# Storytelling with Data

## Module 5: Analyze before you speak—audience analysis

**Scott Spencer**
Faculty and Lecturer
Columbia University

# Agenda

Upcoming deliverable

Today's objectives

Audience: questions, biases

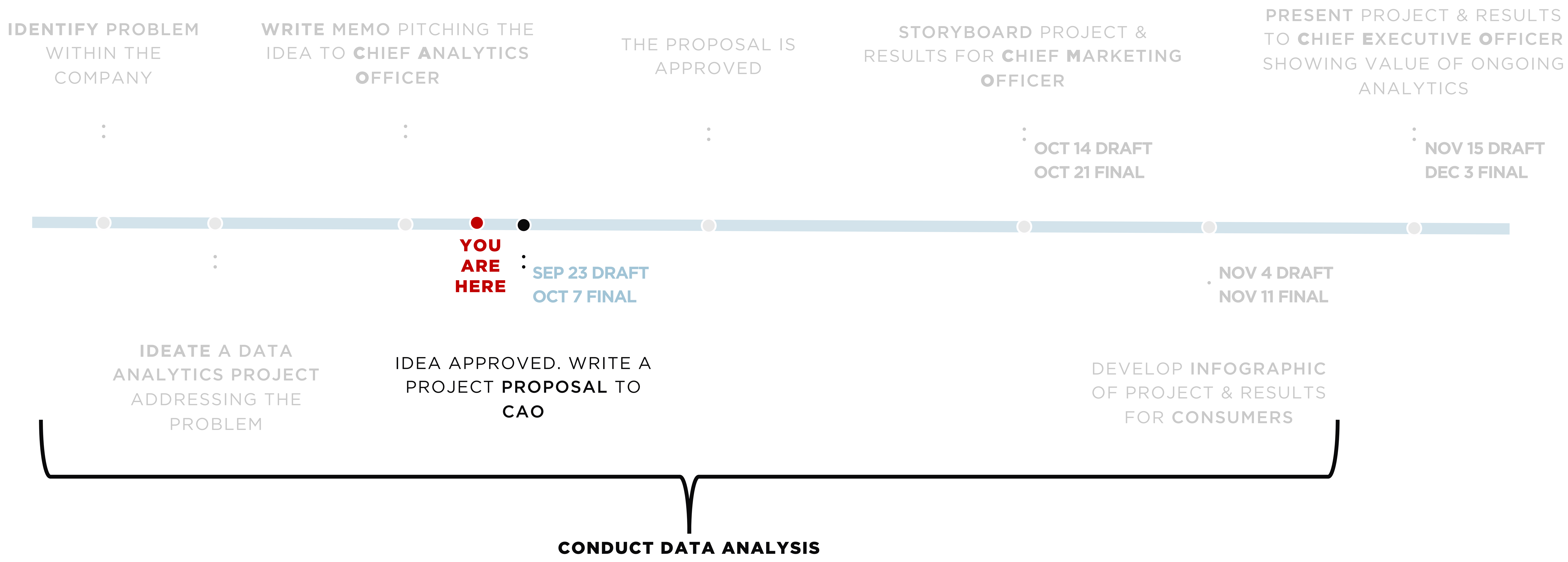
Audience: cultural cues, emotional display

Fair, influential communication

# Upcoming deliverable

# Upcoming deliverable

**750-word brief proposal** — Write a brief proposal to **CAO** detailing your proposed analytics project. Consider background context, problem, data, solution, and impact. At this point you should have data to start an analysis.



# Change generic informational sections,

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| **Introduction** | **Challenge and Opportunity** | **Rationale** | **Cost Analysis** | **Assessment** | **Conclusion** |
| Describe context relevant to the proposal. | Offer necessary details of the problem and identify your solution. | Explain your solution, how the ideas can be applied to the particular need, and why this solution will work. | Consider cost, resources required, and expected return on investment, using quantitative or qualitative language. | Describe how the project will be evaluated both immediately and over time. | Include a clear and complete call to action. |

# into **specific messages** of your project.

# Example brief proposal for a data analytics project. **Constraint—750 words or less in main body.**

## Proposal for exploring game decisions informed by expectations of joint probability distributions


To: Scott Powers, Senior Baseball Analyst, Los Angeles Dodgers
From: Scott Spencer, Faculty and Lecturer, Columbia University


14 February 2019

Our game decisions based on current modeling do not maximize spend per win. We witnessed the mid-market Astros use analytics to overtake us in the 2017 World Series (Luhnow 2018ab). Our efforts also do not maximize expected wins. But we can. To do so, we need to jointly model probabilities of all game events and base decisions on *expectations* of those distributions. With adequate computing emerging, we can be first using the probabilistic programming language Stan and parallel processing. To demonstrate the concept, consider a probability model for decisions to steal second base, below, which suggests teams are too conservative, leaving wins unclaimed. This model allows us to ask, for example—*should Sanchez steal against Sabathia? Or against Pineda?*

### 1 Our current analyses do not optimize expected wins

Seven terabytes of uncompressed data generated per game overshadow the lack of situational data needed for decision-making that maximizes expected utility. Consider that pitchers, on average, only face10 percent of major league batters regardless of game state; the reverse is true, too. Or when deciding whether a base runner should attempt to steal against a specific pitcher and catcher in a state of play, say, we are lucky to have any data. Common analyses and heuristics for these situations are inadequate: they not only overfit the data (if any exist), but also offer no manner of estimating changes in probabilities for maximizing *expected* utility (winning the game).

Accurately quantifying probabilities, and changes thereof, in a given context enable us to answer counterfactuals, from which we can build strategies that maximize our objectives (Parmigiani 2002). This approach is possible at scale using Stan (Carpenter et al. 2017). It's time to jointly model probabilities of all events.

### 2 Modeling probabilities for steal success illustrates a broader benefit

To see the potential of implementing probability models, let's consider, again, the decision to steal bases, given a specific counterfactual:



> In a game against New York Yankees, should Milwaukee Brewers's Lorenzo Cain attempt to steal second base with no one else on base and two outs before the seventh inning, against Gary Sanchez as catcher and Michael Pineda as pitcher? What if against Sanchez and CC Sabathia as pitcher?

More specifically, how can we know the *expectation* that Cain's attempt in each situation increases the probability of expected runs that inning and by how much? Using Stan, I've coded a generative model that along with play outcomes considers various information (runner foot-speed, catcher pop-time) and player characteristics, like pitcher handedness. With the model, we have an answer that also shows the uncertainty. Given 2017 data, this model suggests Cain should steal against Pineda, not Sabathia:

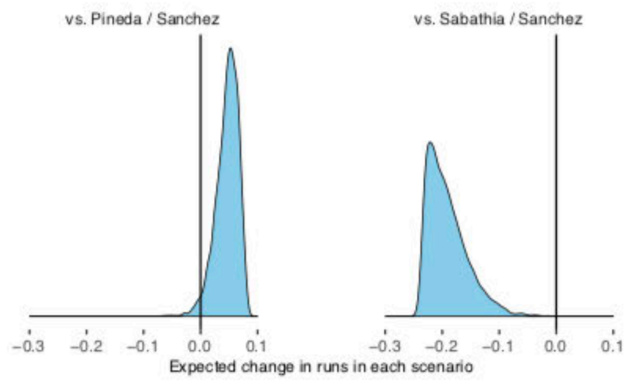



Figure 1. Of the two scenarios, Cain should only attempt to steal against the Sanchez–Pineda duo.

Notably, we get these expectations without multiple trials of either scenario. More generally, this model suggests that on average team managers are too conservative, leaving runs unrealized:

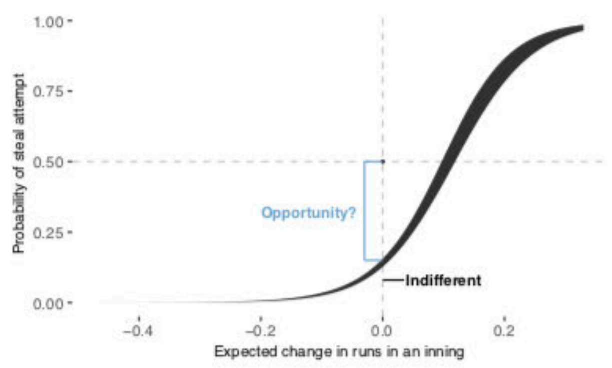



Figure 2. When the change in expected runs is zero, managers should be indifferent to attempted steals, saying go half the time.

The **black band** represents the range of variation across managers' decisions. At the intersection of indifference, managers tend to say steal only 10 percent of the time, leaving opportunity.

The above is but one example of a more general approach that weighs probabilities of all possible outcomes to maximize expected utility. With broad implementation—jointly modeling the conditional probabilities of all relevant events—we can optimize decisions.



### 3 For value, compare an investment to free-agent costs

A fully-realized model will require significant effort from a team with deep experience in baseball, generative modeling, and Stan. To get the talent, we should compare cost to acquiring expected wins from free-agents. Each win above a *replacement-level* player costs about 10 million per year (Swartz 2017). As with free-agent value over replacement player, game-time decisions informed from more accurate probabilities should add wins over a season. The scope of what we can answer, moreover, goes beyond in-game strategy (player acquisitions, salary arbitration). More immediately, however, we can begin to implement this approach for specific events, with a scope closer to the example above, being mindful that information learnt are conditional upon unmodeled context.

### 4 For accuracy, compare model results to betting market odds

Measuring performance of a fully-realized model may seem tricky: *we only see the outcome of our decisions.* But we can, say, compare the accuracy of our estimates against the betting market where interested investors are trying to forecast game outcomes.

### 5 Conclusion

The mid-market Astros show teams can do more with information. Millions in additional revenue—and more wins—await discovery through a joint, probability model of all events from which we can maximize conditional expectations. Let's discuss how to draw the talent for a title worth our spend.

### 6 References


Carpenter, Bob, et. al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1): 1–32.

Luhnow, Jeff. 2018a. "How the Houston Astros are winning through advanced analytics." McKinsey Quarterly 13 June 2018: 1–9.

———. 2018b. "A view from the front lines of baseball's data-analytics revolution." McKinsey Quarterly 5 July 2018: 1–8.

Parmigiani, G. 2002. "Decision Theory: Bayesian." In *International Encyclopedia of the Social Behavioral Sciences*, 3327–34.

Swartz, Matt. 2017. "The Recent History of Free-Agent Pricing." https://www.fangraphs.com/blogs/the-recent-history-of-free-agent-pricing/.


Readability Statistics

| | |
|---|---|
| **Counts** | |
| Words | 720 |
| Characters | 3,997 |
| Paragraphs | 16 |
| Sentences | 35 |
| **Averages** | |
| Sentences per Paragraph | 4.3 |
| Words per Sentence | 18.1 |
| Characters per Word | 5.3 |
| **Readability** | |
| Flesch Reading Ease | 33.2 |
| Flesch-Kincaid Grade Level | 13 |
| Passive Sentences | 0% |



Spencer, Scott. *Proposal for Exploring Game Decisions Informed by Expectations of Joint Probability Distributions*. Columbia University, 2019. Print.

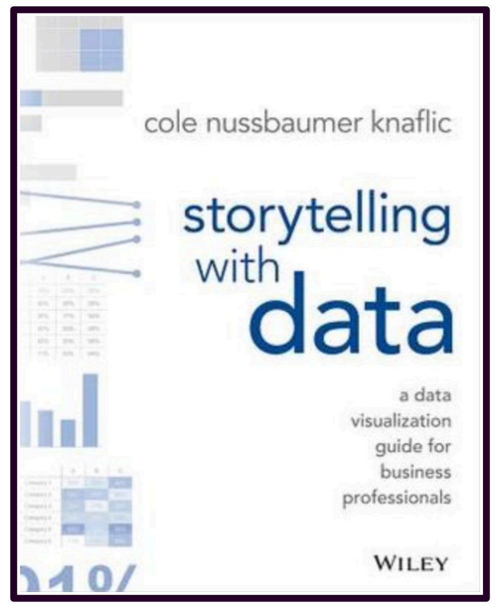
# Today's Objectives

# Objectives

1. Articulate the need for audience analysis and sensitivity in the applied analytics setting.

2. Consider professional, demographic, cultural, and personal issues when building, contributing to, or managing an analytics team or project.

# Audience analysis and biases




# Storytelling with data

*Knaflic*

The author is a consultant focused on visual displays. Her experience arose from human resources in Google where she applied theory learned as a student of Yale's Edward Tufte.

**Audience: questions to specifically answer**

Who is the audience or decision maker?

What is your relationship with them?

What do you need them to know or do?

How will you communicate with them?

What tone do you want your communication to set?

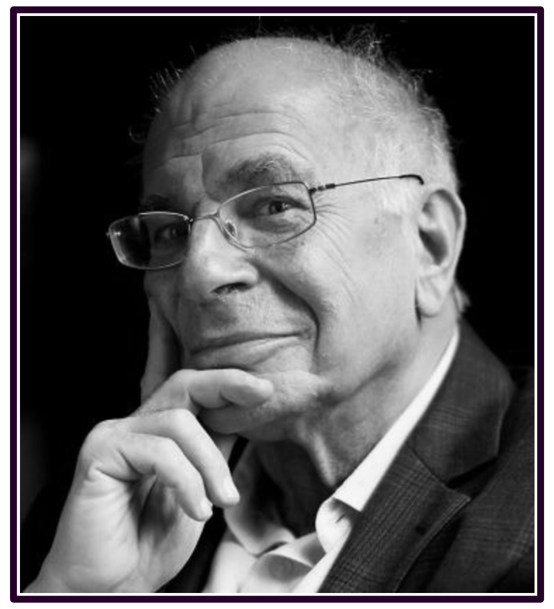
What data are available to help make your point?

Is your audience familiar with these data?

What are audience biases as related to your messages?



Knaflic, Cole Nussbaumer. *Storytelling with Data*. Wiley, 2015. Print.



HBR.ORG

Dangerous biases can creep into every strategic choice. Here's how to find them—before they lead you astray. *by Daniel Kahneman, Dan Lovallo, and Olivier Sibony*

**Before You Make That Big Decision…**

**THANKS TO** a slew of popular new books, many executives today realize how biases can distort reasoning in business. *Confirmation bias,* for instance, leads people to ignore evidence that contradicts their preconceived notions. *Anchoring* causes them to weigh one piece of information too heavily in making decisions; *loss aversion* makes them too cautious. In our experience, however, awareness of the effects of biases has done little to improve the quality of business decisions at either the individual or the organizational level.

Though there may now be far more talk of biases among managers, talk alone will not eliminate them. But it is possible to take steps to counteract them. A recent McKinsey study of more than 1,000 major business investments showed that when organizations worked at reducing the effect of bias in their decision-making processes, they achieved returns up to seven percentage points higher. (For more on this study, see "The Case for Behavioral Strategy," *McKinsey Quarterly,* March 2010.) Reducing bias makes a difference. In this article, we will describe

ILLUSTRATION: LASSE SKARBOVIK

June 2011 Harvard Business Review 51

# Before you make that big decision…

*Kahneman, co-authors*

Awarded the Nobel-Prize in economics and senior scholar at Princeton, Kahneman introduced the idea of cognitive biases, and their impact on decision making.

**Two modes of thinking**

**Intuitive (system one)** thinking, impressions, associations, feelings, intentions, and preparations for actions flow effortlessly. This system mostly determines our thoughts. System one uses **heuristics**, has **biases**.

**Reflective (system two)** thinking is slow, effortful, and deliberate.

Both are continuous, but system two typically monitors things, and only steps in when stakes are high, we detect an obvious error, or rule-based reasoning is required.

It's very hard to remain aware of our own biases, so we need to **develop processes** that identify them and, most importantly, get **feedback from others** to help protect against them.



Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. "Before You Make That Big Decision ...." Harvard Business Review 89.6 (2011): 50–60. Print.


Bias ?.. WHAT Bias ?!

1

## Check for self-interested biases

Is there any reason to suspect the team making the recommendation of errors motivated by self-interest?

Review the proposal with extra care, especially for over optimism.

2

## Check for the affect heuristic

Has the team fallen in love with its proposal?

Rigorously apply all the quality controls on the checklist.

3

## Check for groupthink

Were there dissenting opinions within the team? Were they explored adequately?

Solicit dissenting views, discreetly if necessary.

4

## Check for saliency bias

Could the diagnosis be overly influenced by an analogy to a memorable success?

Ask for more analogies, and rigorously analyze their similarity to the current situation.



Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. "Before You Make That Big Decision ...." Harvard Business Review 89.6 (2011): 50–60. Print.

5

## Check for confirmation bias

Are credible alternatives included along with the recommendation?

Request additional options.

6

## Check for availability bias

If you had to make this decision in a year's time, what inform-ation would you want, and can you get more of it now?

Use checklists of the data needed for each kind of decision.

7

## Check for anchoring bias

Where are the numbers from? Can there be … unsubstantiated numbers? … extrapolation from history? … a motivation to use a certain anchor?

Re-anchor with data generated by other models or benchmarks, and request a new analysis.

8

## Check for halo effect

Is the team assuming that a person, organization, or approach that is successful in one area will be just as successful in another?

Eliminate false inferences, and ask the team to seek additional comparable examples.



Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. "Before You Make That Big Decision …." Harvard Business Review 89.6 (2011): 50–60. Print.

9

## Check for sunk-cost fallacy, endowment effect

Are the recommenders overly attached to past decisions?

Consider the issue as if you are a new executive.

10

## Check for overconfidence, optimistic biases, competitor neglect

Is the base case overly optimistic?

Have a team build a case taking an outside view: use war games.

11

## Check for disaster neglect

Is the worst case bad enough?

Have the team conduct a premortem: imaging that the worst has happened, and develop a story about the causes.

12

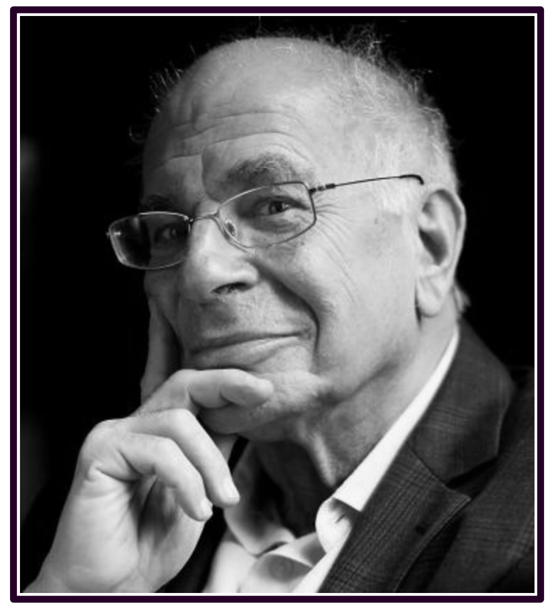
## Check for loss aversion

Is the recommending team overly cautious?

Align incentives to share responsibility for the risk or to remove risk.



Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. “Before You Make That Big Decision ….” Harvard Business Review 89.6 (2011): 50–60. Print.



Dangerous biases can creep into every strategic choice. Here's how to find them—before they lead you astray. *by Daniel Kahneman, Dan Lovallo, and Olivier Sibony*

**Before You Make That Big Decision…**

# Before you make that big decision…

## *Kahneman, co-authors*

Awarded the Nobel-Prize in economics and senior scholar at Princeton, Kahneman introduced the idea of cognitive biases, and their impact on decision making.

**Keeping out the appearance of bias**

Present ideas from a **neutral perspective**. Becoming too emotional suggests bias.

Make **analogies and examples comparable** to the proposal.

Genuinely **admit uncertainty** in the proposal, and **recognize multiple options**.

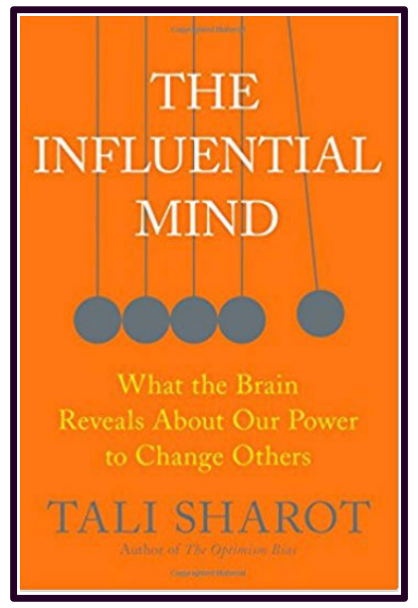
Identify **additional data** that may provide new insight.

Consider **multiple anchors** in the proposal.



Kahneman, Daniel, Dan Lovallo, and Olivier Sibony. "Before You Make That Big Decision ...." Harvard Business Review 89.6 (2011): 50–60. Print.

# Fair, influential communication





# The influential mind

*Sharot*

A London neuroscientist, her research focuses on decision-making, emotion, and influence.

**Learn what your audience is thinking**

If we want to affect the behaviors and beliefs of the person in front of us, we need to first understand what goes on inside their head.

**Formula for changing beliefs**

Four factors come into play when we form a new belief: our **old belief** (this is technically known as the "prior"), **our confidence in that old belief**, the **new evidence**, and **our confidence in that evidence**.

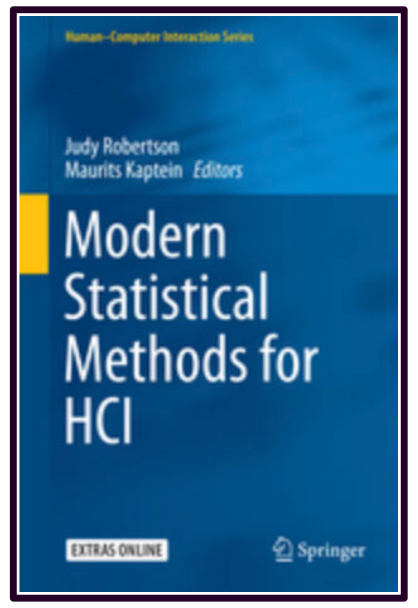
**Find common ground with audience's beliefs**

When you provide someone with new data, they quickly **accept evidence that confirms their preconceived notions** (what are known as prior beliefs) and assess counterevidence with a critical eye.

Focusing on what you and your audience have in common, rather than what you disagree about, enables change.



Sharot, Tali. *The Influential Mind*. Henry Holt and Co. 2017. Print.




# Ch. 13, Fair statistical communication in HCI

*Dragicevic*

He is a researcher, focusing on psychology of data visualization for judgment and decision making, and on transparent statistical communication.

**Report effect sizes, interval estimates, interpretations**

Whether exploring or confirming analyses, show results using an *estimation* approach—use graphs to show **effect sizes** and **interval estimates**, and offer **nuanced interpretations** of results. *Avoid the pitfalls* of dichotomous tests and p-values.

**Tip 19: be creative, use visuals**

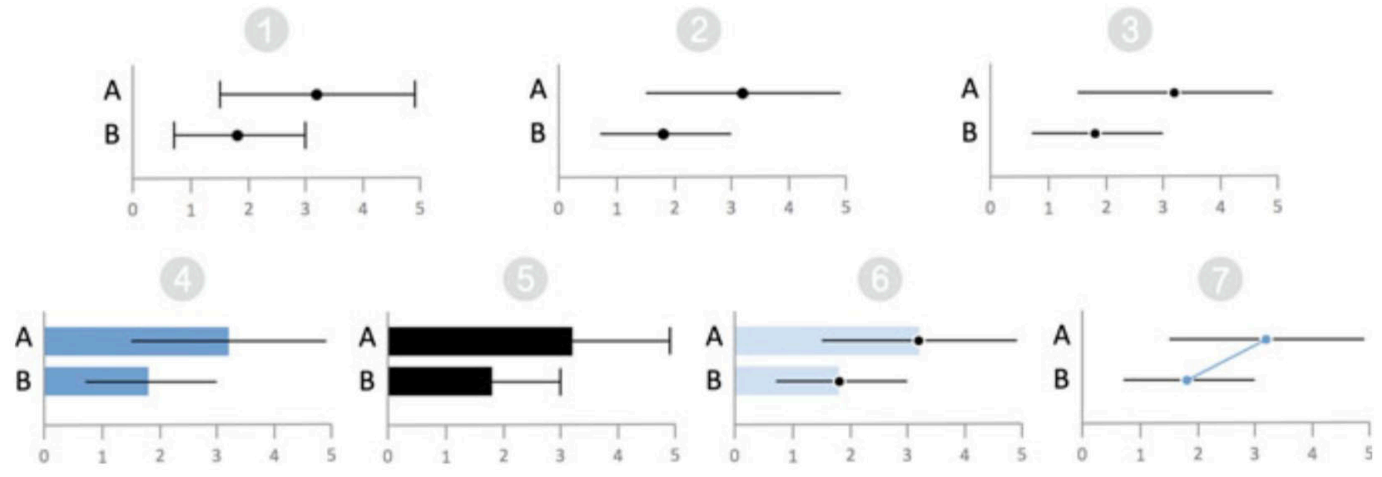


**Fig. 13.9** Seven ways of plotting effect sizes with confidence intervals

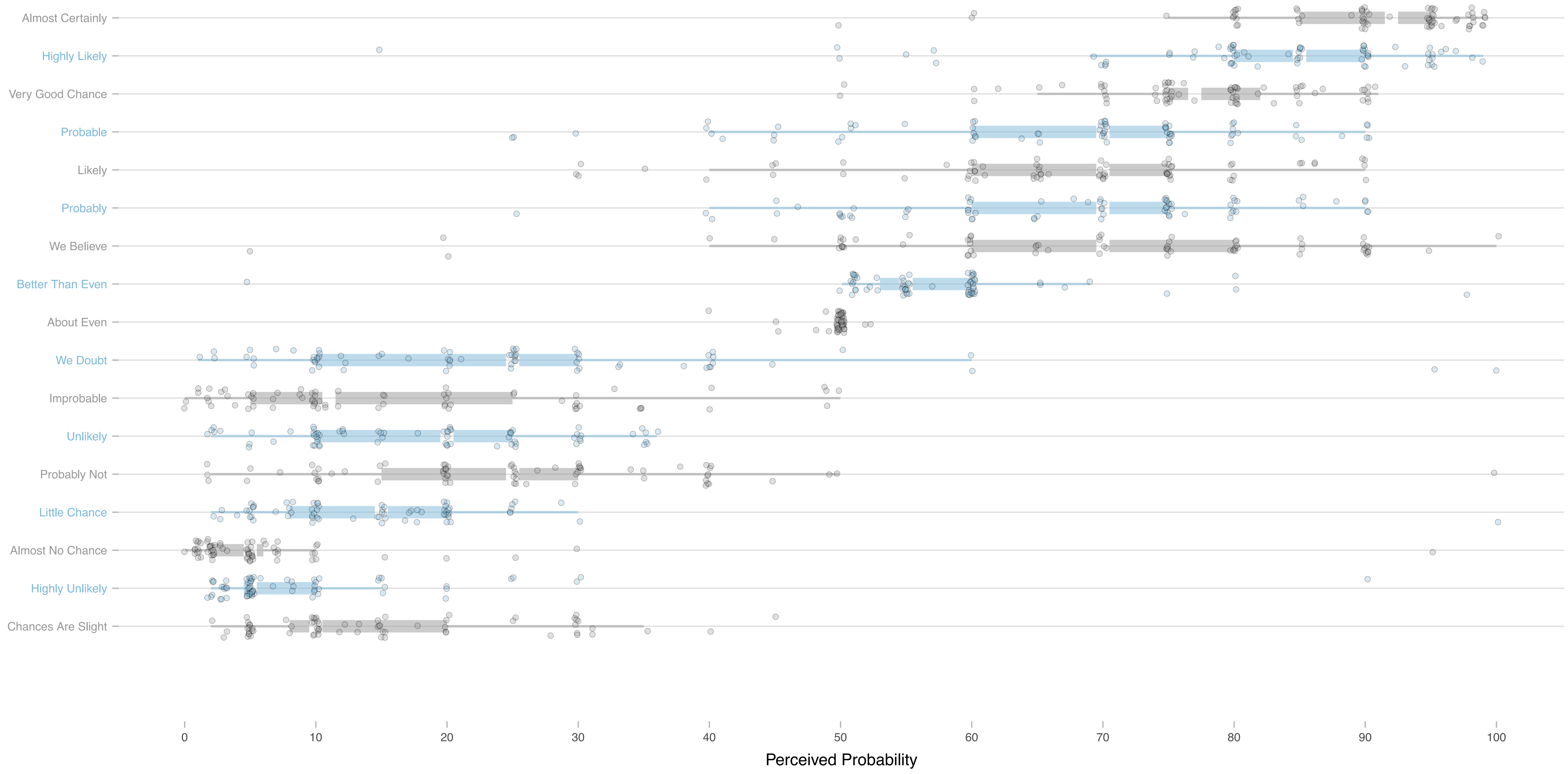
**Tip 26: honestly convey uncertainty**

The use of vague language is necessary for **acknowledging and honestly conveying the uncertainty present in effect size estimates**. Vague language — which is not the same as ambiguous language — plays a key role in reasoning.



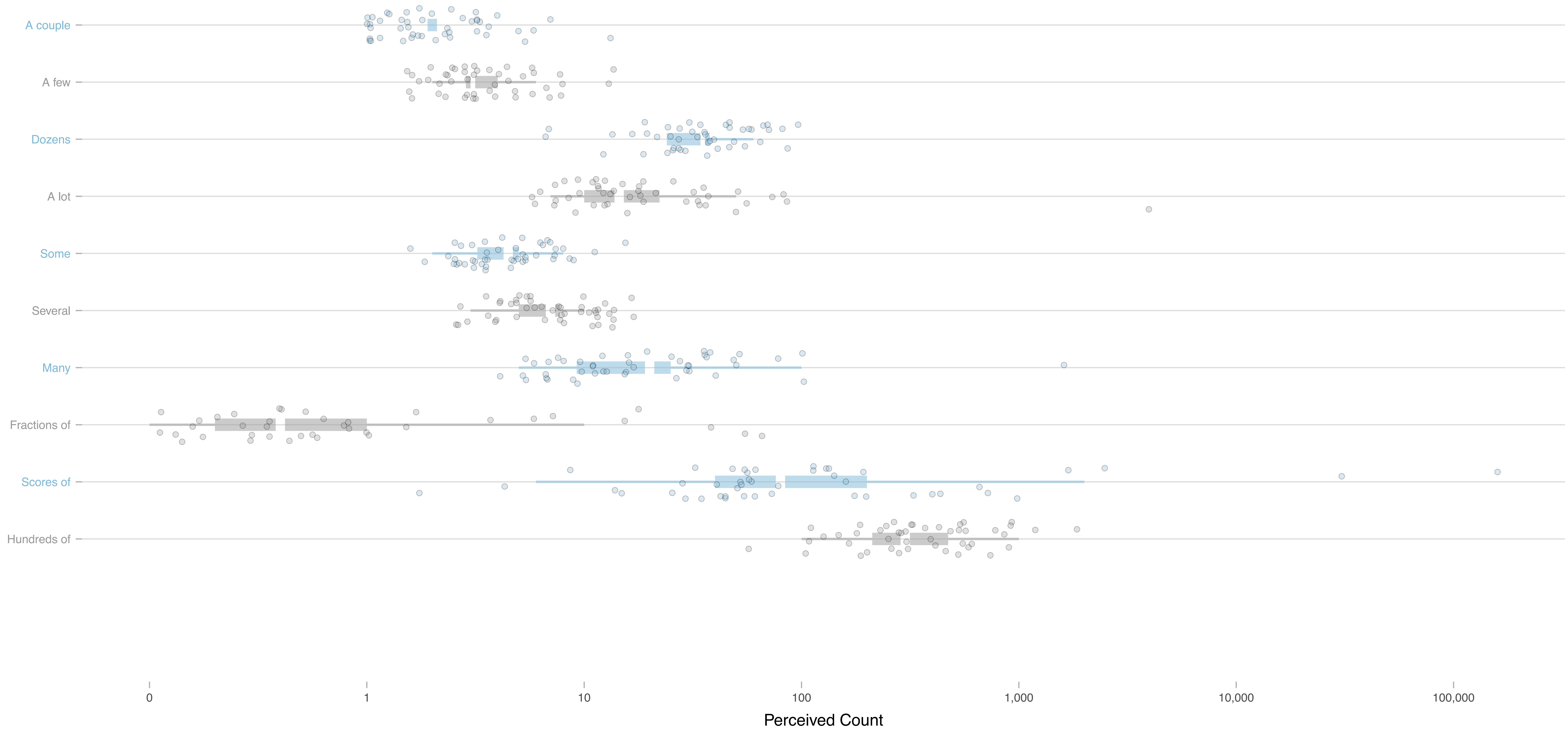
Dragicevic, Pierre. "Fair Statistical Communication in HCI." *Modern Statistical Methods for HCI*. Ed. Judy Robertson and Maurits Kaptein. Springer International Publishing, 2016. 291–330. Print.

# Language describing **probability** are imprecise, depend upon audience and context




Spencer, Scott. *Quantitative persuasion amid uncertainty*. Forthcoming. Print.
Barclay, Scott et al. *Handbook for Decision Analysis*. Decisions and Designs, Inc., 1977. Print.
zonination. "Perceptions of Probability and Numbers." github. github.com, 18 Aug. 2015. Web. 26 Dec. 2018.

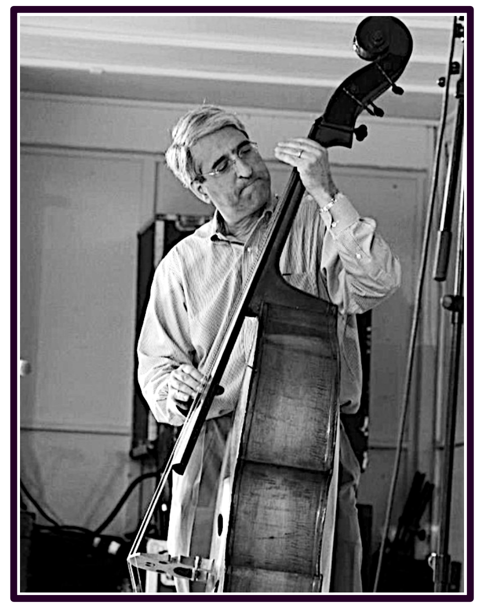
# Language describing **quantities** are imprecise, depend upon audience and context

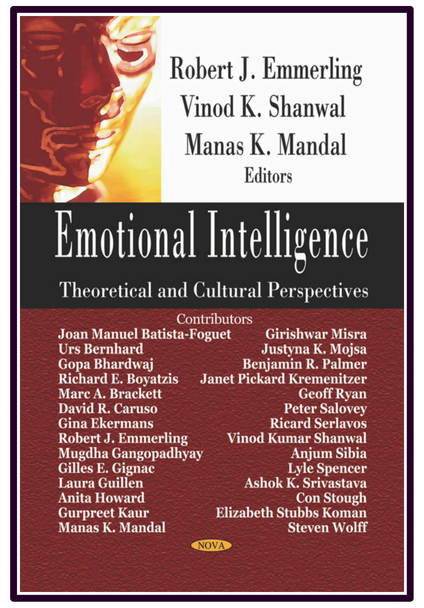





Spencer, Scott. *Quantitative persuasion amid uncertainty*. Forthcoming. Print.
Barclay, Scott et al. *Handbook for Decision Analysis*. Decisions and Designs, Inc., 1977. Print.
zonination. "Perceptions of Probability and Numbers." github. github.com, 18 Aug. 2015. Web. 26 Dec. 2018.

# Cultural cues, emotional display






# Emotional intelligence theoretical and cultural perspectives

*Salovey, co-authors*

Yale president (and double bassist), Salovey introduced the idea of emotional intelligence. This book collects numerous authors' perspectives, examining the **cross-cultural similarity of the concept of emotional intelligence**.

**Emotions are universal, but expression depends on culture**

Basic emotions are **perceived similarly** across the world, but display rules for emotional **expression vary** from culture to culture. That is, the norms pertaining to how—and to what intensity—certain emotions should be expressed within social contexts.

**Emotional Intelligence guides communication**

"Emotional intelligence involves the ability to monitor one's own and **others' feelings and emotions**, to discriminate among them, and to use this information to **guide one's thinking and actions**" including "our **ability to communicate effectively**."

*Caveat: much of the research in this field has originated from Western cultures. This reference includes other perspectives.*

**Higher and lower context cultures**

In **high context cultures**, unlike the US, the internal meaning of a message is usually embedded deeper in the information; **not everything is explicitly written**.



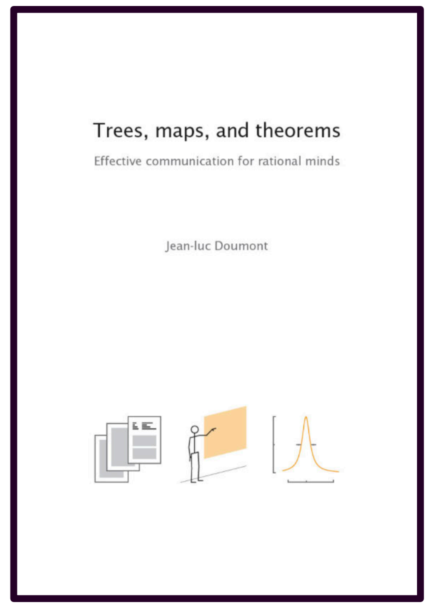
Salovey, Peter et al. *Emotionally Intelligence: Theoretical and Cultural Perspectives*. Ed. Robert J Emmerling, Vinod K Shanwal, and Manas K Mandal. Nova Science Publishers, 2008. Print.

# Questions for discussion

How have your experiences compared with or differed from what we have discussed so far?

If you were the decision maker for approval of a data analytics project, what **type of information** and what **form of communication** would you want to have?

# Group help




# Trees, maps, and theorems

*Doumont*

An engineer from the Louvain School of Engineering and PhD in applied physics from Stanford University, Jean-luc Doumont wrote this book to help engineers, scientists, and managers with business communication.

**Determine purpose(s) of the review**

When reviewing someone else's document, center yourself on the **purpose that was agreed upon**, such as clarity, accuracy, or correctness.

Should this purpose be multiple**, review one aspect at a time, focusing on content first**.

**Reasoning before typos**

Typos are usually more conspicuous than reasoning flaws, but also less important.

**Help, don't judge**

In your comments to the authors, **strive to help, not to judge**.

**Structure the review**

First, provide a **global assessment**, to place further comments in proper perspective. As a rule, point out the **weaknesses**, to prompt improvements, but also the **strengths**, to increase the authors' willingness to revise the document and to learn.



Doumont, Jean-Luc. *Trees, Maps, and Theorems*. Principiæ, 2009. Print.

# Groupwork, feedback from peers

Pair up, share your proposal summary, and **practice giving feedback** on your drafts based on the requirements just discussed.

# Going forward

# For Next Week, Module 6:

## Agenda next week

Next deliverable, **final** 750-word (or less) proposal

The storytelling process

As you review the material, think about how you might storyboard *The Next Rembrandt*.

## The minimum

**Riche, Nathalie Henry et al. Ch. 5 and 9. *Data-Driven Storytelling*. CRC Press, 2018. Print.**

Read for more perspective on the process of data-driven, visual stories.

**Lee, Bongshin et al. "*More Than Telling a Story: Transforming Data Into Visually Shared Stories*." IEEE Computer Graphics and Applications 35.5 (2015): 84–90. Web.**

Read for more perspective on the process of data-driven, visual stories.

**McCloud, Scott. Ch. 6. Show and Tell. *Understanding Comics: the Invisible Art*. Kitchen Sink Press, 1993. Print.**

Read for ideas about how words and pictures combine to tell stories.

**Holtz, Yan, and Conor Healy. *From Data to Viz*. www.data-to-viz.com 2018: web.**

Become familiar with common chart typologies, the types of data structures used for them, and how charts share common underlying attributes.

# Q’s in threes,

**Help on next steps?**

How do you perceive your progress so far, given our timeline? As you get stuck, reach out to both your team (this class) and I for guidance.

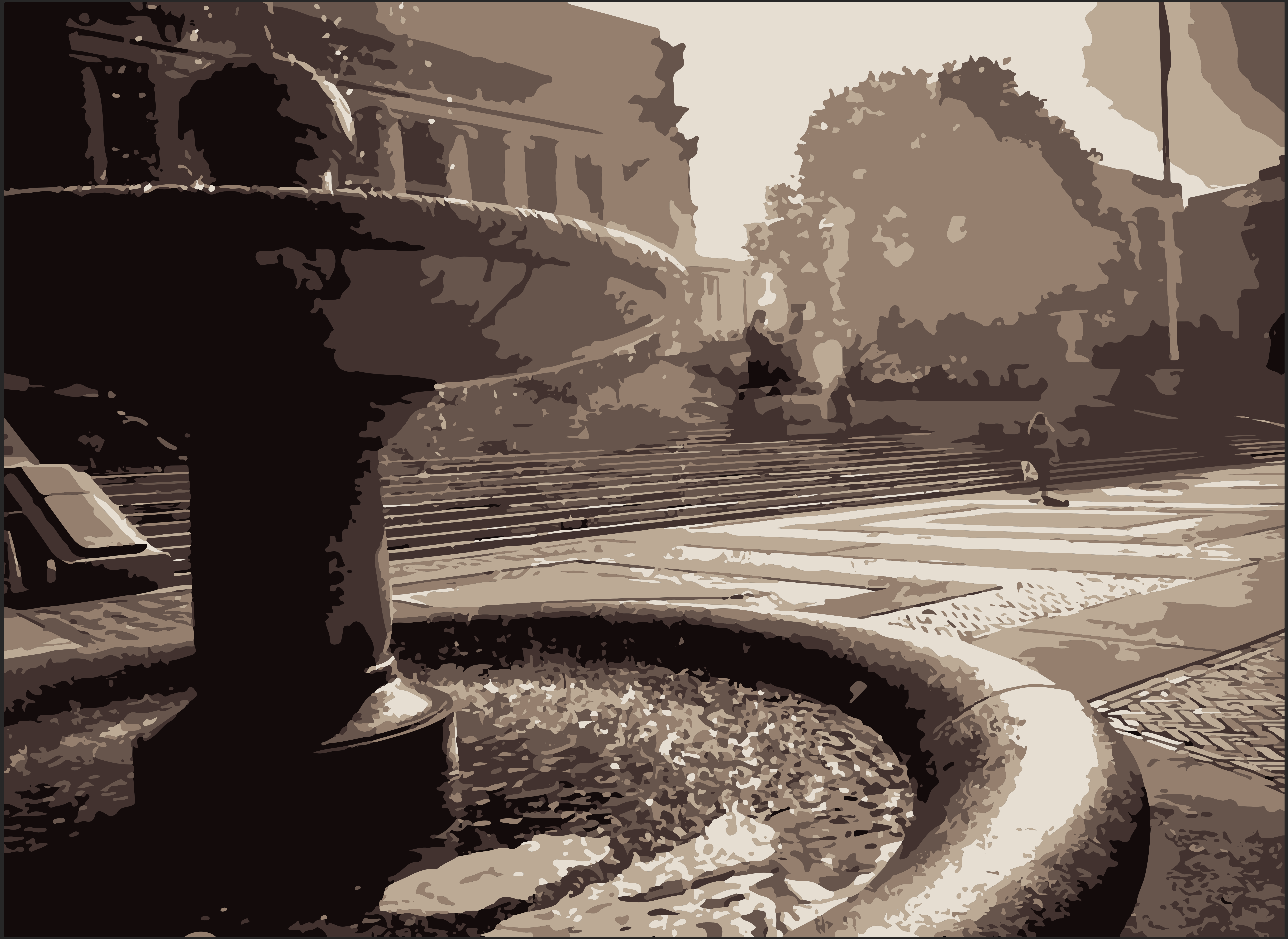
**Too much, too little?**

In the past, I’ve received feedback that the course is quite challenging. As a group, your draft briefs were well done! Tell me where you feel overwhelmed on specific topics, and I’ll try to provide more guidance.

**Of metaphor and story**

Thinking back to last week’s material of metaphor, what metaphors may help you to describe or explain your project and results? Would the metaphor be equally effective with CAO as with CMO?

# See you next week!





Columbia University. *Fountains*. Illustration by Spencer, Scott. 20 June 2019.