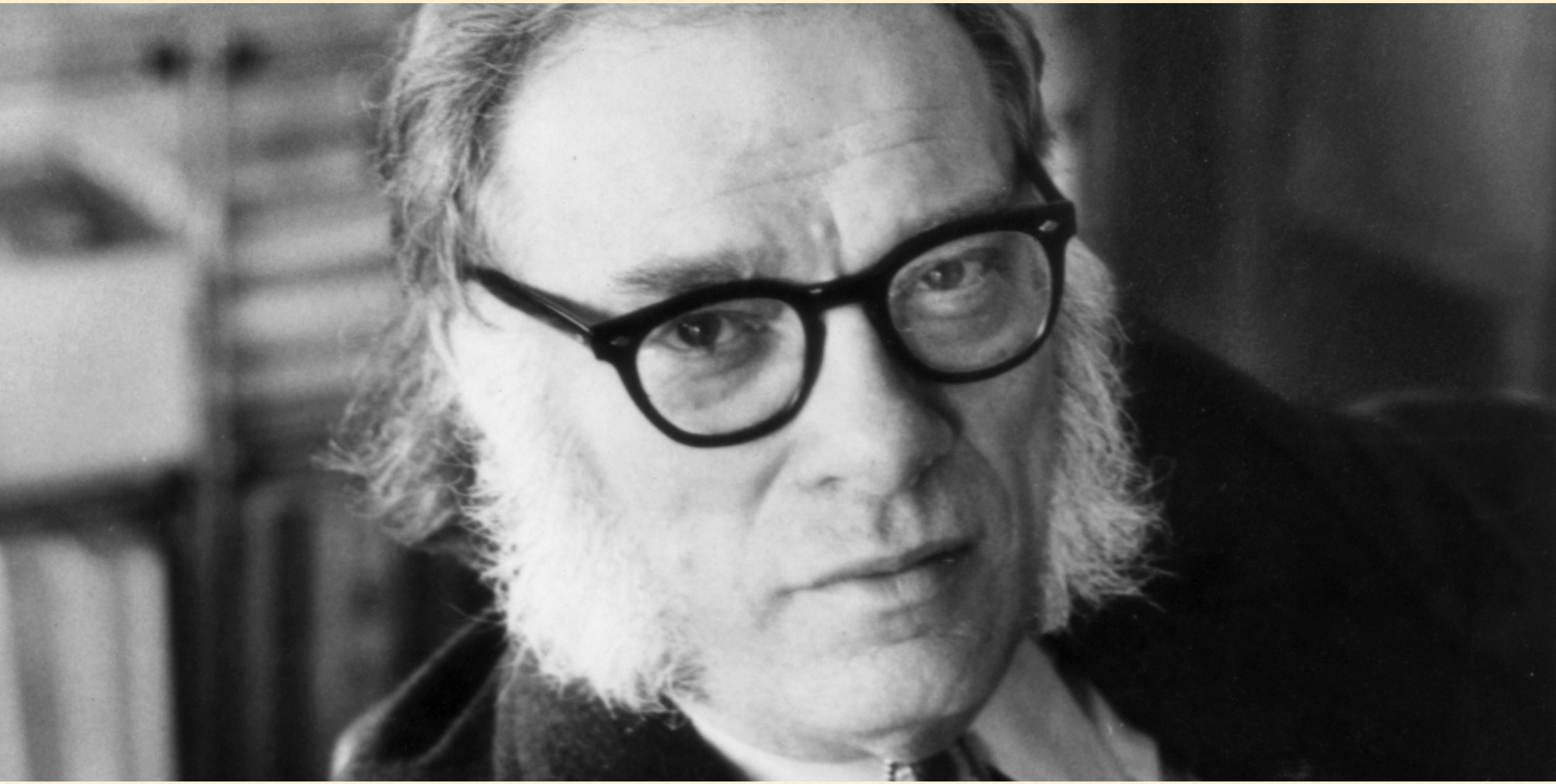
TAR INFORMATION GOVERNANCE AND THE RELATIVITY OF WRONG

Jeremy Pickens
Catalyst Repository Systems

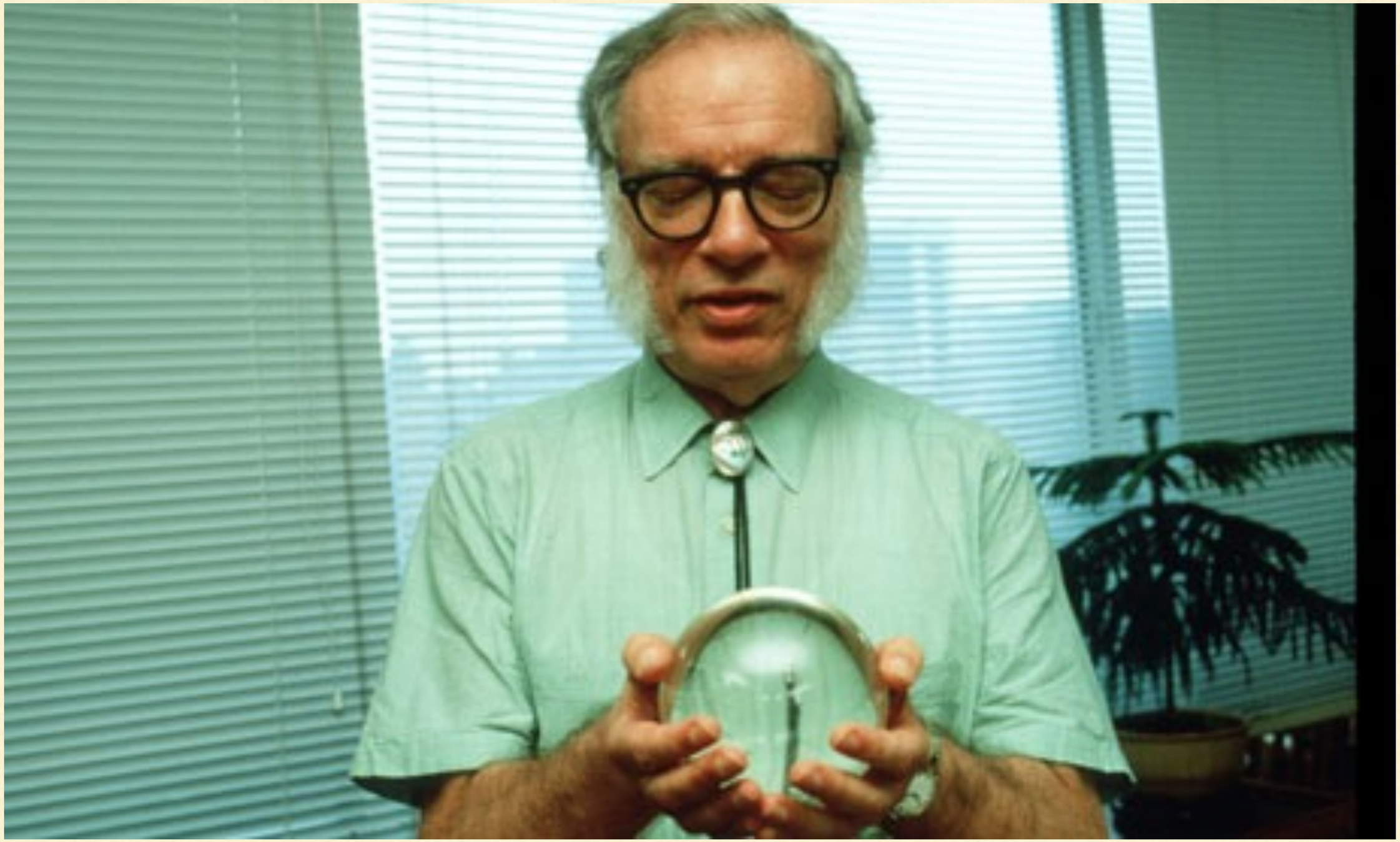
OVERVIEW

- Isaac Asimov and "The Relativity of Wrong"
- Absolutist Thinking in E-discovery
- Psychohistory $\Longleftrightarrow$ Big Data and Information Governance
- Two-Dimensional Model of Information Governance Tasks
- SwarmSketch Thought Experiment
- Managing the Relativity of Wrong

ISAAC ASIMOV



THE RELATIVITY OF WRONG

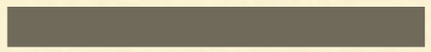


Is the Earth round?

No.

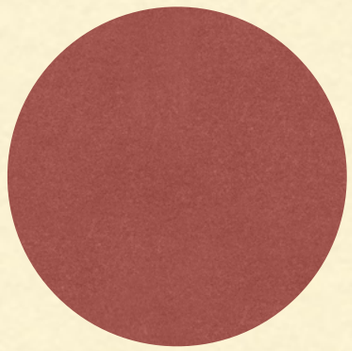
CURVATURE OF THE EARTH

Flat



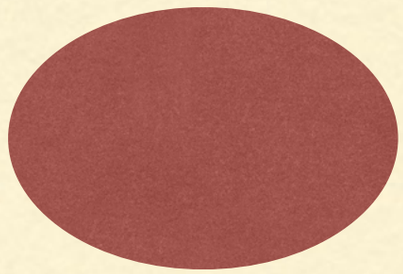
0 inches per mile

Round



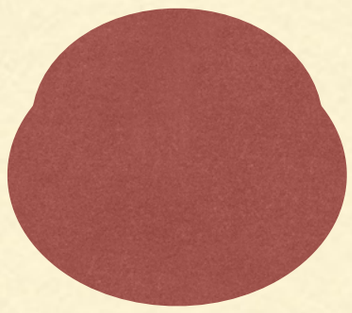
8 inches per mile

Oblate Spheroid



7.937 inches per mile
8.027 inches per mile

Pear



7.936999 inches per mile
8.0270001 inches per mile

"The basic trouble, you see, is that people think that 'right' and 'wrong' are absolute; that everything that isn't perfectly and completely right is totally and equally wrong.."

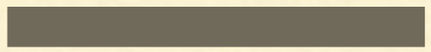
–Isaac Asimov
The Relativity of Wrong
The Skeptical Inquirer, Fall 1989, Vol. 14, No. 1, Pages 35-44

OVERVIEW

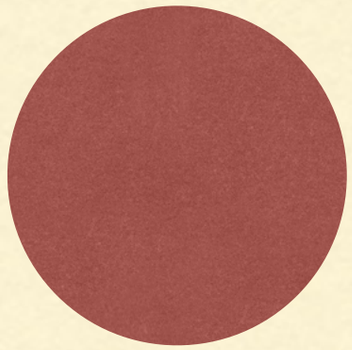
- Isaac Asimov and "The Relativity of Wrong"
- **Absolutist Thinking in E-discovery**
- Psychohistory ⟺ Big Data and Information Governance
- Two-Dimensional Model of Information Governance Tasks
- SwarmSketch Thought Experiment
- Managing the Relativity of Wrong

TAR PROCESSES

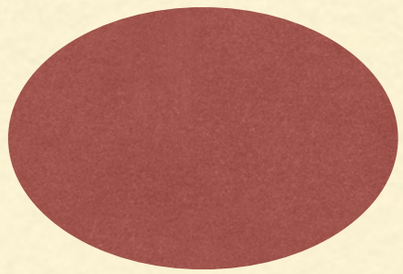
Manual Linear Review



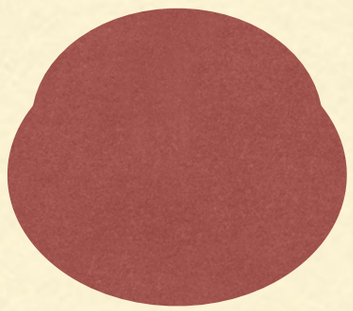
Keyword Search



Supervised Learning with Limited Training (SPL, SAL)



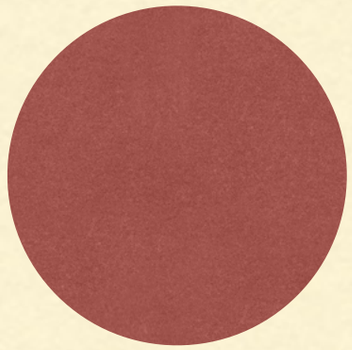
Supervised Learning with Unlimited Training (CAL)



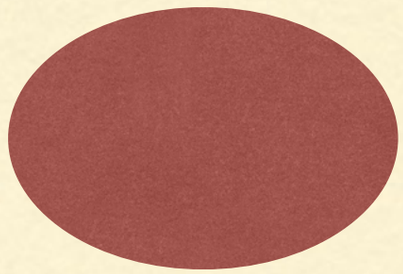
VALIDATION SAMPLING

Size = 0

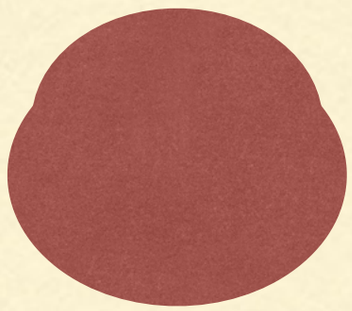
Size = 383



Size = 384

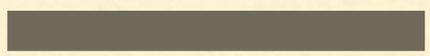


Size = 385

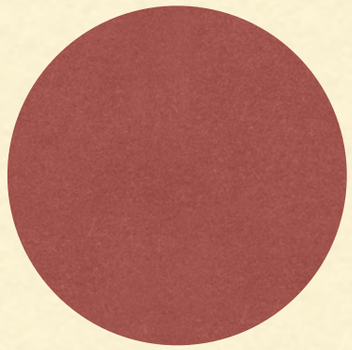


TAR TRAINING

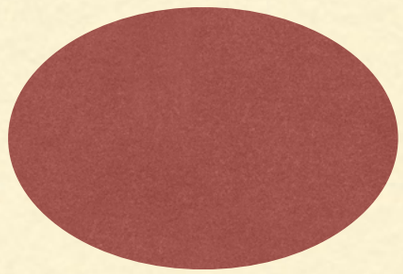
2nd grader
from Kyrgyzstan



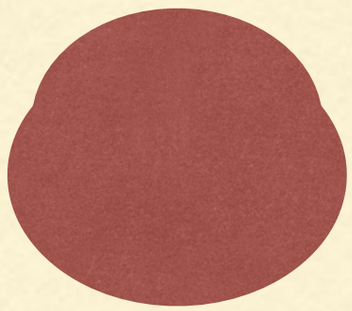
High School Student



Contract Attorney

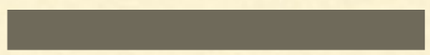


Subject Matter Expert/
Lead Attorney

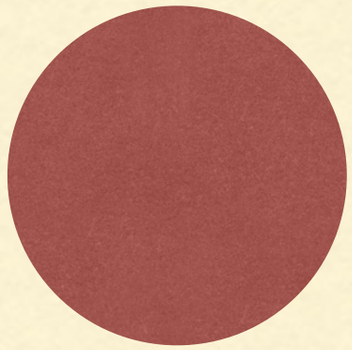


TAR PROGRESS METRICS

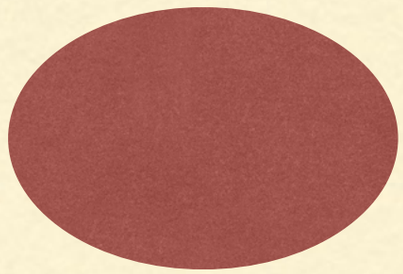
No metric



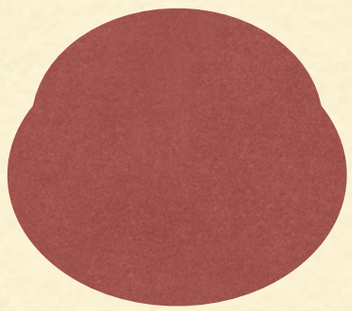
F1

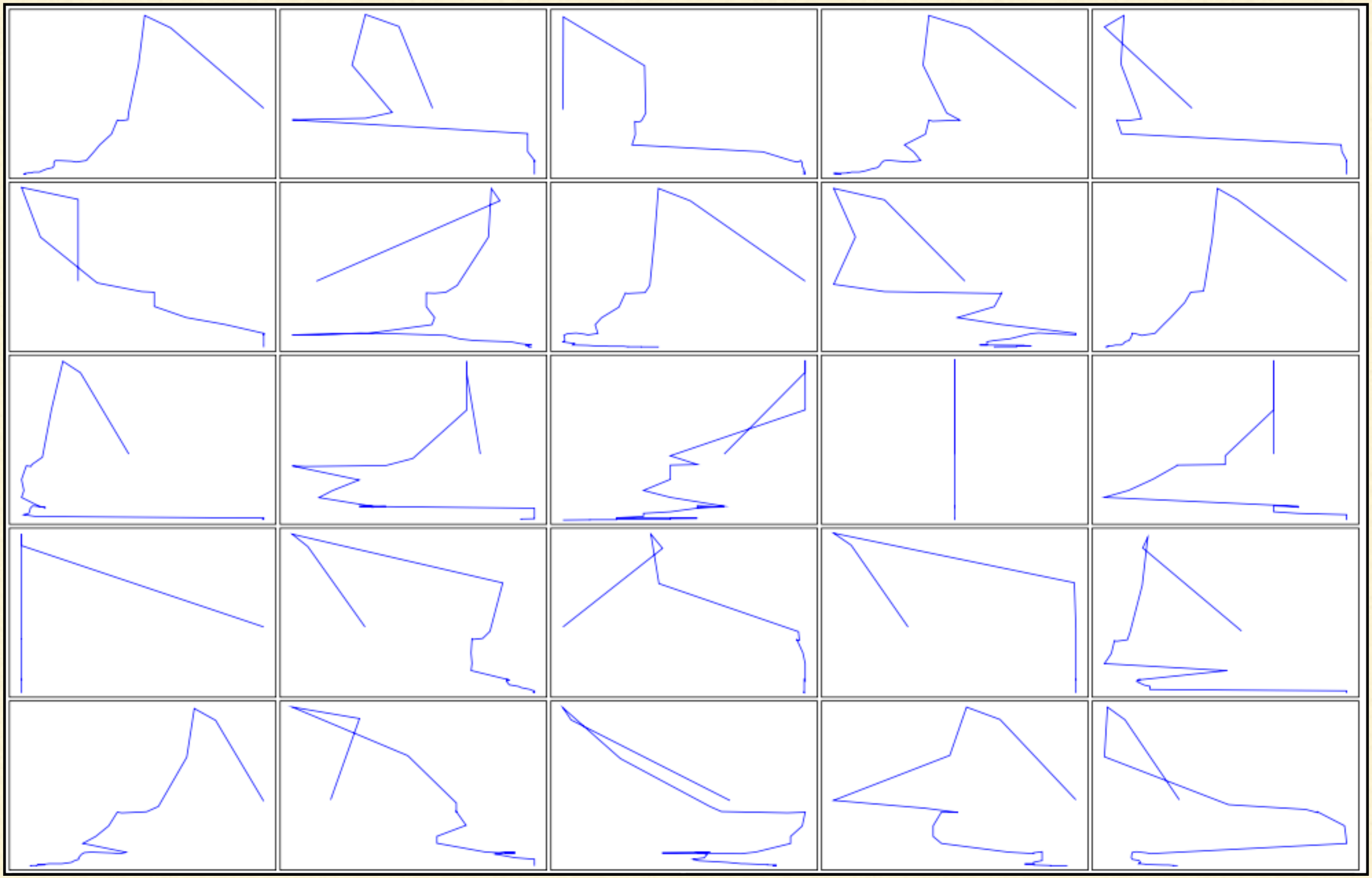


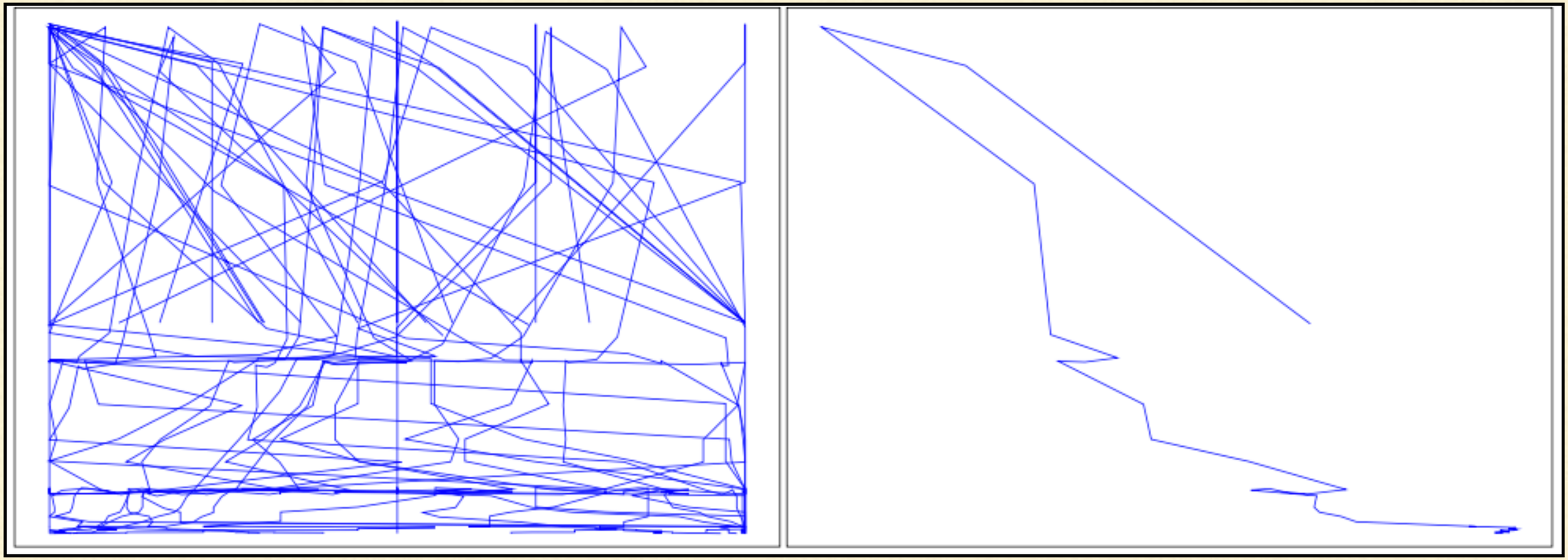
global change

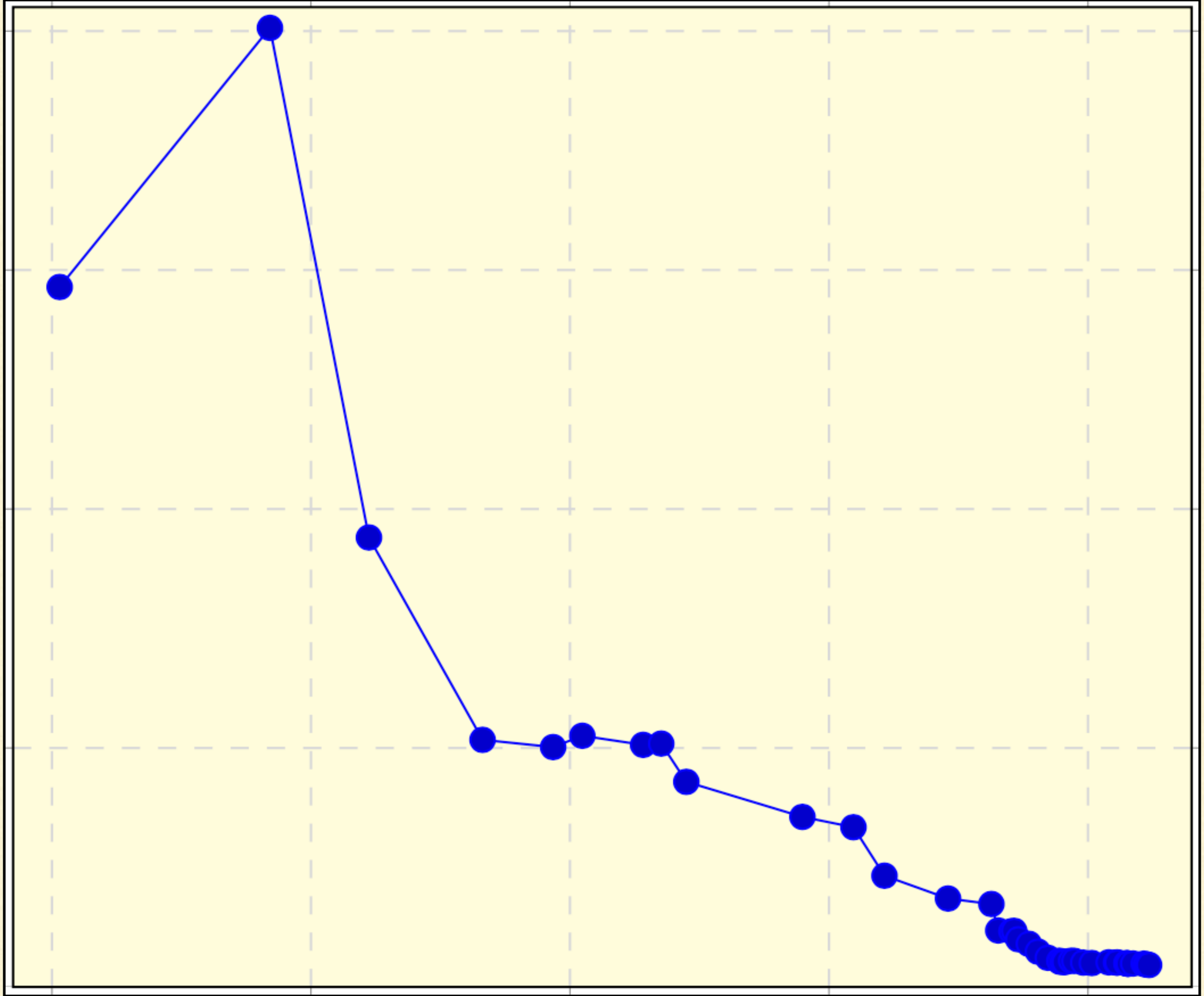


???



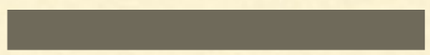




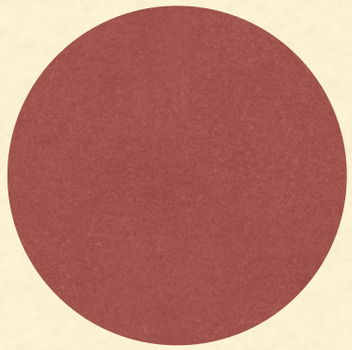


TAR PROGRESS METRICS

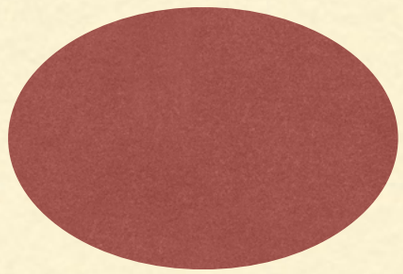
No metric



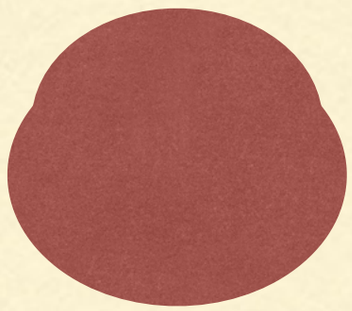
F1



global change



???



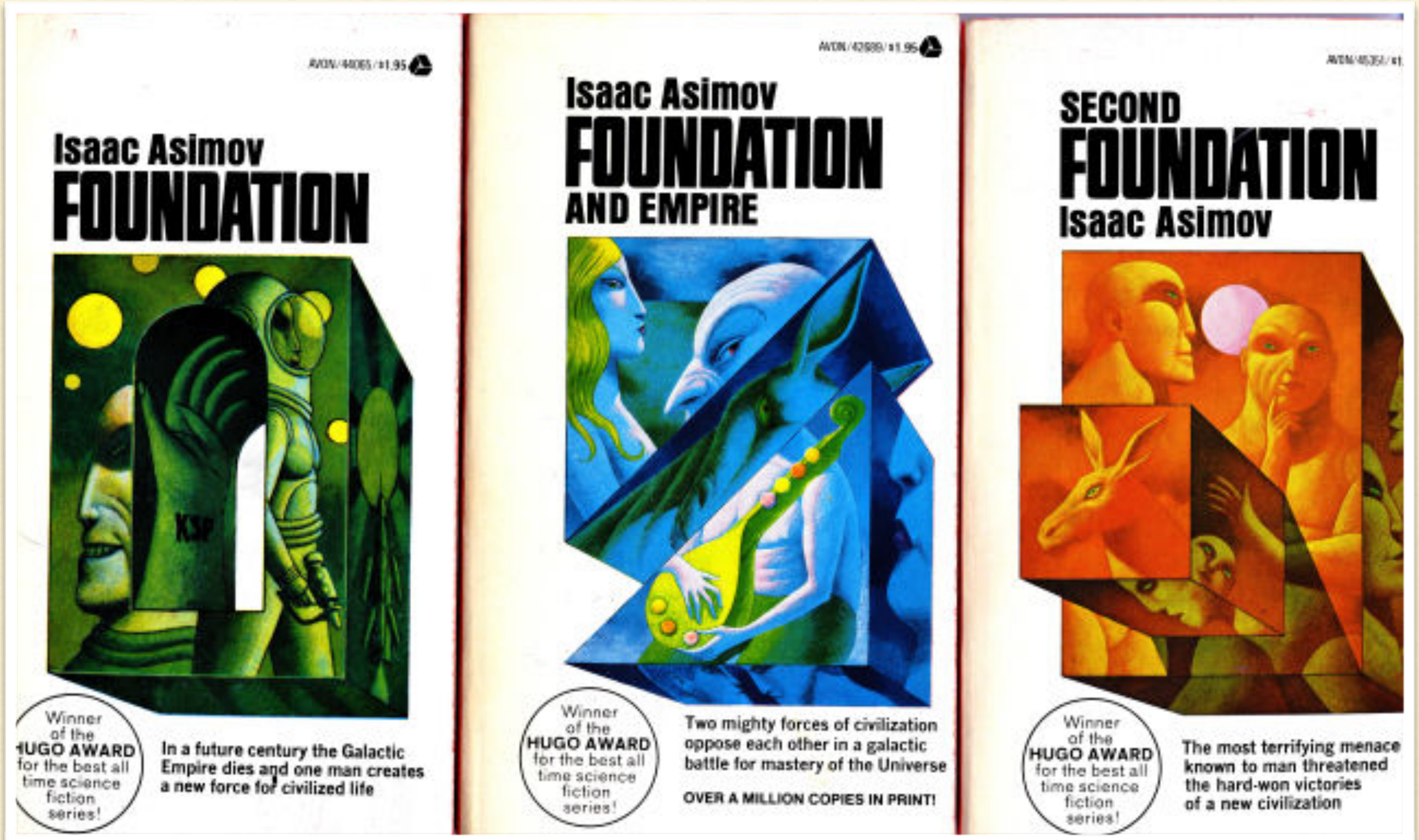
"When people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together."

–Isaac Asimov
The Relativity of Wrong
The Skeptical Inquirer, Fall 1989, Vol. 14, No. 1, Pages 35-44

OVERVIEW

- Isaac Asimov and "The Relativity of Wrong"
- Absolutist Thinking in E-discovery
- **Psychohistory ⟺ Big Data and Information Governance**
- Two-Dimensional Model of Information Governance Tasks
- SwarmSketch Thought Experiment
- Managing the Relativity of Wrong

ASIMOV: PSYCHOHISTORY

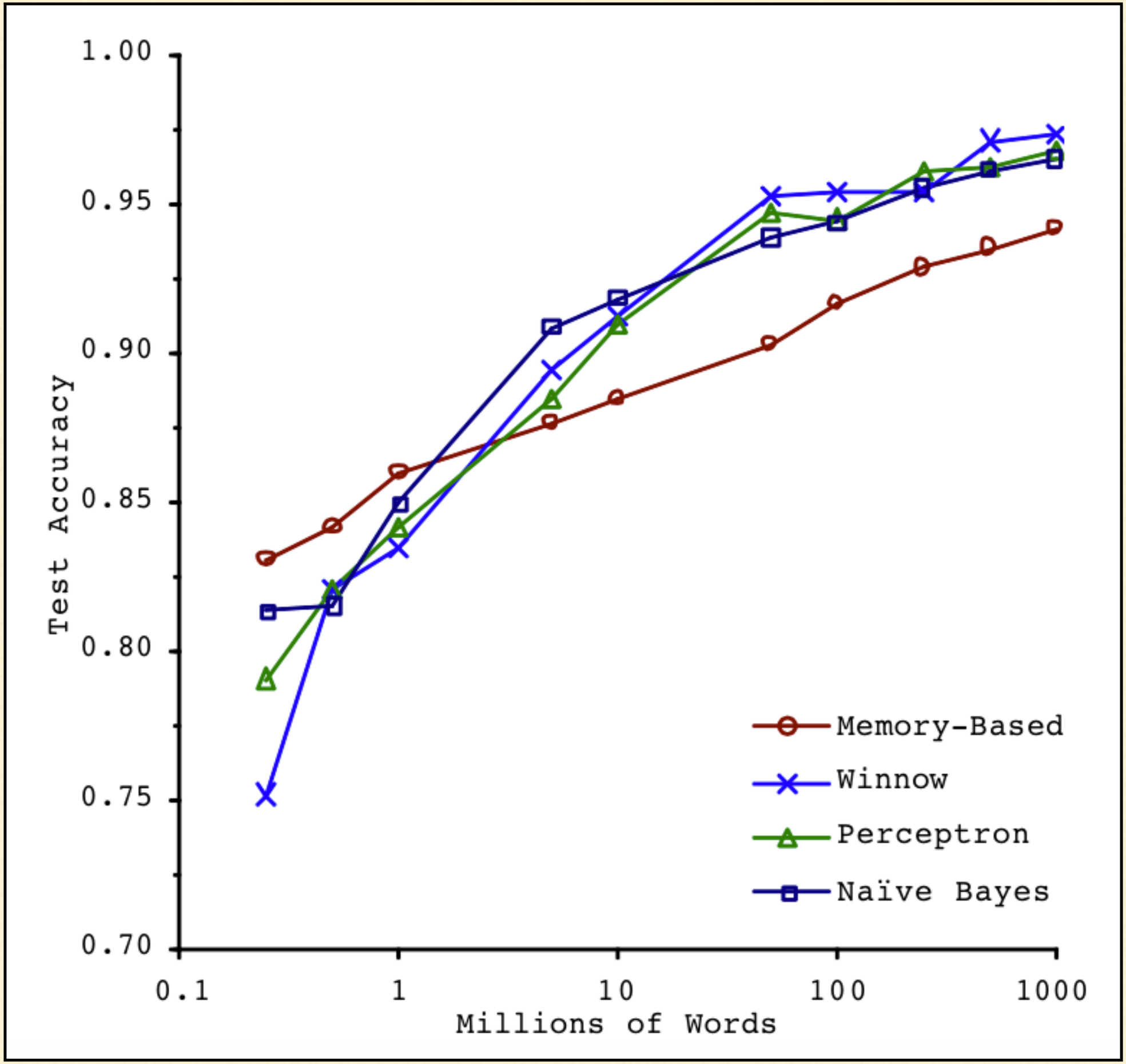


> "There's a key idea in the books that concerns a mathematical theory of human behaviour. 'Psychohistory'…is like a kinetic theory of gases for human beings – gather enough of us together (and the starting premise of the books is that humanity has colonized the entire galaxy so there are lots of people) and the overall, en bloc behaviour becomes statistically predictable."

–*http://johnnymackintosh.com/2011/08/23/influences-on-johnny-mackintosh-isaac-asimov/*

"These definitions seem to get us closer to what makes Big data a new and interesting phenomenon in the world: it is not its volume alone, but the fact that we are able to "mine" large data sets using new and advanced techniques to uncover *unexpected relationships, patterns and categories within these data sets*, that makes the field potentially exciting."

–Bennett B. Borden & Jason R. Baron, *Finding the Signal in the Noise: Information Governance, Analytics, and the Future of Legal Practice*, 20 Rich. J.L. & Tech. 7 (2014), http://jolt.richmond.edu/v20i2/article7.pdf


1.00
0.95
0.90
0.85
0.80
0.75
0.70
Test Accuracy
0.1
1
10
100
1000
Millions of Words
Memory-Based
Winnow
Perceptron
Naïve Bayes

BIG DATA
INFORMATION
GOVERNANCE

OVERVIEW

- Isaac Asimov and "The Relativity of Wrong"
- Absolutist Thinking in E-discovery
- Psychohistory ⟺ Big Data and Information Governance
- **Two-Dimensional Model of Information Governance Tasks**
- SwarmSketch Thought Experiment
- Managing the Relativity of Wrong

INFORMATION GOVERNANCE

DESCRIPTIVE (UNDERSTANDING WHAT IS)

Analysis of due diligence data collection process — Was it done right?

PREDICTIVE (WHAT WILL HAPPEN)

Early Warning System: Are there signs that the company may run into financial troubles?

PRESCRIPTIVE (FUTURE COURSES OF ACTION)

What can be done to increase long term profitability?

PSYCHOHISTORY (ASIMOV)

DESCRIPTIVE (UNDERSTANDING WHAT IS)

Galactic Empire is in decline, despite surface appearances

PREDICTIVE (WHAT WILL HAPPEN)

Galactic Empire will fall within a generation or two,
followed by a 30,000 year period of anarchy (dark ages)

PRESCRIPTIVE (FUTURE COURSES OF ACTION)

By implementing certain courses of action,
the dark ages can be reduced to a mere 1,000 years

past
descriptive
predictive
prescriptive
future

INFORMATION GOVERNANCE

Borden & Baron:

- Internal whistleblower allegations
- Exiting employee regulatory violation allegations (extortion)
- External FCPA whistleblower allegations
- Due diligence for an acquisition

SPECIFIC

INFORMATION GOVERNANCE

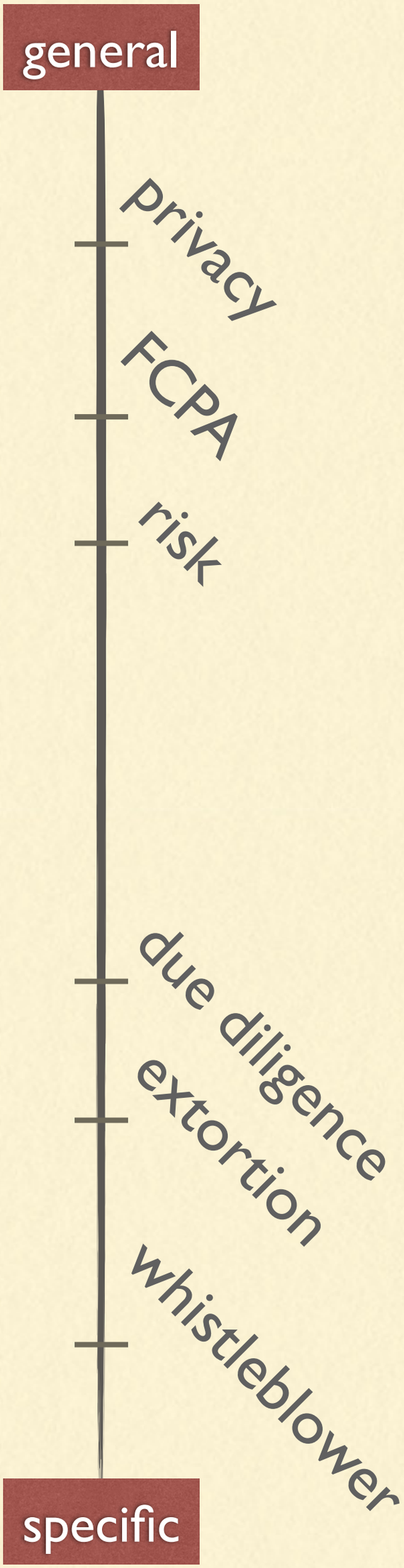
Sedona

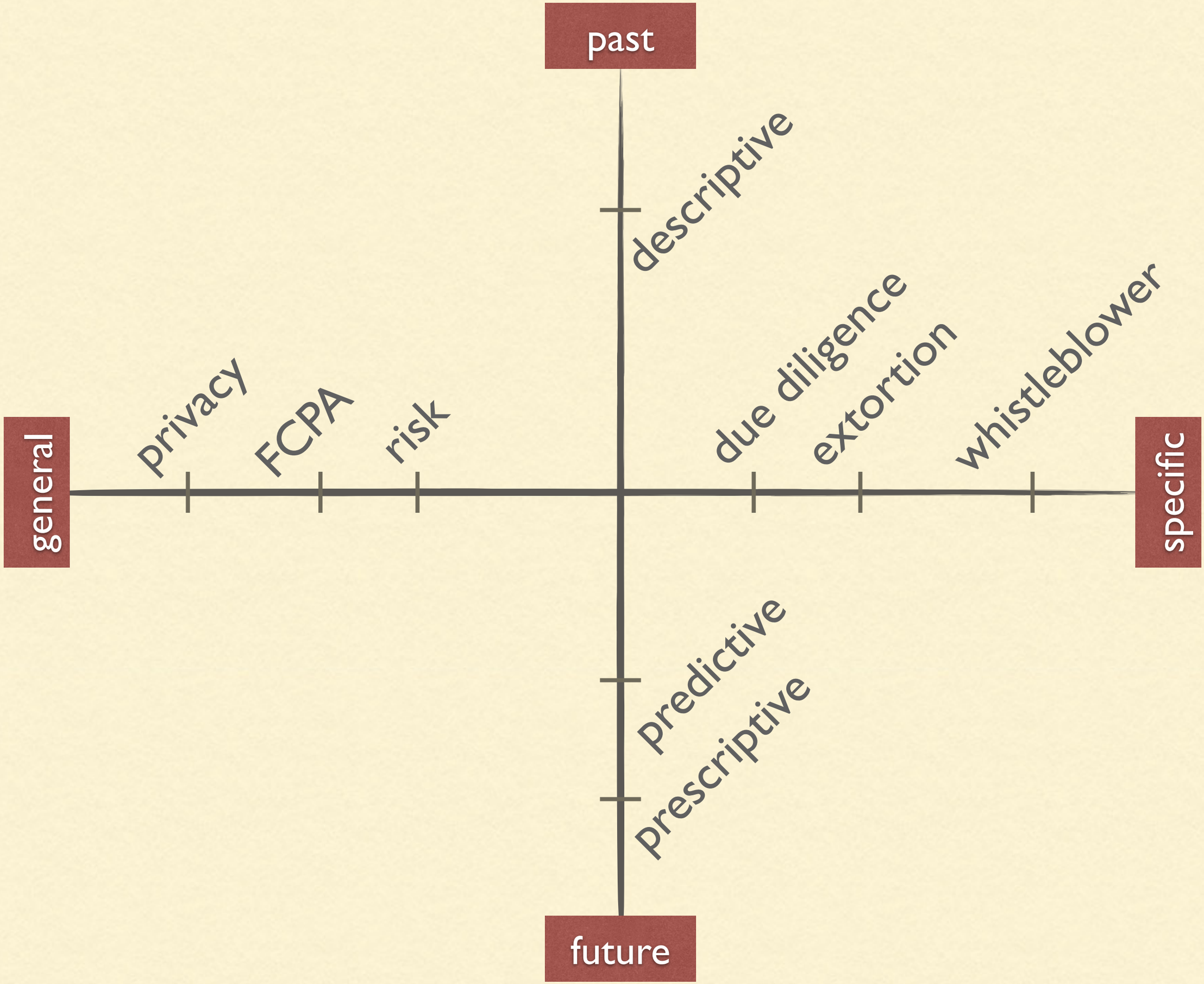
- Satisfy compliance requirements
- Manage Information Risk
- Manage Data Privacy
- Manage Information Security

Borden & Baron:

- Vet candidates for senior leadership positions
- Analyze business divisions to identify good and bad leaders
- Analyze how decisions are made
- Analyze why a division is more successful than another

GENERAL


general
privacy
FCPA
risk
due diligence
extortion
whistleblower
specific


past
descriptive
privacy
FCPA
risk
due diligence
extortion
whistleblower
general
specific
predictive
prescriptive
future

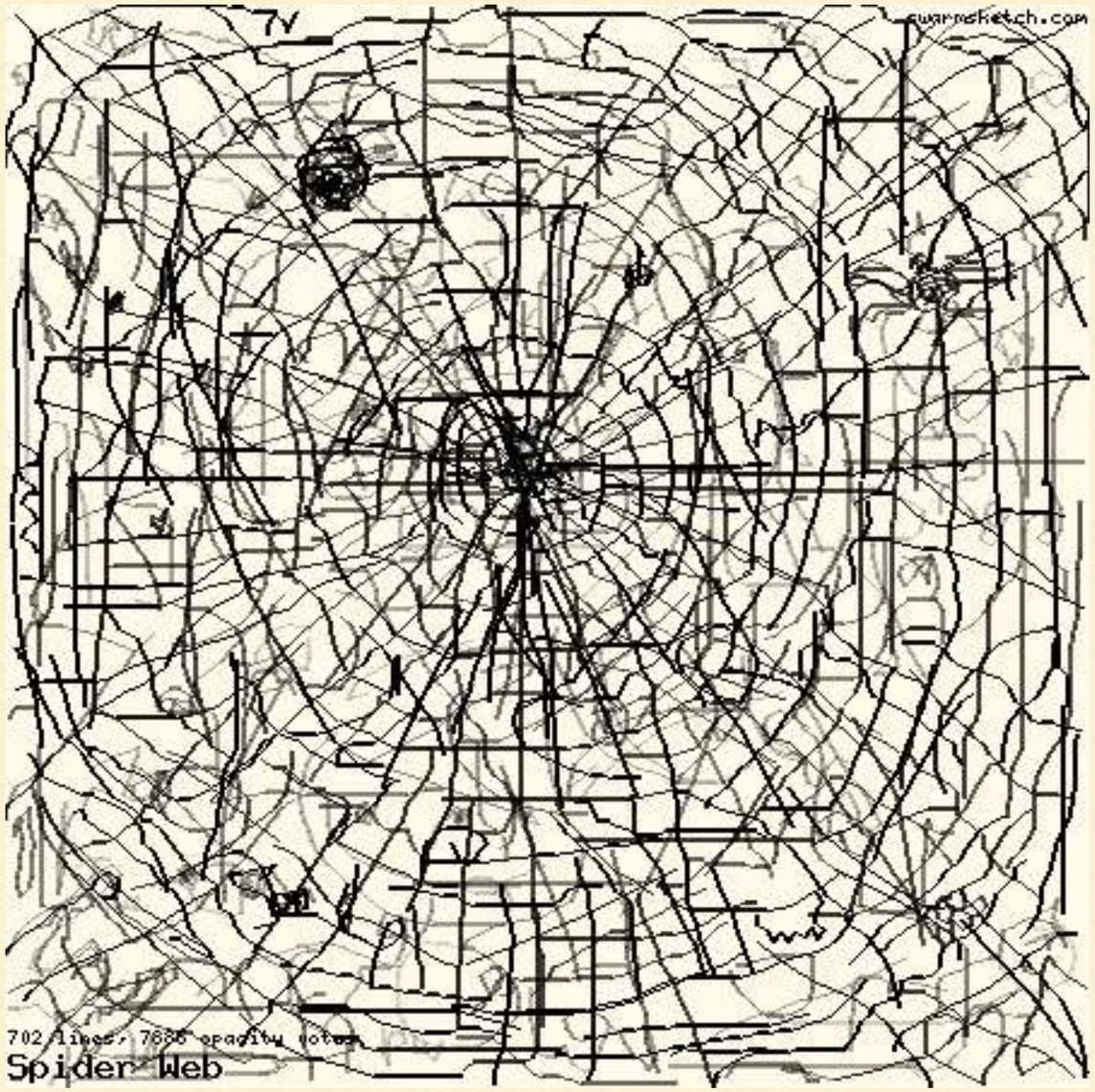
OVERVIEW

- Isaac Asimov and "The Relativity of Wrong"
- Absolutist Thinking in E-discovery
- Psychohistory ⟺ Big Data and Information Governance
- Two-Dimensional Model of Information Governance Tasks
- **SwarmSketch Thought Experiment**
- Managing the Relativity of Wrong

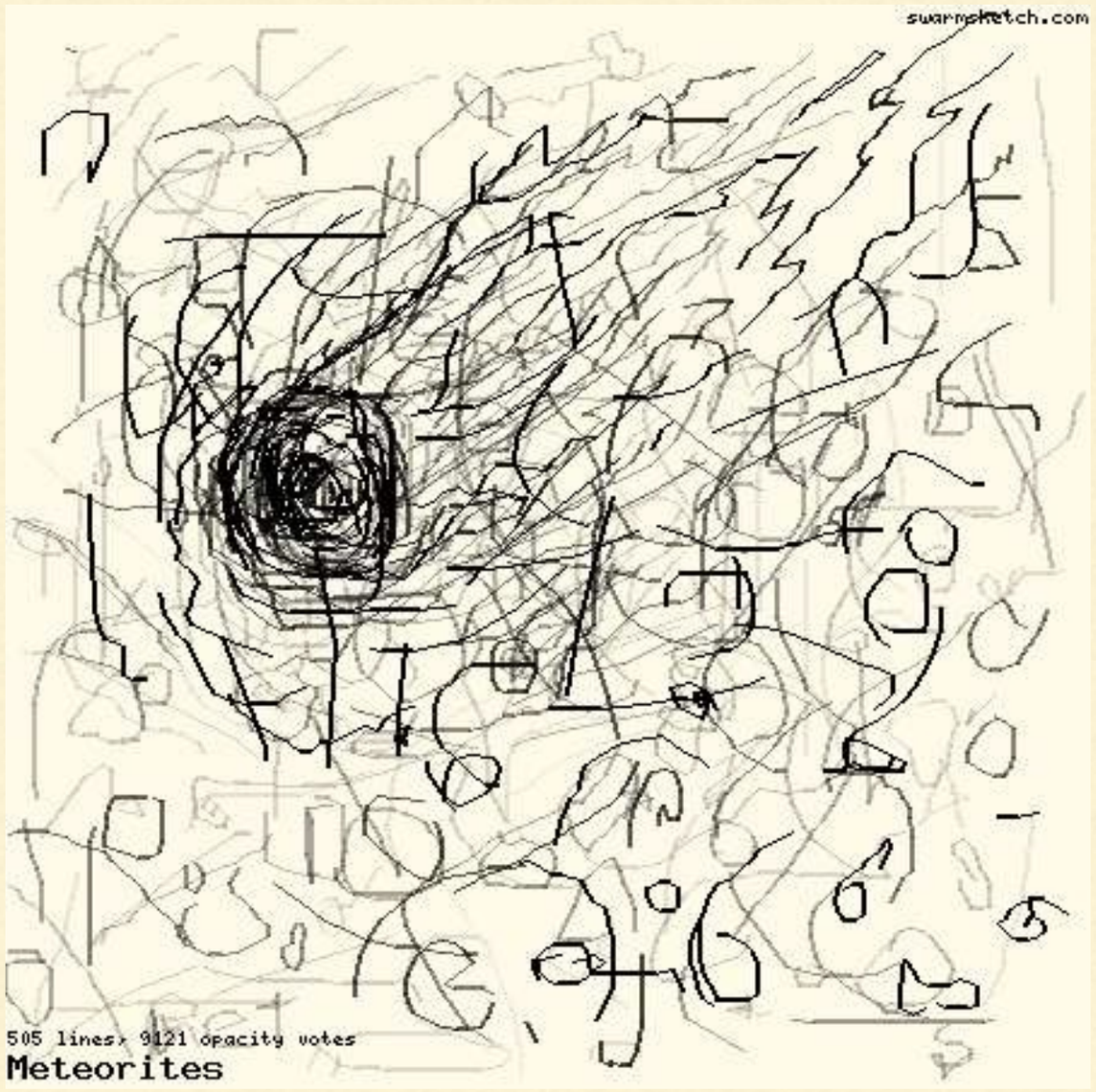
Information Governance	SwarmSketch
An information governance task is chosen, e.g. risk or compliance	A topic or subject is chosen e.g. “giant squid”, “gas prices”
Millions of little pieces of information (documents) = BIG DATA	Millions of little line segments = BIG DATA
Thousands of authors	Thousands of authors
Documents are combined intelligently by system (=machine) to “paint a picture”	Line segments are combined intelligently by system (=humans) to paint a picture
Users train the system by marking documents as relevant/nonrelevant	Users train the system by marking line segments as relevant/nonrelevant

HURRICANES

SPIDER WEB

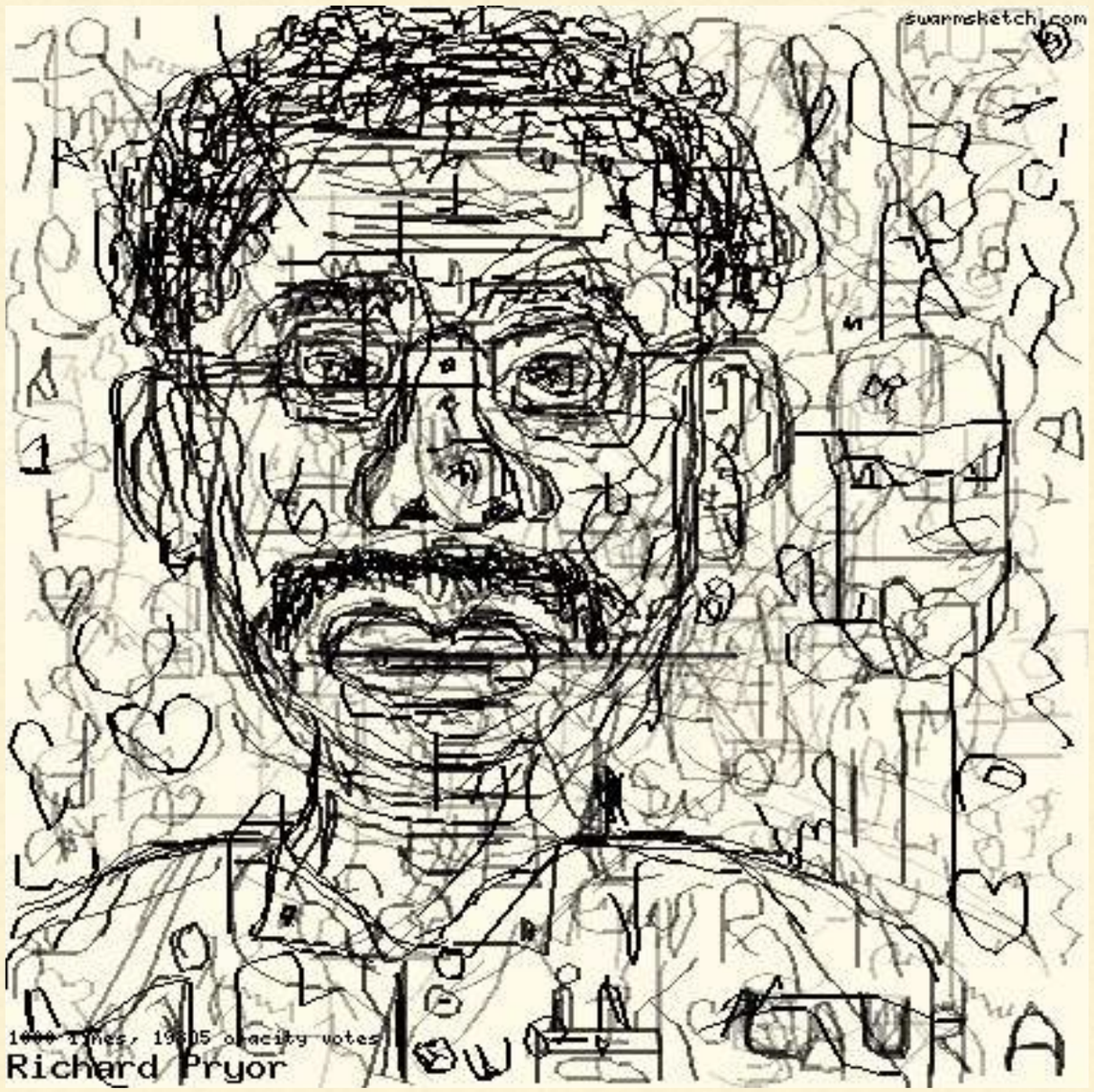


METEORITES



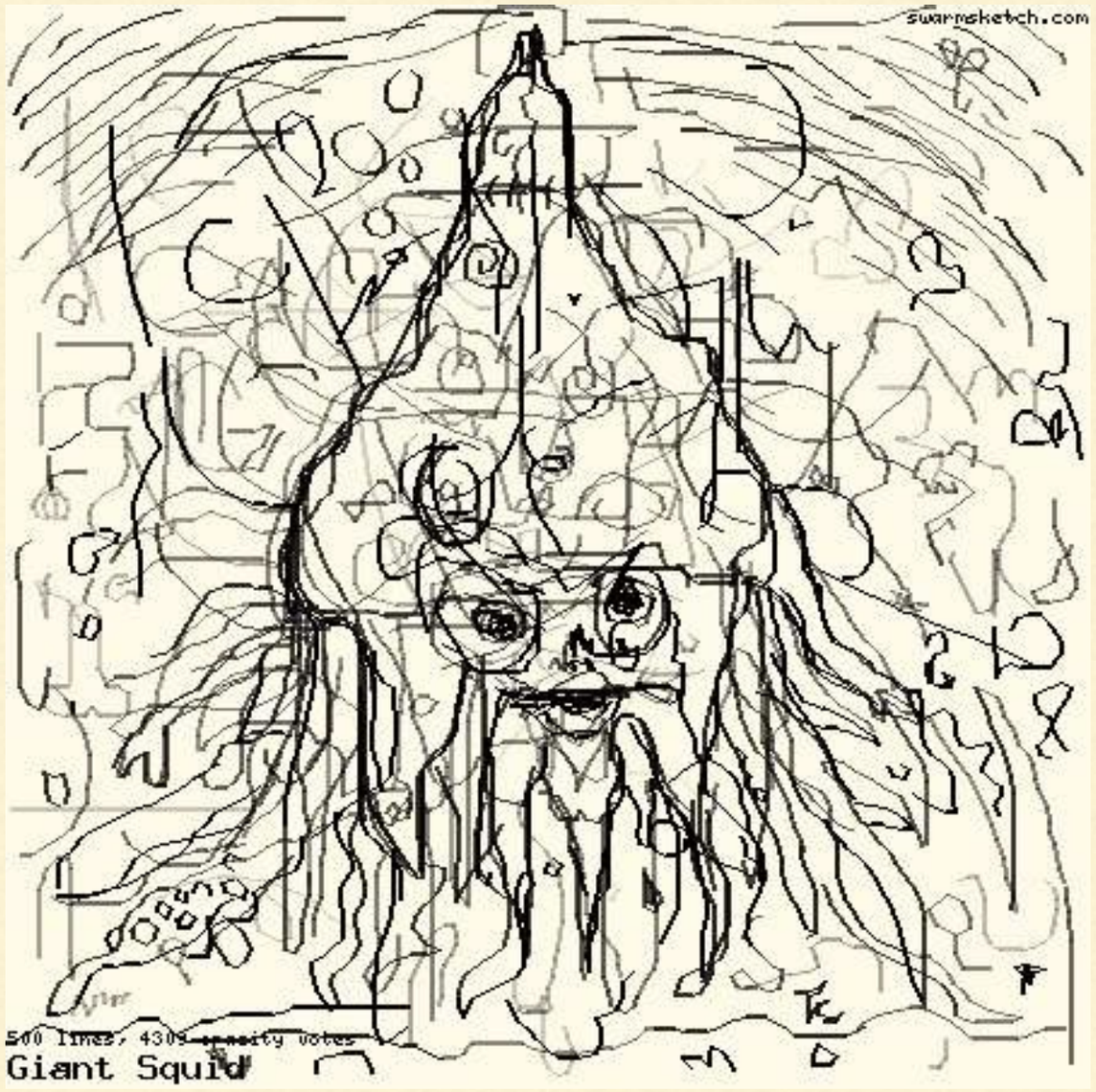
STINGRAY

RICHARD PRYOR



BOEING 787

GIANT SQUID



NAIL BITING

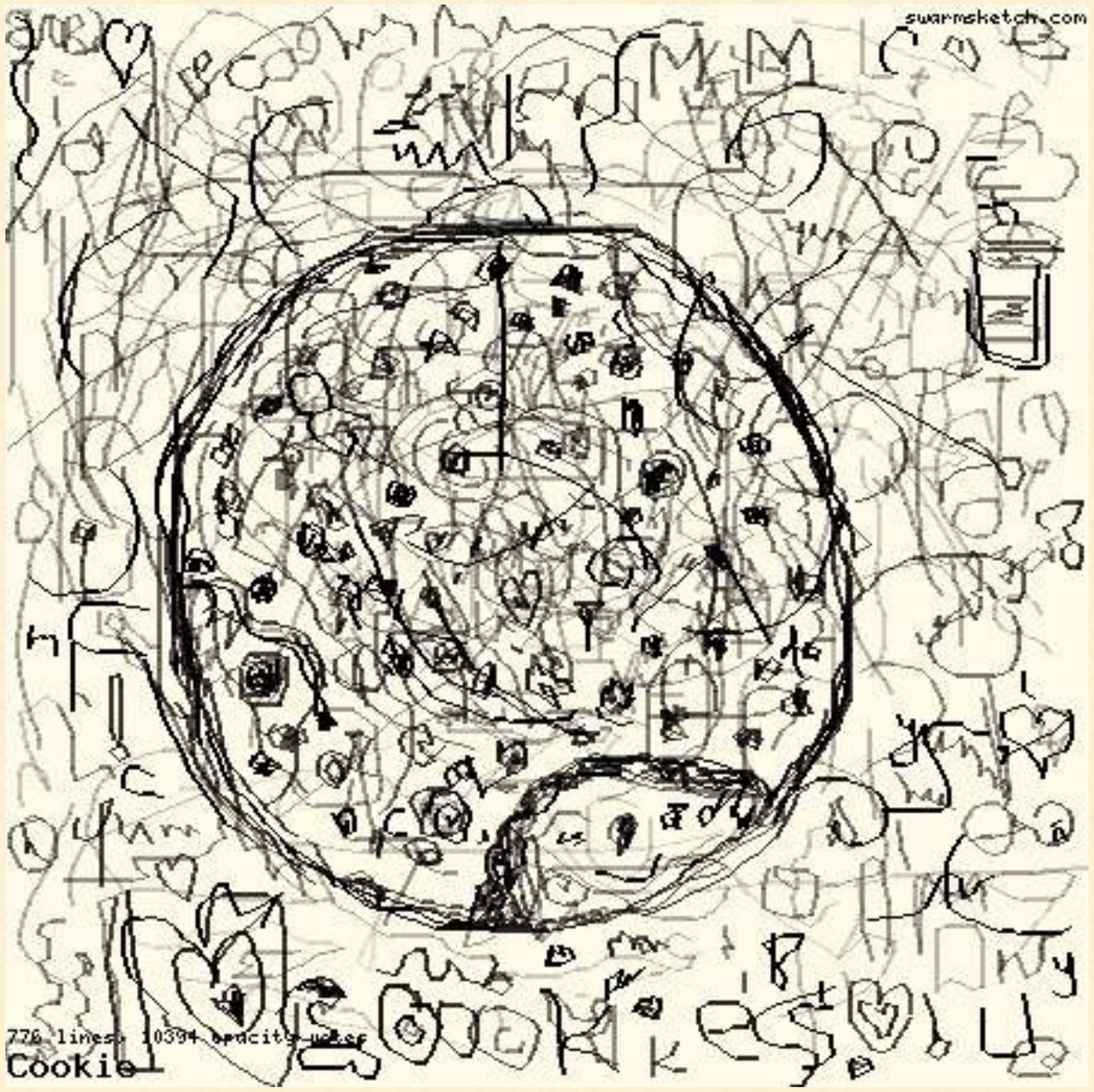
WOOLLY MAMMOTH

KENNEDY SPACE CENTER

DETROIT AUTO SHOW

VENICE

COOKIE



IPHONE

BLACK WIDOW

swarmsketch.com
1000 lines, 11157 opacity votes
Black Widow!

SMOKING GUN

BIRD FLU

DREAM INTERPRETATION

GAS PRICES

CALIFORNIA EARTHQUAKES

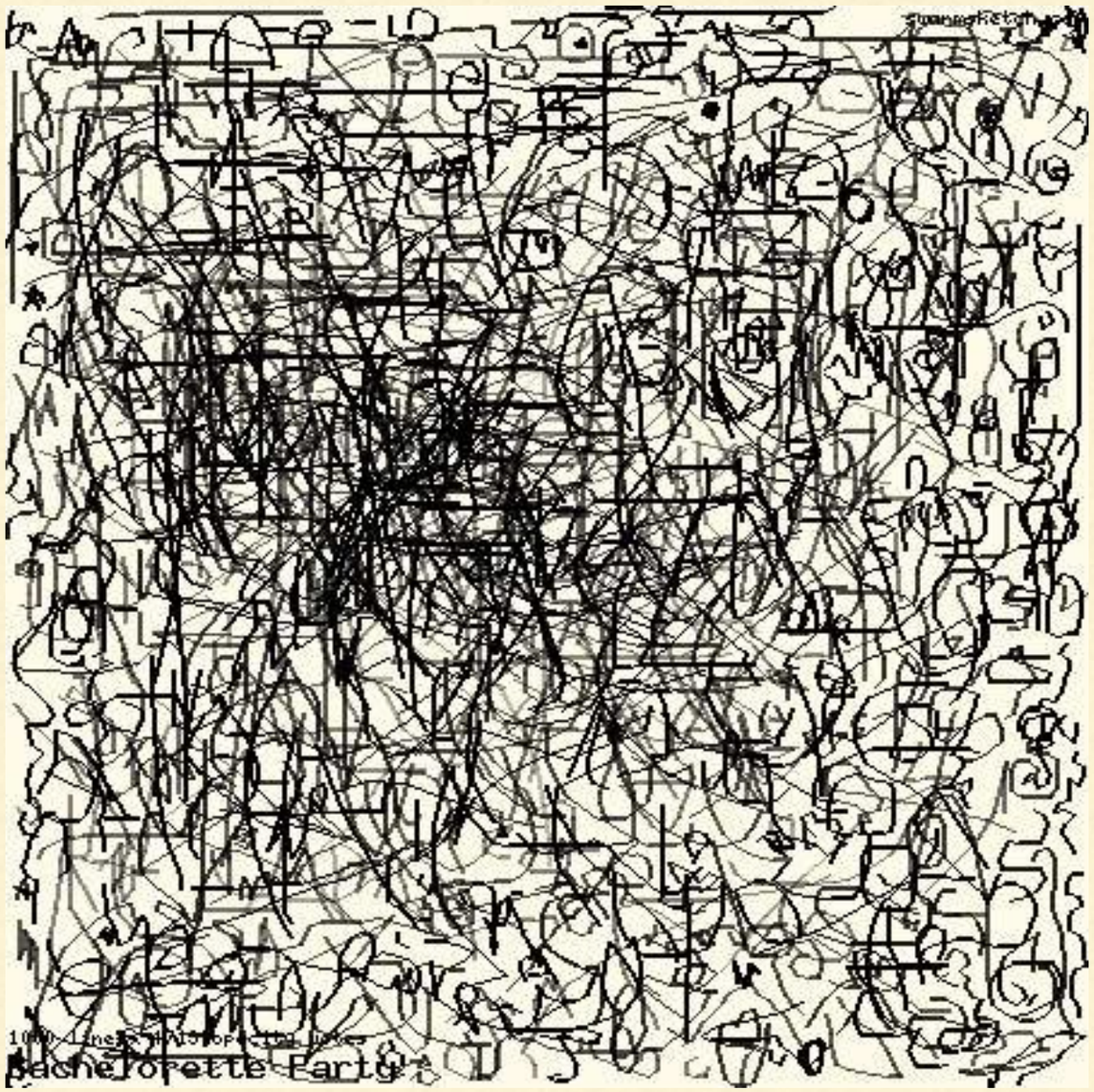
HAWAII EARTHQUAKE

U.S. ECONOMIC OUTLOOK

DAYLIGHT SAVINGS TIME

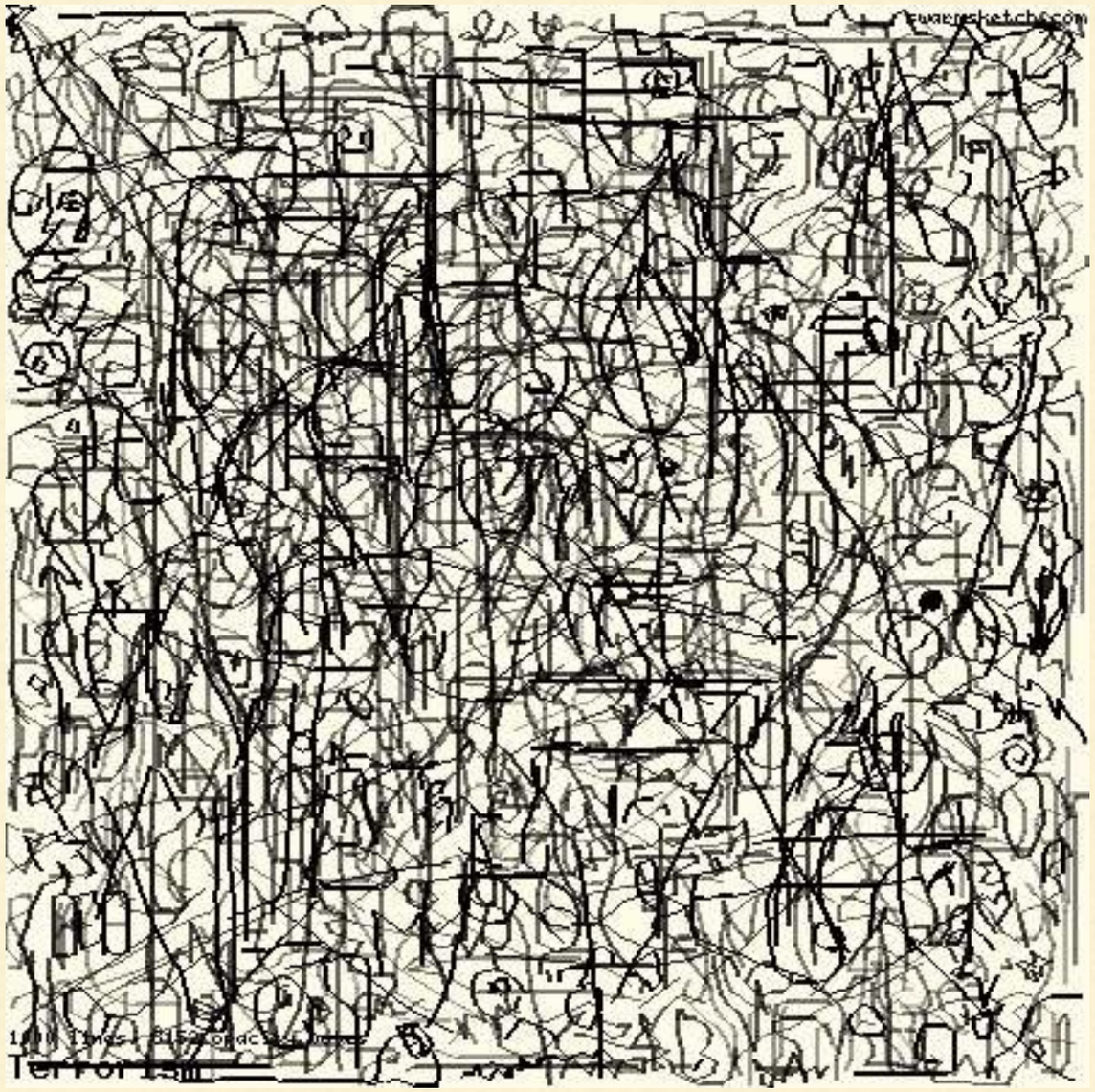
THE SCIENCE OF GENDER TESTING

BACHELORETTE PARTY



JUST MARRIED

TERRORISM



DIVERGENT

CONVERGENT

general

privacy

FCPA

risk

due diligence

extortion

whistleblower

specific

DIVERGENT

CONVERGENT

Do all FCPA violations exhibit similar patterns?
Do all privacy violations exhibit similar patterns?
Do all instances of sexual harassment exhibit similar patterns?

Yes?

No?

Is the earth round?

"The basic trouble, you see, is that people think that 'right' and 'wrong' are absolute; that everything that isn't perfectly and completely right is totally and equally wrong.."

–Isaac Asimov
The Relativity of Wrong
The Skeptical Inquirer, Fall 1989, Vol. 14, No. 1, Pages 35-44

BLACK WIDOW

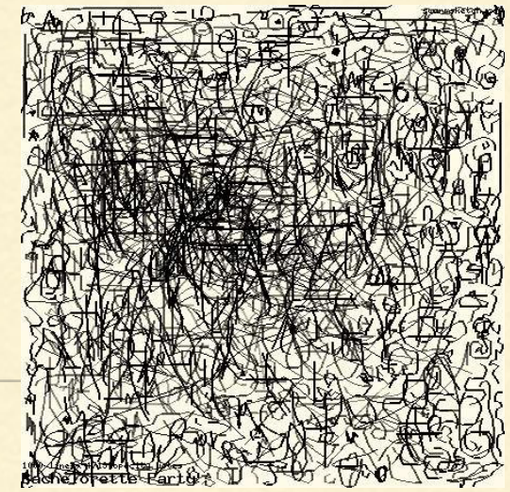
<u>What you may get</u>

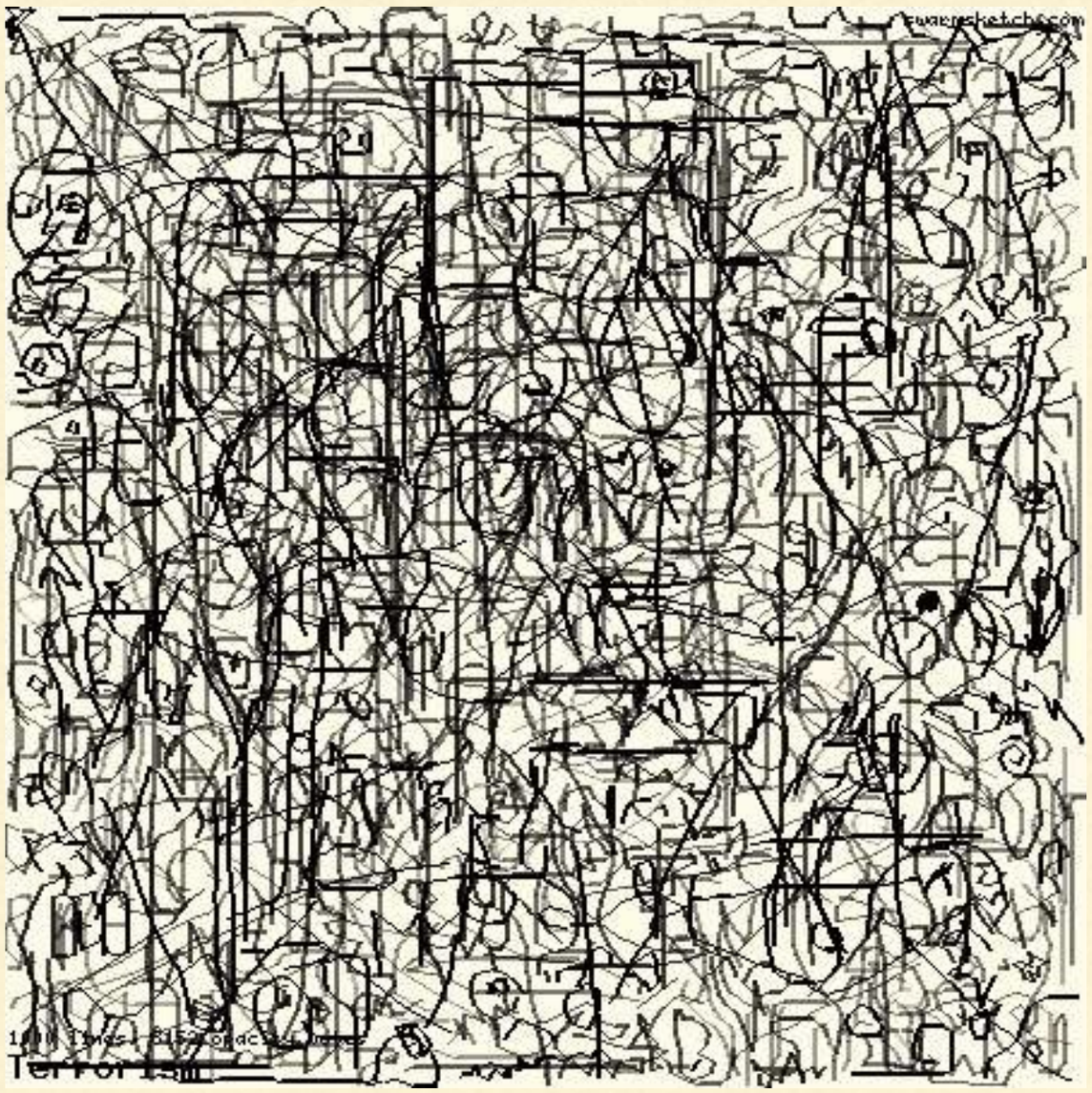
black widow spiders
other spiders
beetles
earwigs (pincher bugs)
centipedes
millipedes

<u>What you probably won't get</u>

basketball hoops
formula-1 race cars
broadway musicals
aluminum recycling plants
digital watches

TERRORISM





<u>What you may get</u>

chaos caused by terrorism
chaos caused by natural disasters
chaos caused by war
chaos caused by bachelorette parties
paintings by Jackson Pollock

<u>What you probably won't get</u>

baby rabbits
summertime swimming ponds
copper mines

OVERVIEW

- Isaac Asimov and "The Relativity of Wrong"
- Absolutist Thinking in E-discovery
- Psychohistory ⟺ Big Data and Information Governance
- Two-Dimensional Model of Information Governance Tasks
- SwarmSketch Thought Experiment
- **Managing the Relativity of Wrong**

MANAGING RELATIVITY OF WRONG

Big Data (Data Science) *alone* is not the answer.

"HCIR aims to empower people to explore large-scale information bases but demands that people also take responsibility for this control by expending cognitive and physical energy." —Gary Marchionini

HCIR→HCI(ta)R

MANAGING RELATIVITY OF WRONG

HCI(ta)R

- no longer only deliver the relevant documents, but must also provide semantic information along with those documents

- increase user responsibility as well as control; that is, information systems require human intellectual effort

- aim to be part of information ecology of personal and shared memories and tools rather than discrete standalone services

- support the entire information life cycle (from creation to preservation) rather than only the dissemination or use phase

- support tuning by end users and especially by information professionals who add value to information resources

http://en.wikipedia.org/wiki/Human–computer_information_retrieval

"Unless we make information governance someone's job, its not going to happen."

–Deb Logan, 11 January 2010

CONCLUSION

(1)
Relativity of Wrong
("Everything is wrong. Some things are less wrong than others.")
(2)
Generality vs Specificity
("Generalization is always wrong")
(3)
Temper Data Science with HCI
("Gotta keep an eye on that wrongness")

THANK YOU!

Questions?

SwarmSketch Animation