Incident response? Let’s do science instead!

Ivan Merrill

Incident response can learn from safety engineering in other domains

A definition...

Incidents are a “set of activities, bounded in time, that are related to an undesirable system behavior.”

Allspaw, J., Cook, R.I. 2018. SRE cognitive work. In Seeking SRE: Conversations About Running Production Systems at Scale, ed. D. Blank-Edelman. O'Reilly Media, 441-465.

Catastrophe is always around the corner

Richard I. Cook, M.D. - How complex systems fail, 1998.

Incident response isn’t easy

An overreliance on dashboards and runbooks

Guesswork

Spending a long time on the wrong hypothesis

Fear of failure

‘History doesn’t repeat itself but it often rhymes’

Mark Twain / Theodore Reik

‘It seems easy to look back at an incident and determine what went wrong. The difficulty is understanding what actually happened and how to learn from it’

Maris R Grayson - Cognitive Work of Hypothesis Exploration During Anomaly Response, 2019

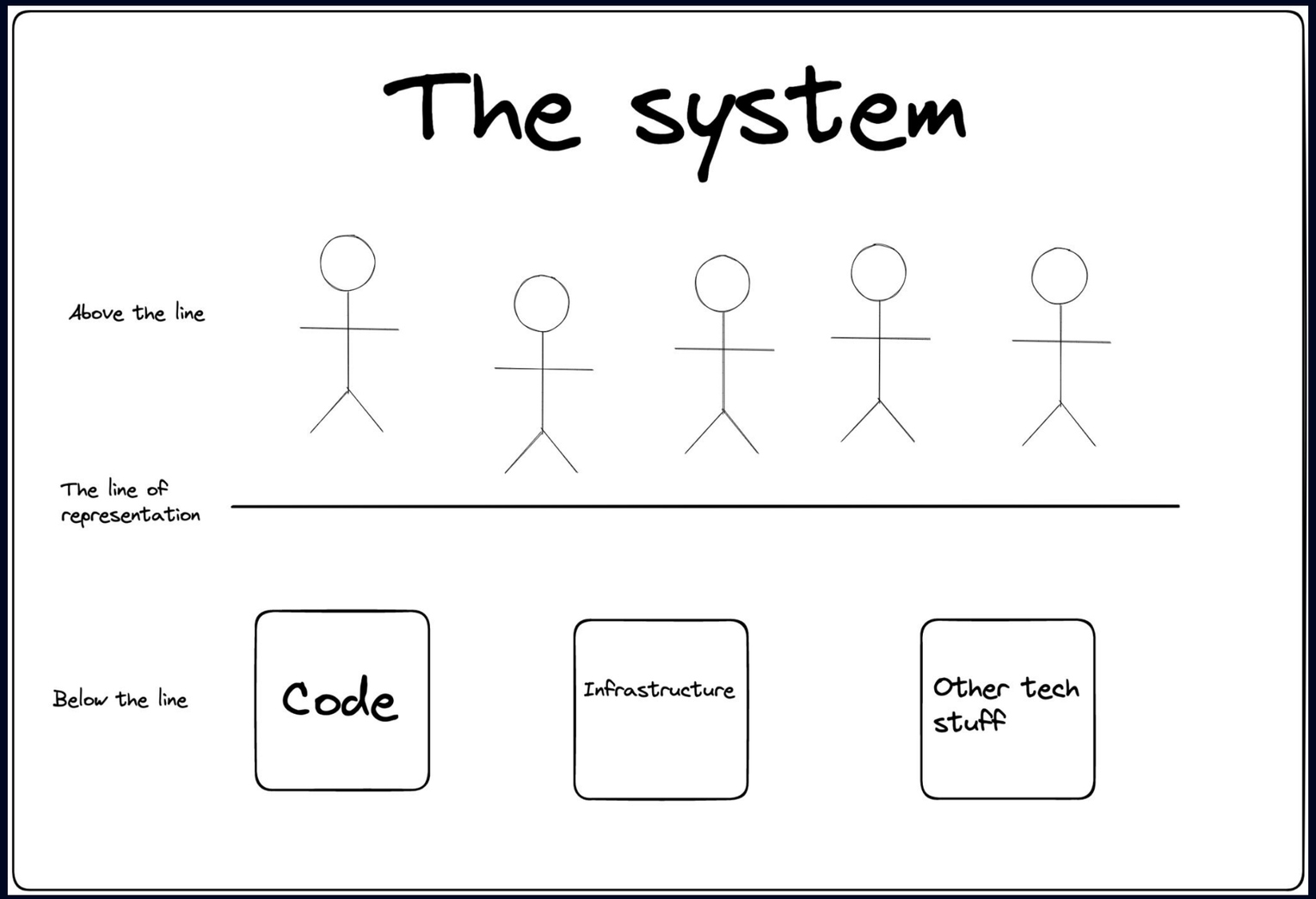
Normative language

‘The team missed this obvious error which they ought to have seen’

Mechanistic reasoning

Above the line, below the line

Richard I. Cook, M.D. 2019



Change introduces new forms of failure

Richard I. Cook, M.D. - How complex systems fail, 1998.

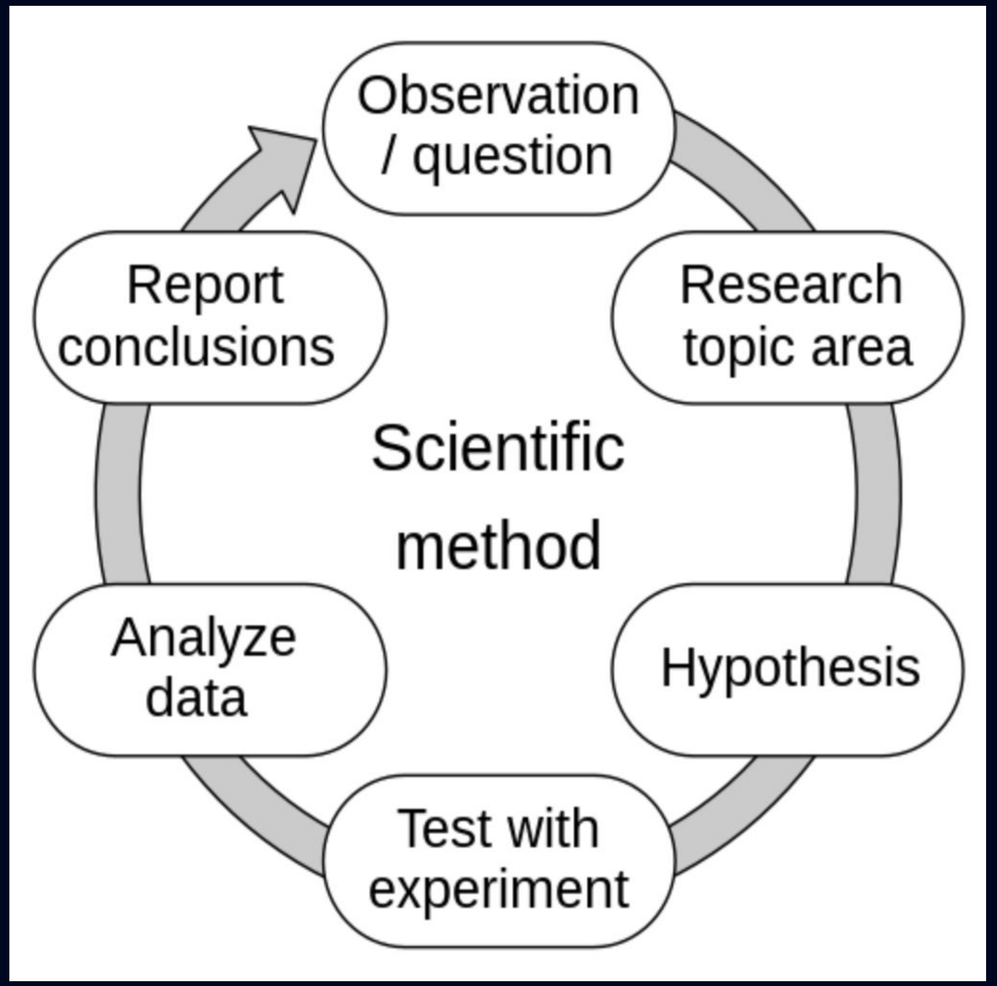
Experienced troubleshooters rely more on case-based strategies

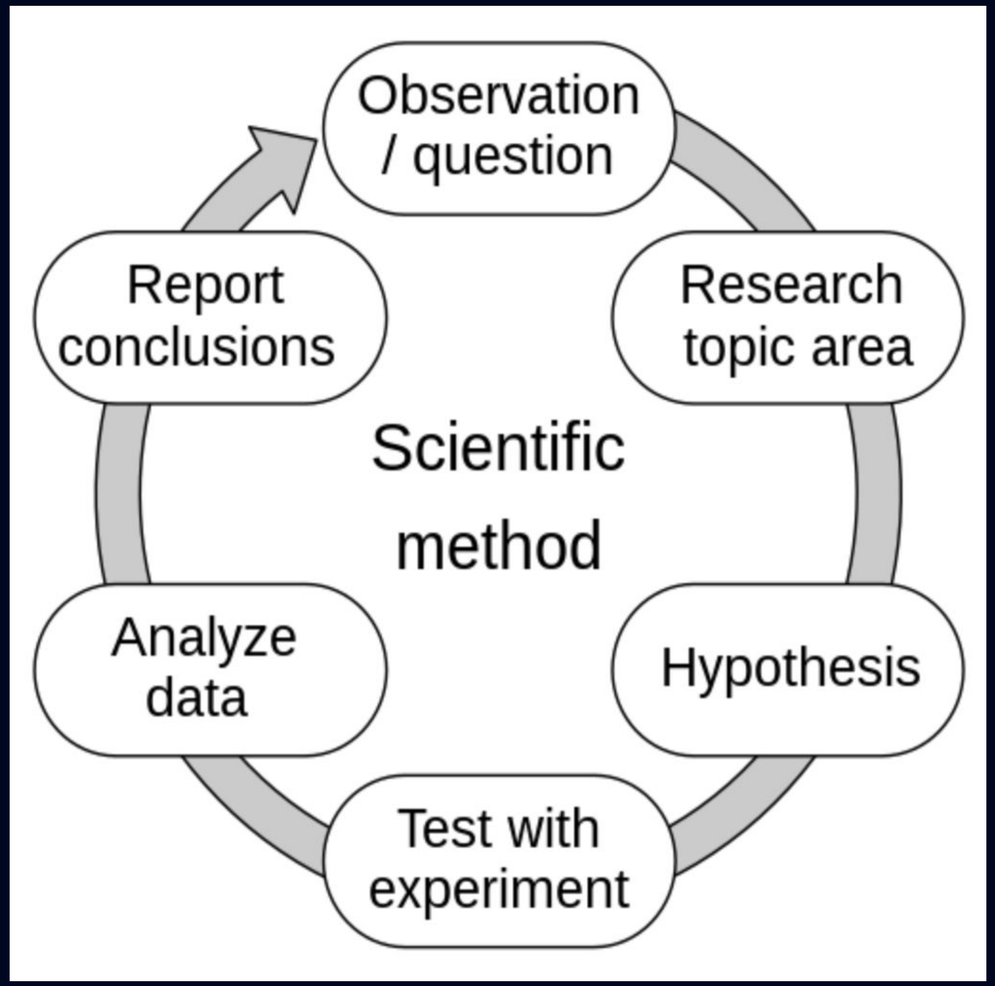
Learning to troubleshoot: A new theory-based design architecture - David H. Jonassen and Woei Hung, 2006

Science is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe.

wikipedia

The Theory of Falsifiability




Observation / question
Research topic area
Hypothesis
Test with experiment
Analyze data
Report conclusions
Scientific method

‘A more scientific, hypothesis-driven, approach to how humans perform and document incident investigations can improve reliability’

Ivan Merrill, Conf42 Platform engineering talk, just now

A possible explanation for ...

A high error rate

Is that...

There is a high database latency creating a backlog of database requests, which are in turn filling our API DB connection pool.

We can disprove this by...

- Looking at the number of available connections in our API DB Connection pool
- Comparing our current database latency to a previous time period

Why bother?

- Remove bad avenues of investigation
- Allows for change as an incident changes
- Formalizes the language used to explain decision making
- Provides a level of safety
- Avoids normative language
- Hindsight bias is reduced as there is context

1. Look for changes
2. Widen the net
3. Occam’s razor

Trade-offs under pressure: heuristics and observations of teams resolving internet service outages, J.Allspaw, 201

All practitioner acts are a gamble

Richard I. Cook, M.D. - How complex systems fail, 1998.