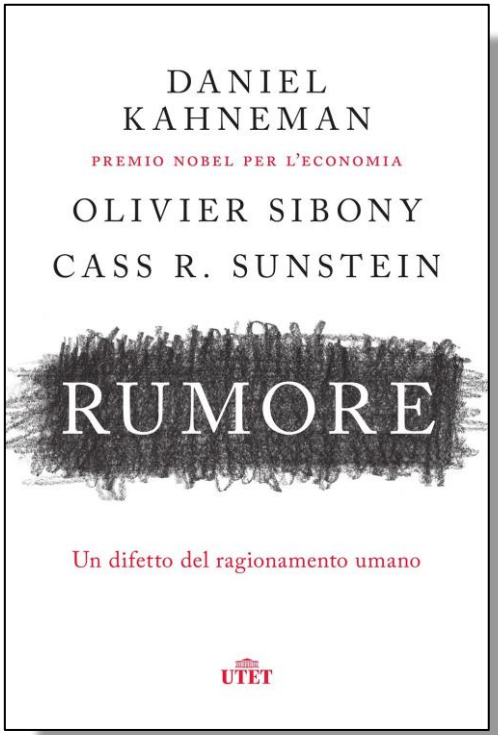
Bias and Noise: Measuring and Managing Two Flaws of Judgment in Organizations

Olivier Sibony

Rome / Online

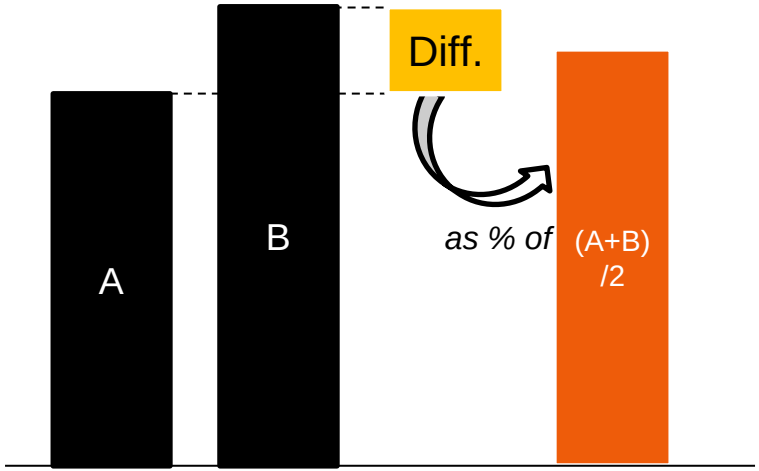
April 2022



A Well-run Insurance Company

If you randomly selected two qualified underwriters or claims adjusters, how different would you expect their estimates for the same case to be?

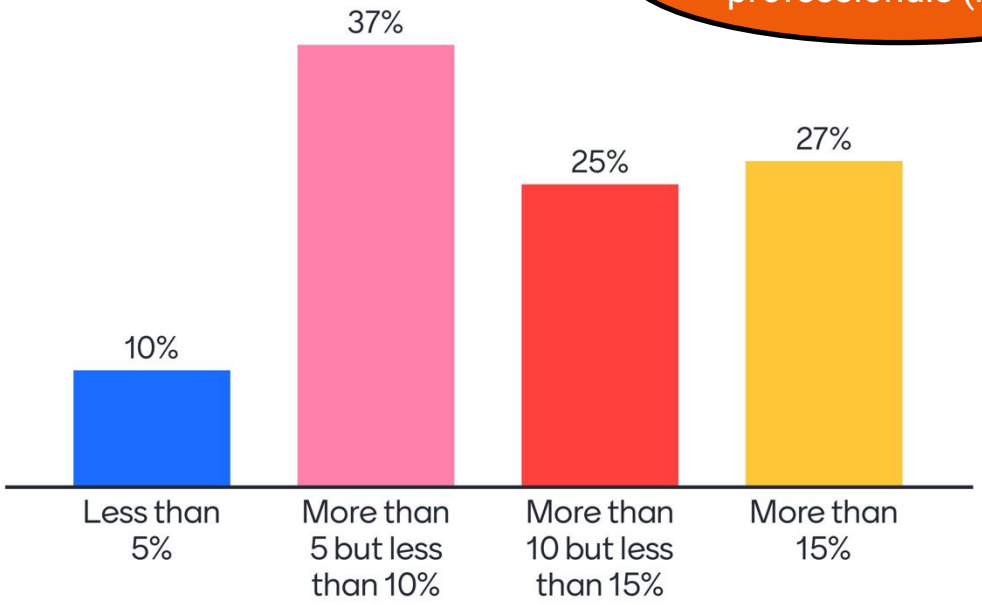
Express your answer as the difference between the two estimates in percentage of their average.



Average difference between two insurance estimates, as a percentage of their average

Mentimeter

Finance professionals (n=51)

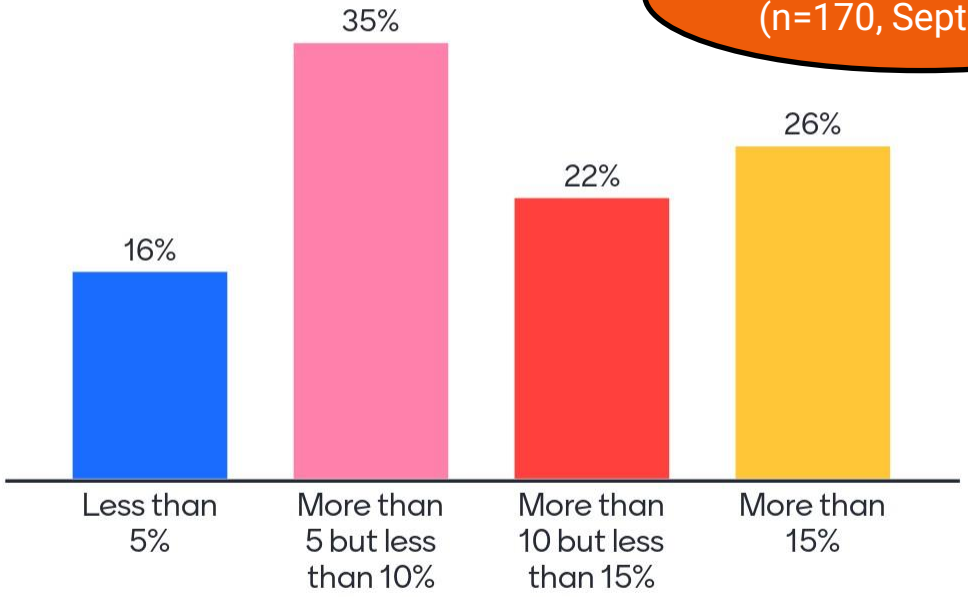


51

How much difference do you expect there will be between two insurance estimates, as a percentage of their average?

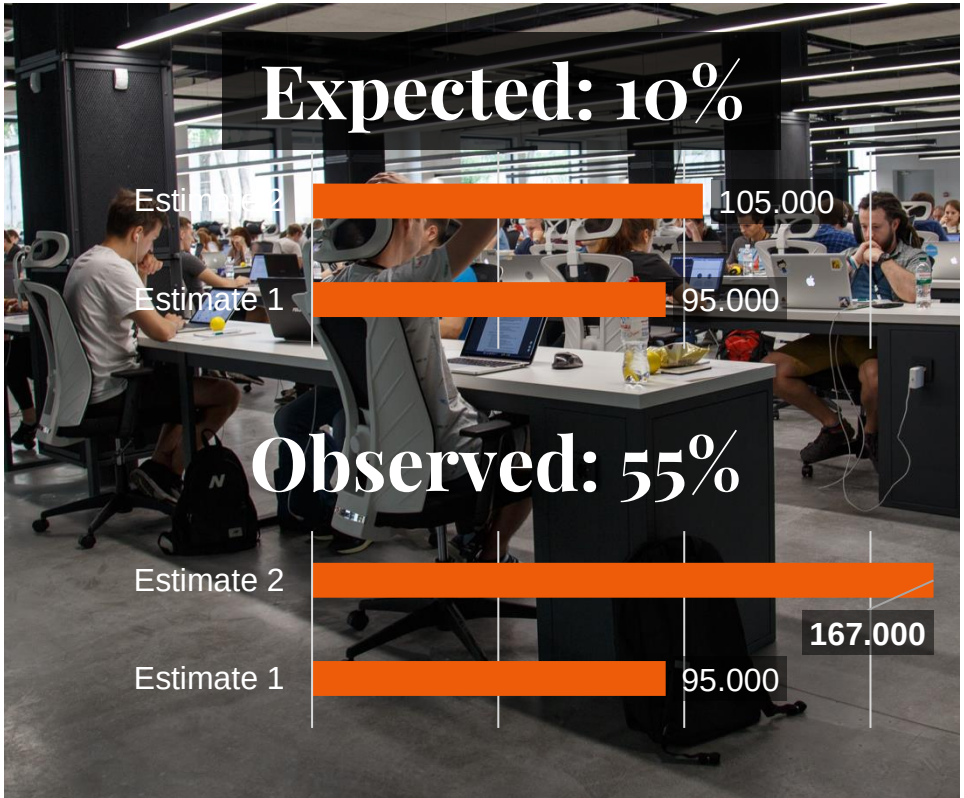
Mentimeter

General public
(n=170, Sept. 2022)



170

Insurance Underwriting: A Noise Audit



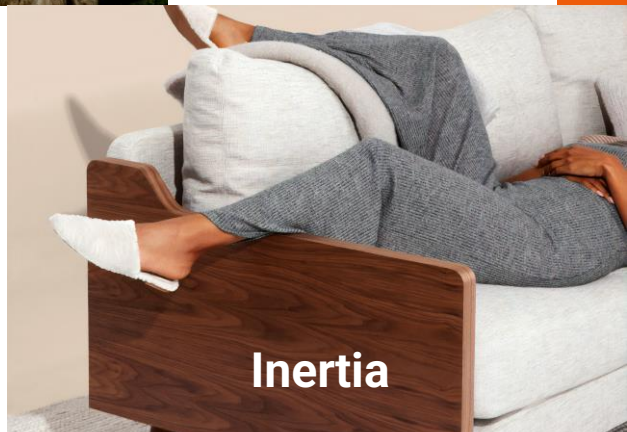
1

Wherever there is judgment, there is 

, and often bias, too.

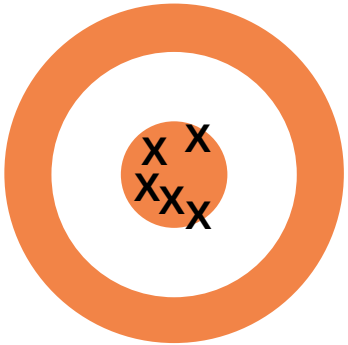
Examples of Bias

Time-inconsistent preferences



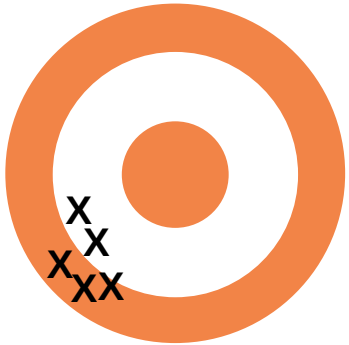
Noise ≠ Bias

A



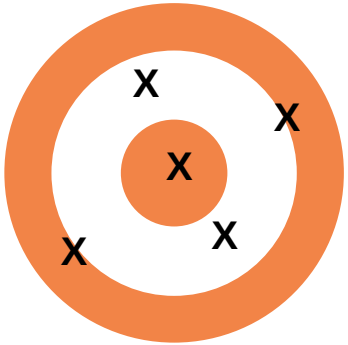
ACCURATE

B



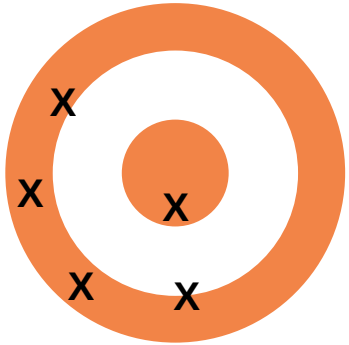
BIASED

C



NOISY

D



NOISY AND BIASED

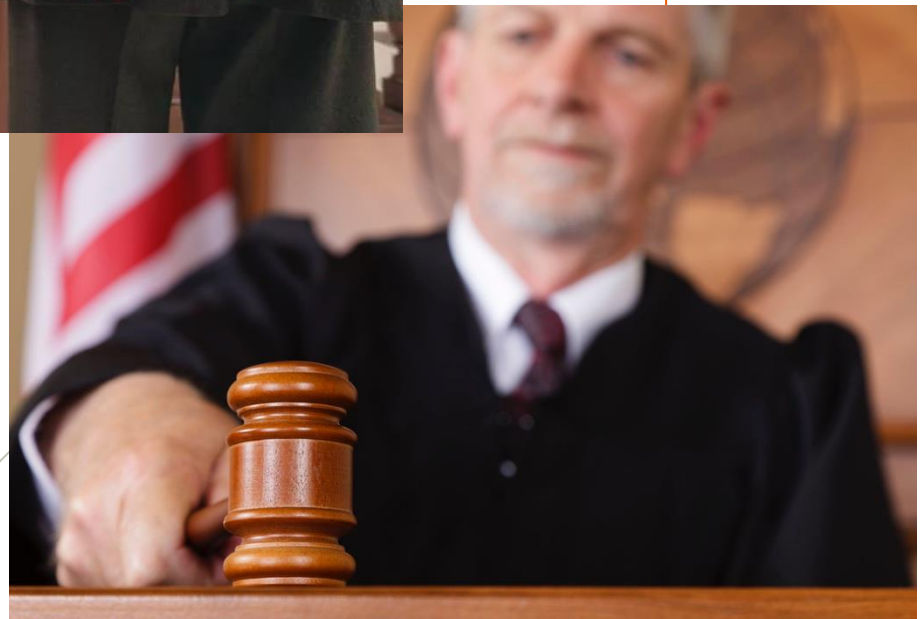
Noise in Criminal Justice

Average sentence:

7.0 years

Mean difference between judges, in the same case:

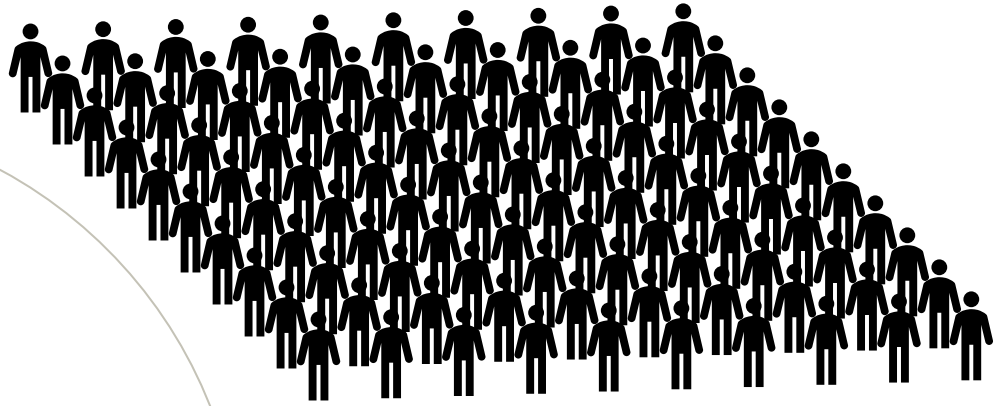
3.8 years



Three Components of Noise

Level Noise

On average, some judges are more severe than others.



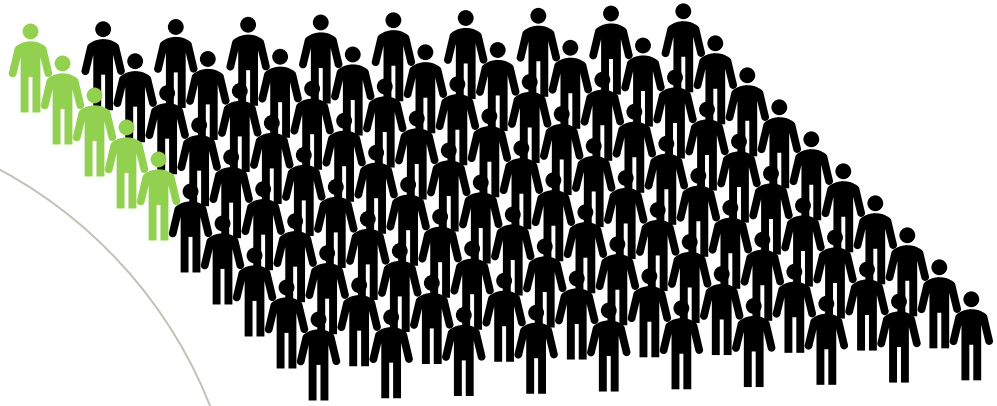
Asylum Decisions

Percent admitted by two judges in the same courthouse

Asylum Decisions

Percent admitted by two judges in the same courthouse

Judge 1: 88%



Asylum Decisions

Percent admitted by two judges in the same courthouse

Judge 1: 88%

Judge 2: 5%

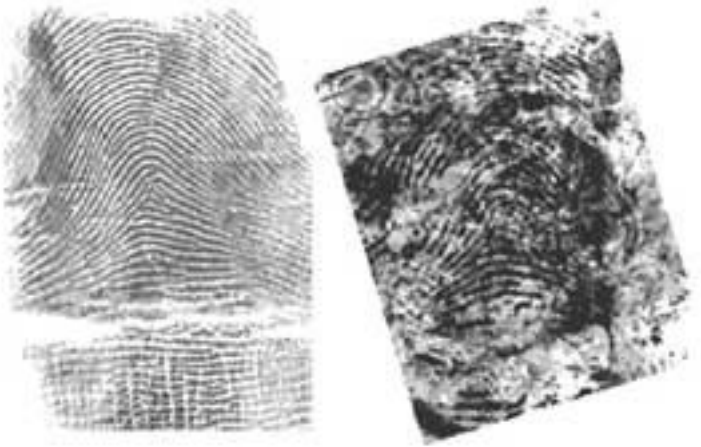
Three Components of Noise

Level Noise

On average, some judges are more severe than others.

Occasion Noise

If faced with the same case twice, a judge will not judge it identically.



Forensic Science

Fingerprint examiners disagree (in 10% of cases, even with themselves).

Faulty forensic science is involved in **45%** of wrongful convictions.

Judicial decisions are more severe

- Before lunch
- After your football team lost
- After several favorable decisions
- On hot days

Medical prescriptions vary with

- Time of day
- Fatigue (number of appts.)
- Day of week
- Salient numbers, e.g., patient age 79 vs 80

Singular decisions

Three Components of Noise

Level Noise

On average, some judges are more severe than others.

Occasion Noise

If faced with the same case twice, a judge will not judge it identically.

Pattern Noise

Each judge has different preferences and views on each case.

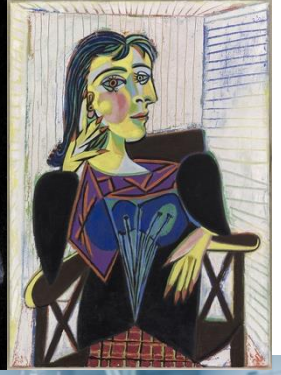
We are all different. So are our judgments.

(Especially if we express our individuality.)

2

Wherever accuracy matters, is costly.

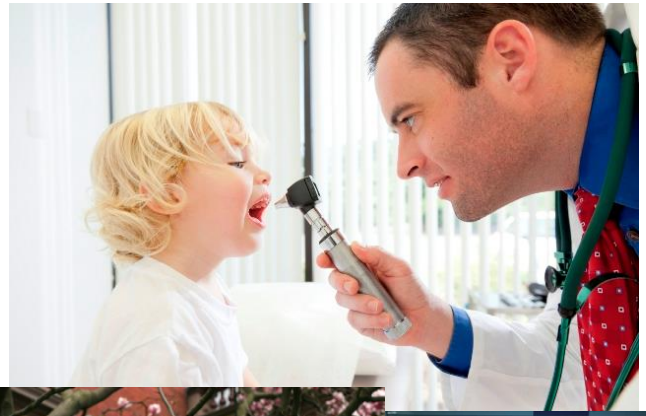
(And we tend to neglect it.)



Sometimes variation is beneficial...

In some situations, disagreement is unproblematic, even welcome:

- Tastes
- Markets and competitions
- Creative endeavors



… but not when professionals make *judgments*,

Defined by:

- *Uncertainty*
- Belief in a *best possible answer*
- Expectation of *bounded disagreement*.

Mean Squared

$$\text{Error} = \text{Bias}^2 + \text{Noise}^2$$

This is *the Error Equation*

Noise Is Damaging.

FAIRNESS

Similarly situated people are not treated similarly

CREDIBILITY

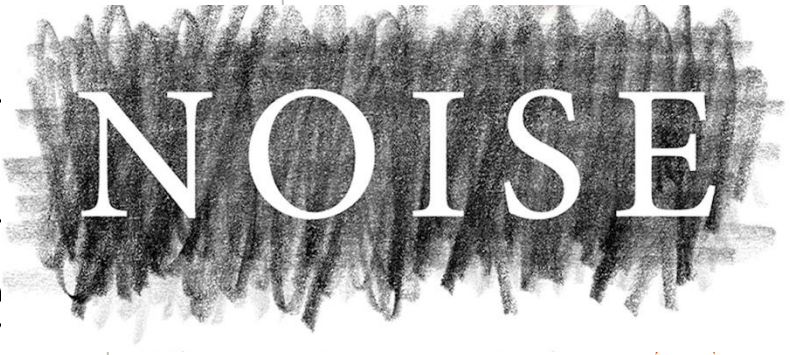
Inconsistency violates expectations

ERROR

Noise causes error – just as bias does

3

Both bias and can be measured and reduced.



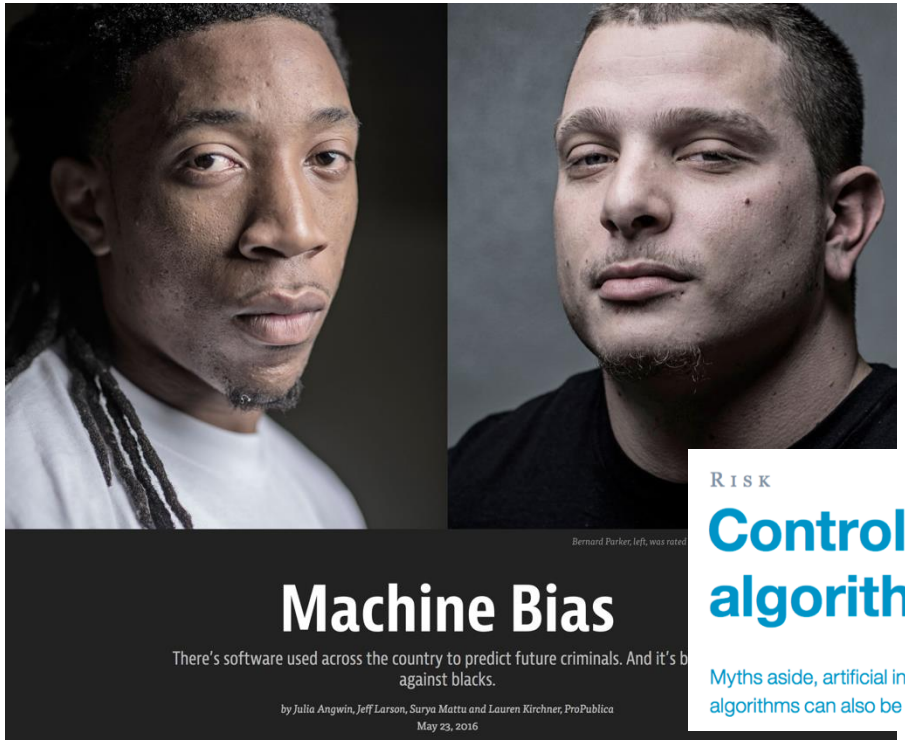
There is a sure way to eliminate Noise (but it may add to bias).

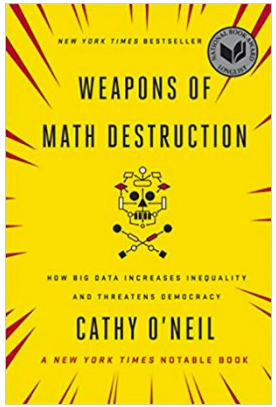
1

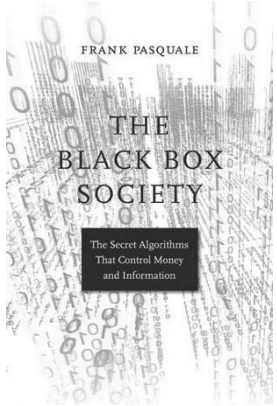
Wherever there is judgment, there is NOISE.

(And more of it than you think.)

Concern about Machine bias is growing







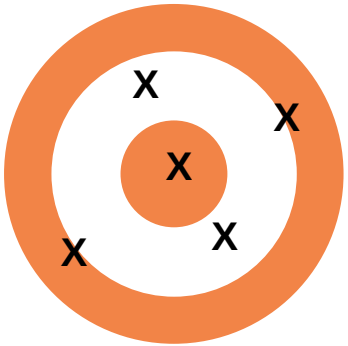
RISK

Controlling machine-learning algorithms and their biases

Myths aside, artificial intelligence is as prone to bias as the human kind. The good news is that the biases in algorithms can also be diagnosed and treated.

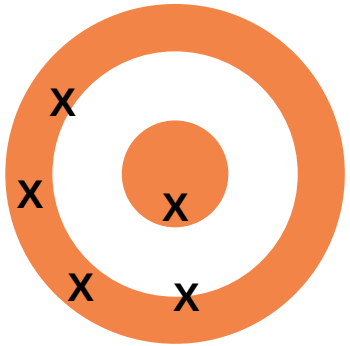
Even when you don't know Bias, you can measure Noise.

C



NOISY

D



NOISY AND BIASED

How can we keep human judgment – without the noise?

Reducing Bias and Noise: “Decision Hygiene”

Decision Hygiene

1

AGGREGATE

3

STRUCTURE

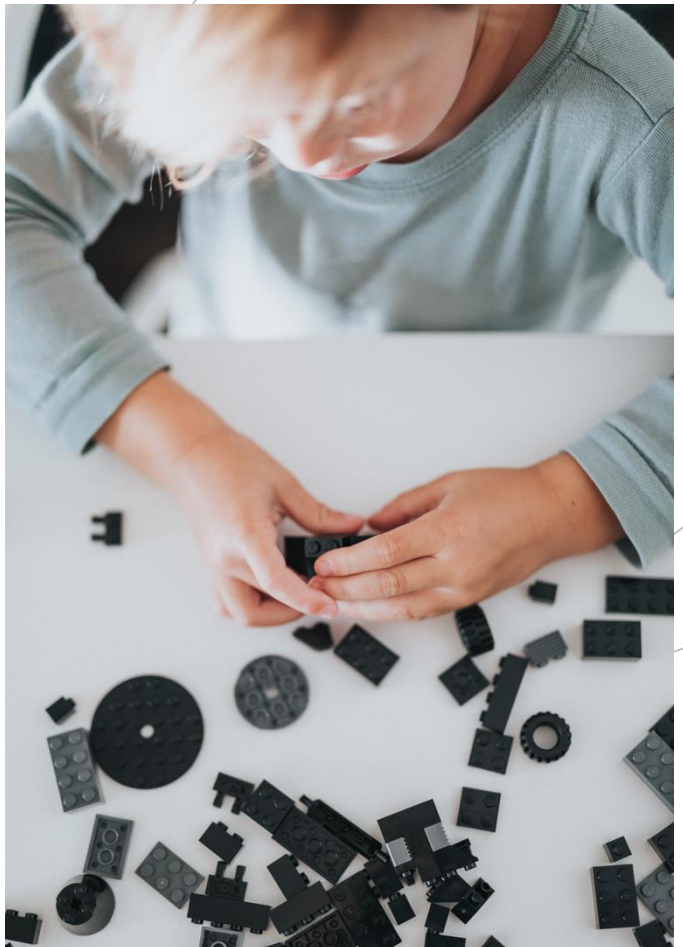
4

KEEP INTUITION FOR THE END

But beware: Discussion ≠ Aggregation

- Because of:
  - Social influence
  - Rational adjustment
- …usually, groups *amplify* noise
- Independence must be managed

You would not let the witnesses influence each other. Why is it different in the office?



3

Structure your judgments

- Divide and conquer: *mediating assessments*
- Quantitative and objective
- Score against frame of reference
- Discuss separately
- Aggregate independent inputs on each score

Structured Judgment: Medical Guidelines

APGAR SCORES EXPLAINED

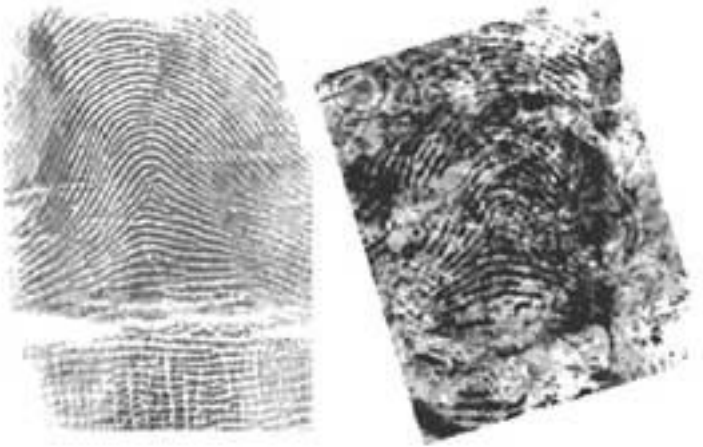
	Indicator	0 Points	1 Point	2 Points
A	Appearance (skin color)	Blue; Pale	Pink Body; Blue Extremities	Pink
P	Pulse	Absent	Below 100 bpm	Over 100 bpm
G	Grimace (reflex irritability)	Floppy	Minimal Response to Stimulation	Prompt Response to Stimulation
A	Activity (muscle tone)	Absent	Flexed Arms and Legs	Active
R	Respiration	Absent	Slow and Irregular	Vigorous Cry

Structured Decisions

“I give everything a numerical rating. So I actually transfer my emotions into a kind of mathematical equation, which helps me look at it from a more objective viewpoint. Because if you stay subjective to it and emotional to it, it’s hard to make a decision because you might be swayed by the emotions.”

- Story
- Written word
- Character
- Director

NYTimes, interview with Kara Swisher, January 19th, 2021



Beware “too much information”

Exogenous information adds noise to judgment…

… even when it is *accurate*

4

Keep intuition for the end

Early intuition adds noise

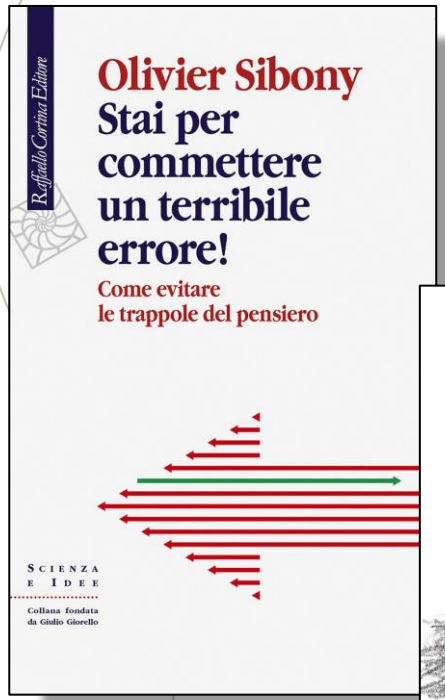
- Selective attention
  - Selective recall
- Excessive coherence

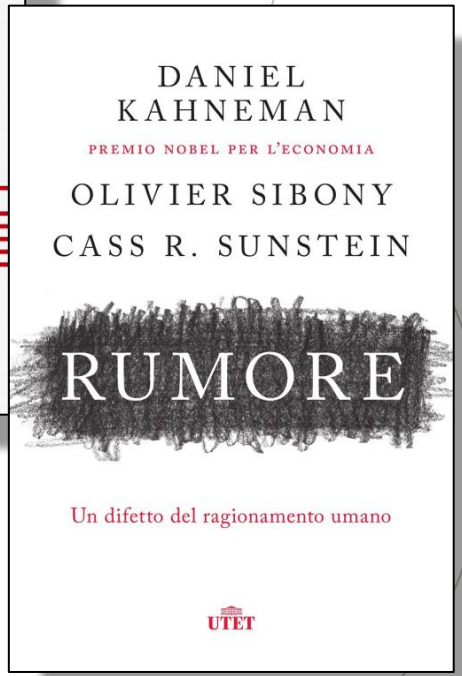
Bias and noise are everywhere.

They are costly.

They can be measured and reduced.

SUMMARY





Thank You

oliviersibony.com

HEC PARIS

Olivier Sibony

@siboliv