



# UNIVERSITY OF MASSACHUSETTS

## DATA SCIENCE AND THE HUMANITIES

---

### Heuristics Worksheet

from Tversky and Kahneman, "Judgment under Uncertainty" 1974

People depend upon **beliefs** to gather, assess, and judge data. Those beliefs are subject to common errors. Certain errors have to do with predicting. Tversky and Kahneman catalogue errors in predicting uncertain data or events.

I. **REPRESENTATIVE HEURISTIC.** You know that  $X$  is part of *Class A*.  $X$  resembles  $Y$ , so you guess that  $Y$  is also part of *Class A*.

1. **Ignores probability.** One falls for stereotyping or typecasting—jock, nerd, Democrat, Republican, humanist, engineer, student, librarian, etc. Note that “when worthless evidence is given, prior probabilities are ignored.”

**ASK:** Is this information relevant to my prediction?

2. **Ignores Sample Size.** Law of large numbers says that the larger the sample size, the closer to class average. Example is the average height of 10 men, 100 men, 1000 men, etc.

**ASK:** What is a significant sample size in this case?

3. **Ignores Probabilities.** There is only an appearance of randomness or order. Global percentages (the average of a college class, for example) are not reproduced locally (the score of an individual student). Chance is not self-correcting (coin tosses do not “even out” to restore an equilibrium!)

**ASK:** What is the sample size? Are the elements actually or only seemingly random?

4. **Ignore Predictability in favor of Description.** The vividness of a description sometimes overwhelms our judgment. For example, a single employee evaluation is not predictive of future behavior.

**ASK:** Am I putting too much stock in this description?

5. **Illusion of Validity.** Like #4, when there is a high degree of “representativeness” (that is, matching a **stereotype**), we have high confidence in predictability. If a job interviewee matches the stereotype of a successful employee, an employer will be

confident of success. An internal consistency in a pattern (**repetition**) increases a feeling of confidence in predictability. For example, the GPA of a student who receives [B, B, B, B] seems easier to predict than the GPA of a student who receives [A, C, A, C]. It isn't!

6. **Misunderstand Regression.** In many areas of life, events and data regress to the mean. A poor performance is often followed by improvement, and a good performance is followed by deterioration. But people do not expect such regression. When we see it, we invent excuses for it. People regress to the mean; arbitrary events (like coin tosses) do not.

**ASK:** Is this human action a regression to the mean?

- II. **AVAILABILITY HEURISTIC.** Something close at hand or recently in memory has a stronger effect on our judgment than something far away or long past. We rely on our own experiences first, then the experiences of friends and family, and then if at all, experiences of strangers. The easier it is to recall, the more worth we give it.

1. **Retrievability.** If you recall it more easily, it appears more numerous or more influential. If it happened to you, it's more salient than if it happened to someone else.

**ASK:** Is my personal experience affecting my assessment of probability?

2. **Effectiveness of a Search Set.** Ease implies relevance. It's easier to search your memory for words beginning with "R" than it is to search for words that have "R" as their third letter. If it's easy to imagine disaster, we think disaster is more likely. If it's hard to imagine disaster, we think success is more likely. It's easier to imagine abstractions than concrete objects or events—events/data described abstractly are easier to remember than those described concretely.

**ASK:** Is it too easy?

3. **Illusory Correlation.** Paranoia, conspiracy theories, etc. A false or corrupted belief that Event X is correlated with Event Y. "Extremely resistant to contradictory data." And long-standing, illusory associations of paired people or events prevents us from seeing relationships that are actually there. For example, they say bad things come in threes.

**ASK:** What is fact, what is a likely probability, and what is speculation?

III. **ADJUSTMENTS AND ANCHORING.** Tversky and Kahneman write that “different starting points yield different estimates, which are biased toward the initial value.”

1. **Insufficient Adjustment.** We don't adjust our guesses based on all the data. So people who are asked to guess the product of  $[1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8]$  will guess low because the series is anchored on “1” while those asked to guess  $[8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1]$  will guess higher because the series is anchored on “8”.

ASK: What if I reorder the data?

2. **Conjunctive and Disjunctive Events.** A disjunctive event has lots of moving parts and is complex. For example, the development of a new product. People tend to “underestimate the probabilities of failure in a complex system.”

ASK: In a chain of events, what is the probability of each event on its own?

3. **Probability Distributions.** People tend to express greater confidence than is justified when asked about probability distributions.

ASK: Would my assessment change if I used my best guess as an anchor?