Week 2: Heuristics and Biases

Dr Christian Engels
ce50@st-andrews.ac.uk

FI5614 Behavioural Finance

Department of Finance
University of St Andrews Business School

University of St Andrews

Outline

Heuristics and Biases Approach

Representativeness Heuristic

- Law of Small Numbers
- Misunderstanding of Randomness
- Base-Rate Neglect
- Conjunction Fallacy
- Regression to the Mean

Anchoring Heuristic

Availability Heuristic

Review Quiz

Interactive quiz on Vevox

Quick review of Week 1 material

Please open Vevox and enter the session code

Today's Lecture

Are people good intuitive statisticians?

The counties with the lowest rates of kidney cancer in the US are small, rural, and Republican-leaning.

The counties with the highest rates are also small, rural, and Republican-leaning.

How can the same factor explain opposite outcomes?

Required Readings:

- Kahneman, Part II
- Ackert and Deaves, chap. 5

Heuristics and Biases Approach

Two Systems Produce Predictable Errors

Our behaviour is not consistent with fully rational Bayesian agents—but it is not random either.

System 1 — Fast

- Automatic, effortless
- Associative
- Prone to systematic errors

System 2 — Slow

- Deliberate, effortful
- Rule-based
- Monitors System 1 (lazily)

People's intuitive probability judgements are *largely functional* for speedy decisions—but they fail in predictable ways.

Simple Heuristics Create Predictable Biases

Judgement heuristic — a simple, efficient rule of thumb used to form judgements and make decisions (“fast and frugal”).

Bias — a systematic deviation from rational behaviour and the predictions of classical probability theory.

Today we study three foundational heuristics:

1. **Representativeness** — judging by similarity
2. **Anchoring** — relying on initial values
3. **Availability** — judging by ease of recall

Tversky & Kahneman (1974), “Judgement under Uncertainty: Heuristics and Biases,” *Science*.

Representativeness Heuristic

We Judge Probability by Similarity to Prototypes

Representativeness — the tendency to judge the probability of an event by how well observable data *resemble* a known category or population.

We compare the likeness of a sample—even a very small sample—with the features we expect the population to have.

Law of Small Numbers

Misunderstanding of Randomness

Base-Rate Neglect

Conjunction Fallacy

Regression to the Mean

A Puzzle: Kidney Cancer in America

A study of kidney cancer incidence across 3 141 US counties:

- The counties with the *lowest* incidence are mostly **rural, sparsely populated**, and in traditionally Republican states.

A Puzzle: Kidney Cancer in America

A study of kidney cancer incidence across 3 141 US counties:

- The counties with the *lowest* incidence are mostly **rural, sparsely populated**, and in traditionally Republican states.
- Intuitive story: clean air, clean water, fresh food, active lifestyle.

A Puzzle: Kidney Cancer in America

A study of kidney cancer incidence across 3 141 US counties:

- The counties with the *lowest* incidence are mostly **rural, sparsely populated**, and in traditionally Republican states.
- Intuitive story: clean air, clean water, fresh food, active lifestyle.
- Now consider: the counties with the *highest* incidence are **also rural, sparsely populated**, and Republican-leaning.

A Puzzle: Kidney Cancer in America

A study of kidney cancer incidence across 3 141 US counties:

- The counties with the *lowest* incidence are mostly **rural, sparsely populated**, and in traditionally Republican states.
- Intuitive story: clean air, clean water, fresh food, active lifestyle.
- Now consider: the counties with the *highest* incidence are **also rural, sparsely populated**, and Republican-leaning.
- Intuitive story: poverty, high-fat diet, alcohol, tobacco, poor medical access.

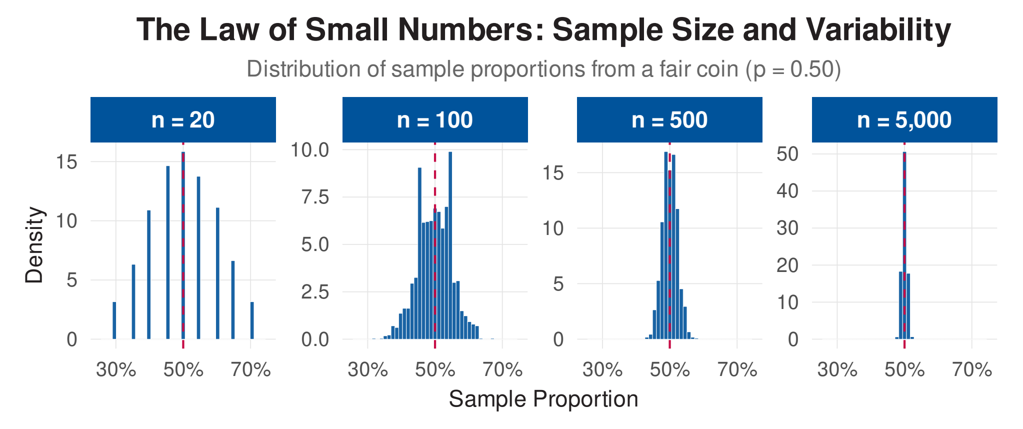
A Puzzle: Kidney Cancer in America

A study of kidney cancer incidence across 3 141 US counties:

- The counties with the *lowest* incidence are mostly **rural, sparsely populated**, and in traditionally Republican states.
- Intuitive story: clean air, clean water, fresh food, active lifestyle.
- Now consider: the counties with the *highest* incidence are **also rural, sparsely populated**, and Republican-leaning.
- Intuitive story: poverty, high-fat diet, alcohol, tobacco, poor medical access.
- Something is wrong. The rural lifestyle cannot explain *both* extremes.

Small Samples Show Extreme Variance, Not True Differences

The key factor is not lifestyle—it is sample size: $\mathrm{Var}(\bar{X}) = \sigma^2/n$. Small counties land at *both* extremes.



We Expect Small Samples to Mirror Populations—They Don't

Law of Large Numbers — the sample mean converges in probability to the population mean as $n \to \infty$.

"Law" of Small Numbers — people *believe* that even small samples should closely mirror the population. They expect $\bar{X} \approx \mu$ even when n is tiny.

- System 1 automatically constructs a causal story for any extreme outcome
- System 1 is inept when faced with "merely statistical" facts

Small Schools Aren't Better—They're Just More Variable

A study of the most successful US schools found they tend to be **small**. This prompted large investments in creating small schools.

- Causal story: small schools give students more individual attention and encouragement.

Wainer & Zwerling (2006), "Evidence That Smaller Schools Do Not Improve Student Achievement."

Small Schools Aren't Better—They're Just More Variable

A study of the most successful US schools found they tend to be **small**. This prompted large investments in creating small schools.

- Causal story: small schools give students more individual attention and encouragement.
- But the *worst-performing* schools are also disproportionately small.

Wainer & Zwerling (2006), "Evidence That Smaller Schools Do Not Improve Student Achievement."

Small Schools Aren't Better—They're Just More Variable

A study of the most successful US schools found they tend to be **small**. This prompted large investments in creating small schools.

- Causal story: small schools give students more individual attention and encouragement.
- But the *worst-performing* schools are also disproportionately small.
- The pattern is the same as the cancer counties: small samples $\Rightarrow$ large variance $\Rightarrow$ more extreme outcomes in *both* directions.

Wainer & Zwerling (2006), "Evidence That Smaller Schools Do Not Improve Student Achievement."

We Reject Genuinely Random Sequences as “Not Random”

Consider six consecutive births at a hospital (B = boy, G = girl, $p = 0.5$):

1. BBBGGG
2. GGGGGG
3. BGBBGB

- Which sequence is most likely?

We Reject Genuinely Random Sequences as "Not Random"

Consider six consecutive births at a hospital (B = boy, G = girl, $p = 0.5$):

1. BBBGGG
2. GGGGGG
3. BGBBGB

- Which sequence is most likely?
- Most people choose 3. But all three are equally likely $\left(p = \frac{1}{64}\right)$.

We Reject Genuinely Random Sequences as “Not Random”

Consider six consecutive births at a hospital (B = boy, G = girl, $p = 0.5$):

1. BBBGGG
2. GGGGGG
3. BGBBGB

- Which sequence is most likely?
- Most people choose 3. But all three are equally likely $\left(p = \frac{1}{64}\right)$.
- We reject patterns (**GGGGGG**) and orderly alternation (**BBBGGG**) as “not representative” of a random process.

After a Streak, We Expect Reversal—But Coins Have No Memory

Gambler's fallacy — after observing a short run of the same outcome, people believe the *opposite* outcome becomes more likely.

The intuition: a small sample "should" look like the population (≈50/50), so after three heads in a row, tails is "due."

In finance: the *disposition effect*—investors hold losers expecting a reversal.

Odean (1998), "Are Investors Reluctant to Realize Their Losses?"

We See Momentum in Random Streaks

Hot hand fallacy — after observing a streak of the same outcome, people believe the process is *positively correlated*, that streaks are self-reinforcing.

- Basketball players who hit several shots are said to have a “hot hand”
- Fund managers with a few years of outperformance are rated “excellent” (cf. S&P Business Week awards)

Both fallacies stem from the same root: we expect small samples to mirror the population, and over-interpret deviations.

Gilovich, Vallone & Tversky (1985), “The Hot Hand in Basketball.”

The Cab Problem

A cab was involved in a hit-and-run accident at night. Two cab companies operate in the city: **Green** and **Blue**.

- 85% of cabs in the city are Green; 15% are Blue.
- A witness identified the cab as Blue. The court tested the witness under similar conditions: she correctly identifies each colour **80%** of the time.

The Cab Problem

A cab was involved in a hit-and-run accident at night. Two cab companies operate in the city: **Green** and **Blue**.

- 85% of cabs in the city are Green; 15% are Blue.
- A witness identified the cab as Blue. The court tested the witness under similar conditions: she correctly identifies each colour **80%** of the time.

What is the probability the cab was Blue?

Most People Say 80%—Ignoring the Base Rate

The most common answer is $p = 80\%$.

This equals the witness's accuracy—it ignores the base rate, i.e. the prior probability that the cab is Blue (only 15%).

The correct answer requires Bayes' rule.

Bayes' Rule Updates Prior Beliefs with New Evidence

From the definition of conditional probability:

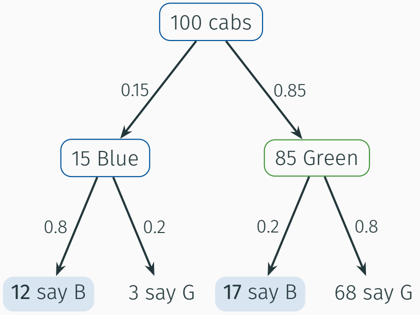
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Since $P(A \cap B) = P(B \mid A)\, P(A)$, we obtain:

Bayes' Rule

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c)}$$

Bayes' rule tells us how to *update* a prior belief when new evidence arrives.

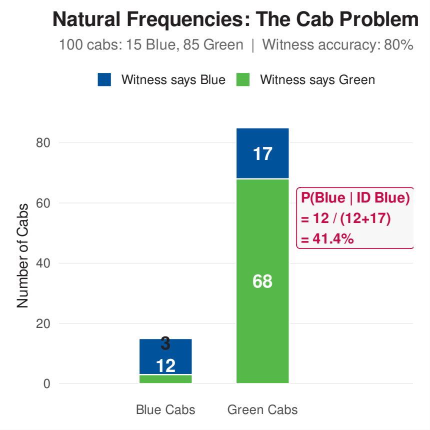


The Correct Answer Is 41%—Far Below 80%

Let B = cab is Blue, W = witness says Blue.

- $P(W \mid B) = 0.80 \quad \text{(correct ID)}$
- $P(W \mid G) = 0.20 \quad \text{(false ID)}$
- $P(B) = 0.15 \quad \text{(base rate)}$

$$P(B \mid W) = \frac{0.80 \times 0.15}{0.80 \times 0.15 + 0.20 \times 0.85} = \frac{0.12}{0.29} = 41.4\%$$



We Ignore Prior Probabilities When Judging Events

Base-rate neglect — the tendency to ignore or heavily discount prior information about the frequency of an event when judging its probability.

People confuse $P(A \mid B)$ with $P(B \mid A)$: the witness is 80% accurate, so the cab "must be" 80% likely to be Blue.

The prior ($P(B) = 0.15$) pulls the true answer far below 80%.

Causal Stories Make Base Rates Matter—Pure Statistics Don't

Rephrase the problem:

"Green cabs are involved in 85% of accidents; Blue cabs in 15%."

- With this framing, people give much more accurate answers.

Causal Stories Make Base Rates Matter—Pure Statistics Don't

Rephrase the problem:

"Green cabs are involved in 85% of accidents; Blue cabs in 15%."

- With this framing, people give much more accurate answers.
- Why? The new phrasing supplies a *causal* story: Green cabs *cause* more accidents.

Causal Stories Make Base Rates Matter—Pure Statistics Don't

Rephrase the problem:

"Green cabs are involved in 85% of accidents; Blue cabs in 15%."

- With this framing, people give much more accurate answers.
- Why? The new phrasing supplies a *causal* story: Green cabs *cause* more accidents.
- **Statistical** base rates are ignored; **causal** base rates are not.

Causal Stories Make Base Rates Matter—Pure Statistics Don't

Rephrase the problem:

"Green cabs are involved in 85% of accidents; Blue cabs in 15%."

- With this framing, people give much more accurate answers.
- Why? The new phrasing supplies a *causal* story: Green cabs *cause* more accidents.
- **Statistical** base rates are ignored; **causal** base rates are not.
- System 1 naturally processes causal narratives but struggles with abstract frequencies.

The Linda Problem

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

Which is more likely?

1. Linda is a bank teller.
2. Linda is a bank teller **and** is active in the feminist movement.

- 85–90% of respondents choose option 2.

The Linda Problem

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

Which is more likely?

1. Linda is a bank teller.
2. Linda is a bank teller **and** is active in the feminist movement.

- 85–90% of respondents choose option 2.
- The description of Linda is more *representative* of a feminist bank teller—but option 2 is a **subset** of option 1.

Conjunction fallacy — the intersection of two events is mistakenly judged more probable than one of the events alone.

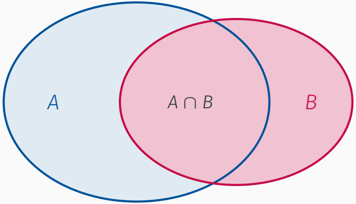
A basic axiom of probability:

$$P(A \cap B) \;\leq\; P(A)$$

Yet respondents ranked: $P(B) > P(A \cap B) > P(A)$

where A = bank teller, B = feminist.

Tversky & Kahneman (1983).



$P(A \cap B) \leq P(A)$

The Ski Jump Puzzle

Unusually good performance is likely to be followed by worse, and unusually bad performance is likely to be followed by better. Two stories:

- Commentator A: “Norway had a great first jump; he’ll be tense, trying to protect his lead—he’ll probably do worse.”

The Ski Jump Puzzle

Unusually good performance is likely to be followed by worse, and unusually bad performance is likely to be followed by better. Two stories:

- Commentator A: “Norway had a great first jump; he’ll be tense, trying to protect his lead—he’ll probably do worse.”
- Commentator B: “Sweden had a bad first jump and now has nothing to lose: she’ll be relaxed and do better.”

The Ski Jump Puzzle

Unusually good performance is likely to be followed by worse, and unusually bad performance is likely to be followed by better. Two stories:

- Commentator A: “Norway had a great first jump; he’ll be tense, trying to protect his lead—he’ll probably do worse.”
- Commentator B: “Sweden had a bad first jump and now has nothing to lose: she’ll be relaxed and do better.”
- Both stories sound plausible. Both are wrong—or at least, they miss the real reason.

The Ski Jump Puzzle

Unusually good performance is likely to be followed by worse, and unusually bad performance is likely to be followed by better. Two stories:

- Commentator A: “Norway had a great first jump; he’ll be tense, trying to protect his lead—he’ll probably do worse.”
- Commentator B: “Sweden had a bad first jump and now has nothing to lose: she’ll be relaxed and do better.”
- Both stories sound plausible. Both are wrong—or at least, they miss the real reason.
- The truth is simpler:

$$\text{Performance} = \text{Talent} + \text{Luck}$$

The Ski Jump Puzzle

Unusually good performance is likely to be followed by worse, and unusually bad performance is likely to be followed by better. Two stories:

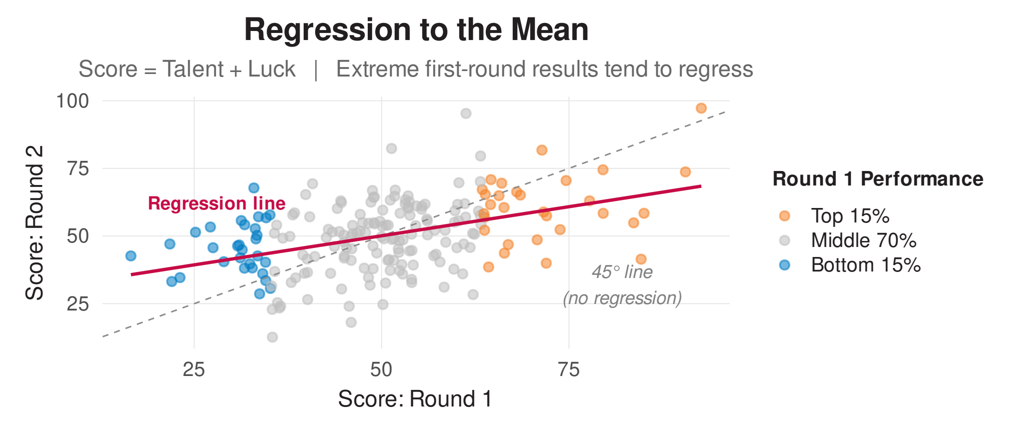
- Commentator A: "Norway had a great first jump; he'll be tense, trying to protect his lead—he'll probably do worse."
- Commentator B: "Sweden had a bad first jump and now has nothing to lose: she'll be relaxed and do better."
- Both stories sound plausible. Both are wrong—or at least, they miss the real reason.
- The truth is simpler:

$$\text{Performance} = \text{Talent} + \text{Luck}$$

- Extreme (bad) luck in Round 1 is unlikely to repeat in Round 2.

Extreme Performances Revert Toward the Average

Regression to the mean — if a variable is extreme on its first measurement, it will tend to be closer to the average on its second.



Intuitive Predictions Ignore Regression to the Mean

Julie read fluently at age four. She is now in college. What is her GPA?

- **Intuitive prediction** (substitution + intensity matching): early reading is top 5%, so GPA should also be top 5% $\Rightarrow \widehat{\text{GPA}} = 3.80$.

Intuitive Predictions Ignore Regression to the Mean

Julie read fluently at age four. She is now in college. What is her GPA?

- **Intuitive prediction** (substitution + intensity matching): early reading is top 5%, so GPA should also be top 5% $\Rightarrow \widehat{\text{GPA}} = 3.80$.
- **Corrected prediction** — account for imperfect correlation ($\rho = 0.30$):

$$\widehat{\text{GPA}}_{\text{Julie}} = \underbrace{3.40}_{\bar{X}} + 0.30 \times (3.80 - 3.40) = 3.52$$

Intuitive Predictions Ignore Regression to the Mean

Julie read fluently at age four. She is now in college. What is her GPA?

- **Intuitive prediction** (substitution + intensity matching): early reading is top 5%, so GPA should also be top 5% $\Rightarrow \widehat{\text{GPA}} = 3.80$.
- **Corrected prediction** — account for imperfect correlation ($\rho = 0.30$):

$$\widehat{\text{GPA}}_{\text{Julie}} = \underbrace{3.40}_{\bar{X}} + 0.30 \times (3.80 - 3.40) = 3.52$$

- The weaker the correlation, the more the prediction should regress toward the average.

Award-Winning CEOs Underperform After Recognition

Malmendier & Tate (2009) study CEOs who win prestigious awards (Best Manager, Business Week; CEO of the Year, Financial World; Best Performing CEO, Forbes).

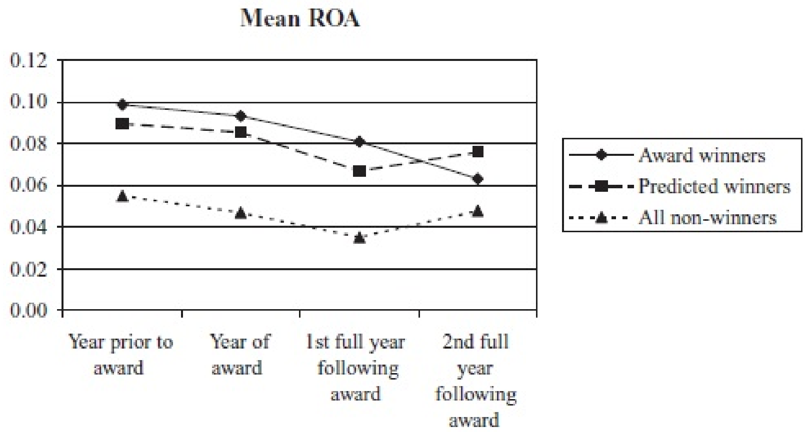
After the award, winners:

- **Underperform** relative to prior results and a matched sample of non-winning CEOs
- Extract **higher compensation**
- Spend more time on outside activities (boards, books, golf)

Effects are strongest in firms with *weak* corporate governance.

⇒ Is this regression to the mean, CEO distraction, or both?

Superstar CEOs: Performance After Awards



Mean ROA for award winners (Malmendier & Tate, 2009).

Representativeness in Finance: Key Applications

Representativeness distorts financial decisions across many domains:

Phenomenon	Reference
Company/fund name changes	Cooper, Dimitrov & Rau (2001); Cooper, Gulen & Rau (2005)
Good stocks vs good companies	Shefrin & Statman (1995); Shefrin (2007)
Value vs growth stock returns	Lakonishok, Shleifer & Vishny (1994)
Analyst recommendations	Mokoaleli-Mokoteli, Taffler & Agarwal (2009)
Fund manager selection	Goyal & Wahal (2008)
Chartism / technical analysis	Park & Irwin (2007)

Anchoring Heuristic

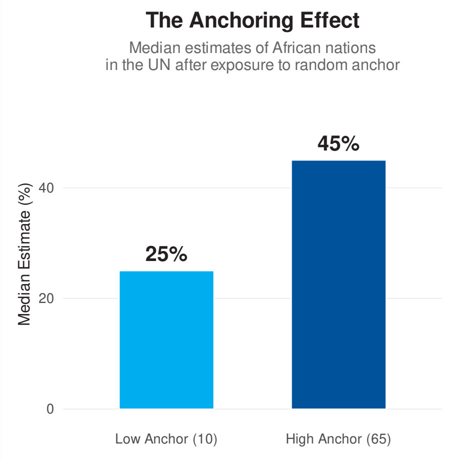
The Wheel of Fortune Experiment

A wheel of fortune—rigged to land on either **10** or **65**—was spun in front of participants.

They were then asked:

1. Is the percentage of African nations in the UN *larger or smaller* than the number on the wheel?
2. What is your *best guess* of the actual percentage?

The wheel number is *obviously* irrelevant—yet it profoundly shifted estimates.



Initial Values Pull All Subsequent Estimates

Anchoring effect — the tendency to rely too heavily on the first piece of information encountered (the “anchor”) when making numerical estimates.

- People consider a particular value before estimating an unknown quantity

Initial Values Pull All Subsequent Estimates

Anchoring effect — the tendency to rely too heavily on the first piece of information encountered (the "anchor") when making numerical estimates.

- People consider a particular value before estimating an unknown quantity
- Estimates stay close to the anchor, even when it is *transparently irrelevant*

Initial Values Pull All Subsequent Estimates

Anchoring effect — the tendency to rely too heavily on the first piece of information encountered (the "anchor") when making numerical estimates.

- People consider a particular value before estimating an unknown quantity
- Estimates stay close to the anchor, even when it is *transparently irrelevant*
- Participants exposed to random anchors confidently deny that the number influenced them

Anchoring Mechanics: Two Pathways

1. Insufficient Adjustment
(System 2)

- Start from the anchor
- Adjust up or down
- Stop too early—when you are “no longer sure” you should move further

Try it: What is the boiling point of water at the summit of Mount Everest? (Normally 100°C at sea-level; goes down with height.)

2. Priming Effect
(System 1)

- The anchor activates compatible memories
- System 1 builds a coherent picture around the anchor
- No conscious adjustment occurs

Try it: Was Gandhi older or younger than 144 at death? How old was he?

Anchors Shift Estimates by 40–55%

Is the tallest redwood taller or shorter than x_0 feet? What is your best guess?

Results: $x_T^H = 844$ when $x_0^H = 1200$; $\quad x_T^L = 282$ when $x_0^L = 180$.

Anchoring Index

$$AI = \frac{x_T^H - x_T^L}{x_0^H - x_0^L} = \frac{844 - 282}{1200 - 180} = 55\%$$

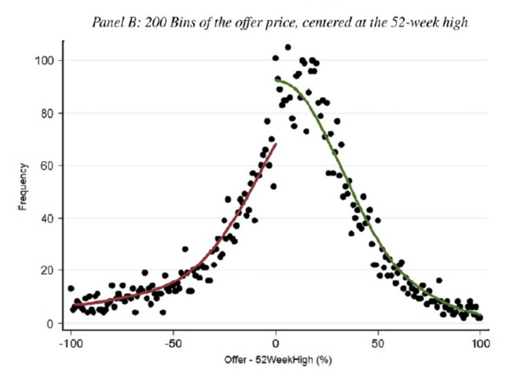
House prices: Real-estate agents assessed a house after seeing either a high or low asking price. Anchoring index: 41%. They insisted the asking price did *not* influence their judgement.

The 52-Week High Anchors M&A Offer Prices

The 52-week high is widely reported and psychologically salient—yet it should be largely irrelevant for valuation.

Baker, Pan & Wurgler (2012) find:

- The 52-week high exerts a **positive effect** on bid prices in M&A
- Acceptance probability **jumps discontinuously** when the offer exceeds the 52-week high



Baker, Pan & Wurgler (2012), JFE.

Availability Heuristic

A Quick Exercise

In four pages of a novel (≈2 000 words), which pattern appears more often?

`_ _ _ _ _ n _` vs `_ _ _ _ i n g`

- Most people say `-ing` words are more frequent.

A Quick Exercise

In four pages of a novel (≈2 000 words), which pattern appears more often?

_ _ _ _ _ n _ vs _ _ _ _ i n g

- Most people say `-ing` words are more frequent.
- This is false: every `-ing` word is also a _ _ _ _ _n_ word. The first set is strictly larger.

A Quick Exercise

In four pages of a novel (≈2 000 words), which pattern appears more often?

`_ _ _ _ _ n _`   vs   `_ _ _ _ i n g`

- Most people say `-ing` words are more frequent.
- This is false: every `-ing` word is also a `_ _ _ _ _n_` word. The first set is strictly larger.
- `-ing` words come to mind more easily—they are more *available*.

Ease of Recall Distorts Probability Judgments

Availability heuristic — the tendency to judge the frequency or probability of an event by the ease with which examples come to mind.

- We judge frequency by how easily instances are *retrieved from memory*

Ease of Recall Distorts Probability Judgments

Availability heuristic — the tendency to judge the frequency or probability of an event by the ease with which examples come to mind.

- We judge frequency by how easily instances are *retrieved from memory*
- *Salience* governs ease of retrieval

Ease of Recall Distorts Probability Judgments

Availability heuristic — the tendency to judge the frequency or probability of an event by the ease with which examples come to mind.

- We judge frequency by how easily instances are *retrieved from memory*
- *Salience* governs ease of retrieval
- *Question substitution*: "How frequent is X?" becomes "How easily can I think of X?"

Ease of Recall Distorts Probability Judgments

Availability heuristic — the tendency to judge the frequency or probability of an event by the ease with which examples come to mind.

- We judge frequency by how easily instances are *retrieved from memory*
- *Salience* governs ease of retrieval
- *Question substitution*: "How frequent is *X*?" becomes "How easily can I think of *X*?"
- This is often a sensible shortcut—but it systematically distorts probability judgements

What Makes Events More Available?

Salience

Events that attract attention are easily retrieved.

Example: We overestimate the frequency of celebrity divorces because they dominate the news.

Drama and Vividness

A single vivid experience outweighs dry statistics.

Example: A plane crash temporarily changes your feelings about flying—even though the base rate is unchanged.

Personal Experience

First-hand events are more available than events that happen to others.

Example: A judicial error you experience undermines your faith in the system more than reading about one.

Recency

Recent events dominate over distant ones.

Example: Insurance purchases spike after an earthquake, then gradually decline.

Vivid Events Are Systematically Overestimated

Availability bias systematically distorts our perception of risk:

- Strokes cause **twice** as many deaths as all accidents combined—yet 80% of respondents judge accidental death more likely.
- Tornadoes are perceived as more lethal than asthma, although asthma kills **20 times** more people.
- After each earthquake, insurance purchases spike—then gradually decline as the memory fades.

Dramatic, newsworthy events are systematically overestimated; common, undramatic causes are systematically underestimated.

$\Rightarrow$ Media coverage and the *affect heuristic* amplify these distortions.

Conclusions

Summary: Three Heuristics, Many Biases

Heuristic	Core Mechanism	Key Manifestations
Representativeness	Judge probability by similarity to a prototype	Law of small numbers, gambler's fallacy, base-rate neglect, conjunction fallacy, regression neglect
Anchoring	Estimates stay close to an initial reference point	Insufficient adjustment (System 2), priming (System 1), 52-week high in M&A
Availability	Judge frequency by ease of mental retrieval	Salience, vividness, recency; distorted risk perception

Key Takeaways

1. "Statistical thinking" is **hard**. When interpreting outcomes affected by chance, we make systematic errors.
2. Heuristics are not irrational—they are **efficient shortcuts** that work well most of the time but fail predictably in specific contexts.
3. These failures have direct consequences for financial decision-making: asset pricing, portfolio choice, corporate governance, and market efficiency.
4. The antidote is not to abandon intuition but to **know when to distrust it**—and to apply formal tools (Bayes' rule, regression analysis) where they matter.

Closing Quiz

Interactive quiz on Vevox

Test your understanding of:
Representativeness, anchoring, and availability heuristics

Please open Vevox and enter the session code