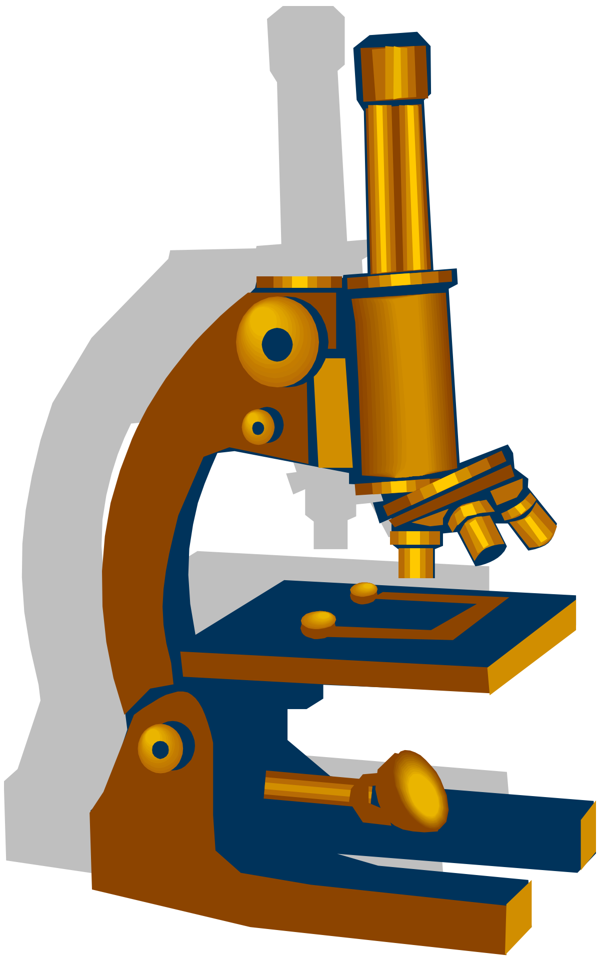
Addressing a Nobel Challenge to Modeling and Simulation

Bob Bordley
Booz Allen Hamilton

Kahneman's Nobel Prize in Economics

<u>Extensive Work Documenting Human Biases</u>

- Tendency to be overoptimistic
- Tendency to be influenced far too much by recent experience
- Tendency to be influenced by whether a problem is framed in terms of gains or losses
- Tendency for assessments to be biased by self-interest (or pride)



The Symptom

- Model-Predictions are Overoptimistic regardless of whether the model outputs are
  - *Project Completion Times*
  - *Project Costs*
  - *Product Sales*
  - *Product Performance*

Lovallo and Kahneman's Diagnosis

- These models infer outputs from input factors
- Who estimates these input factors?
  - *Sometimes Historical Data*
  - *Sometimes human beings*

  Human Bias is the problem

Lovallo and Kahneman's Solution

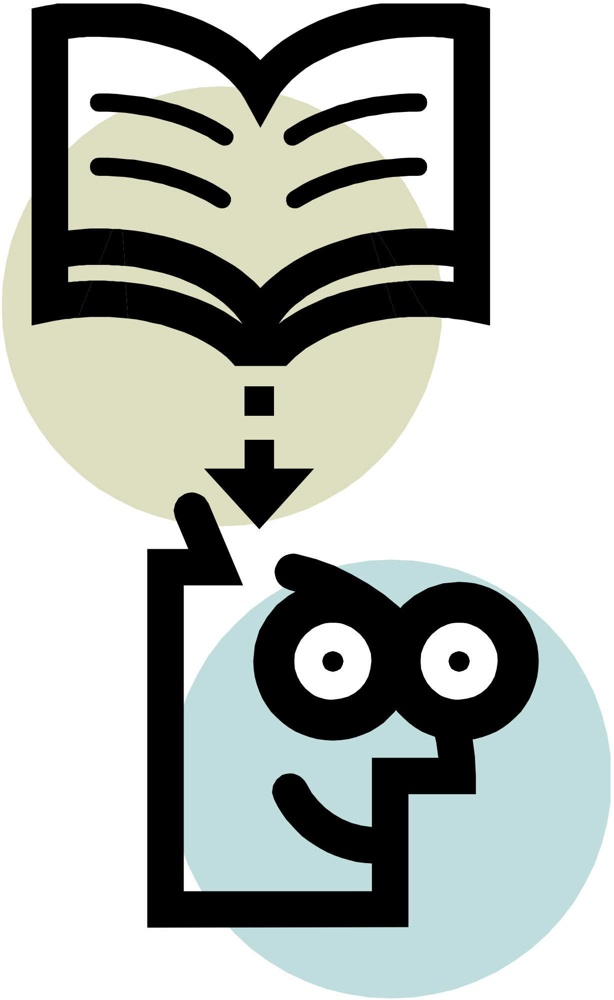
- Identify projects in the past that are `similar' to the project of interest
- Determine the output value (variable of interest) for those projects
- Weight each project by the degree of similarity
- Use the frequency chart of past project output values to estimate the probability distribution of the current project's output value

Their Conclusions

- Modeling is an
  "Inside-Out Approach":
  - Predicts OUTPUT from INPUT
- Their approach is an
  "Outside-In Approach"
  - Predicts OUTPUT from past data on OUTPUT

- They favor Outside-in Approach

Another Example



- Output of interest: Product Sales
- A model expressed OUTPUT as a function of INPUTS
  - *INPUTS were `demand-price elasticities' :*
    - *Don't worry if you don't know what this is*
  - *Bayesian methods define uncertainty about INPUTS*
  - *Model made inferences about OUTPUTS (sales) from INPUTS*
- The company reaction
  - *Model received serious attention*
  - *But no one felt confident about INPUT values*
  - *Company eventually followed the Garbage-In/Garbage-Out principle and did not use the model*

The modeler was following the standard recipe

- Create a model defining
  - OUTPUT in terms of INPUTS
  - Classical Physics says: ***Give me the state of the system now and I can predict its future***
- Quantify Uncertainty about INPUTS
  - Use Bayesian ideas if there’s extreme uncertainty
- Infer an uncertain forecast of OUTPUTS

Why didn’t the standard recipe work?

Clairvoyant Test

- An individual who is uncertain about the answer to question should at least understanding the question
- There should be some way, possibly hypothetical, of determining what the answer to the question is
- This is the clairvoyant test: a clairvoyant robot who sees the future and the past can determine what the answer to the question is
- Does `price impact’ pass the clairvoyant test?
  - It is technically defined as an elasticity which is a derivative of a function
  - That is a mathematical abstraction which is not directly observable

Established Rules of Assessment

- Do not
  - Ask people to provide estimates of abstract quantities
  - Ask people to estimate non-abstract quantities about which they have no experience or intuition
- Do
  - Ask people about concrete quantities about which they may have some intuition
  - It is OK to infer estimates of abstract quantities from intuitions about concrete quantities
- So individuals could not use their intuition to estimate the INPUTS to the model

Summarizing these Two Examples

- Model specifies OUTPUTS as a function of INPUTS
- Garbage-In/Garbage-Out Example
  - Model INPUTS aren't intuitive
  - Reject Model:
    - Rely on experienced intuitions about OUTPUTS
- Lovallo and Tversky
  - Model INPUTS are biased
  - Reject Model:
    - Reply on `reference class' historical data on OUTPUTS

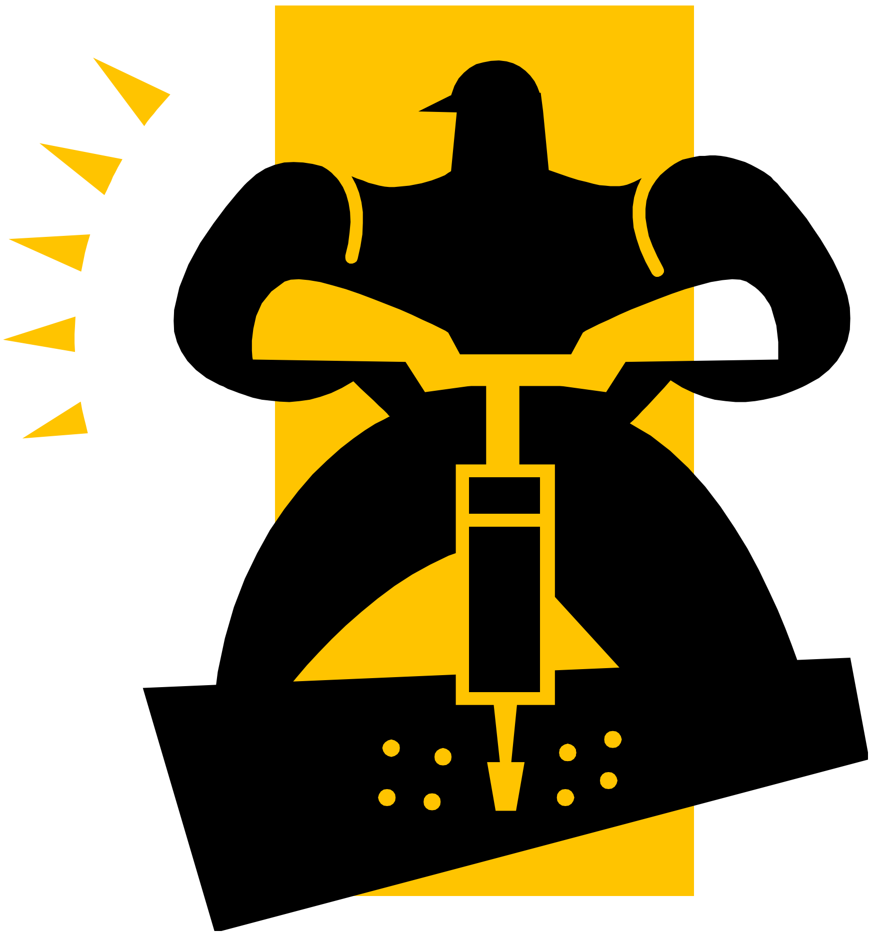
The problem is there are two recipes

The two recipes

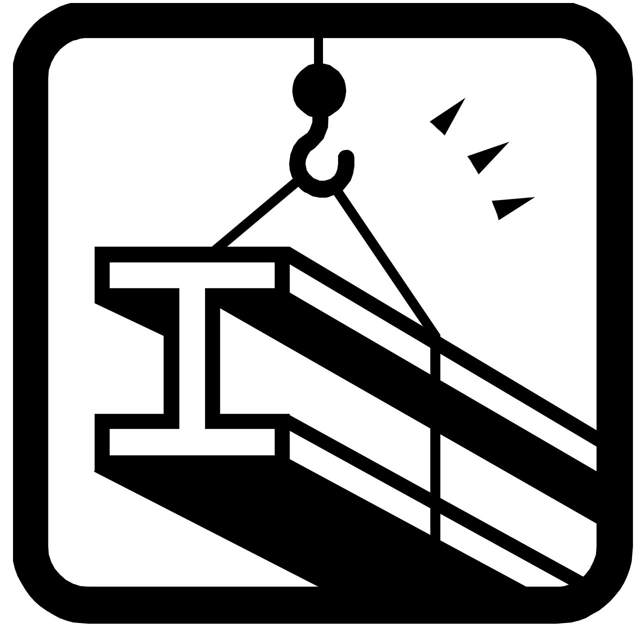
- *Model recipe: specifies OUTPUT from uncertain beliefs about INPUTS*
- *Experience recipe: specifies OUTPUT from direct information (experience) about that OUTPUT*

How about an Integrated recipe? which balances INPUT and OUTPUT information

My challenge: Rigorously building an integrated recipe



Key Steps in Recipe



- Assume
  - Uncertainty in
    - measurement of INPUTS
  - Uncertainty in
    - measurement of OUTPUTS
  - Model relates true values of INPUTS to true values of OUTPUTS
- CROMWELL'S RULE: BE OPEN TO BEING WRONG ABOUT ANYTHING!

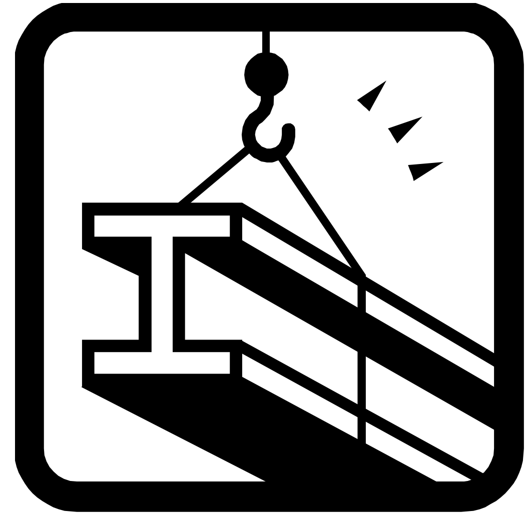
Motivation for next assumption

- There are three kinds of information
  - Information about outputs
    - Often based on manager intuition or experience
  - Information about inputs
    - Often based on measurements by manager’s subordinates
  - Information about how inputs and outputs are related
    - This is the information in the model
    - Often reflects scientific or logical principles

How does the model contribute to the understanding of outputs?

- In the beginning
  - Beliefs about both inputs and outputs
  - People may have felt they were unrelated
- Then comes the model
  - The model provides new information on how inputs and outputs are related
- This leads to beliefs about BOTH inputs and outputs being updated
  - The model might convincedthe manager that his beliefs about outputs are wrong or
  - The model might convince the manager that his beliefs about inputs are wrong
  - Or both

Focus on Joint Information about both Inputs & Output



- Prior to knowing how the model relates inputs and outputs
  - We express our joint information with a probability distribution
- Once we learn about the model

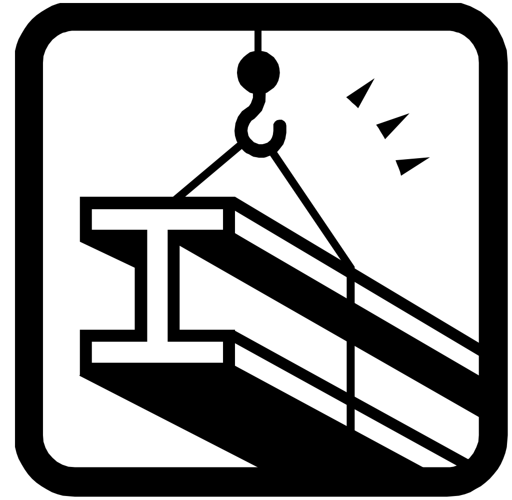
  We update this probability distribution

What is the information provided by a model?

- It is common to suppose you have several models
  - One of them is right
  - The others are wrong

  You have different beliefs about which one is right
- In reality
  - All models are wrong
  - So you can't make it about which model is right
- So how do we think about model?

Summarize model as list of of simulation outcomes



- (X1,Y1): If you INPUT values X1 into model, you get Y1
- (X2,Y2): If you INPUT values X2 into model, you get Y2
- (X3,Y3): If you INPUT values X3 into model, you get Y3
- (X4,Y4): If you INPUT values X4 into model, you get Y4

- (XK,YK): If you INPUT values XK into model, you get YK

What does Bayes Rule tell you?

- Your probability distribution over OUTPUTS and INPUTS (given the model) is the product of
  - Your initial probability distribution over OUTPUTS and INPUTS
  - The LIKELIHOOD of getting these simulation results given the model, if the INPUT AND OUTPUT values were known

Problem!

We know how to estimate the likelihood of an INPUT/OUPUT combination occurring in reality randomly

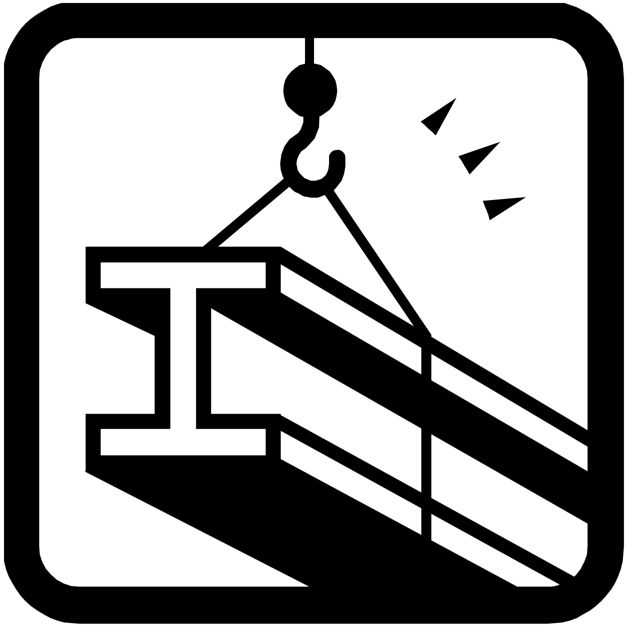
- But the INPUT was made up based on our prior beliefs

- So how do we estimate the likelihood?

One Approach

- We could assume model tells you nothing about the values of the inputs
- But that’s not right
  - If your input values consistently gave rise to bizarre output values
  - You could begin to question your input values
- So model does provide you information about the input values
  - If you also know the output values associated with those inputs

ASSUMPTION



Assume that the Simulation Inputs are randomly drawn from decision maker's prior beliefs about what values of INPUTS are reasonable

- Assume that knowing only the INPUT Values

--- without knowing anything about OUTPUT values ---

gives you no new information

These assumptions imply

THE LIKELIHOOD OF THE INPUT/OUPUT COMBINATIONS IMPLIED BY THE MODEL

EQUALS

THE LIKELIHOOD OF THESE INPUT/OUPUT COMBINATIONS *occurring in reality* / **THE LIKELIHOOD OF THE INPUTS** *occurring in reality*

Model Probability over OUTPUT

MODEL PROBABILITY FORECAST OVER OUTPUT

Equals

Average of Model Likelihood over all INPUT Values

Overall Probability

=

Reference Class Probability X Model Probability

CONCLUSION

- When you have no intuition about outputs,
  - Reference class probability is constant
  - Overall probability is model probability
- When you have no intuition about inputs,
  - Model probability is constant
  - Overall probability is reference class probability
- When you have some intuition about both, multiply the two probabilities

Special Cases

- Everything is Gaussian
- Then

  Treat reference class information

  on OUTPUT as a forecast

  Treat model information using INPUTS

  as a second FORECAST
- Final Forecast is weighted average of both forecasts

But reference class information on Y will typically not be Gaussian

- You need to calculate formula numerically

Conclusion

- Battle between
  - Experienced intuition
    Outside-In
  - Models
    Inside-Out
- Existing recipes force
  you to choose one
  or the other
- New recipe allows
  you to combine both.