Newsroom Sensemaking

A Process Model for Epistemic Humility in Data and Computation

Charles Berret
Stephen Kasica
Tamara Munzner

University of British Columbia

February 19, 2021

A process model for data journalism

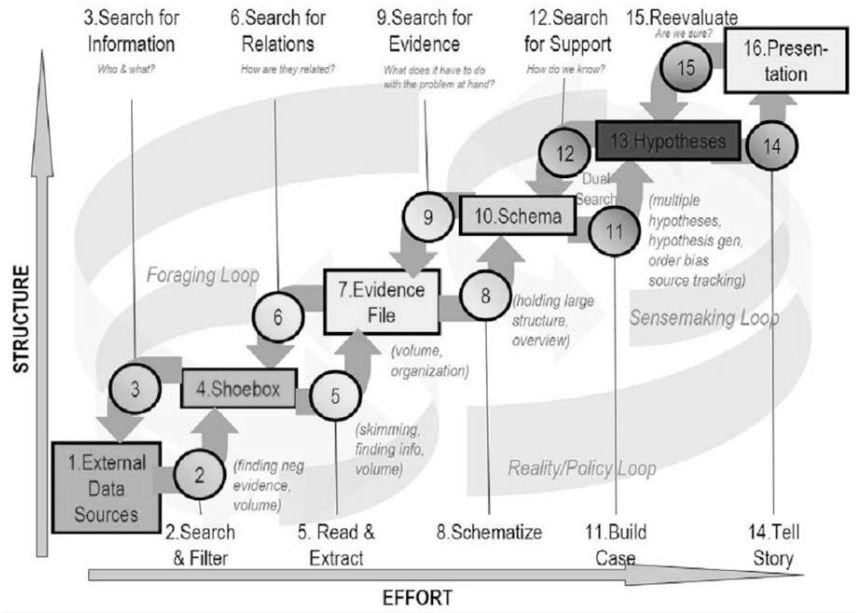
Models of sensemaking capture the “how” of data analysis.

Classic models from data science are missing key aspects of journalistic practice.

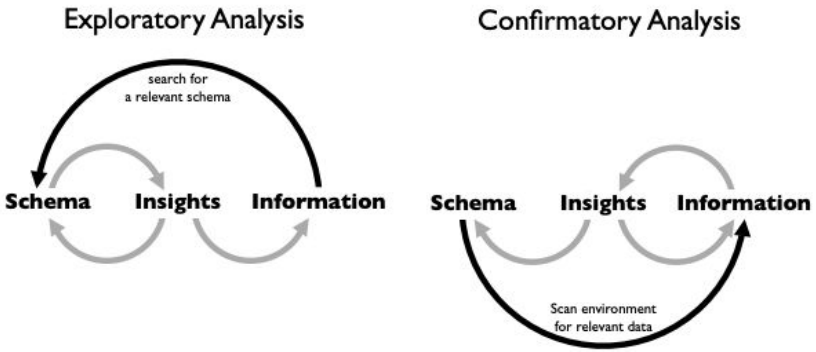
Why design a sensemaking model for data journalism?

- Provide practical, stage-by-stage guidance for data-driven reporting.
- Reveal differences between data analysis practices in journalism and other fields.
- Highlight journalistic values and practices that may be beneficial elsewhere.

Previous Models of Data Sensemaking

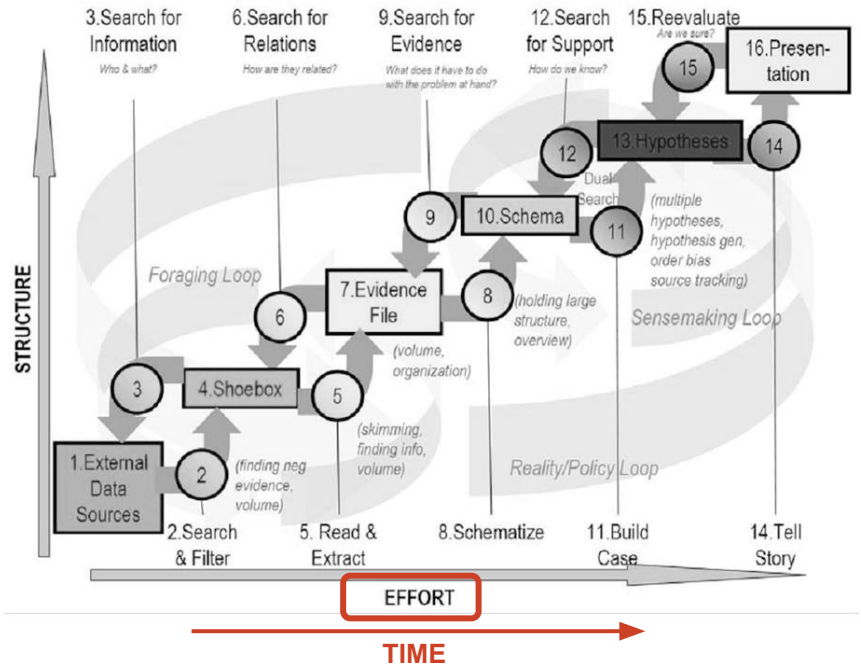


Pirolli, Peter, and Stuart Card. “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis.” Palo Alto: Xerox PARC, 2005.

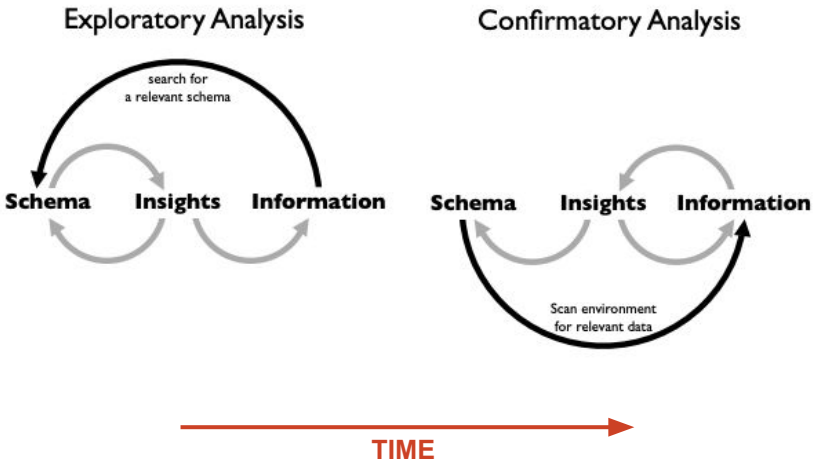


Grolemund, Garrett, and Hadley Wickham. “A Cognitive Interpretation of Data Analysis.” International Statistical Review 82, no. 2 (August 2014): 184–204.

Previous Models of Data Sensemaking

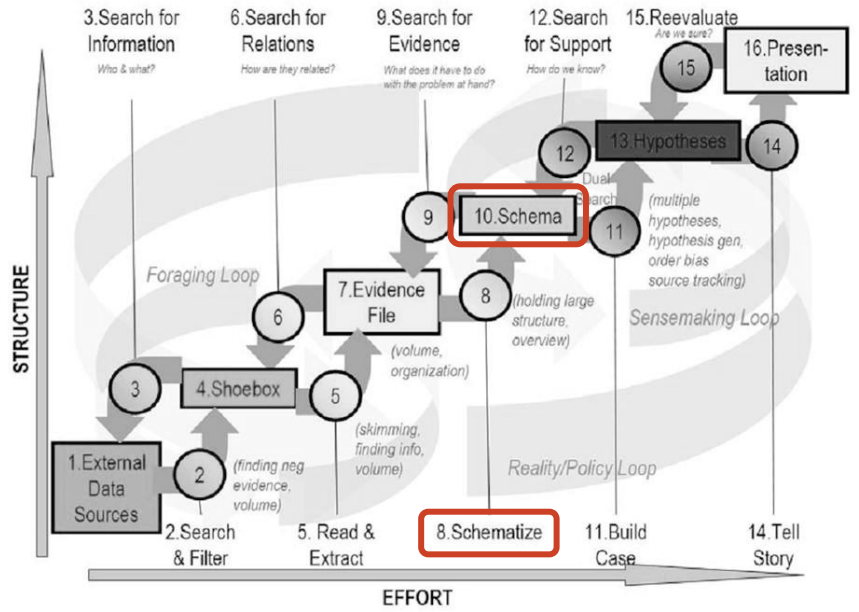


Pirolli, Peter, and Stuart Card. “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis.” Palo Alto: Xerox PARC, 2005.

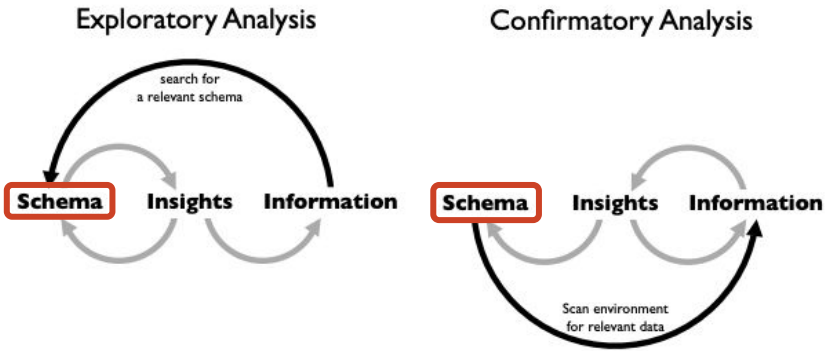


Grolemund, Garrett, and Hadley Wickham. “A Cognitive Interpretation of Data Analysis.” International Statistical Review 82, no. 2 (August 2014): 184–204.

Previous Models of Data Sensemaking



Pirolli, Peter, and Stuart Card. “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis.” Palo Alto: Xerox PARC, 2005.



Grolemund, Garrett, and Hadley Wickham. “A Cognitive Interpretation of Data Analysis.” International Statistical Review 82, no. 2 (August 2014): 184–204.

Schemas and Sensemaking

A schema is a mental model, frame, hypothesis, argument, perspective.

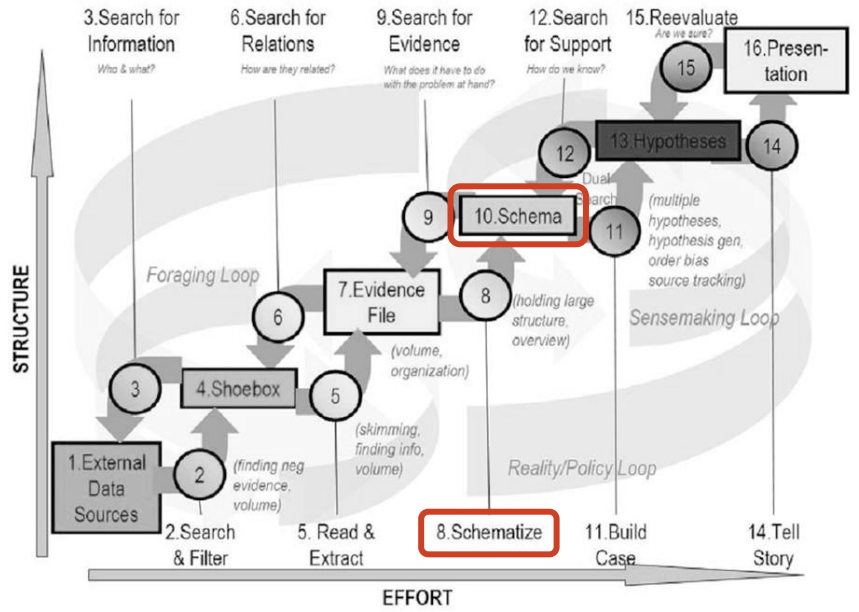
- *not* a database schema.

Example of schemas in journalism — *What's causing low public school performance?*

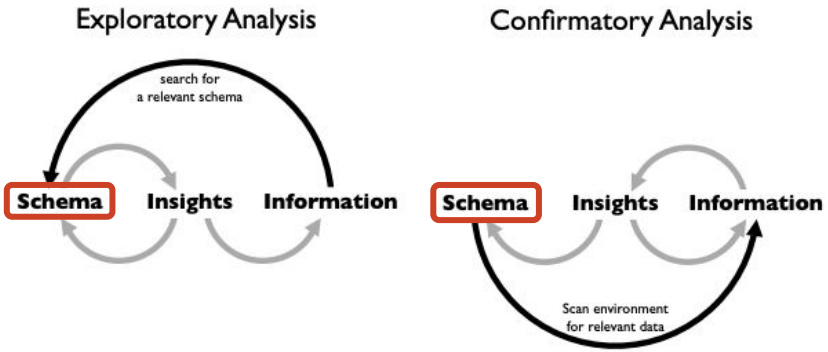
- Is it underpaid or poorly trained teachers? Outdated curriculum or textbooks? A biased testing system? Rising poverty? Lead poisoning?
- Each angle uses a different schema and requires a different approach.

Frame analysis in journalism, communication, and media studies amounts to critique of schemas represented in the news. Ditto critiques of bias in the tech industry.

Previous Models of Data Sensemaking



Pirolli, Peter, and Stuart Card. “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis.” Palo Alto: Xerox PARC, 2005.



Grolemund, Garrett, and Hadley Wickham. “A Cognitive Interpretation of Data Analysis: A Cognitive Interpretation of Data Analysis.” International Statistical Review 82, no. 2 (August 2014): 184–204.

Schematic Multiplicity in Journalistic Practice

Previous models do not address the danger of rejecting viable schemas due to heuristic biases and other prejudices.

Journalists have developed techniques to mitigate this problem by considering multiple schemas (or frames) in parallel.

- *Is low school performance due to underpaid or poorly trained teachers? Outdated curriculum or textbooks? A biased testing system? Rising poverty? Lead poisoning?*
- A reporter needs to seek out and check out various possible explanations.

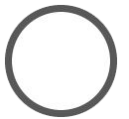
Our model insists on schematic multiplicity to better address this problem common to both journalists and data analysts.

Our Model — Schemas First

The process always begins with a schema.

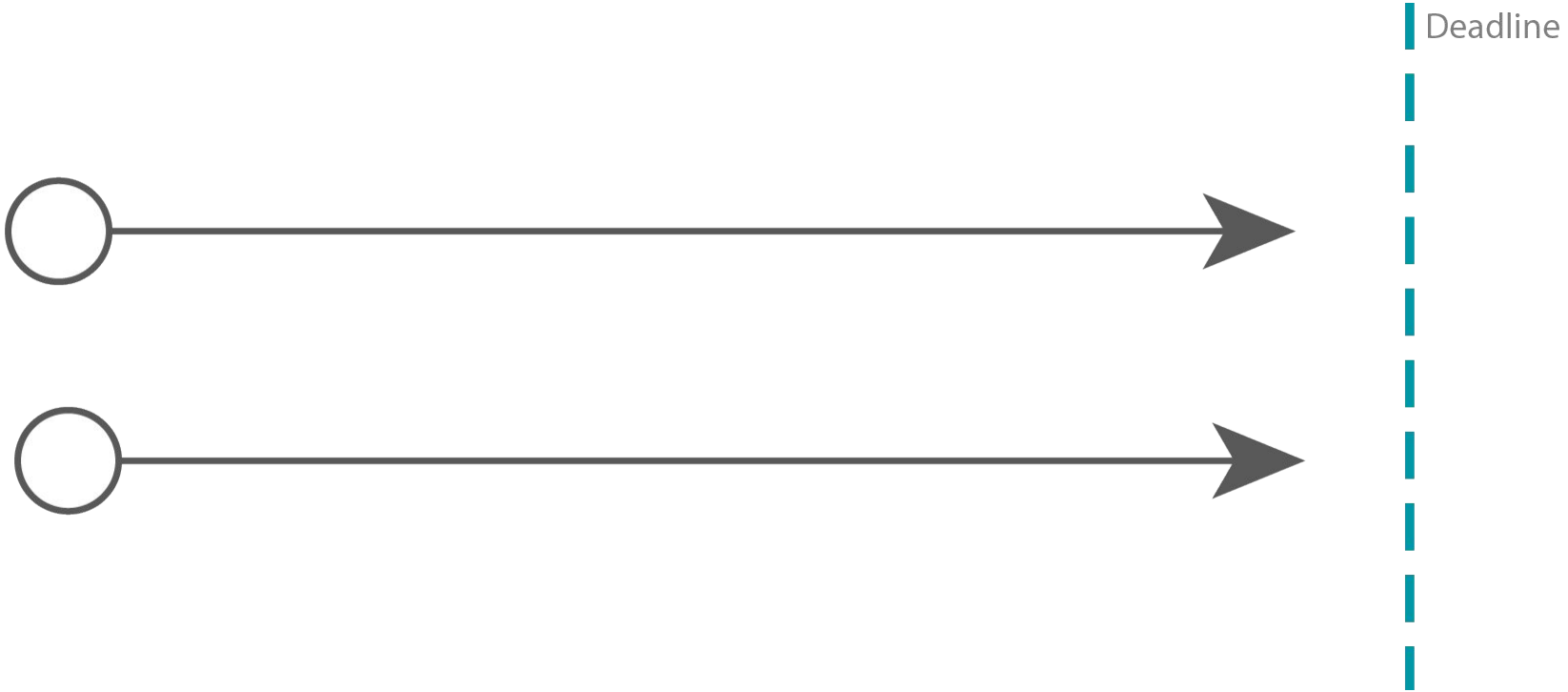
Our Model — Schematic Multiplicity

The process must proceed with consideration of additional schemas.



Our Model — Deadline-Driven

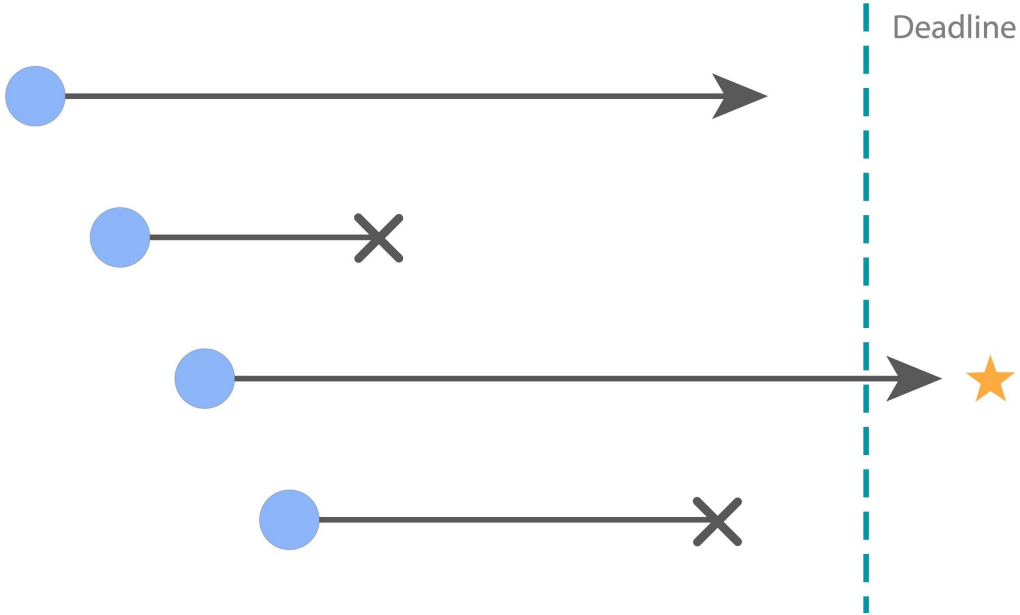
The story you publish is the most defensible schema on deadline.



Our Model — Adding and Dropping Schemas

New schemas are added over time.

Some may be discredited even before deadline.

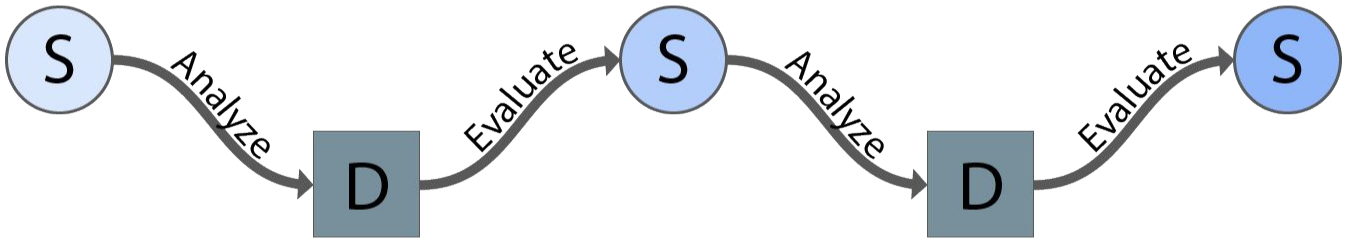


Our model — Sensemaking as Casemaking Over Time

Analyze data to evaluate defensibility of a schema.

Over time, strengthen or weaken case.

Deadline



Defensibility

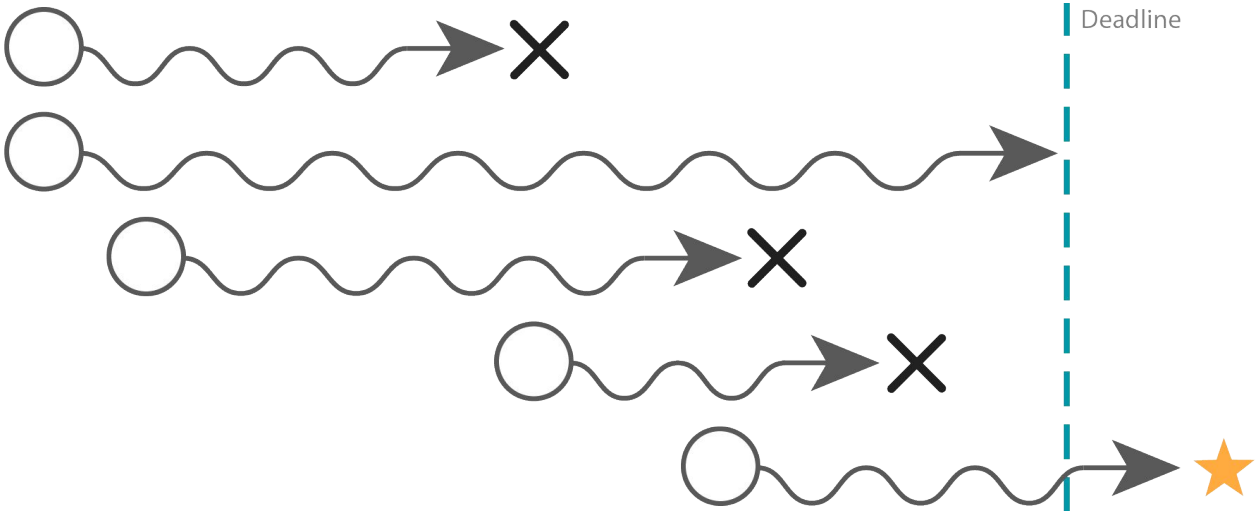
Data D

Schema S

Our Model — Sensemaking as Casemaking Over Time

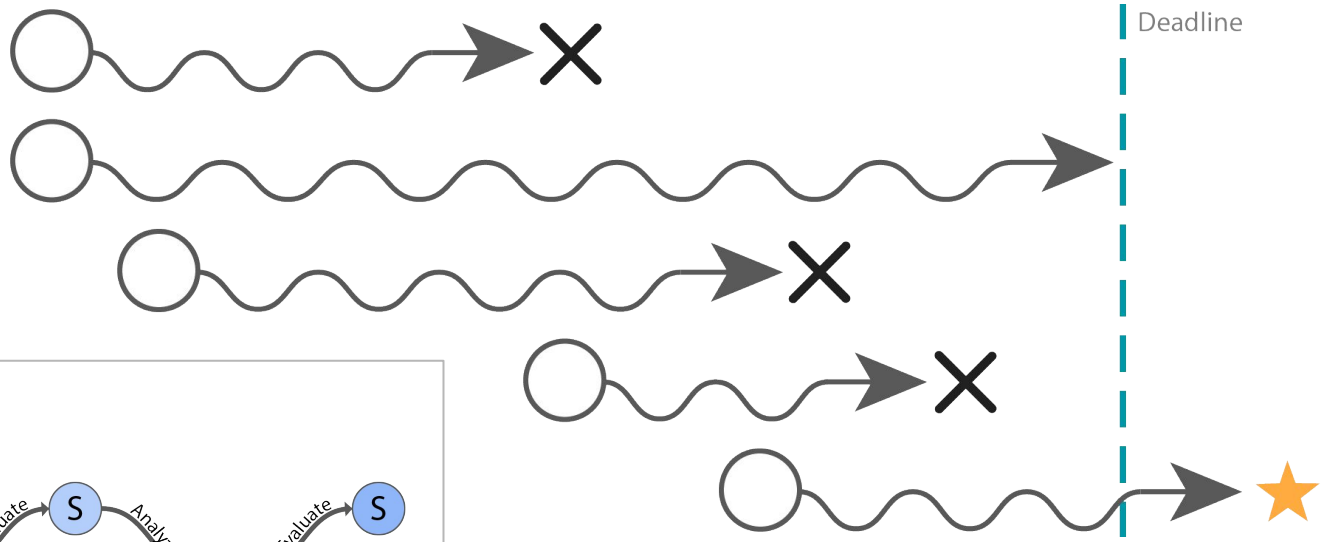
Possible outcomes of analysis stage:

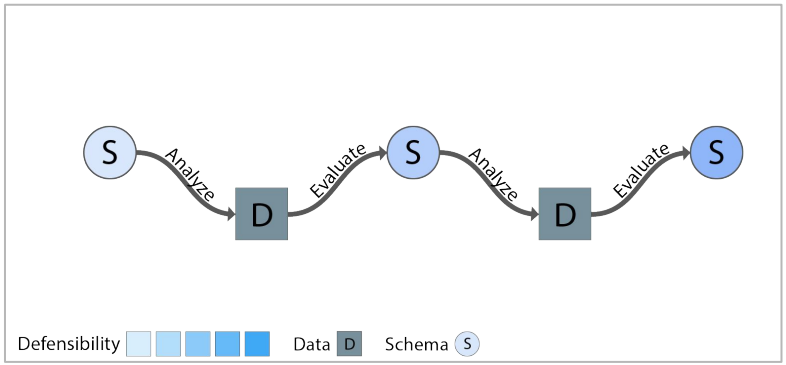
- Refine existing schema
- Abandon indefensible schema
- Form new schema



Our Model — Multi-Schema Sensemaking on Deadline

the story you publish reflects the most defensible schema on deadline





Our Model — Key Features **(ORIGINAL)**

1. **Schemas first**

   The entry point for any sensemaking process is a schema.

2. **Schematic multiplicity**

   Sensemaking requires pursuit and consideration of competing schemas.

3. **Temporal progression**

   Time is a first-class citizen in our model.

4. **Sensemaking as casemaking through data analysis on deadline**

   Relative defensibility of schemas assessed through progressive analysis as data accumulates.

   The story you publish reflects the schema that is most defensible on deadline.

Key Features of the Epistemic Humility Model

1. **Schemas first**

   The entry point for any sensemaking process is a schema.

2. **Schematic multiplicity**

   Sensemaking requires pursuit and consideration of competing schemas.

3. **Data analysis on deadline**

   Time is a first-class citizen: temporal progression is built into the model

   Termination condition is deadline pressure not certainty of conclusions

4. **Sensemaking as Casemaking**

   Journalistic sensemaking is an intrinsically collaborative process: writer is making a case to an editor

   (Multiple people rather than collaborative? Discursive?)

Relative defensibility of the multiple schemas assessed through progressive analysis, publish story reflecting most defensible schema at time of deadline

Our Model — Implications for Journalism

- Offer a model of data analysis tailored to journalism.
  - Develop tools and workflows that accurately reflect professional practices, values, norms, and pressures of data-driven reporting.
- Model with explanatory + prescriptive power
  - Explanatory power to identify potential hazards of the reporting process.
    - *Biased/Incomplete*: insufficient variety of schemas considered.
    - *Inaccurate*: faulty judgment of schema's defensibility.
    - *Under-resourced*: inadequate reporting time before deadline.
  - Prescriptive power to mitigate those hazards
    - Explicitly consider multiple schemas -> to avoid biased/incomplete
    - Explicitly consider defensibility of schema -> to avoid inaccurate reporting
- Support journalism instruction and editorial guidance.
  - Teach students and direct working journalists to seek out multiple schemas and continuously weigh their defensibility as deadline approaches.
- +when should you do data analysis with info gathered through traditional journalistic practices?
  - PDFs thru FOIA? yes. Interview data? maybe.

Our Model — Implications for Data Science

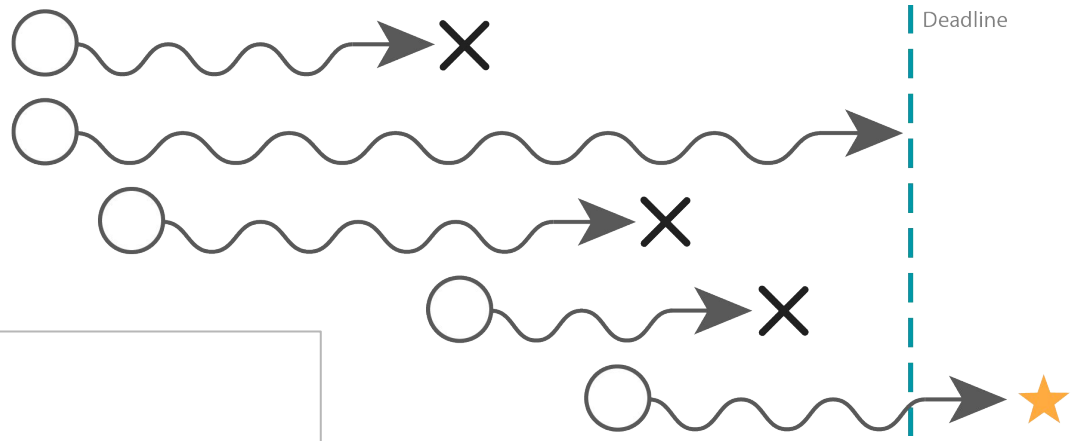
- In our **schema-first** model, every project begins with an initial schema.
  - Making this explicit encourages attention to our own biases.
- **Schematic multiplicity** underlines the importance of considering alternative explanations and perspectives.
  - Current models of data analysis imply that analysts work with a single schema throughout.
- **Deadline-driven** model foregrounds the role of time pressure.
  - Current models account for the cost-structure of time spent, but not the effect of deadlines on the conclusions we reach.
- Everyone who works with data could benefit from the **epistemic humility** at the heart of sound, ethical journalism.
  - Journalism is more than storytelling chops.

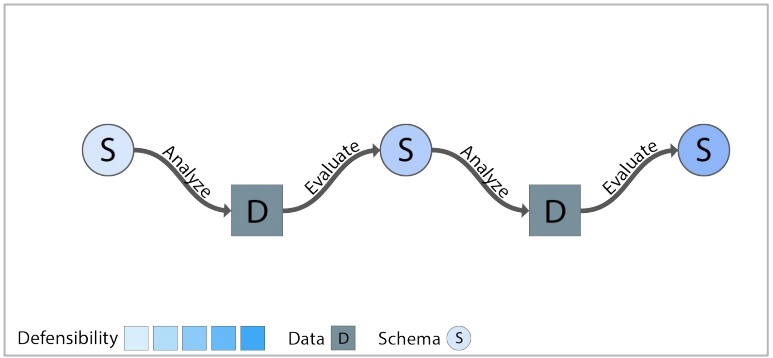
Newsroom Sensemaking

A Process Model for Epistemic Humility in Data and Computation

Charles Berret
Stephen Kasica
Tamara Munzner

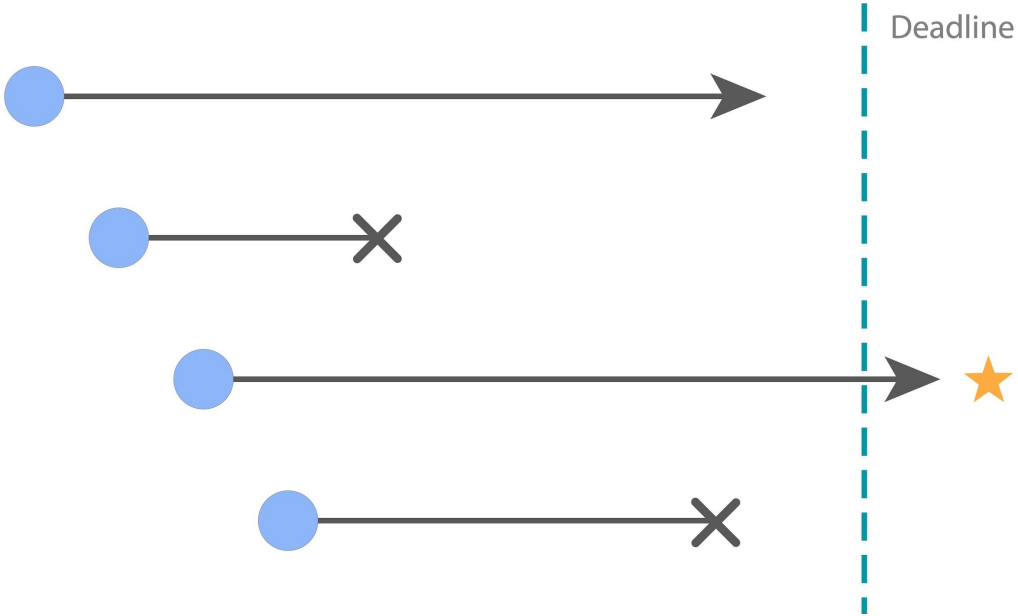
University of British Columbia



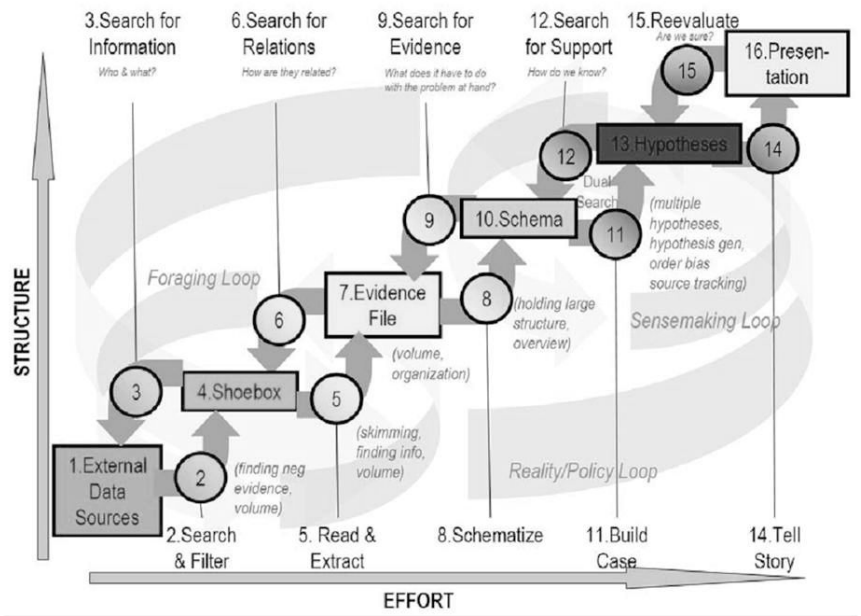


Multiple schemas and journalistic practice

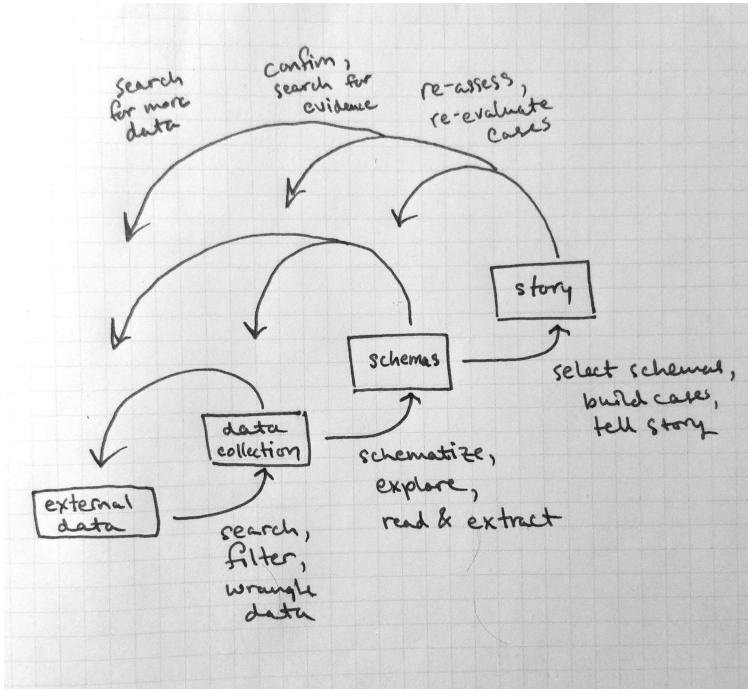
A process model for journalistic sensemaking must account for the consideration of multiple competing schemas in sound reporting.



Previous Models of Data Analysis



Pirolli & Card (2005)



Adapted for data journalism

What is journalistic objectivity?

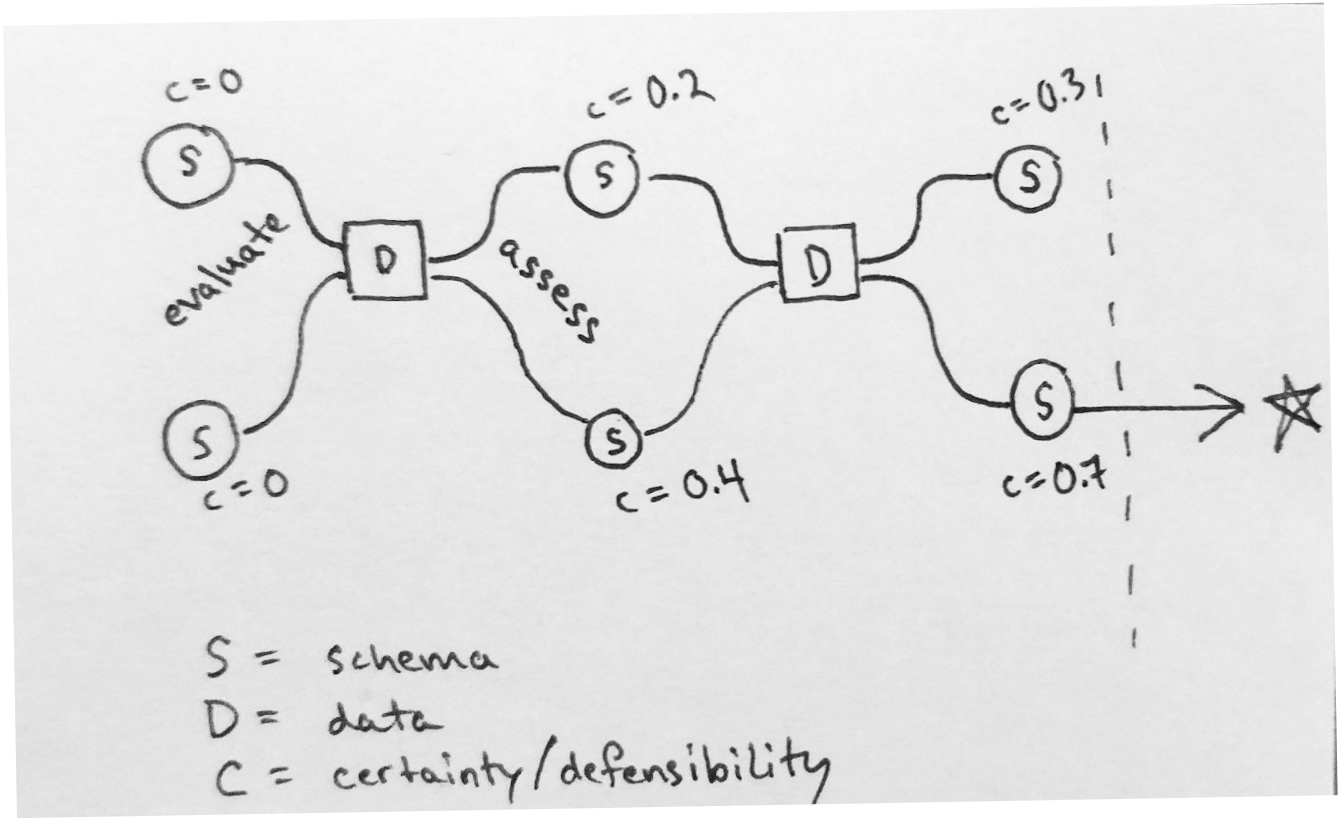
Bill Keller and Glenn Greenwald

The New York Times. October 27, 2013.

What journalistic objectivity **does not mean**

- that journalists can reveal objective truth, or that facts speak for themselves
- that journalists should be completely impartial, telling both sides even if one is misleading or clearly untrue
- that journalists are blank slates without personal opinions.

What is journalistic objectivity?

- journalists should operate with **epistemic humility**, attentive to the many ways they can get the story wrong, and they must back up claims with evidence drawn from many sources.
- journalistic objectivity is motivated by distrust of human biases, limitations, errors of judgment — but above all, the fabrications of self-interested parties.

Our model (alternative)