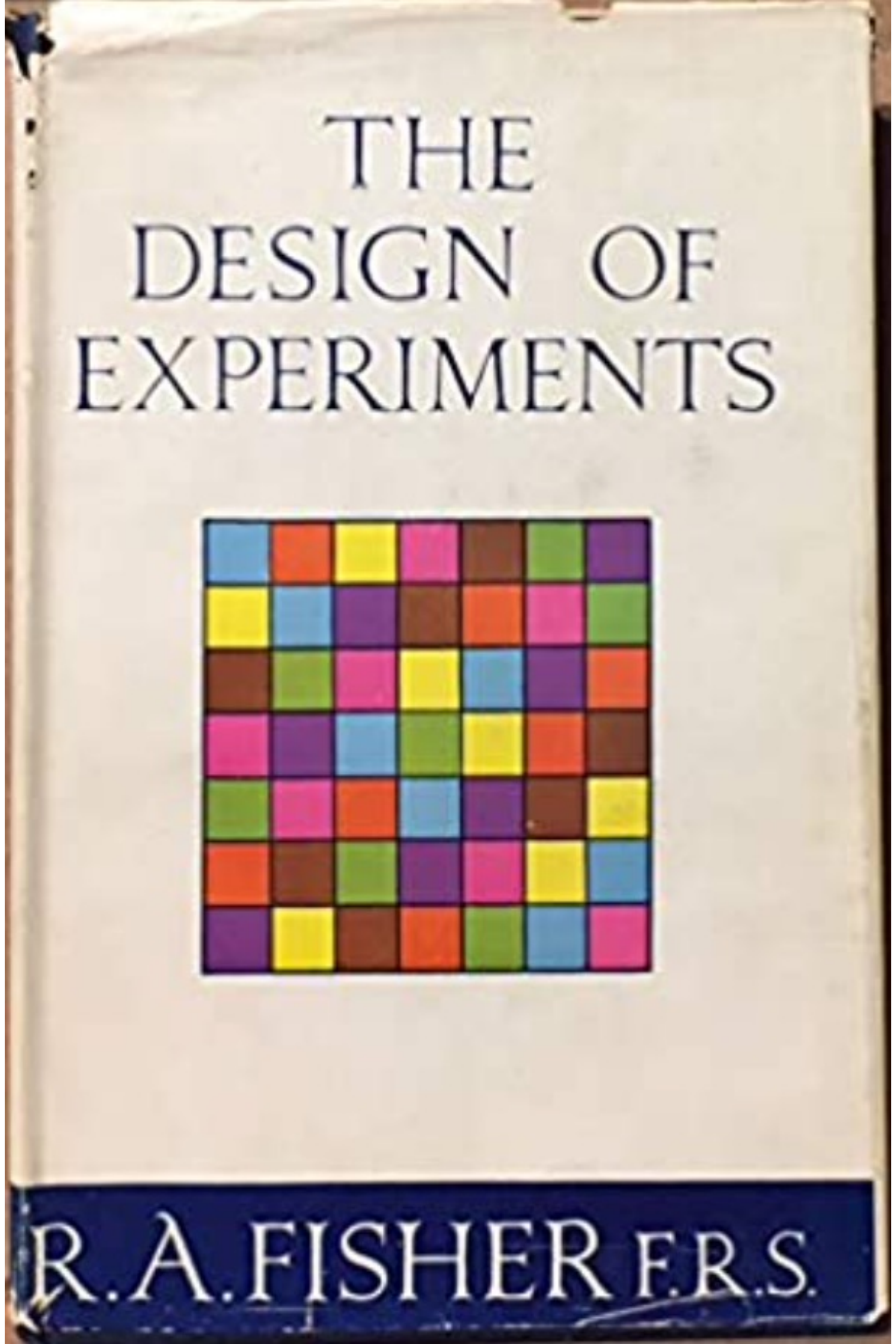


Two big ideas

"Factorial Design"

(vs. OFAT)

Randomized Assignment



Two big ideas

"Factorial Design"

(vs. OFAT)

Randomized Assignment

Both experiments have *replicates*

Thinking like an experimentalist

Shift in how we think about platforms

AMT

- Participant knows, deception is hard
- Close to laboratory experiment
- Purely discretionary

Facebook

- Participant may not know
- Field experiment
- May be necessary

WSDM 2016 Workshop on the Ethics of Online Experimenta...

Home

Agenda

Attending

Organizers

Resources

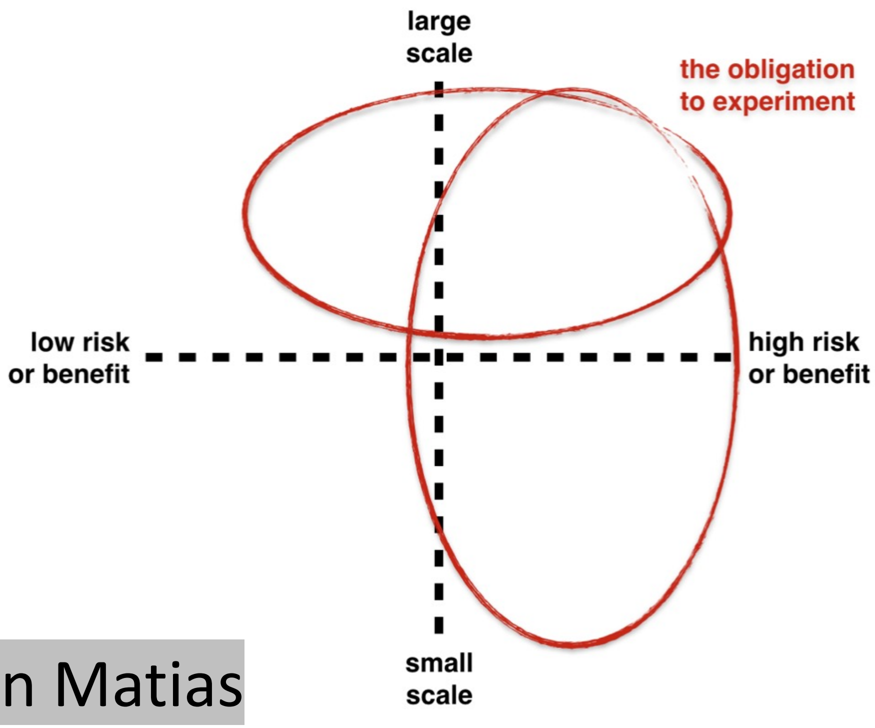
Topics

it okay to experiment?

Who Has The Obligation To Experiment?

The obligation to experiment should vary by a platform or service's ability to manifest significant risks to its users. For example, the obligation to experiment might apply if an entity:

- legally collects (or has the ability to collect) extensive data about people
- attempts to influence people's behavior
- functions as a key node in infrastructure, through its ubiquity in public life, or through substantial market saturation
- is a common carrier, or is broadly understood and expected to uphold objectivity, neutrality, or public goods



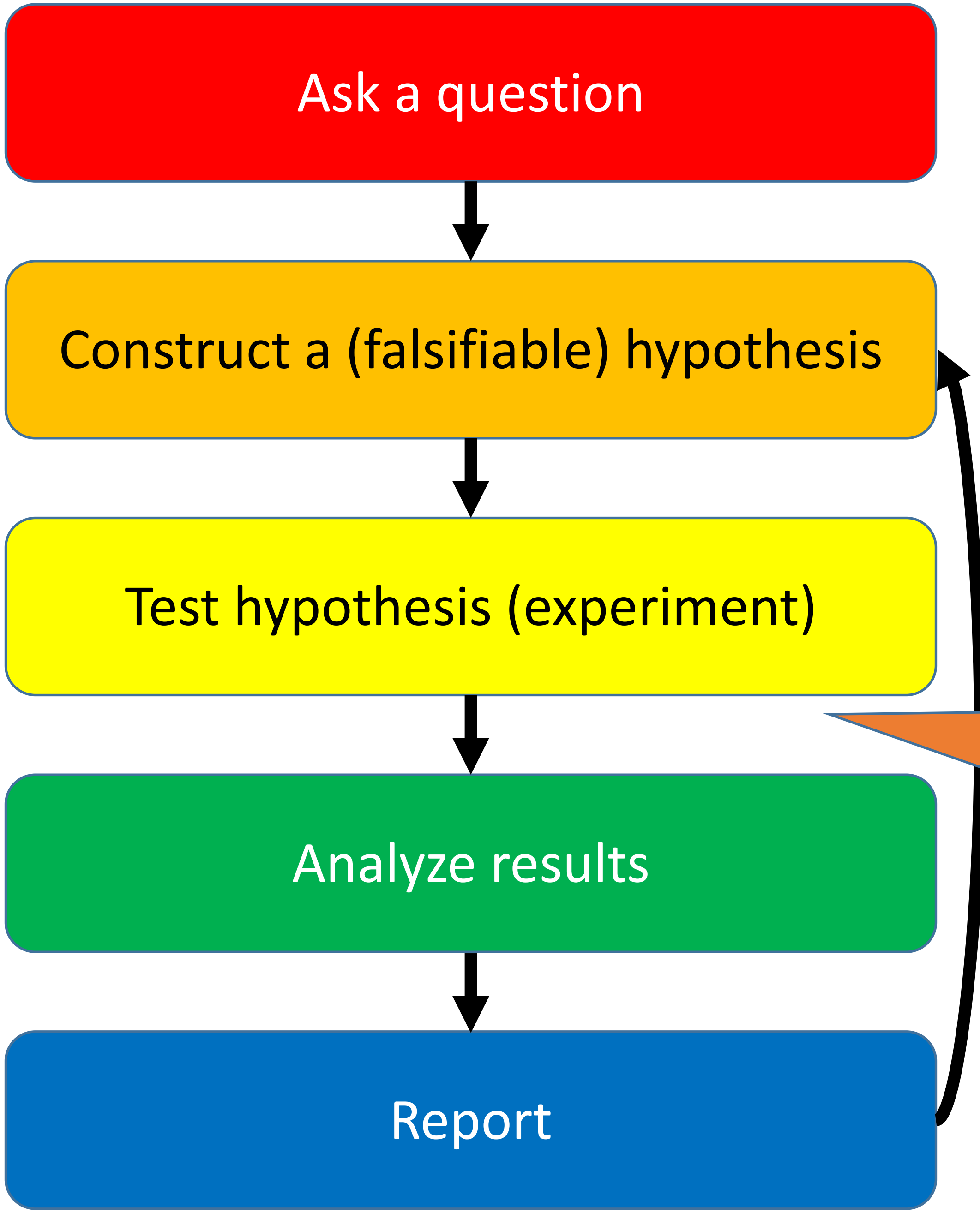
Solon Barocas is a P
department of Medi
with the Data & Soci
emerging computati
raise. His recent wor
Scholars Conference
Center for Internet a
Impact program. He
University 2014), an
2015).

eton University. He completed his doctorate in the
the Information Law Institute. Dr. Barocas also works
Big Data, Ethics, and Society. His research focuses on
plores the ethical and epistemological issues that they
won the Best Paper Award at the 2014 Privacy Law
its Algorithmic Living research theme, the Berkman
of Business at New York University on its Social
l Web Privacy and Transparency (Princeton
nsparency in Machine Learning (NIPS 2014 and ICML

Fernando Diaz is a s
scientist at Yahoo Re
retrieval approaches
cross-lingual inform
and a B.A. in Politica
best paper awards a
2013. He is a co-orga
Social Media During

ior to joining Microsoft, Fernando was a senior
earch experience includes distributed information
ng of temporal patterns from news and query logs,
ble corpora. He received a B.Sc. in Computer Science
chusetts Amherst. His work on federation won the
cs has received awards at SIGIR 2011 and ISCRAM
DM 2014. He is also co-organized workshops on
and Reproducibility of Results (SIGIR 2015).

Allan Ko, Merry Mou, J. Nathan Matias



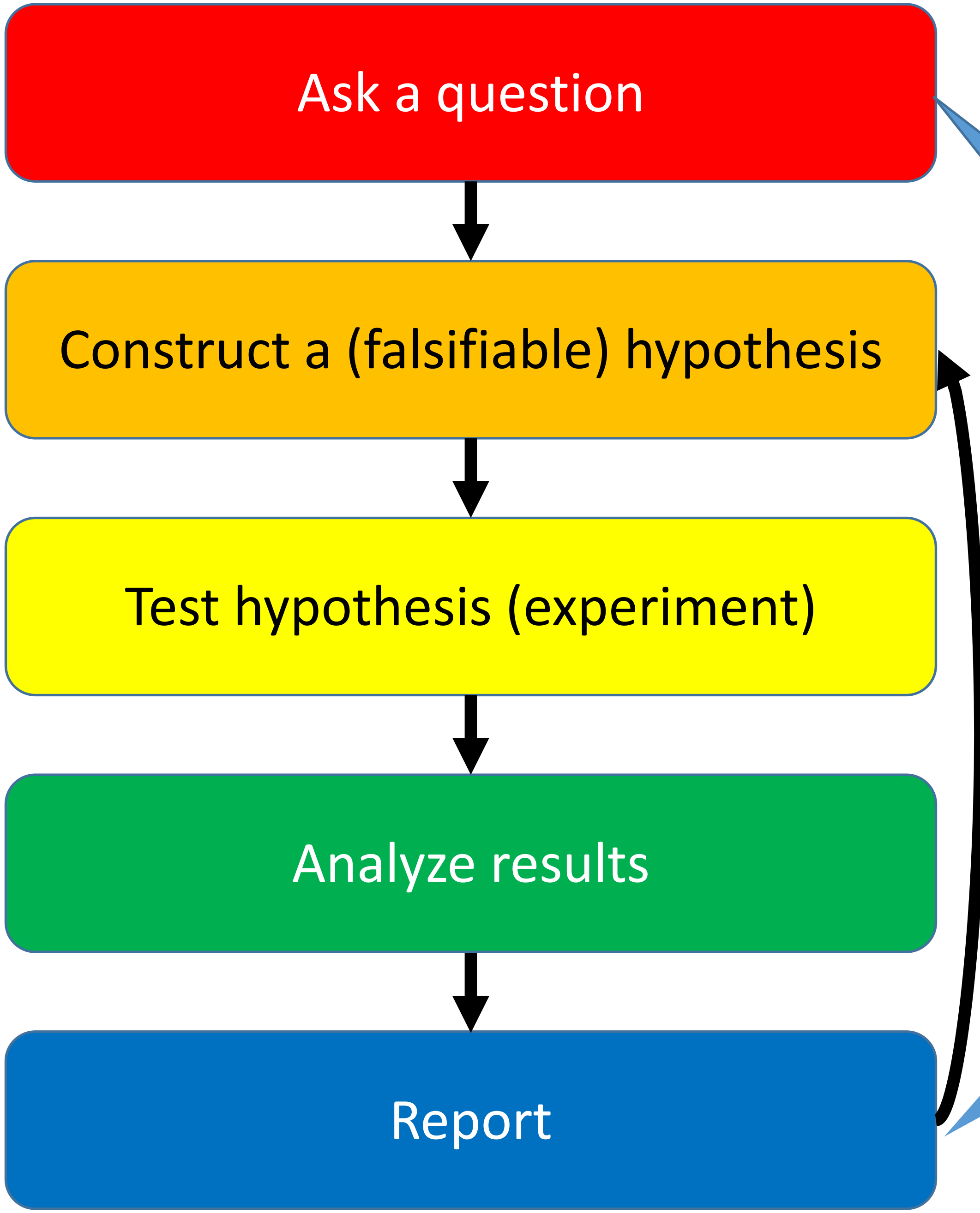
Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized

Formalization:
Explicit rules

(usu. mathematical notation, bounded ambiguity)

Draw diagram of Qs on b



Scientific Method

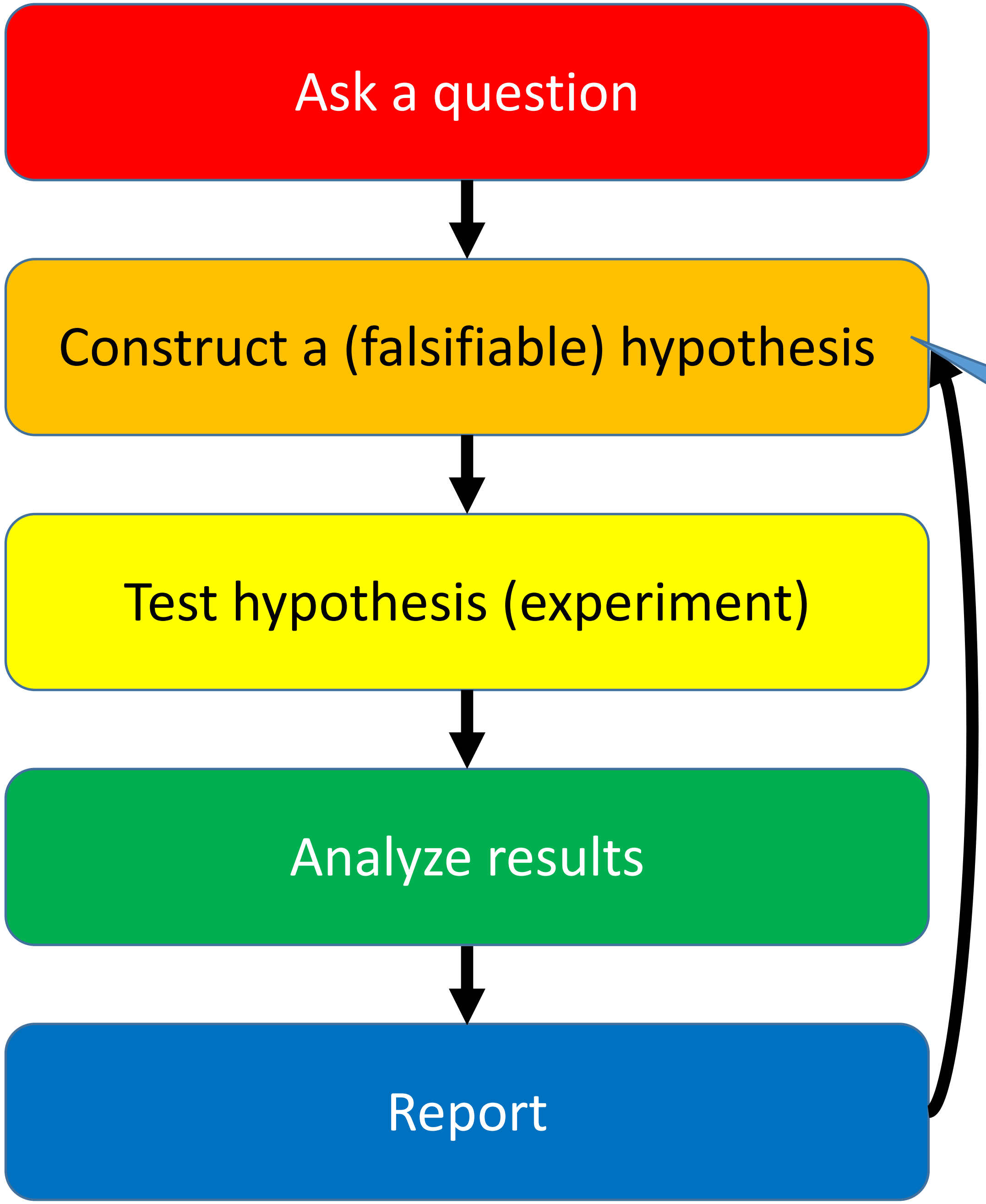
Science is a subset of knowledge

- tific method under-formalized

Questions & Reports don't need formalization

(About communication)

Draw diagram of RQs on board



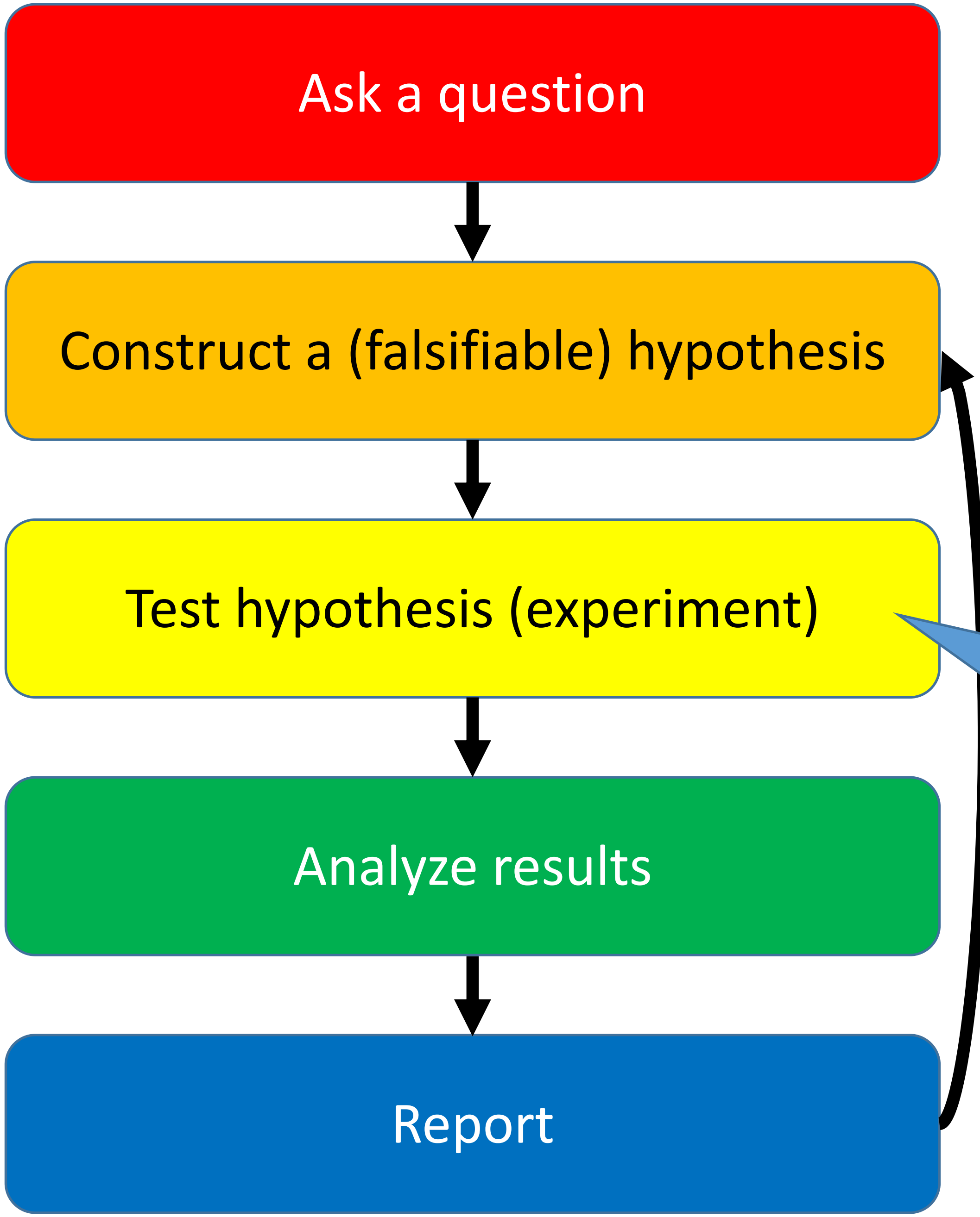
Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized

Are hypotheses under-formalized?
No: hypotheses are candidate causal models

(can be expressed as CGMs)

Draw diagram of RQs on board



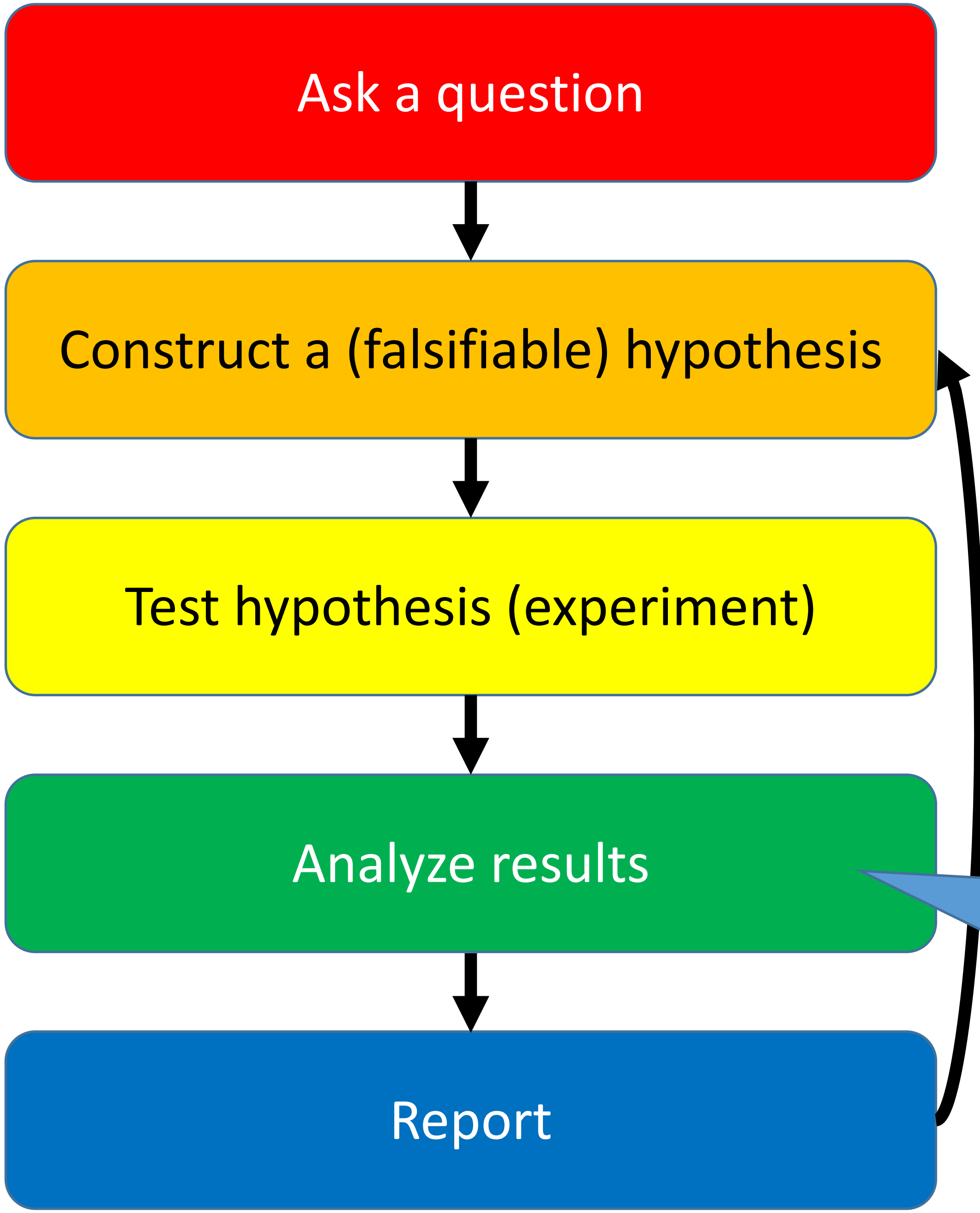
Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized

Are experiments under-formalized?
Yes! (often treated as a black box)

- Variability based on context
- Operationalization not yet generalized

Draw diagram of RQs on board



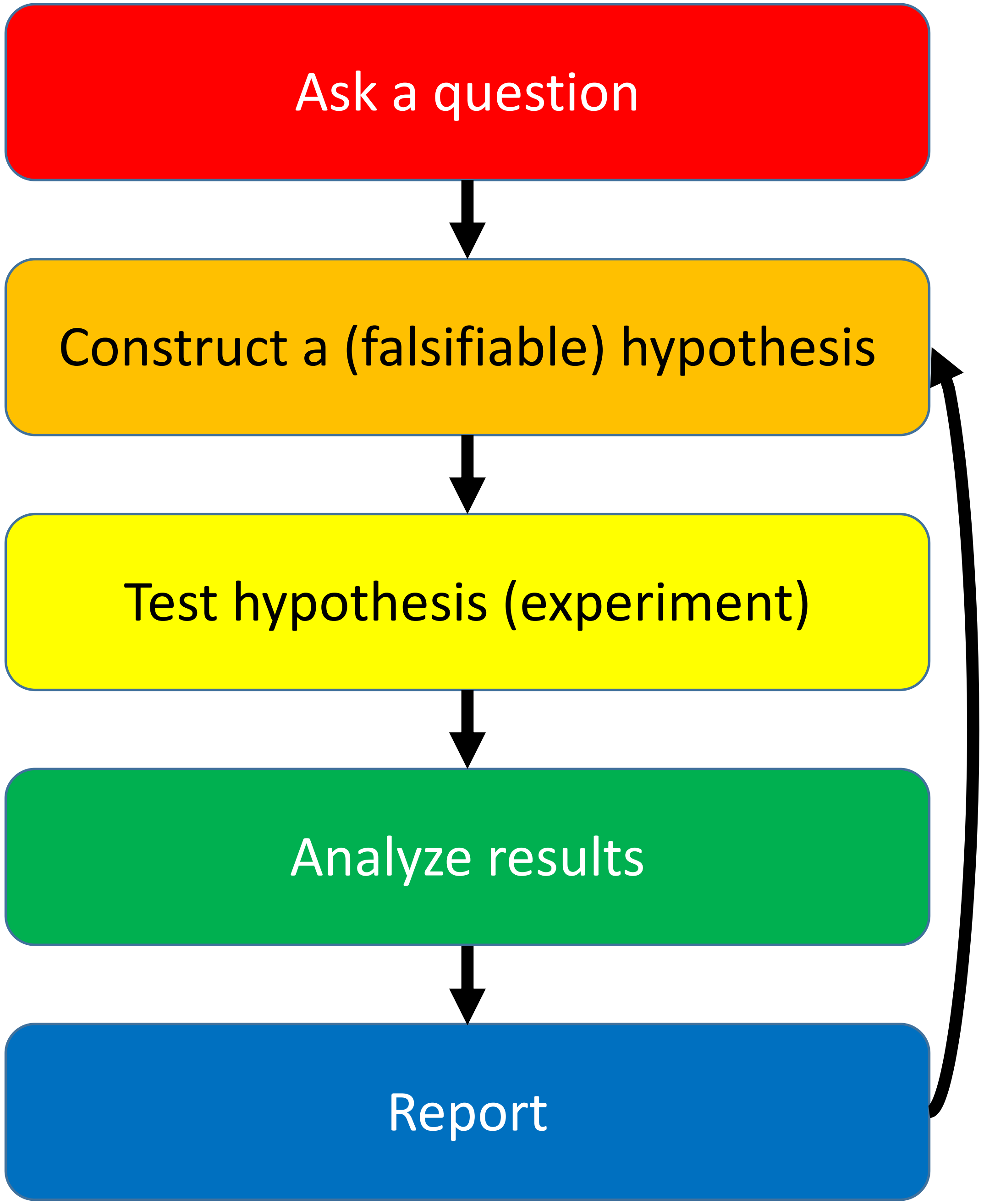
Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized

Are analyses under-formalized?
Yes…ish…

- See: work on formal methods for statistical software
- Key problems: assumptions, definitions of correctness

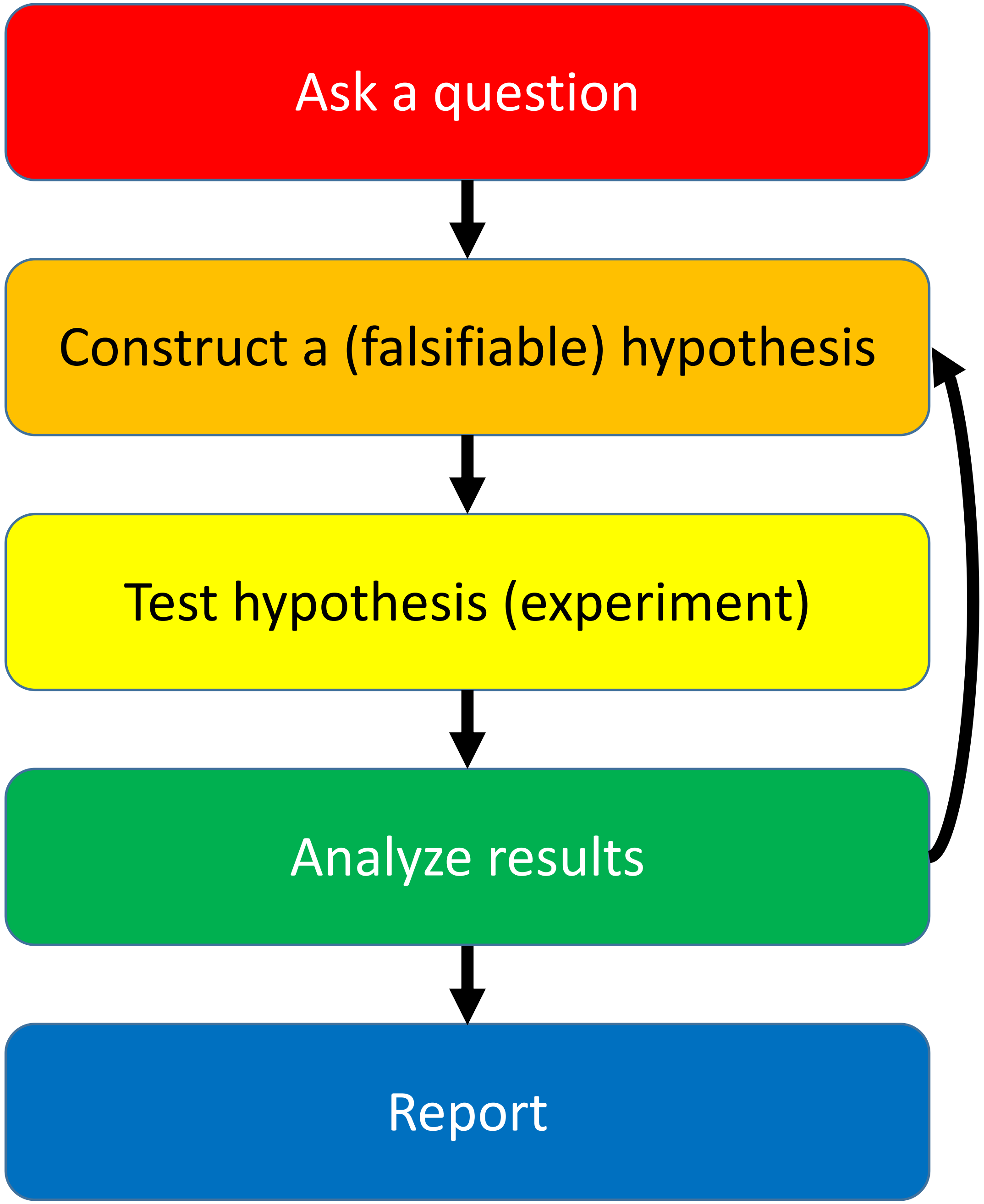
Draw diagram of RQs on board



Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized
- Diagram not always true to practice

Draw diagram of RQs on board

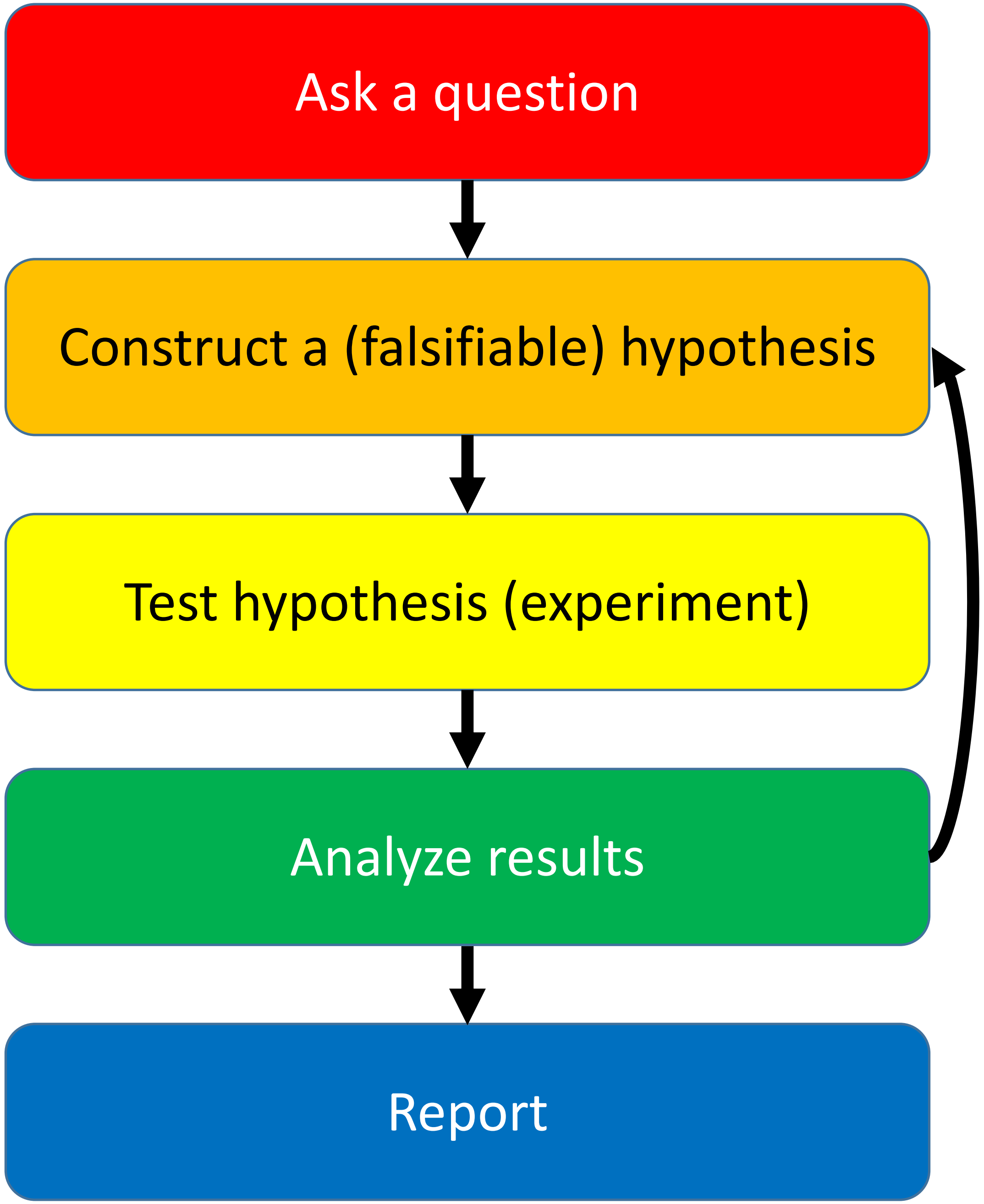


Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized
- Diagram not always true to practice

Not just about negative results!

Draw diagr



Scientific Method

- Science is a subset of knowledge
- Scientific method under-formalized
- Diagram not always true to practice

Byproduct of limitations of software systems

Draw diagr

Ask a question

Construct a (falsifiable) hypothesis

Test hypothesis (experiment)

Analyze results

Report

Necessity is the mother of invention

- Typical: limited interventions

Draw diagram of RQs on board

Test hypothesis (experiment)

Ask a question

Construct a (falsifiable) hypothesis

Analyze results

Report

Necessity is the mother of invention

- Typical: limited interventions

Draw diagram of RQs on board

Design intervention

Ask a question

Construct a (falsifiable) hypothesis

Analyze results

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes

Draw diagram of RQs on board

Design intervention

Analyze results

Ask a question

Construct a (falsifiable) hypothesis

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Ask a question

Construct a (falsifiable) hypothesis

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome

Hypothesis model always shallow

Draw diagram

Design intervention

Define measurable outcome

Ask a question

Construct a (falsifiable) hypothesis

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Ask a question

Deploy & collect data

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Deploy & collect data

Ask a question

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Deploy & collect data

Ask a question

Report

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection
- Bundled reports & RQs, add back analysis

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Deploy & collect data

Report on RQs

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection
- Bundled reports & RQs, add back analysis

Draw diagram of RQs on board

Design intervention

Define measurable outcome

Deploy & collect data

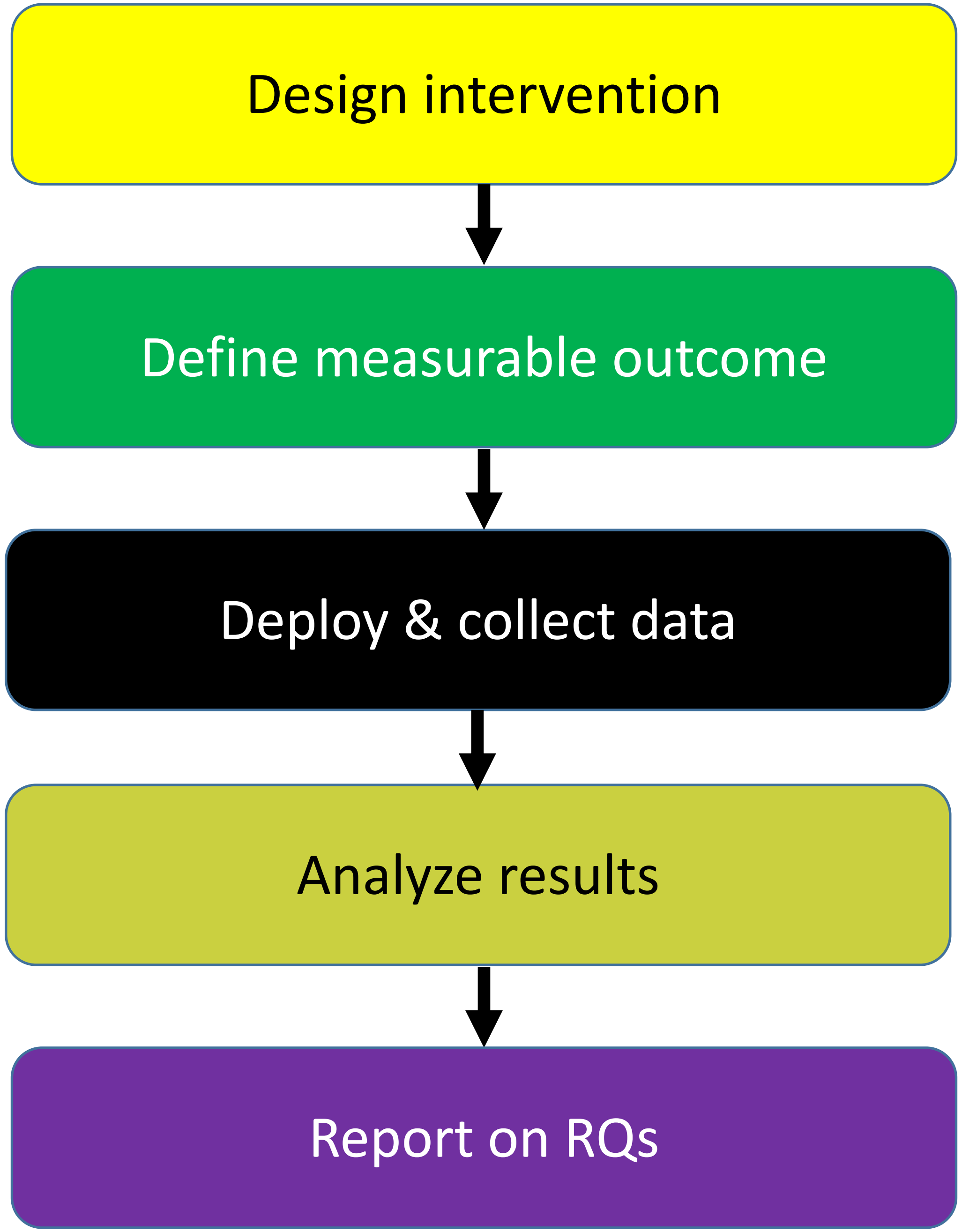
Analyze results

Report on RQs

Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection
- Bundled reports & RQs, add back analysis

Draw diagram of RQs on board



Necessity is the mother of invention

- Typical: limited interventions
- Typical: limited measurable outcomes
- Hypothesis falls out of intervention/outcome
- Instead: bundle experiment + logging as data collection
- Bundled reports & RQs, add back analysis

Draw diagram of RQs on board

Discussion:
Computer experiments vs. field experiments

Where does variability come from?