Causality and Hypotheses

Review

- Science is not science fiction. We evaluate our work by correspondence to physical reality. Experiments formally evaluate that correspondence.
- Naming, describing, or giving context are less useful than providing causal explanations of underlying function.
- More rapid technical progress can be achieved by seeking an understanding of fundamental principles rather than by using a "just build it" approach.

What is causality?

"The paradigmatic assertion in causal relationships
is that manipulation of a cause
will result in the manipulation of an effect…
Causation implies that
by varying one factor,
I can make another vary."

- Cook & Campbell (1979)

Why care about causality?

- Explanation
  - Association provides prediction, but not explanation.
  - Identifying causal mechanisms produces *more satisfying explanations.*
- Control
  - Understanding causality allows us to predict *the effects of actions* without ever having to perform them.
  - This allows *more efficient exploration* of design space of possible algorithms.

Development of Western science is based on two great achievements: the invention of the formal logic system (in Euclidean geometry) by the Greek philosophers, and the discover of the possibility to find out causal relationships by systematic experiment (during the Renaissance).

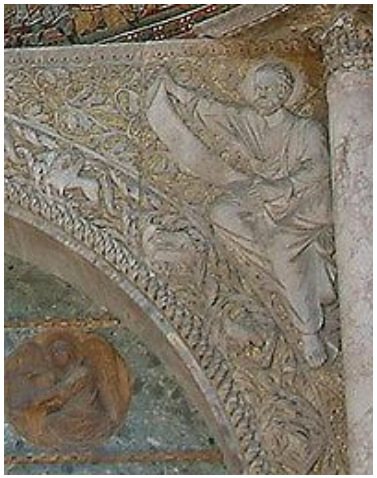
- Albert Einstein

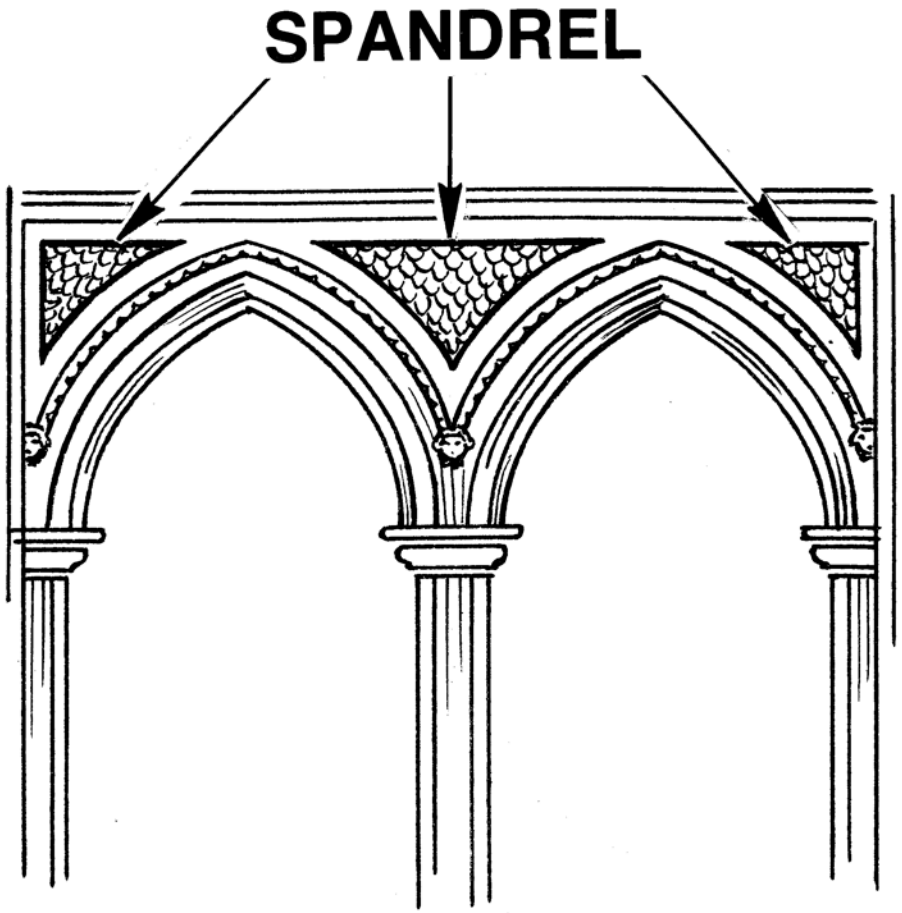
Isn't causality *easier* in designed systems?

As computer scientists, we design our systems, and we leave out extraneous components. Isn't everything causal?

- Task/environment: We don't get to design many of the aspects that produce behavior.
- Complexity: We don't always understand the underlying causes of system behavior.
- Spandrels: Design constraints create incidental features, some of which can appear to be (or actually become) causal.

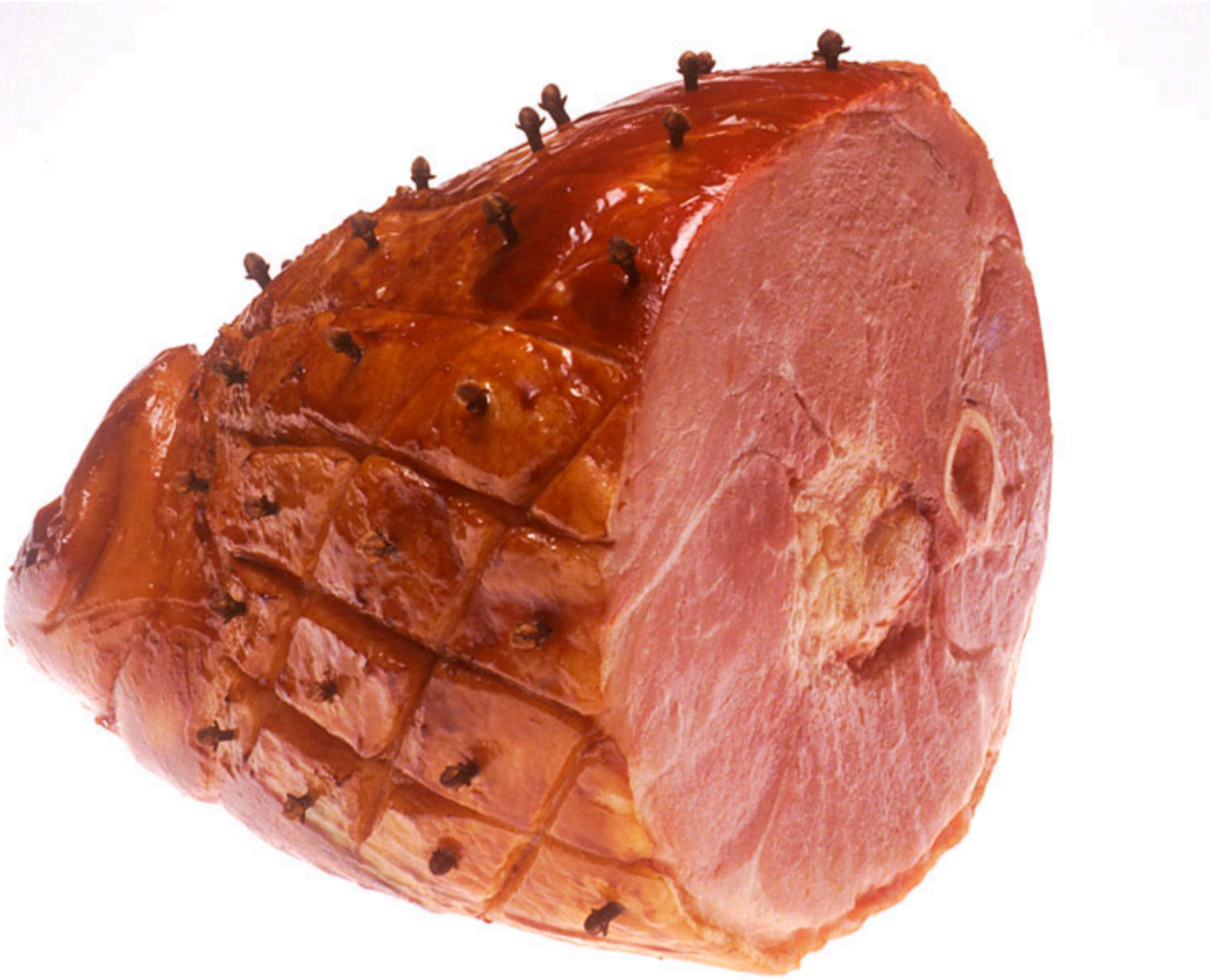
Spandrels and designed systems





Spandrels in CS

- Many features of computer systems are
  - entirely incidental ("Well, we had to choose something.")
  - results from constraints in the initial design space ("Oh, the Z328 compiler couldn't handle what we wanted to use, so we…")
  - selected because of beliefs at the time ("That was when we thought dynamic routing wasn't really feasible.")
- ….and other design choices are contingent on them…
- …so those features come to appear essential over time.



Conditions for causal inference

- Association
- Direction
- Elimination of potential common causes

What's hard about association?

"Almost every known algorithm for effective Internet routing uses confabulo-martingale technology. Using them is associated with high performance."

- Variation of potential cause
  - have alternatives been tried?
- Size of effect
  - what does "almost" mean?
- Sample size
  - two algorithms, or twenty?
- Independence
  - do all the algorithms derive from a common source?

Are feathers associated with flight?

Why isn't association enough?

- Correlation with expected causation
  - "Smoking causes cancer"
- Correlation with inverse of expected causation
  - "Cancer causes smoking"
- Correlation with latent causation
  - "Specific genes cause cancer and smoking"

Eliminating common causes

- Control
  - Hold potential common causes constant so they cannot affect outcomes.
- Randomization
  - Assign treatments randomly so potential common causes cannot systematically affect outcomes.
- Modeling
  - Measure, model, and mathematically remove effects of potential common causes.

Control

- "Explicitly set the values of all other potential causes in the experiment, so only the assumed cause can exhibit an affect."
- Examples:
  - "all participants in the study were college juniors…"
  - "at the beginning of each trial, the agent had an empty knowledge base…"
  - "each protocol was run on each of k randomly generated networks…"
- What can go wrong?
  - Missed causes
  - Failed control
  - Confounded variables

Randomization

- "…attempt to ensure that the effects of other potential causes are equivalent in aggregate. 'Average out' other potential causes."
  - Assigns values of independent variable for no reason
  - Allows elimination of potential causes that we do not know about and cannot model
- Examples
  - "We randomly assigned participants to either the training or no-training group…"
  - "Incoming web requests were handled with either the old or new protocol, based on a random number…"
- What can go wrong?
  - Non-random assignment
  - Confounded variables

Modeling

- "…explicitly adjust the measured effect for the effects of other potential causes."
  - often accomplished with a joint model of all potential causal variables
- Example: "We included all known causes besides x in our linear regression model, and the effect of x remains statistically significant."
- What can go wrong?
  - Missed causal factors
  - Specification error in model structure
  - Bias in model parameters

Hypotheses

Hypothesis

- A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.
- *Also* – theory, theorem, thesis, conjecture, supposition, postulation, postulate, proposition, premise, assumption

Types

- Existential:
  - An entity or phenomenon exists (perhaps with a specified frequency)
  - "Atoms contain uncharged subatomic particles."
- Compositional
  - An entity or phenomenon consists of a number of components (perhaps with specified frequency)
  - "Atoms consist of protons, neutrons, and electrons."
- Correlational
  - Two measurable quantities have a specified association.
  - "An element's atomic weight and its properties are correlated."
- Causal
  - A given behavior has a specified causal mechanism.
  - "The low reactivity of the noble gases is caused by their full outer shell of valence electrons."

Popper's "conjectures and refutations"

- Confirmations of a theory are typically easy to find when sought.
- True confirmations should be surprising --- unenlightened by the theory, we should have expected an event which was incompatible with it.
- Good scientific theories are prohibitions --- they forbid certain things from happening. The more a theory forbids, the better it is.
- A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue, but a vice.

"No amount of experimentation can ever prove me right; a single experiment can prove me wrong."

-Albert Einstein

"A theory which cannot be mortally endangered cannot be said to be alive."

-W. A. H. Rushton

Falsifiability

- Falsifiability is the logical possibility that an assertion can be shown to be false by evidence.
- "Falsifiable" does not imply "false." Instead, if a falsifiable proposition is false, then its falsehood can be shown by experiment, theorem, or simulation.
- There are degrees of falsifiability --- some theories are more falsifiable than others.
- Falsified theories can be rescued by introducing some *ad hoc* change, but only by lowering their apparent validity.

What makes hypotheses unfalsifiable?

- Vagueness --- the theory that does not predict any particular experimental outcome
  - "IDE x leads to better, more focused, and cleaner design."
- Complexity --- a theory that can "explain" any experimental result
  - "C6.5's performance is a function of the characteristics of the data, the existing knowledge base it starts with, the settings provided by the user, and the state of the random number generator. So it's understandable for performance to go up or down when you do x."
- Special pleading --- traditional experimental methods are claimed not to apply
  - "The thought processes promoted by our IDE are disrupted in laboratory settings, so we can't really run experiments on it. You just have to experience it for yourself."

Degree of falsifiability

- Number of experiments
- Probability that a given experimental result would be produced given that the theory is false
  - How many other outcomes were possible?
  - In the absence of the theory, what is the probability distribution over those outcomes?
- Diversity of attempts
  - Potential unknown causal factors
  - Potential experimental flaws

"Science is what we have learned about how to keep from fooling ourselves."

-Richard Feynman

Multiple working hypotheses

Why use multiple working hypotheses?

- **Personal investment** --- helps to separate you from your hypothesis; shifts your personal investment from the hypothesis to the hypothesis test.
- **Focus** --- reinforces a focus on falsification rather than confirmation.
- **Efficiency** --- allows experiments and proofs to be designed to distinguish among several competing hypotheses rather than evaluating a single one.
- **Harmony** --- limits the potential for professional conflict and rejection because all hypotheses are considered rather than only one.