Decision-Driven Data Management

A Strategy for Better Data-Driven Decision Making

Presentation Overview

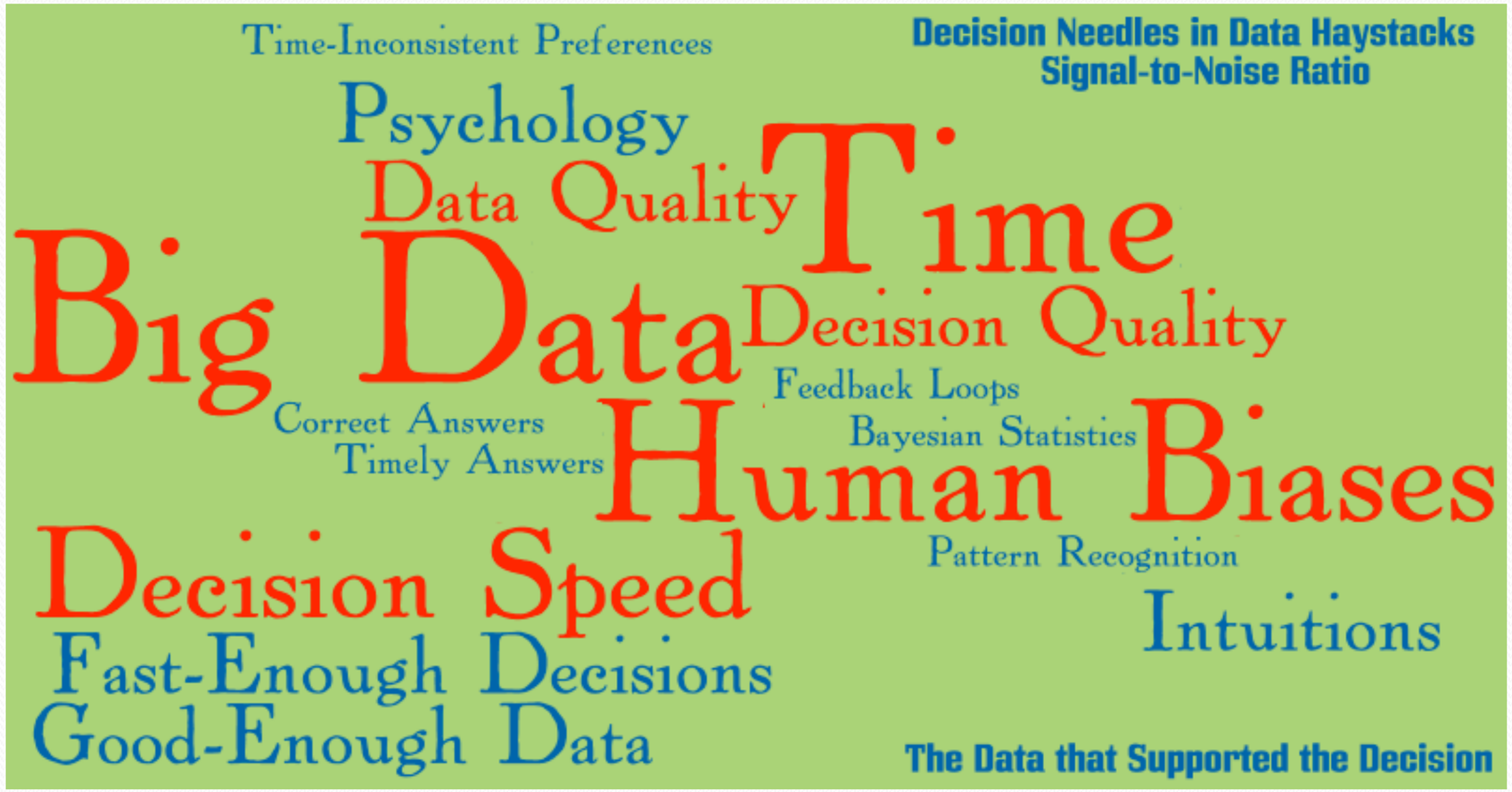
Data-Driven Decision Making:

- Exists at the intersection of data quality and decision quality, where quality data supports quality business decisions
- Happens at all levels of the organization on a daily basis

During this presentation, we will examine:

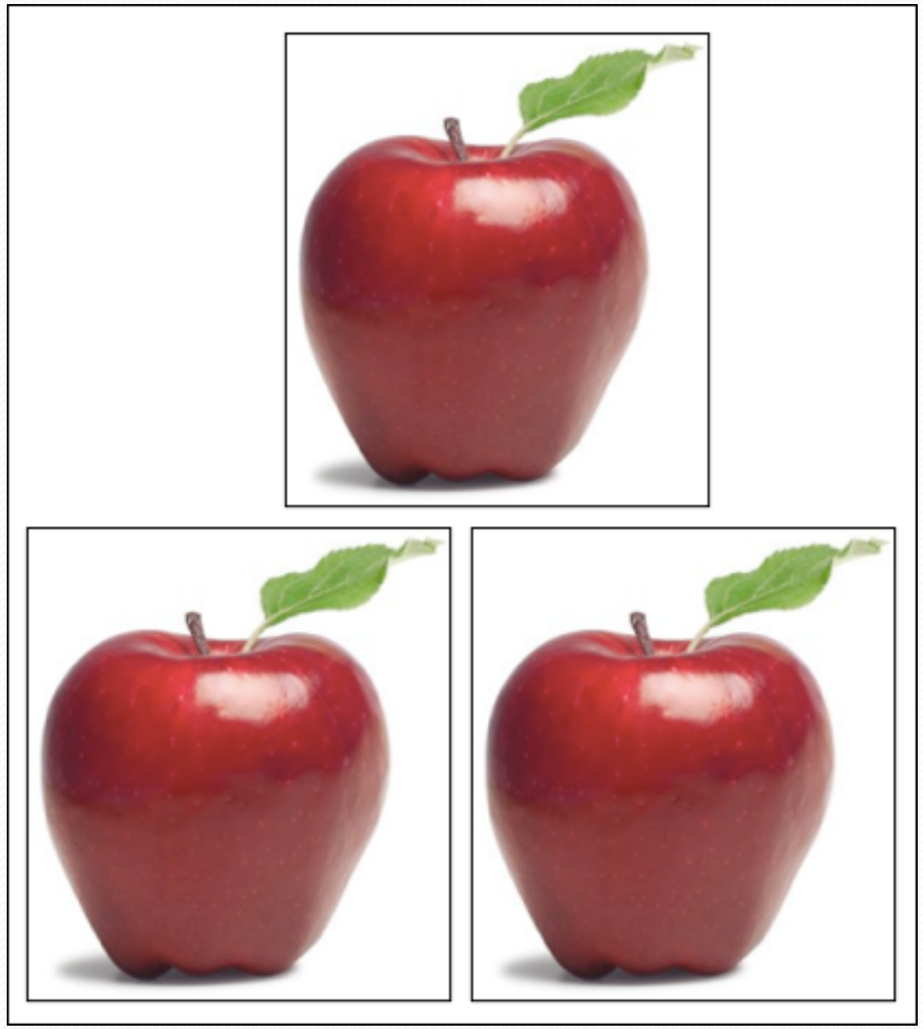
- Variables that can disrupt data-driven decision making
- Why decision-driven data management is a strategy for better data-driven decision making
- Why closing the decision-data feedback loop is the most critical success factor in data-driven decision making

Data-Driven Decision Disrupters



Our *Data Trek* will include Data Quality Oranges, Big Data Haystacks, Jimmy Stewart in a Cowboy Hat, three Nobel Prizes (two from 1978), the Big Bang Theory (no, not *that* one), and Getting Really Loopy

Thaler's Apples



Richard Thaler created a behavioral economics thought experiment known as **Thaler's Apples**

Which would you prefer:
(A) 1 **Apple** in 1 year, or
(B) 2 **Apples** in 1 year + 1 day?

Now, which would you prefer:
(C) 1 **Apple** today, or
(D) 2 **Apples** tomorrow?

Most people preferred B over A ***and*** C over D, illustrating ***time-inconsistent preferences***

Data Quality Oranges

The more **Data Quality Oranges** you have, the better the decision-making quality of your data is

Which would you prefer:
(A) 1 **Data Quality Orange** in 1 month, or
(B) 2 **Data Quality Oranges** in 1 month + 1 day?

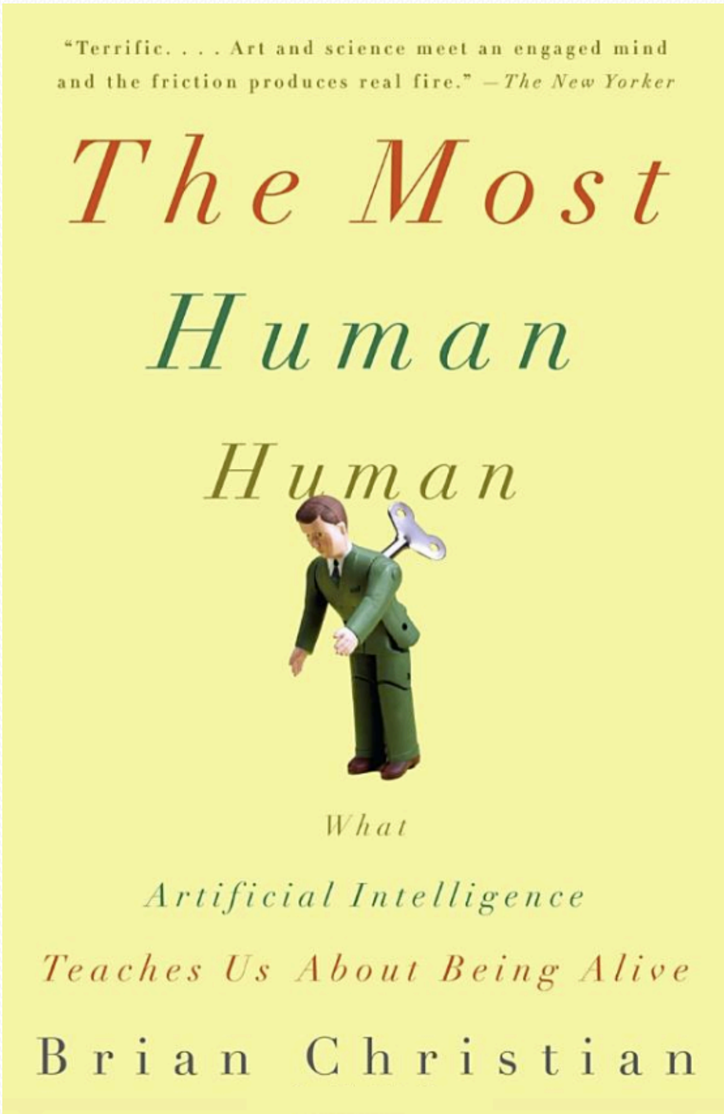
Now, which would you prefer:
(C) 1 **Data Quality Orange** today, or
(D) 2 **Data Quality Oranges** tomorrow?

In my data-driven decision making experience, most people prefer B over A ***and*** C over D

Time-Inconsistent Data Quality Preferences

- Time constraints provide a framing effect for our decision criteria
  - Decisions that must be made within 30 seconds are very different than decisions that can be made within 30 minutes or 30 days
- The speed at which the business world changes requires real-time decisions, which is why **decision speed** *is sometimes more important than* **data quality**
- We can't sacrifice data quality for faster business decisions, and high-quality data is preferable, but data quality concerns can not delay a critical decision
- We often need **good-enough data** for **fast-enough decisions**
- Data-driven decision making requires **decision-specific data quality thresholds**

Computability versus Complexity

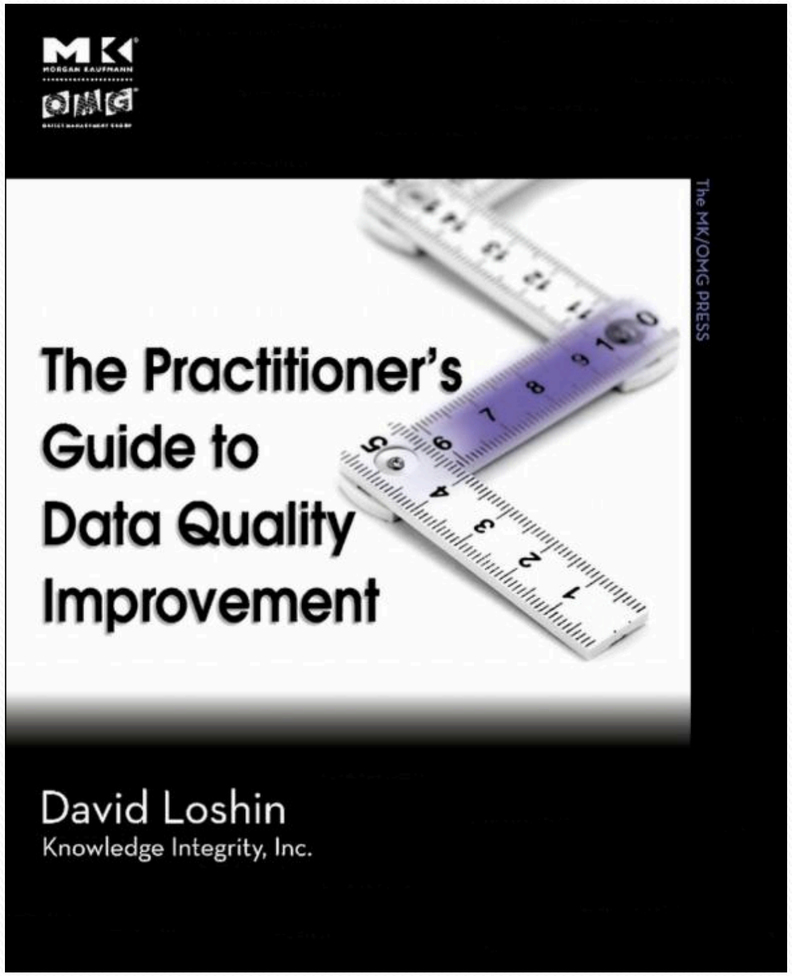


- Computability, the first branch of computer science, didn't care how long a computation would take, only whether or not the correct answer could be computed
- Complexity, the second branch of computer science, recognized that certain problems were *intractable*, meaning that the correct answer could be computed, but not quickly enough to be of any practical use
- Computability Theory:
  Produce correct answers – quickly if possible
- Complexity Theory:
  Produce timely answers – correctly if possible

Moore's Law versus Andy and Bill's Law

- **Moore's Law** (Intel co-founder Gordon Moore in 1965): Number of transistors that can be placed inexpensively on an integrated circuit, thereby increasing processing speed and memory capacity, doubles approximately every two years
- **Andy and Bill's Law** (Grove of Intel and Gates of Microsoft): Computers got faster but not faster to use, because the software system resource demands increased as the hardware improved: *"What Andy giveth, Bill taketh away"*
- But these advancements in computational power, increased network bandwidth, parallel processing frameworks (e.g., Hadoop), scalable and distributed models (e.g., cloud computing), and other techniques (e.g., in-memory computing) make real-time data-driven decisions more technologically possible than ever before

Time-Related Data Quality Dimensions



- **Currency** is the degree to which data is current with the world that it models, measuring how up-to-date data is, asserting limits to the lifetime of a data value, indicating when it possibly needs to be refreshed
- **Timeliness** refers to the time expectation for the accessibility of data, measuring the time between when data is expected and when it's readily available for use
- You may not be able to ask for more time since doing so could delay making a critical business decision
- But *how much data* is really needed to make a decision?

More than just Data Volume

- **Data Volume** – Exponential growth, availability, and use of information in today's data-rich landscape
- **Data Variety** – Structured, Semi-structured, Unstructured, and other types (e.g., sensor data from the Internet of Things)
- **Data Velocity** – How fast data is being produced, as well as how fast data must be processed to meet demand

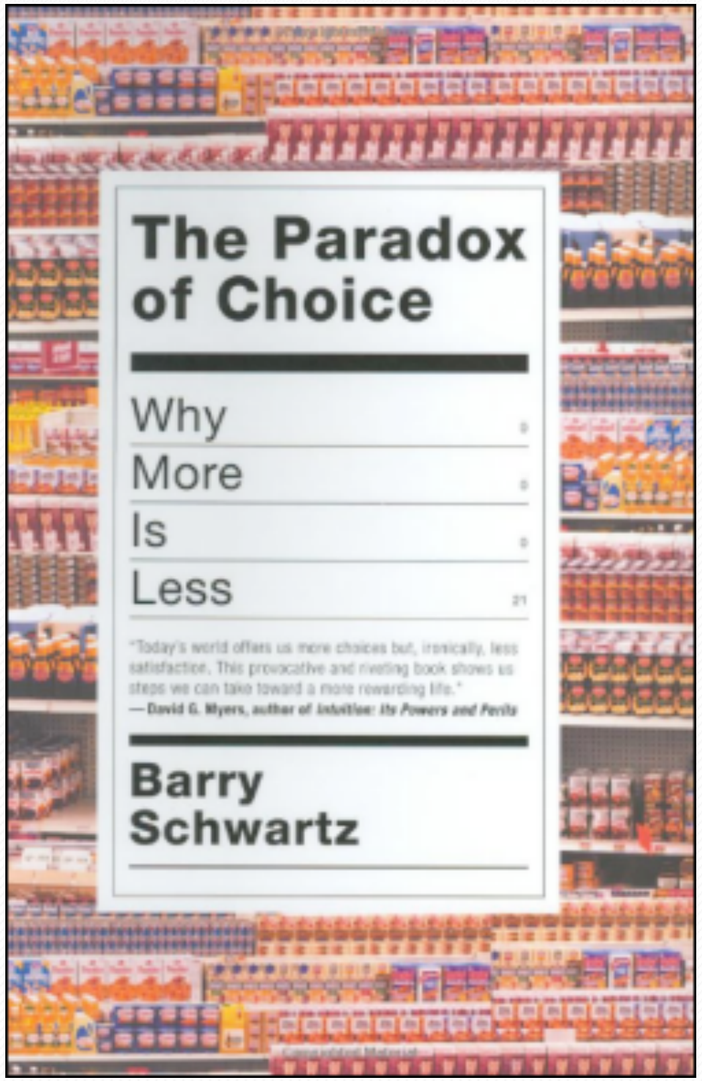
Big Data el Memorioso

- *Funes el memorioso* is a short story by Jorge Luis Borges, which describes a young man named Ireneo Funes who, as a result of a horseback riding accident, has lost his ability to forget (In Spanish, the word *memorioso* means "having a vast memory")
- Funes has a vast memory, but he's so lost in the details of everything he knows that he's unable to convert the information into meaningful insights about his world
- If your organization becomes so lost in the details of everything big data delivers, you're unable to convert it into the insight needed for data-driven decision making
- Without insight, big data has, as Borges said of Funes, "a stammering greatness," that amounts to nothing more than, as Shakespeare said in *The Tragedy of Macbeth*, "A tale told by an idiot, full of sound and fury, signifying nothing."

Decision Needles in Data Haystacks

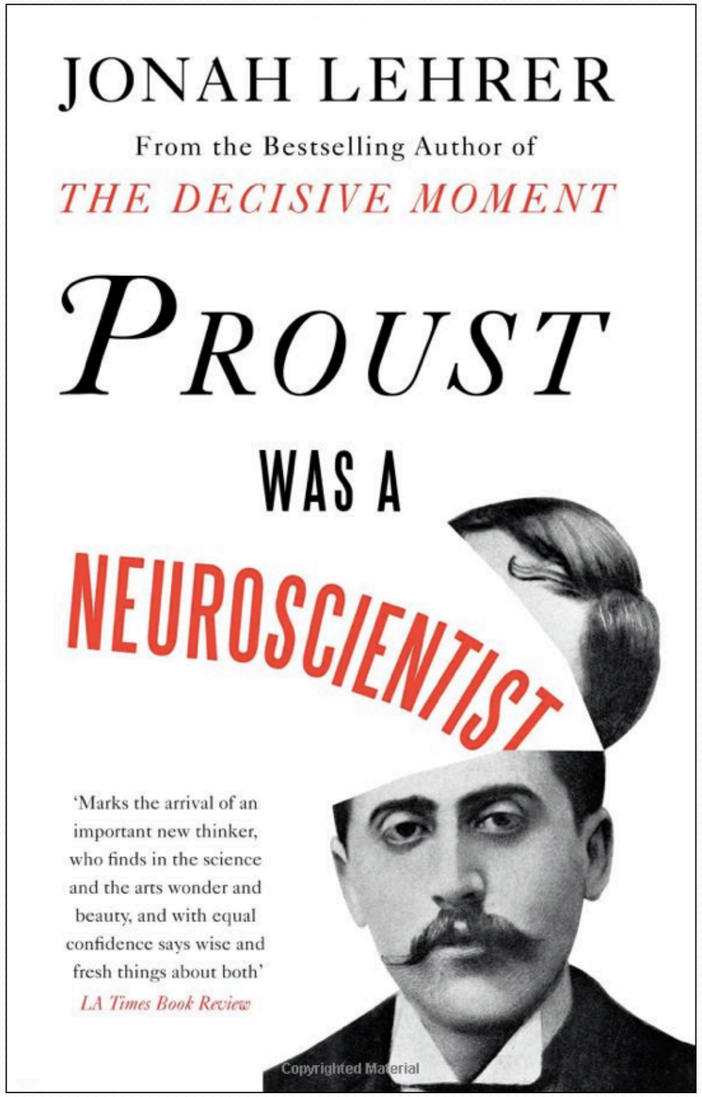
- The decision-making value of big data comes from discovering *the little data within big data* that helps data-driven decision making, as if we are trying to find ***Decision Needles*** in ***Data Haystacks***

The Paradox of More Data



- Sometimes more data only disables (**analysis paralysis**) or overwhelms (**information overload**) our decision making
- Clay Shirky: *"It's not information overload, it's filter failure."*
- Herbert Simon won the 1978 Nobel Prize in Economics, coined **satisficing** — combines **satisfying** with **sufficing**
- *Maximizers* are perfectionists trying to consider all the data, but *Satisficers* determine data criteria for decision adequacy
- Focus on determining the sufficient amount and quality of data needed to satisfy the execution of a business decision
- Filtering out data noise until we detect decision signal is very similar to how we experience music . . .

The Music of the Data

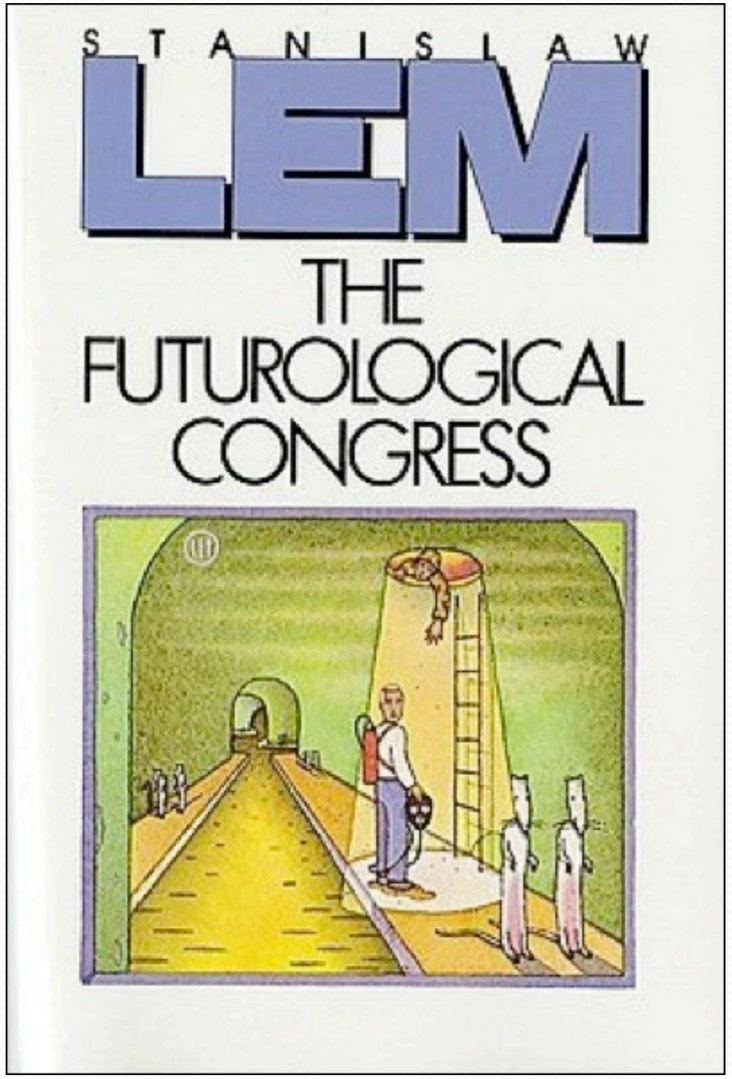


- "Music is a sliver of sound that we have learned how to hear. Instead of representing the full spectrum of sound waves vibrating inside the ear, the auditory cortex focuses on finding the note amid the noise. We tune out the cacophony we can't understand."
- *"Music is the pleasurable overflow of information . . .* Whenever a noise exceeds our processing abilities, we stop trying to understand the individual notes and seek instead to understand the relationship between the notes. It is this psychological instinct, this desperate neuronal search for a pattern, any pattern, that is the source of music."

The Data-Decision Symphony

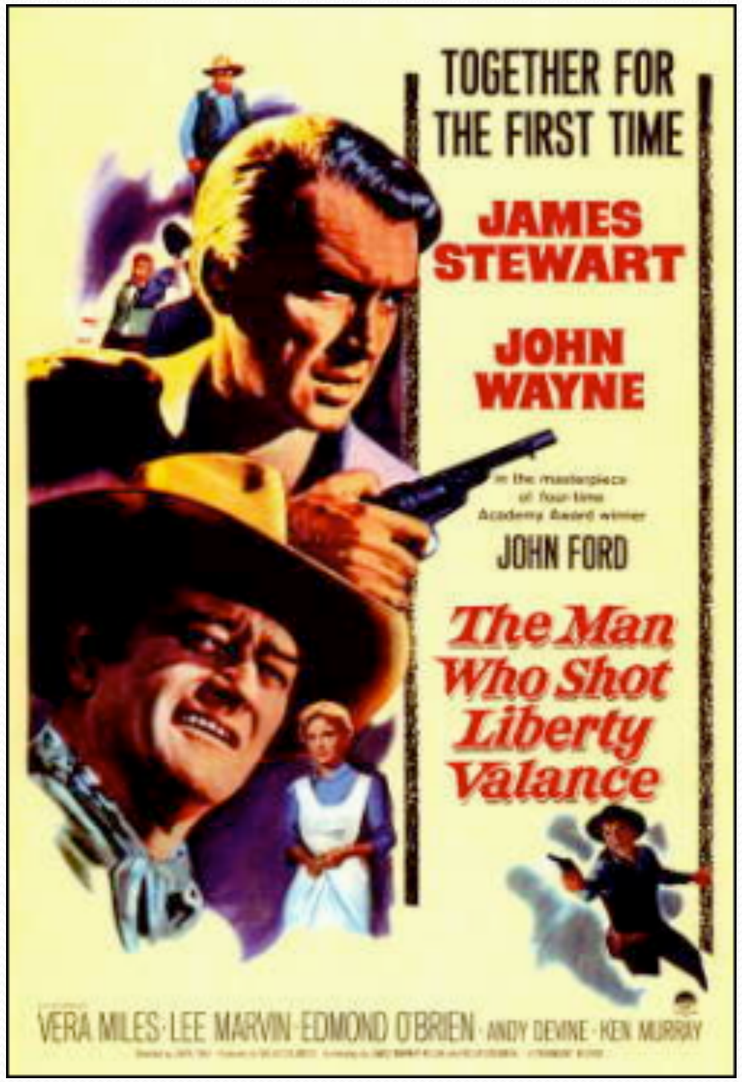
- "We invent patterns to keep pace with the onrush of noise. Once we find a pattern, we start to make predictions, projecting imaginary order into the future, turning the scraps of sound into the ebb and flow of a symphony." ~ Jonah Lehrer
- We invent decision patterns to keep pace with the onrush of data. Once our brain finds a decision pattern, we start making predictions, projecting imaginary order into the data stream, turning it into the ebb and flow of the data-decision symphony.
- Jonah Lehrer: *"Music is the sound of art changing the brain."*
- But sometimes the music of the data is the sound of pattern recognition changing our brain such that our search for decision consonance among data dissonance biases us with comforting, but false, conclusions, perhaps facilitated by a trip to . . .

The Data Psychedelicatessen



- Hallucinogenic drugs from the *Psychedelicatessen*:
  - **Amnesol** — Removes real, but unwanted memories
  - **Authentium** — Creates believable, but false memories
- Hallucinogenic drugs from the *Data Psychedelicatessen*:
  - **Gastroflux** — Aids digestion of "going with your gut"
  - **Datamine** — Convinces the decision-maker the only reliable data is their own personally cultivated data
  - **Selectium** — Renders invisible any data that doesn't support the decision-maker's personal preferences
  - **Qualitol** — Convinces the decision-maker that their data quality doesn't impact their decision quality at all

The Man Who Shot Liberty Valance



- Liberty Valance (Lee Marvin) was an outlaw terrorizing the frontier town of Shinbone
- Ransom Stoddard (James Stewart), despite his lack of skills, kills Valance with one shot during a gunfight
- Stoddard later tells a reporter the man who really shot Liberty Valance was Tom Doniphon (John Wayne)
- But when he hears the truth, the reporter responds:

"This is the West, sir.

When the Legend becomes Fact,

Print the Legend!"

The Data that Supported the Decision

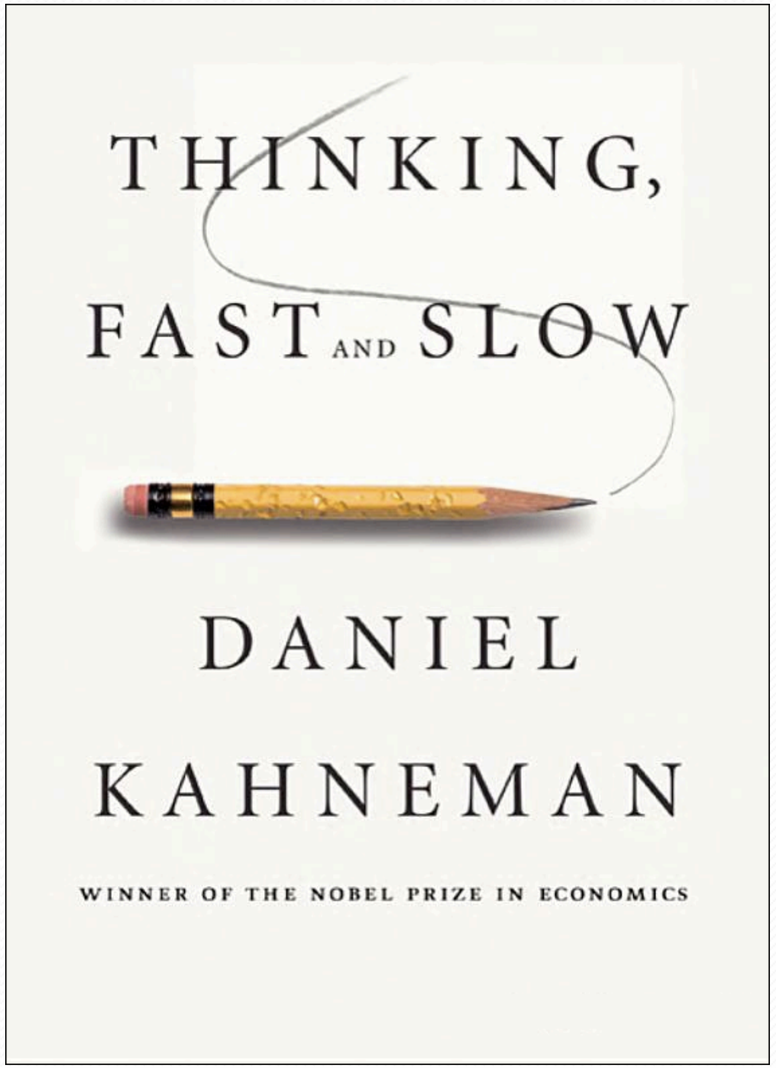
- Sometimes, **post-decision data bias** is used to construct a story to support the intuition-driven decision that has already been made (aka **confabulation**)
- When intuition is used instead of data, decision makers will declare:

"This is the Business, sir.

When Intuition makes the Decision, Confabulate the Data!"

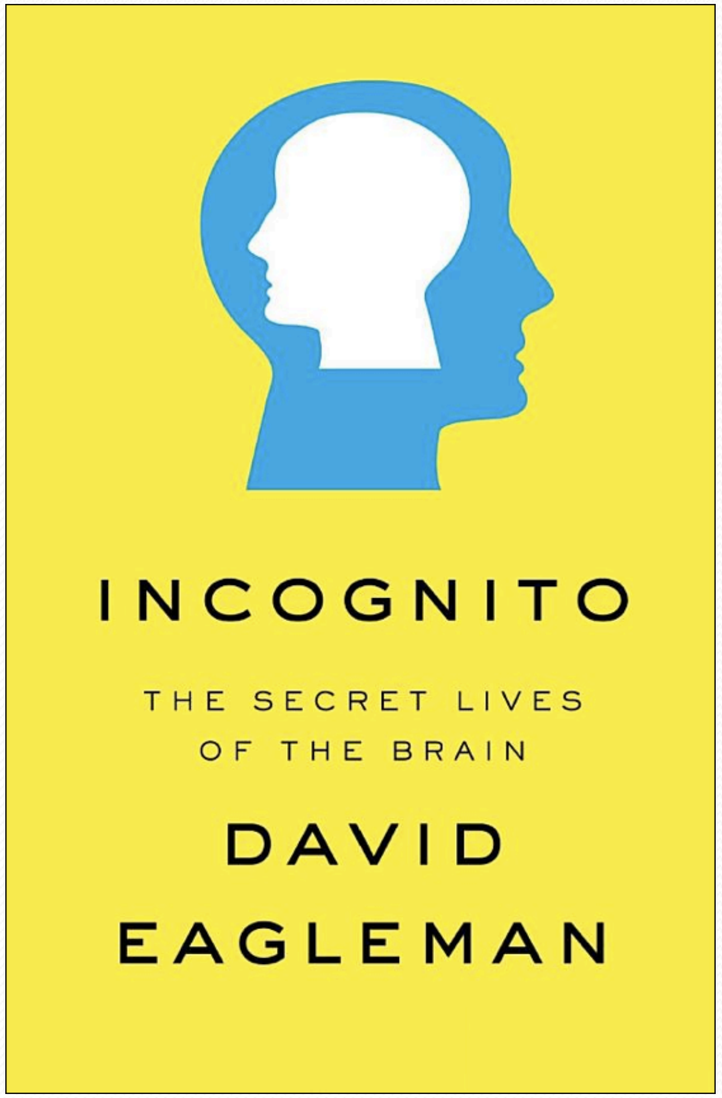
- Other times, **pre-decision data bias** is used to confirm preconceptions, thus masking an intuition-driven decision as data-driven (aka **confirmation bias**)
- The moral of this confabulation is to always confirm the basis of your organization's data-driven decision making, and pay particular attention to the human biases of your decision makers, including yourself because . . .

What You See Is All There Is (WYSIATI)



- Although intuition is valuable, we are often radically insensitive to both the quality and the quantity of the information that gives rise to impressions and intuitions
- We're often looking for, not good data, but a good story, making consistency matter more than completeness
- **WYSIATI** facilitates the coherence and cognitive ease allowing us to make sense of a complex world, because, most of the time, the coherent story we put together is close enough to reality to support reasonable action
- Sometimes, we see data not as it is, but as we want it to be
- However, what you *choose to see* **is not** *all there is*

Data-Driven Decision Making Incognito

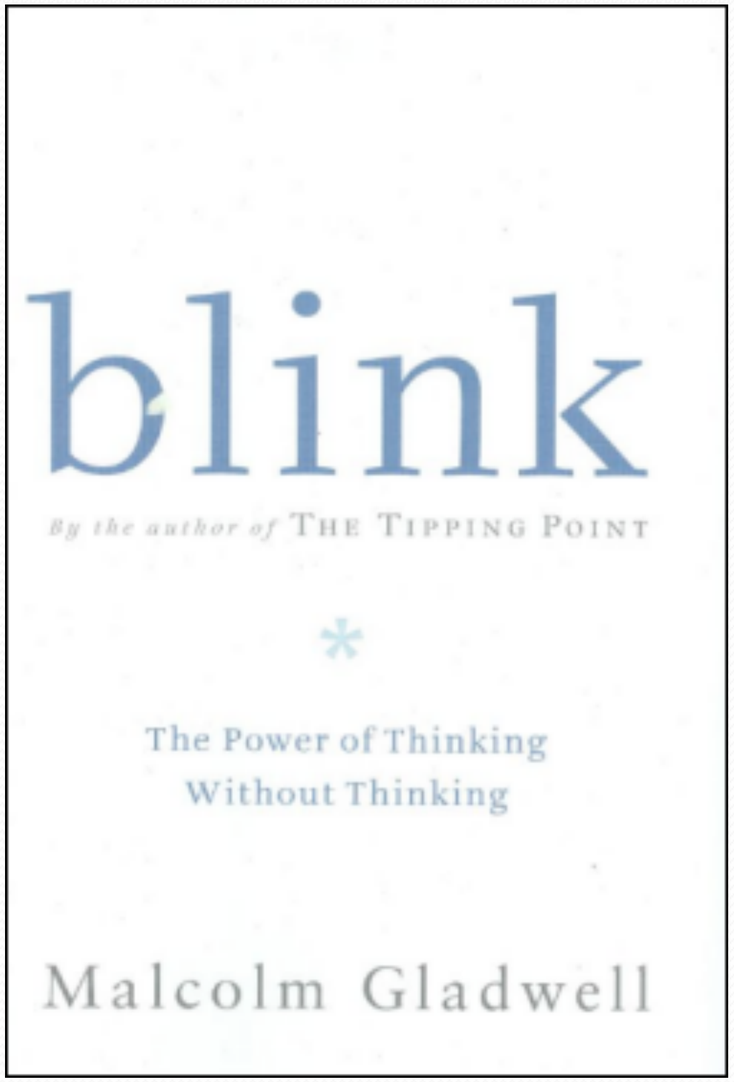


- Our brain samples just a small bit of the world, making time-saving assumptions, seeing only as well as it needs to
- As our eyes interrogate the world, they optimize their data strategy, arbitrating any battles between conflicting data, so we see not what is really there, but a moment-by-moment version of which perception is winning over the others
- Perceptions work not by building up bits of captured data, but by *matching our expectations* to incoming sensory data
- Our brain's internal data is *not generated* by this external sensory data, but is instead *merely modulated* by it

Timeliness and Availability Bias

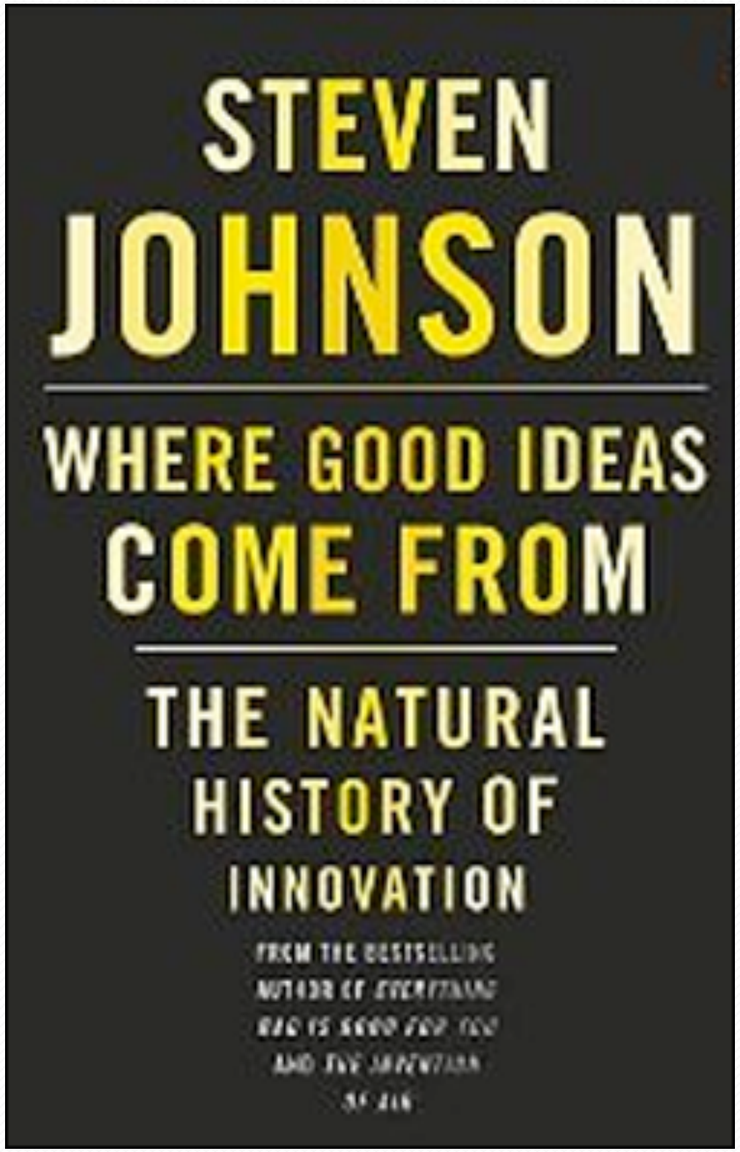
- **Timeliness** — refers to the time expectation for the accessibility of data, measuring the time between when data is expected and when it's readily available for use
- **Availability Bias** — We are affected more strongly by how easy data is to retrieve than we are by either the quality or quantity of the data
- Imagine two decision support systems, one fast, but with a higher error rate, and one slow, but with a lower error rate — which system would get used more often?
- Errors of intuitive thought are often difficult to prevent, and even when we are more aware of our potential biases, they can still undermine our decision making
- However, this doesn't mean that our intuition is always wrong . . .

A Partial Defense of Intuition



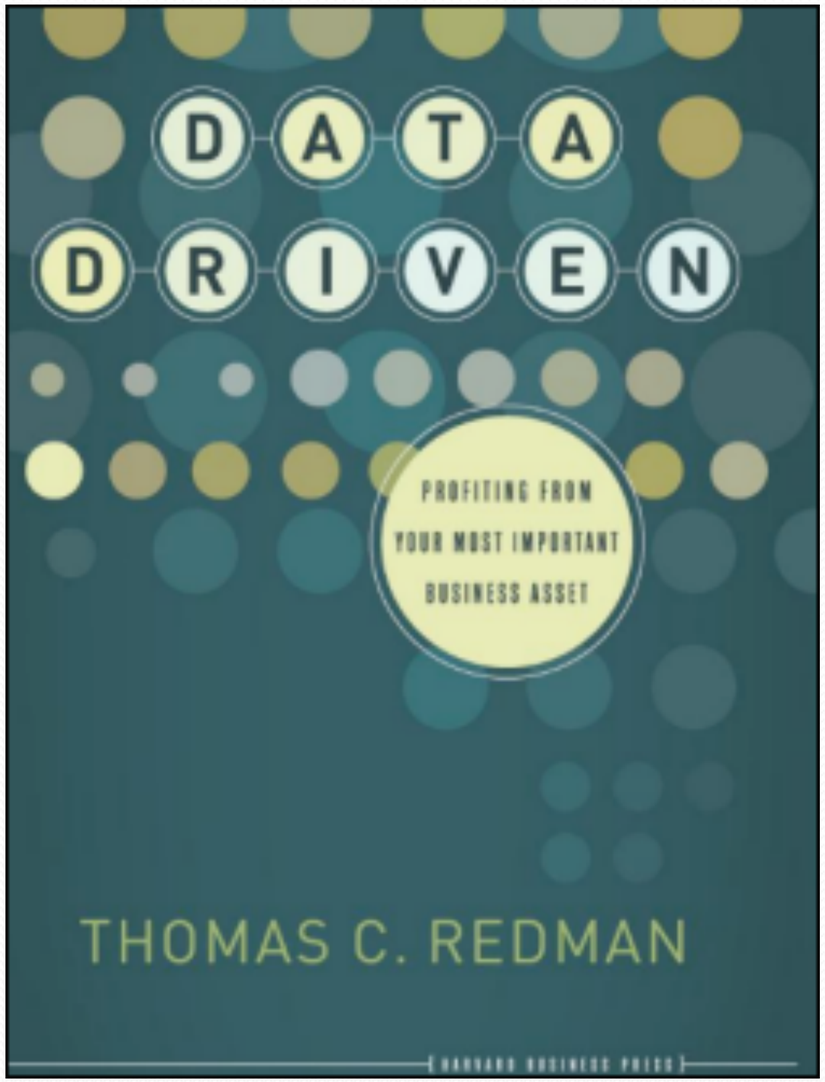
- Occasionally, there are times when *gathering less data* allows us to *make better decisions*
- **Thin-slicing** is a psychological term, which says we're capable of making sense of complex situations based on thin slices of experience or data
- Cook County changed its emergency room heart attack diagnosis procedure by **measuring only critical data** (blood pressure and ECG), and now they're considered one of the best hospitals at diagnosing chest pain
- Intuitions can prove erroneous, but that doesn't mean they're inherently bad — to understand why, we need to look at the role error plays in good decision making . . .

Where Good Decisions Come From



- Error creates a path away from your assumptions
- But we have a natural tendency to dismiss error, since we assume the error is noise, not signal
- One **mistaking signal for noise** example comes from the winners of the 1978 Nobel Prize in Physics, Arno Penzias and Robert Wilson, who discovered cosmic microwave background radiation, but for almost a year they thought that all the static they heard simply meant their telescope was broken
  - In the Big Data raining down from Big Sky, they managed to hear the remnants of the Big Bang
- Transforming error into insight is a characteristic of statistically good decision making . . .

Bayesian Data-Driven Decision Making

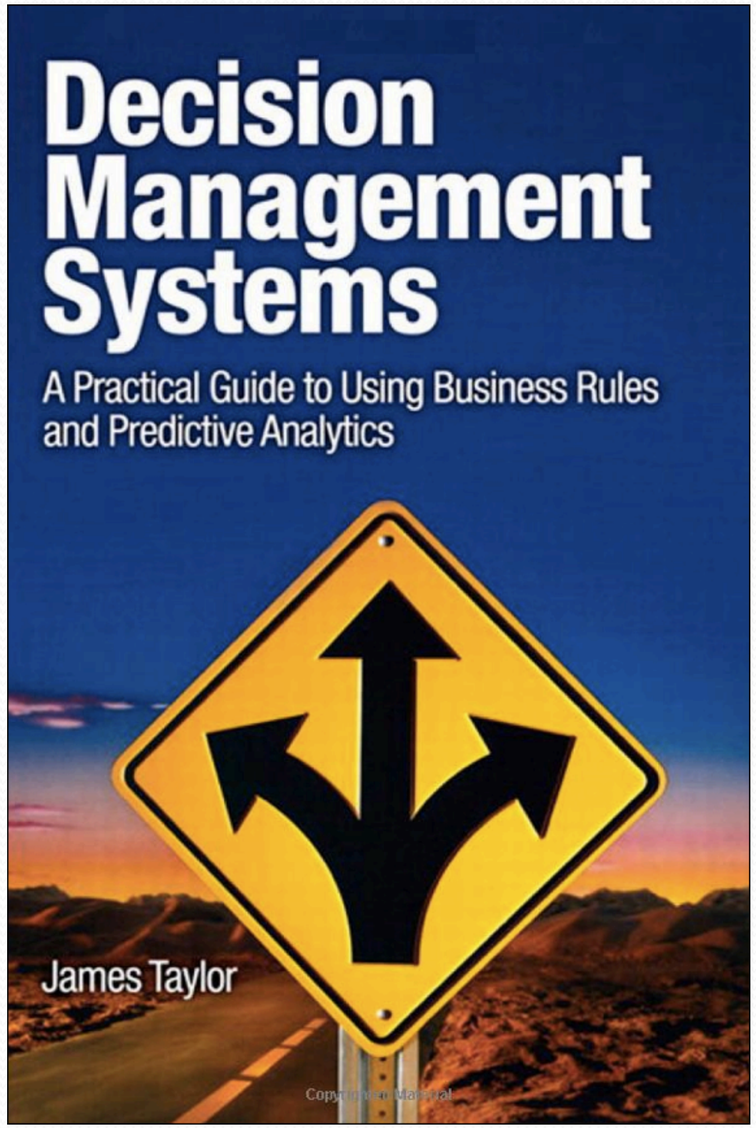


- When John Maynard Keynes was asked what he does when new data is presented that does not support his earlier decision, responded: *"I change my opinion. What do you do?"*
- We need to regularly and systematically evaluate how well a decision is proving itself in practice
- Central to **continuous improvement** is the concept of **closing the feedback loop** that allows a process to monitor itself, learn from its mistakes, and make adjustments when necessary
- Building feedback loops into our data-driven decision making is essential, but too often ignored

Summary of the Key Points

- Data-driven decisions are made at all levels of the organization on a daily basis
- Changing business world demands **good-enough data** for **fast-enough decisions**, which is why **decision speed** *is sometimes more important than* **data quality**
- Data-driven decision making requires **decision-specific data quality thresholds**
- The potential for pre-decision and post-decision **human biases** means sometimes we see data not as it is, but as we want data to be (i.e., *"what you see is all there is"*)
- **Error-driven learning** is essential to data-driven decision making, transforming error into business insight is a characteristic of statistically good decision making
- We need to regularly evaluate how well a decision is proving itself in practice, which requires that we **build feedback loops** into our data-driven decision making

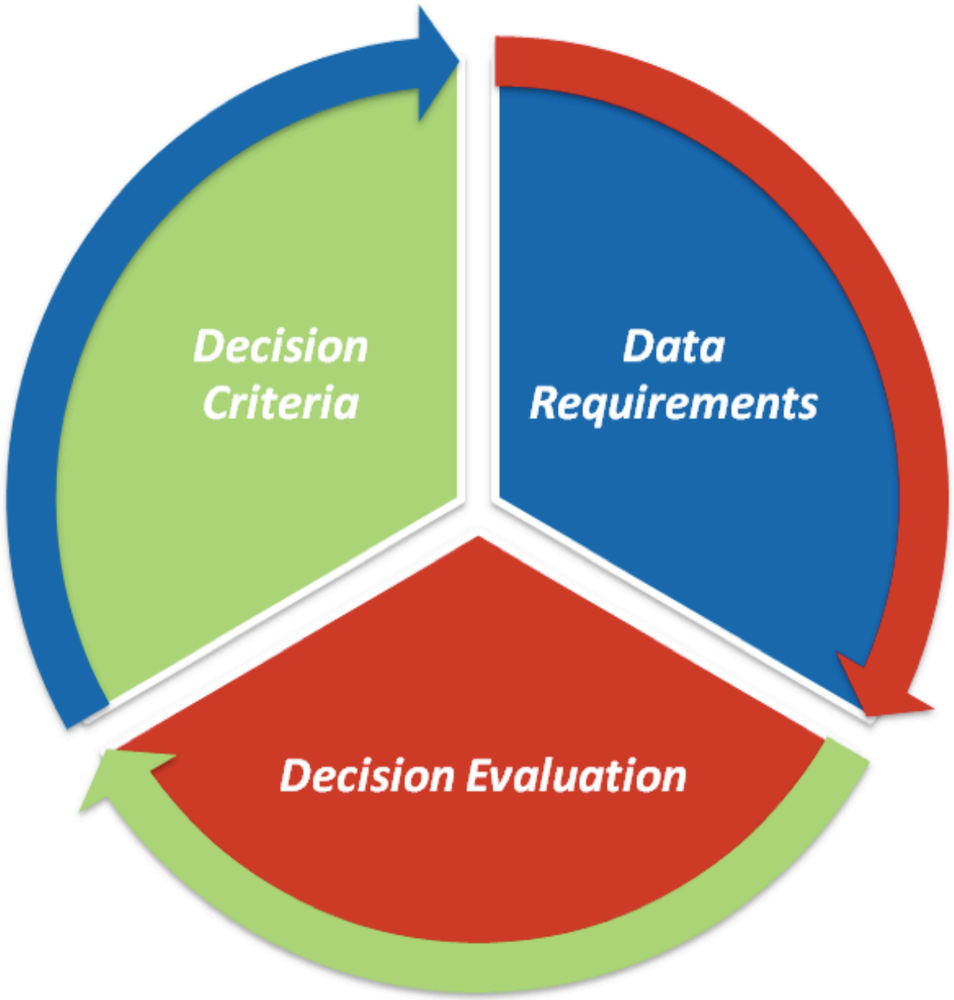
"Begin with the Decision in Mind"



- The Four Principles of Decision Management Systems:
  - Begin with the Decision in Mind
  - Be Transparent and Agile
  - Be Predictive, not Reactive
  - Test, Learn, and Continuously Improve

- "Begin with the decision in mind" means understand **decision criteria** before gathering **data requirements** to support efficient and effective decision execution, and enable **decision evaluation** for continuous improvement

Decision-Driven Data Management



Decision Criteria

- Time Constraints and Preferences
- Business Impacts and Risks
- Human Intuitions and Biases

Data Requirements

- Data Volume, Variety, and Velocity
- Data Management and Governance
- Data Quality Thresholds

Decision Evaluation

- Evaluate Business Results
- Decision Quality Thresholds
- Re-evaluate Decision Criteria
- Re-evaluate Data Requirements

Decision Criteria

- Achieving **better decisions** with **better data** begins with **better decision criteria**, by acknowledging each business decision's unique characteristics, including:
  - **Time Constraints** and **Time-Inconsistent Data Quality Preferences**
  - Business **Impacts of a Good Decision** and **Risks of a Bad Decision**
  - Human **Intuitions** and **Biases** (e.g., Confabulation and Confirmation Bias)
- Not every business decision is equal — the business impacts of a decision vary, as will the potential business risks associated with making a bad decision
  - For example, buying jam at the grocery store (where you go with your gut) has less impact and risk than choosing the mutual funds for your 401(k) plan
- Document any pre-decision or post-decision **data biases** to better evaluate the effectiveness of the data requirements, and the business results of the decision

Data Requirements

- **Different decisions** will have **different data requirements**, including the data **volume**, **variety**, and **velocity** necessary for data-driven decision making
- Requirements **align** operational **data management** and **data governance** for better decision support and establish **decision-specific data quality thresholds**
- Not every business decision will require the same amount and quality of data, but this *doesn't mean that data quality is irrelevant to decision quality*
- Better decisions come from better data quality thresholds, which will deliver **better data to the decisions that require it**, and provide **business justifications** for data quality improvements that will truly improve business performance
  - Measuring data quality independently of any business context is by far the leading reason why data quality metrics get ignored by business stakeholders

Decision Evaluation

- **Implement decision quality thresholds** to **close the feedback loop** on the business impacts and risks associated with your data-driven decision making
  - Decision quality is far more difficult to define and measure than data quality
- **Quality of a decision** is determined by its **business results**, not who made the decision, the quality of the supporting data, or the decision-making technique
- Result of **every decision**, even an intuition-driven decision, **produces data**, which you can **use to re-evaluate** decision criteria and data requirements
- Even though evaluating decision quality only establishes a **correlation**, and **not a causation**, between the execution of a decision and its business results, this is why you must regularly evaluate our data-driven decision making

Getting Really Loopy

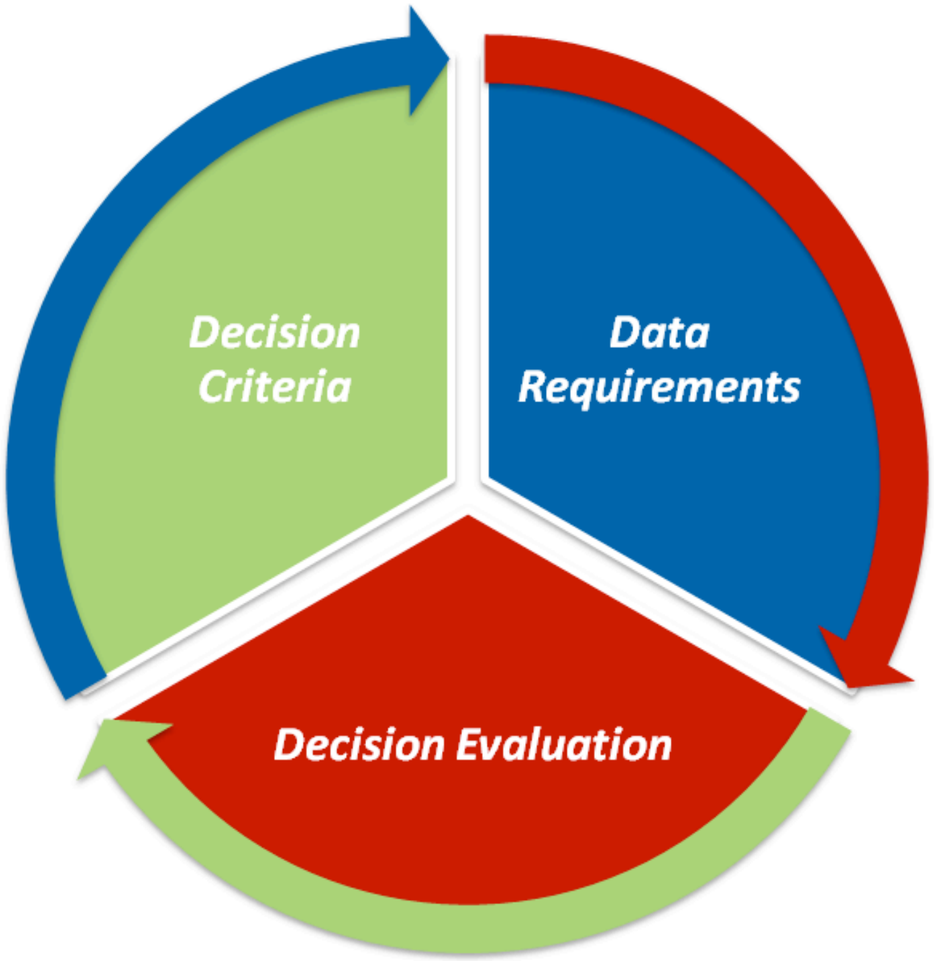
(A) **Better Business Performance** is often correlated with

(B) **Better Decisions**, which are often correlated with

(C) **Better Data**, which is precisely why *Better Decisions with Better Data is foundational to Business Success* – however . . .

- *We can't draw* **lines of causation** between (C) and (A), (C) and (B), or (B) and (A)
  - If **bad data** was **the cause** of bad decisions and/or bad performance, *every organization* **would never** *be profitable*
  - If **good data** was **the cause** of good decisions and/or good performance, *every organization* **could always** *be profitable*
- We need to connect the dots of (A), (B), and (C) by drawing **loops of correlation**
- We need to get Really Loopy with our Data-Driven Decision Making

The Decision-Data Feedback Loop



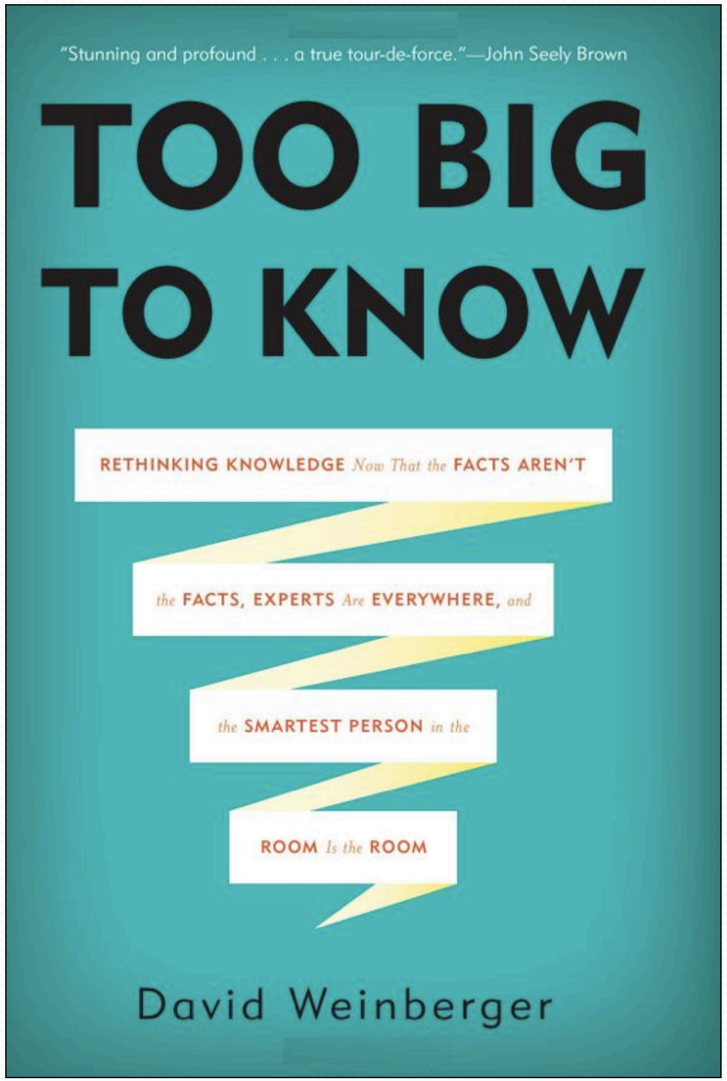
- Understand decision criteria before evaluating the data requirements needed to support decision execution
- **Data quality thresholds** deliver better data to the decisions that require it
- **Decision quality thresholds** evaluate the business results of decisions
- Continuous improvement enables **better decisions** with **better data**, which drives **better business** performance

Outcome Bias and Hindsight Bias

- *"Our comforting conviction that the world makes sense rests on a secure foundation: Our almost unlimited ability to ignore our ignorance."* ~ Daniel Kahneman
- **Outcome Bias** — Blaming decision makers for good decisions that failed and giving too little credit for successful decisions that appear obvious after the fact
- **Negativity Bias** — Bad evokes a stronger reaction than good in the human mind, just compare an insult and a compliment, which one do you remember more often?
- **Probability Neglect** — We imagine the numerator while ignoring the denominator (one bad data-driven decision out of a hundred is 1/100: 1% Bad, but 99% Good)
- **Hindsight Bias** — We cannot suppress the powerful intuition that what makes sense in hindsight today was predictable yesterday, and this illusion that we understand the past fosters overconfidence in our ability to predict the future . . .

OCDQ Blog Obsessive-Compulsive Data Quality by Jim Harris

Why Can't We Predict the Weather?



- Every day we receive a continuous real-time stream of meteorological data from space satellites, ocean buoys, and Wi-Fi-enabled sensors in rain forests, measuring temperatures, rainfall, wind speeds, and CO^2 levels
  - So . . . *why can't we predict the weather?*
  - Does meteorological data have *data quality issues?*
  - Do meteorologists have *decision quality issues?*
- An old statistics joke says the best predictive variable of whether it will rain is a spike in umbrella sales
  - It's a joke because people buy more umbrellas when *it's already raining* — umbrella sales *do not forecast* rain

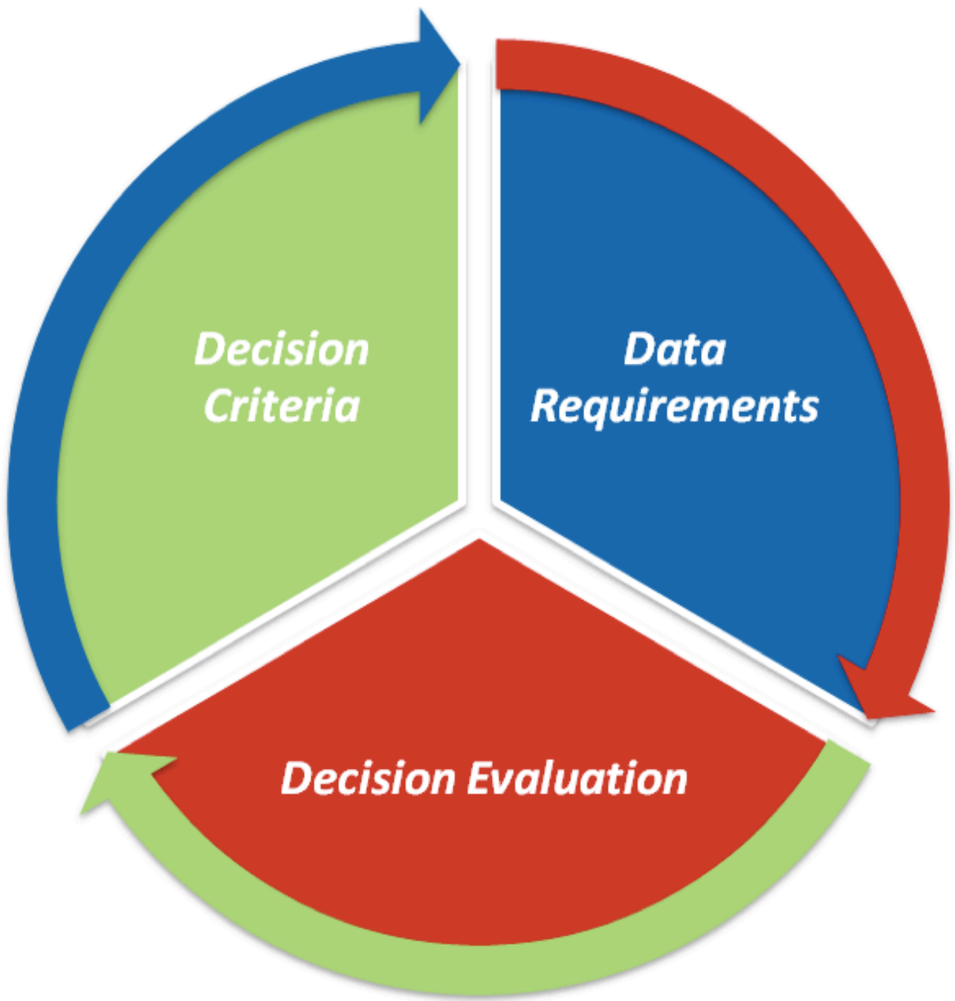
OCDQ Blog Obsessive-Compulsive Data Quality by Jim Harris

Business is as Predictable as the Weather

- The **business world** will forever remain **as predictable as the weather**, which is why you need a strategy for better data-driven decision making
- But the **harsh reality** is that **no strategy can eliminate** the potential for poor data quality and decision quality – nor the potential for poor business results even despite better data quality and decision quality
- Closing the feedback loops and practicing continuous improvement make data-driven decisions **more transparent through better monitoring**, allowing you to learn from decision-making mistakes, and adjust when necessary
- Although it cannot predict business success, continuous improvement can reliably forecast **better business performance** for your organization

Decision-Driven Data Management

A Strategy for Better Data-Driven Decision Making



Decision Criteria

- Time Constraints and Preferences
- Business Impacts and Risks
- Human Intuitions and Biases

Data Requirements

- Data Volume, Variety, and Velocity
- Data Management and Governance
- Data Quality Thresholds

Decision Evaluation

- Evaluate Business Results
- Decision Quality Thresholds
- Re-evaluate Decision Criteria
- Re-evaluate Data Requirements