

Data-Driven Reasoning

Joseph Nathan Cohen

*Department of Sociology & Program in Data Analytics and Applied Social
Research*

City University of New York, Queens College

www.josephnathancohen.info

23 August, 2019

Abstract

DRAFT IN DEVELOPMENT. DO NOT CITE OR CIRCULATE. This discussion explores how data analytics can support decision-making and problem-solving.

Data-driven reasoning is the practice of using empirical data in the process of reasoning, problem-solving, or decision-making. It can be contrasted with reasoning based on non-data grounds (like intuition, logical reasoning, pure theory, anecdotal reasoning, or common sense).

One cannot *exclusively* forge ideas, make decisions, or solve problems through data-driven reasoning. This method has some clear limitations (see below). It is more squarely involved with one particular facet of our reasoning: factual beliefs.

Beliefs

Data analytics strengthen our sense- and decision-making by sharpening a particular category of belief that influences us strongly: our *factual beliefs*.

Beliefs can be understood as pieces of information that we retain and use in our thinking. They are mental representations that purport to describe what is real, correct, or right. For example, I believe that vaccinations do not cause autism, that Montreal bagels are better than New York bagels, and that Bon Iver is a great band. These are three distinct ideas in which I believe, which is to say that I think these three pieces of information are true. Some of these beliefs might not actually be true, and it seems impossible to render truth judgements on some beliefs.

Beliefs can take different forms. For example, they can take the form of a moral, aesthetic, or factual belief.

Moral beliefs articulate people's views on good *versus* evil, or right *versus* wrong. Moral beliefs often involve questions about what people *ought to* do, as opposed to what they actually do. For example, one might believe that capital punishment is always wrong. Others feel that it is permissible under a limited set of circumstances. Some people strongly support capital punishment and wish that society used it more often. These are all moral beliefs,

as they involve classifying a thing or action's goodness, righteousness, rectitude, or propriety.

Data analysis cannot adjudicate the truth of moral beliefs. Data *can* tell us what moral beliefs are prevalent, but not which morals are better or more true. Sometimes, most people harbor bad moral beliefs, at least in my opinion.

Aesthetic beliefs are concerned with the amount of pleasure that someone feels upon sensing something. It is concerned with what is beautiful *versus* ugly, delicious *versus* disgusting, or attractive *versus* repulsive. Some people like the smell of perfume, while others hate it. These are aesthetic beliefs, as they pertain to whether or not someone likes the experience of sensing perfume.

Again, data analysis cannot easily adjudicate aesthetic beliefs. We can use data to make guesses about what people are likely to find pleasant, beautiful, or compelling, but we can't *prove* one aesthetic judgment to be objectively correct. How often does everyone love something that you think is bad or gross?

Factual beliefs are concerned with the objective facets of reality. ***Objective reality*** describes existence that is independent of people's thoughts or judgments. The existence of the sun is an objective fact. If all human beings disappeared tomorrow, the massive burning sphere around which our planet revolves would continue to exist. In contrast, moral and aesthetic beliefs are considered to be ***subjective realities***, which fundamentally exist in people's minds. Beauty and moral righteousness largely exist in people's minds, and can be understood as articulating preferences. If all humans disappeared tomorrow, would there still be good and evil in the world? Divorced from human sensibilities, who is to say that a dung beetle is less beautiful than a peacock?

Factual beliefs are not necessarily true. Do not treat these terms as synonymous. A factual belief is defined by its *attempt* to describe objective reality. This attempt can fail. For example, I might believe that the sun revolves around the earth. This is a factual belief in that it purports to describe objective reality. The problem is that we have good reasons to believe that this description of reality is inaccurate.¹

Moving forward, we describe a factual belief that successfully describes reality as having "truth" or "accuracy." A factual belief's failure to accurately describe reality (and thus misrepresent reality) can be understood as having some degree of "inaccuracy" or "falsity." Notice that I discuss truth in terms of degrees – ideas can be more or less true, as opposed to be absolutely true or false.

Discussion #1

Are these factual, moral, or aesthetic beliefs?

1. David Bowie's music is great. *Aesthetic*

¹For a presentations of reasons to believe in a heliocentric model of space, see Rhett Allain (2014) "How Do We Know the Earth Orbits the Sun?" *Wired*, April 14. <http://www.wired.com/2014/04/how-do-we-know-the-earth-orbits-the-sun/>

2. People should wash their hands after using the restroom. *Moral*
3. Many people do not wash their hands after using the restroom. *Factual*
4. New York City is the capital of the United States. *Factual*
5. Pizza Hut is terrible. *Not clear. Depends on what you mean by “terrible.” Terrible tasting? Terribly unhealthy food? Terribly unethical corporation?*

Discussion #2

When people express beliefs about dieties, which of these kinds of beliefs are they grappling? *It can be any of them. Questions about whether a diety exists is wrestling with factual beliefs. Those concerned with whether this diety embodies a force for good in the universe, or whether or not the rule systems ascribed to a diety truly articulate what is right or wrong, are grappling with moral questions. People wrestle with how to see a diety in their mind’s eye, and will grapple with whether or not that imagery evokes attraction, pleasure, or some other positive feeling.*

Factual Beliefs Shape Choices

Factual beliefs structure our thoughts and action. They shape what we deem to be “realistic” ideas, choices, or plans. Our thoughts and actions are strongly shaped by our sense of “objective realities”. If I believe that no one will say “yes” if I ask them to dance, then I might not ask anyone to dance. The belief may be wrong, but it still shapes the choices made by the person possessing it. Factual beliefs structure our expectations about the risks and rewards of our choices.

The quality of our decision-making process is going to be influenced by the quality of the information upon which it is premised. It presses us to conform to “realities” that do not really exist. It can distort our sense of risk and reward. For example, consider the importance of the factual information that a young person employs in making the decision to try cigarettes. Studies suggest that young people often mistakenly underestimate nicotine’s addictivity, their susceptibility to addiction, or the ease of breaking a substance addiction. The (faulty) belief that nicotine is not strongly addictive, or that addictions are easy to break, influences the perceived risks of trying cigarettes. In this case, faulty information seems likely to raise a young person’s risk of health damage as a result of smoking.

Our world is replete with faulty information. Faulty information can produce faulty decisions. For this reason, we have good cause to try to improve the quality of the information we employ when making choices.

Discussion

John is deciding whether or not to go to graduate school after finishing his Bachelor’s degree. How would each of these beliefs influence his post-undergraduate plans?

- People with Master’s and Doctoral degrees earn much more money than people without these degrees.
- People with graduate degrees are phoney snobs.

- Education is sexy!
- People who study hard and achieve academically deserve respect for their intelligence and hard work.
- People with graduate degrees have been duped by a deceptive higher education industry, who suckers people into worthless educational programs
- The graduate school campus is gorgeous!
- People with graduate degrees are more likely to marry other people with graduate degrees.

Do you share all, some, or none of these beliefs? Do you know anyone who disagrees with your beliefs on these specific topics? Who is right, and why?

Do any of these statements clearly and directly express a factual belief? Do any seem to be premised on implicit factual beliefs? Can you identify these implicit factual beliefs?

Probing Factual Beliefs

Scientific Thinking

Science is one of society's most trusted institutions. It has a track record of immense accomplishment. Data-driven reasoning is a reasoning method that basically asks people to think scientifically using quantitative data. It is worth taking a brief look at what scientific thinking entails.

Many people conflate science with truth. For example, they might believe that a scientifically-derived fact is a true fact. They are distinct. Facts can be true, even if they aren't scientifically-derived. Scientifically-derived facts can be untrue.

Science is best understood as a *style* of reasoning or thinking. It is an approach to building and testing our beliefs. To *think scientifically* is to employ a style of thinking that has at least four distinguishing characteristics:

1. **Skepticism:** Withholding belief until given adequate reason
2. **Empiricism:** Requiring observable proof that a fact is true
3. **Objectivity:** Prioritizing the truth over other agenda
4. **Replicability:** Being open and explicit about your research, making it easy for other to repeat your study to see if they get the same results.

Science cannot answer all of life's questions. It deals strictly with knowledge that concerns objective, observable events that have already occurred. Many kinds of beliefs deal with the unobservable, the non-objective, and events that have not yet transpired. Some ideas that science cannot adjudicate:

- **Moral and Aesthetic Beliefs**, as discussed above
- **Beliefs Concerning Unobservable Phenomena.** Science cannot comment on what cannot be observed. The most obvious case in which this limit applies involves religious questions, like the existence of a diety or other supernatural force that governs reality.

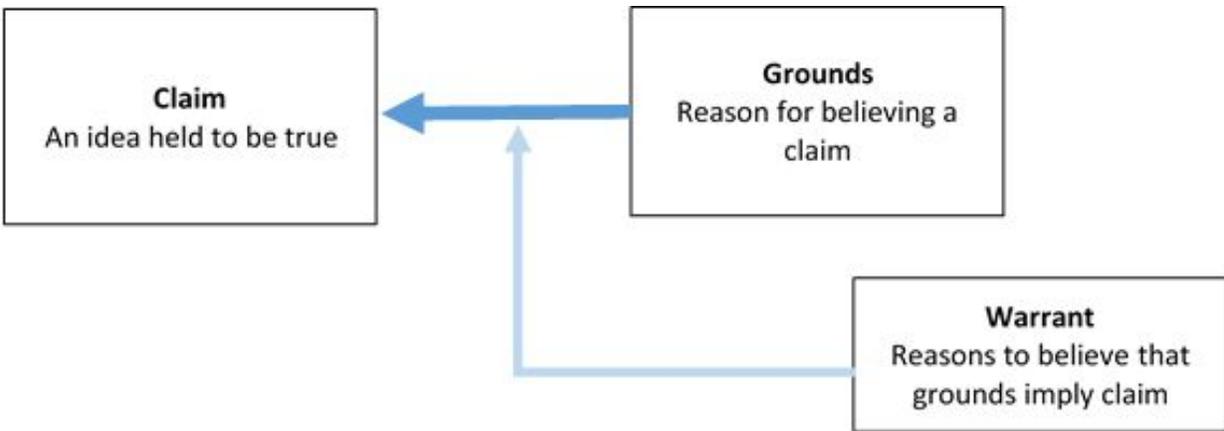


Figure 1: The Toulmin Model

- **Future Prediction.** Science cannot predict the future We can use the fruits of scientific studies, extrapolating from observations of past events to make guesses about the future. However, science itself cannot deduce the future because it is not yet observable.

Discussion

How does scientific thinking differ from:

- Religious faith?
- Ideas justified by pure reasoning?
- Ideas justified by their acceptance by authority figures?
- Ideological thinking?
- Fortune-telling?

The Toulmin Model (Simplified Version)

We develop our beliefs in discussions with others or with ourselves. Generally, this process involves issuing *arguments*, which are appeals to accept a claim as fact. If I want you to believe that New York is the world’s greatest city, I will make this appeal by giving you reasons to accept my view as true. These are *arguments* in favor of the belief that I am trying to get you to accept.

The Toulmin Model is a model that illustrates the parts of an argument. It is useful because it reminds us that arguments have multiple parts. Each of these parts must hold for the argument to hold. The model is depicted below in Figure 1.

The Toulmin Model is often described as having four to six parts. This is a simplified version with only three parts:

1. **Claim.** This is the belief that I want you to accept as true.
2. **Grounds.** This is the reason that I give you for accepting the claim as true.

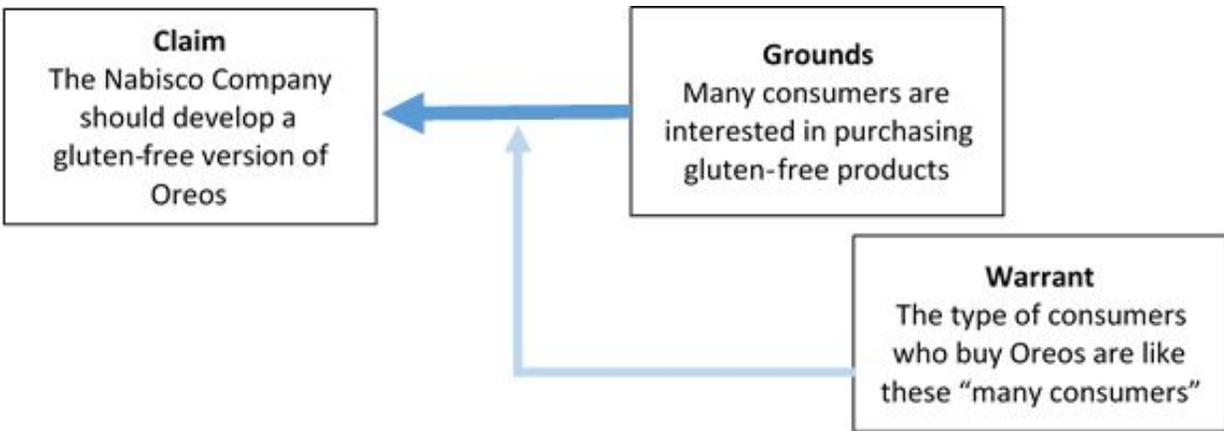


Figure 2: Example Application of the Toulmin Model

3. **Warrant.** This is the implicit rationale for believing that the grounds imply the claim. This part is often not explicit, but must be accepted for the overall argument's acceptance.

When faced with an argument. The data-driven thinker employs a form of scientific reasoning to judge whether or not to accept the grounds and warrant. They will be *skeptical* unless presented with *empirical evidence* suggesting their truth. They will try to retain an *objective* disposition, and focus on whether the argument's parts seem to reflect the true state of the world. They will reserve acceptance of empirical evidence that is not *verifiable*.

For example, consider the argument depicted below in Figure 2. The claim being advanced is that the Nabisco Company should create a gluten-free version of Oreos. Why? The person making the claim might tell you that gluten-free products are very popular now, which may or may not be true. Even if it is true, the argument is premised on the idea that the types of people who want gluten-free products are the same kind of people who buy Oreos. A data analyst might be the one who pushes the team to look for data suggesting that gluten-averse people are interested in buying gluten-free cookies, or that they'd like the new gluten-free Oreo recipe.

Discussion

What is your opinion on this question: *Should our MA program require applicants to write the GRE?* Why or why not? Use a Toulmin model to deconstruct three or four reasons for your opinion.

Can your arguments' grounds or warrants be verified using data? How? What would you expect to find if your factual claims are *wrong*?

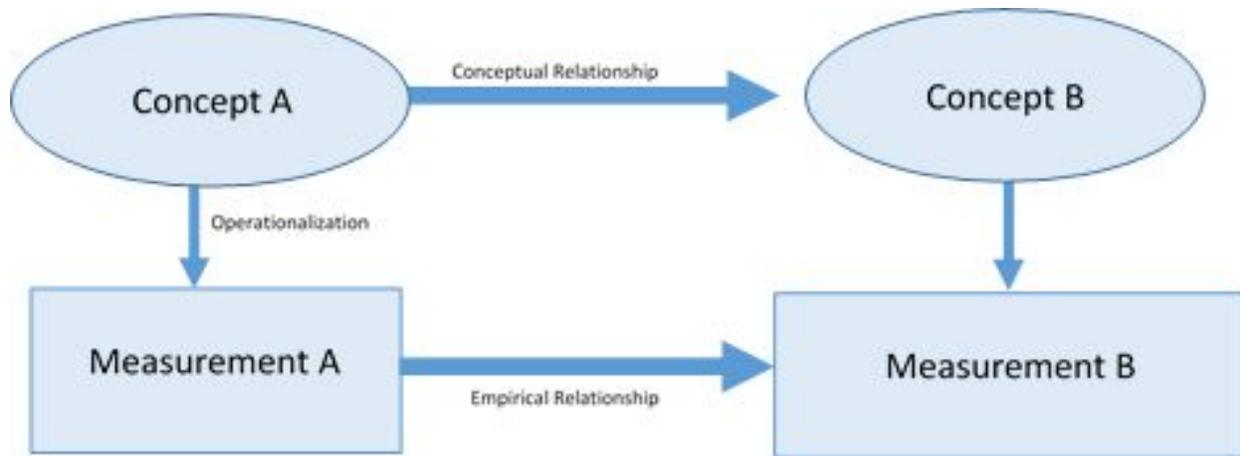


Figure 3: A Model of a Theory

Theories

Theories are mental models that describe how things work. They articulate relationships between people, things, or phenomena. Theories often shape our decision-making by shaping our expectations. For example, if policy-makers think that lowering business taxes will create more jobs, they might have reason to lower business taxes. The relationship is theoretical, though. Lowering business taxes might not result in job creation. We need some way to probe our theories to ensure that our decision-making is shaped by well-justified ones.

One way to think about theories is depicted below.

This model of a theory has four parts:

1. **Concepts:** Descriptors, characteristics, or characterizations of people, things, or phenomena.
2. **Conceptual Relationship:** Hypothesized ways in which concepts are related
3. **Measurement:** The process of converting a concept into a quantifiable measurement. This process is called *operationalization*.
4. **Empirical Relationship:** The degree to which measures of the related concepts are correlated.

Discussion

Sketch out a theory about love and marriage. How would you test it using data?

Evaluating Causal Theories

When we talk about *causality*, we are discussing how certain we are that something causes something else. For example, when the AIDS epidemic broke out, epidemiologists worked hard to find out what *caused* AIDS transmission. People were worried about toilet seats

and hand-shakes. Eventually, research refined our views of what causes people to catch the disease.

Of course, you can remember your undergrad stats prof going on that “CORRELATION IS NOT CAUSATION”. It’s a classic go-to-bit in stats teaching. It’s also very true, in the sense that correlation does not prove causation in and of itself. However, it’s not irrelevant. It’s just one of several important factors:

When examining theories using data, remember to look for three conditions:

1. **Coincidence:** Cause and effect must appear in the same subjects
2. **Precedence:** Cause must precede effect
3. **Eliminating Possibility of Spurious Causation:** Eliminate outside factors that may be deceiving us into believing a causal relationship exists.

The final condition is generally the sticking point. No matter how sound our tests, there is always questions about whether something is going on that we don’t yet see or understand.

Further Reading

Eugene Bardach (xxx) *A Practical Guide to Policy Analysis*

Andrew Gelman (2011) “Causality and Statistical Learning” *American Journal of Sociology*, 117 (3): 955 – 966.

Ian Hacking (xxx) *An Introduction to Probability and Logic*

James March (xxx) *A Primer on Decision-Making*.

Metcalfe, Mike and Saras Sastrowardoyo (2013) “Complex Project Conceptualisation and Argument Mapping” *International Journal of Project Management*, 31: 1129 – 1138.

Morgan, Stanley and Christopher Winship (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. New York: Basic Books.

Provist, Foster and Tom Fawcett (2013) *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking* O’Reilly.

Rex, Lesley A., Ebony Elizabeth, and Steven Engel (2010) “Applying Toulmin: Teaching Logical Reasoning and Argumentative Writing” *The English Journal*, 99 (6): 56 - 62

Max Schron (xxx) *Thinking with Data*.

Eric Siegel (xxx) *Predictive Analytics*.

James E. Warren (2010) “Taming the Warrant in Toulmin’s Model of Argument” *The English Journal*, 99 (6): 41 – 46.