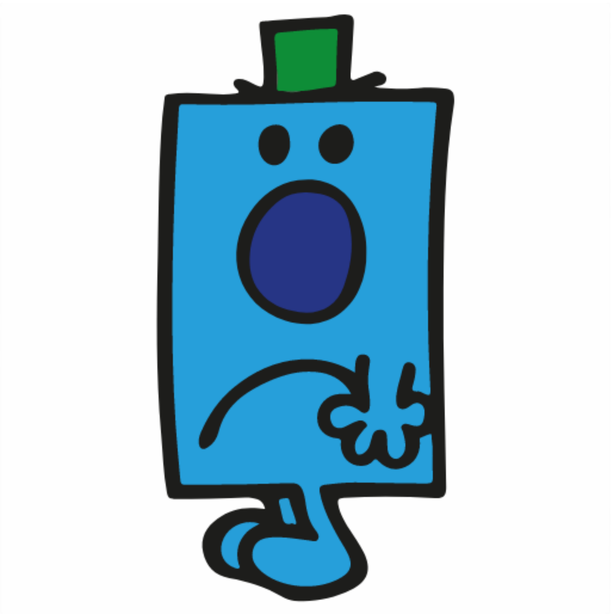
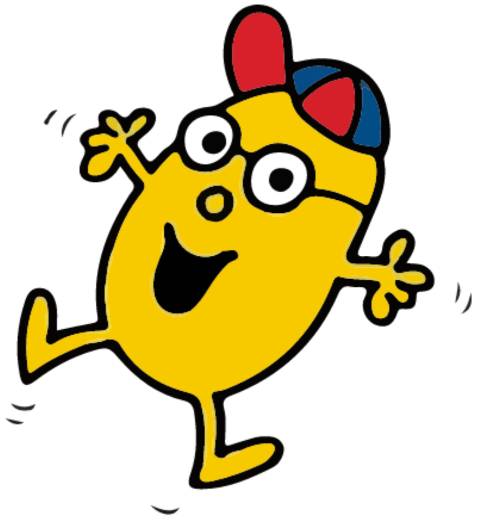
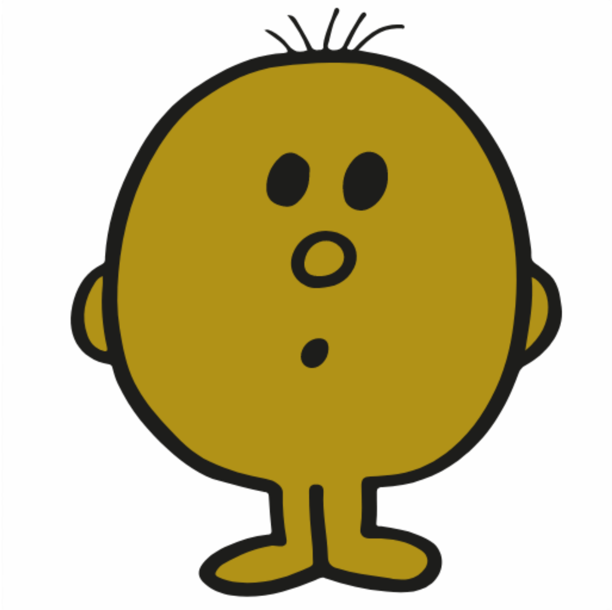
How expert are your experts?

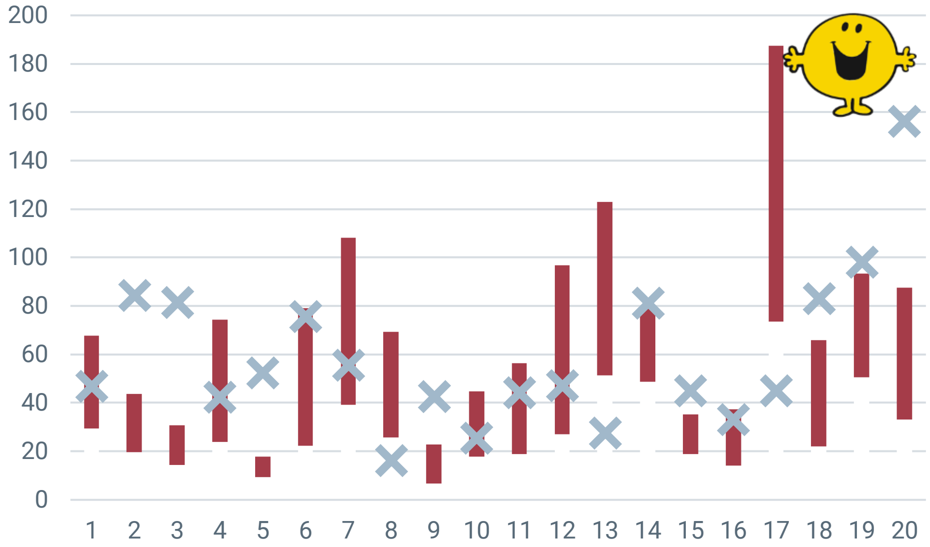
Bias scores and aggregation in subject matter expert elicitation

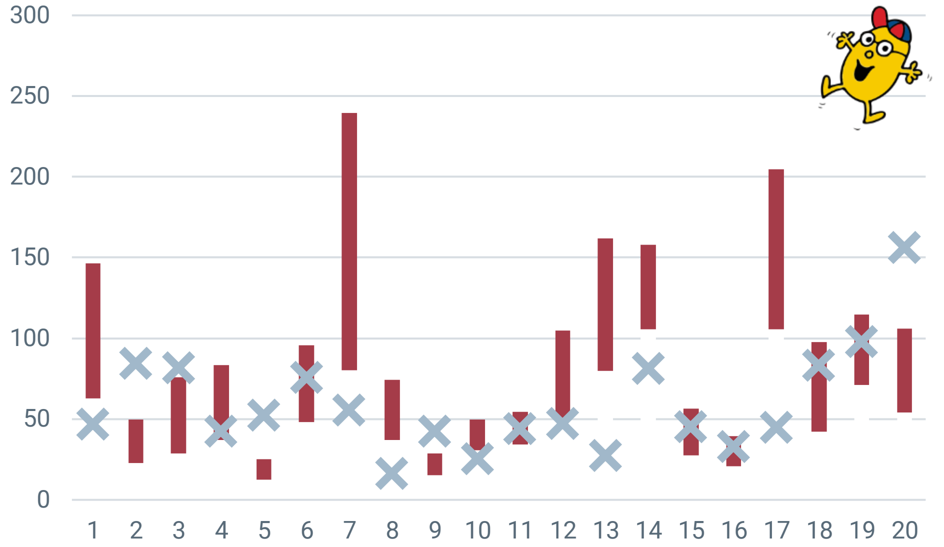
I have four subject matter experts…

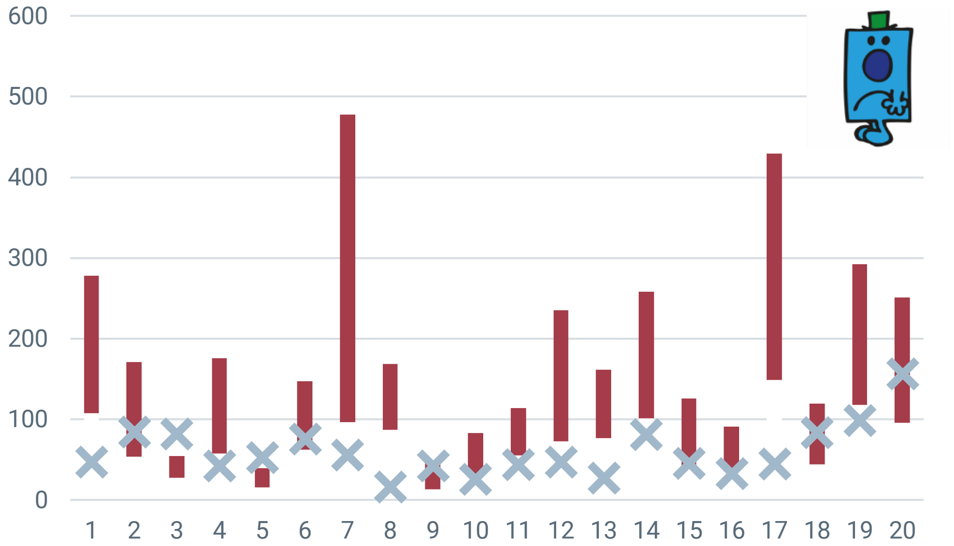


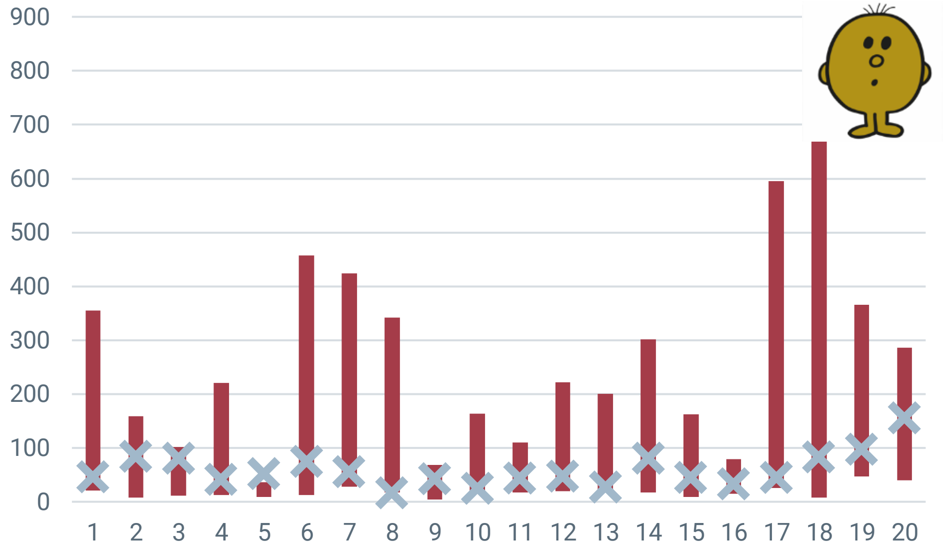



Each SME has assessed ranges for a variety of different project costs.
These ranges have been recorded together with the actual costs eventually incurred

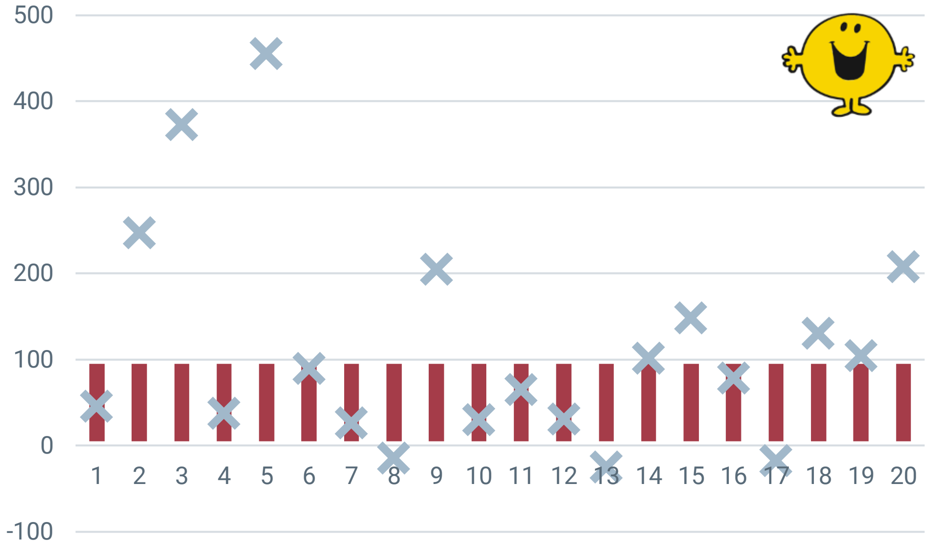


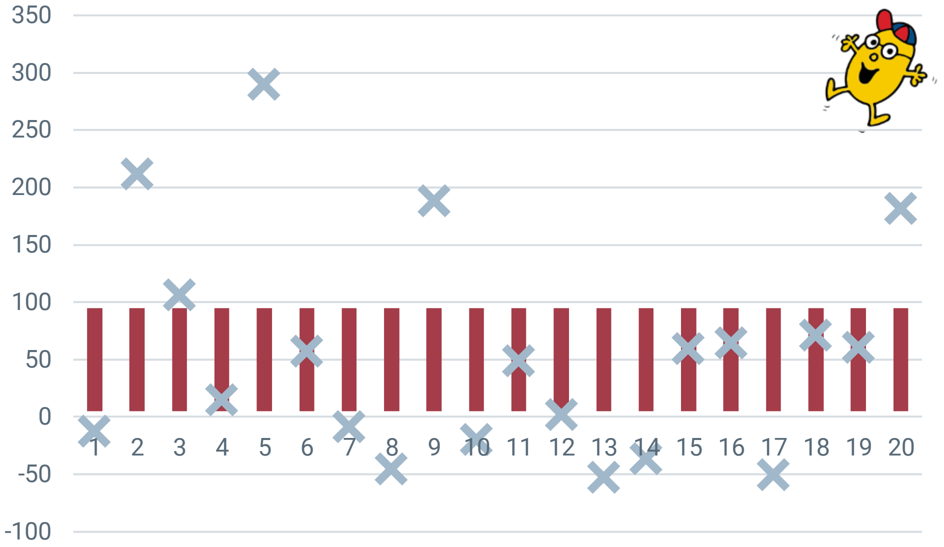


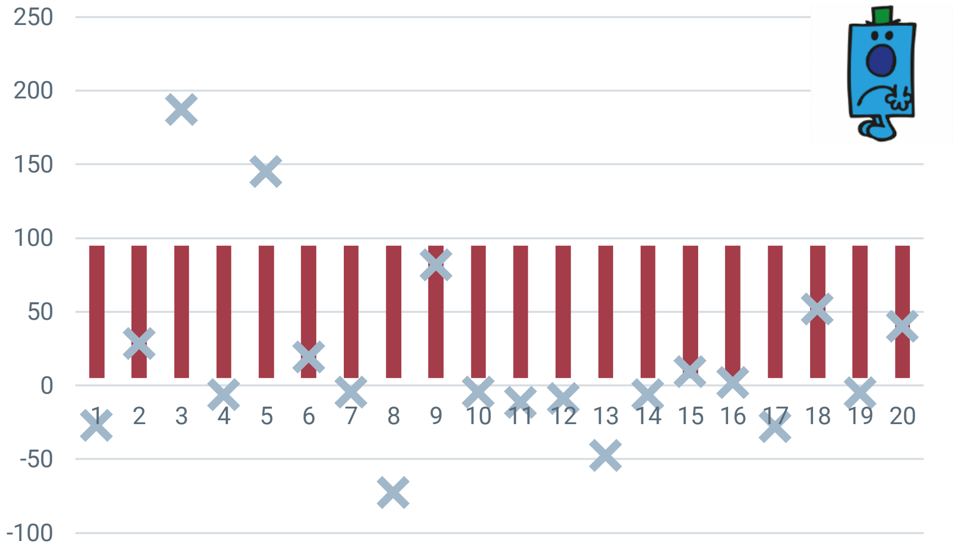


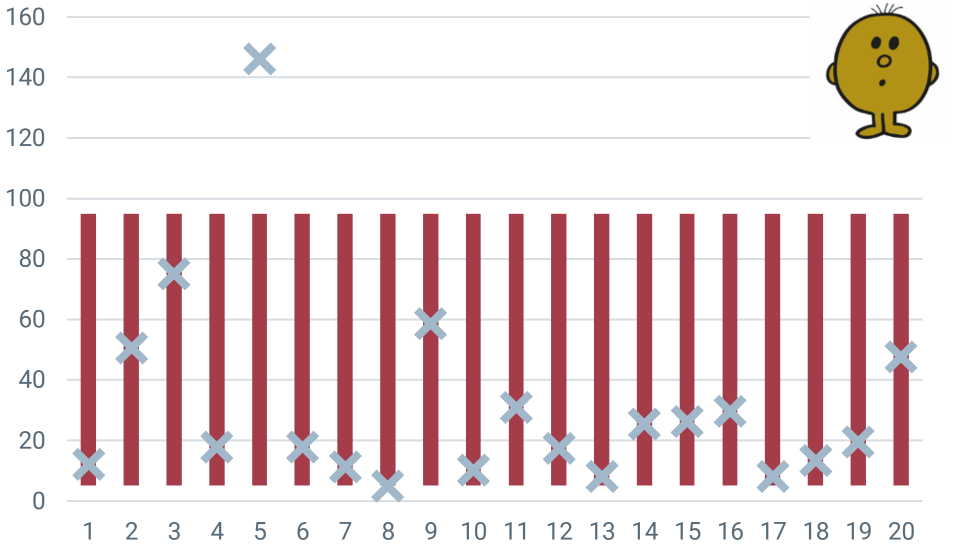


Rescale so that P5-P95 range is 5-95









Two fundamental lines of enquiry

How good are those assessments?

- Can we score them?

Can we characterize SMEs by their biases?

- Are they optimistic / pessimistic?
- Are they over or under confident

Can we combine assessments?

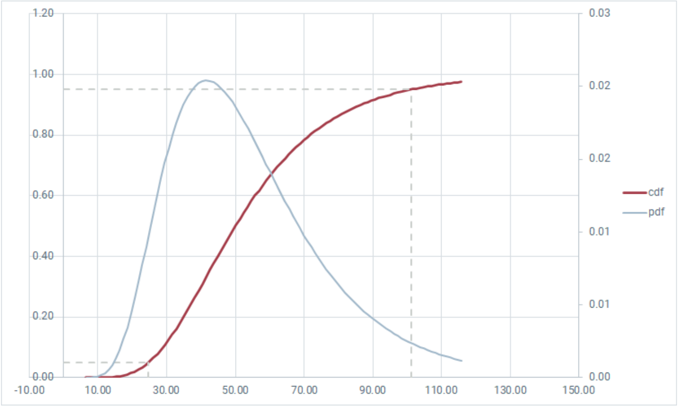
- Given a handful of SMEs, how do we combine their assessments?

Hubbard Decision Research

This work is carried out in proud collaboration with Hubbard Decision Research

Bias scores

Postulate that each SME has a “true” or faithful uncertainty distribution corresponding to the data they have to hand



lower

upper

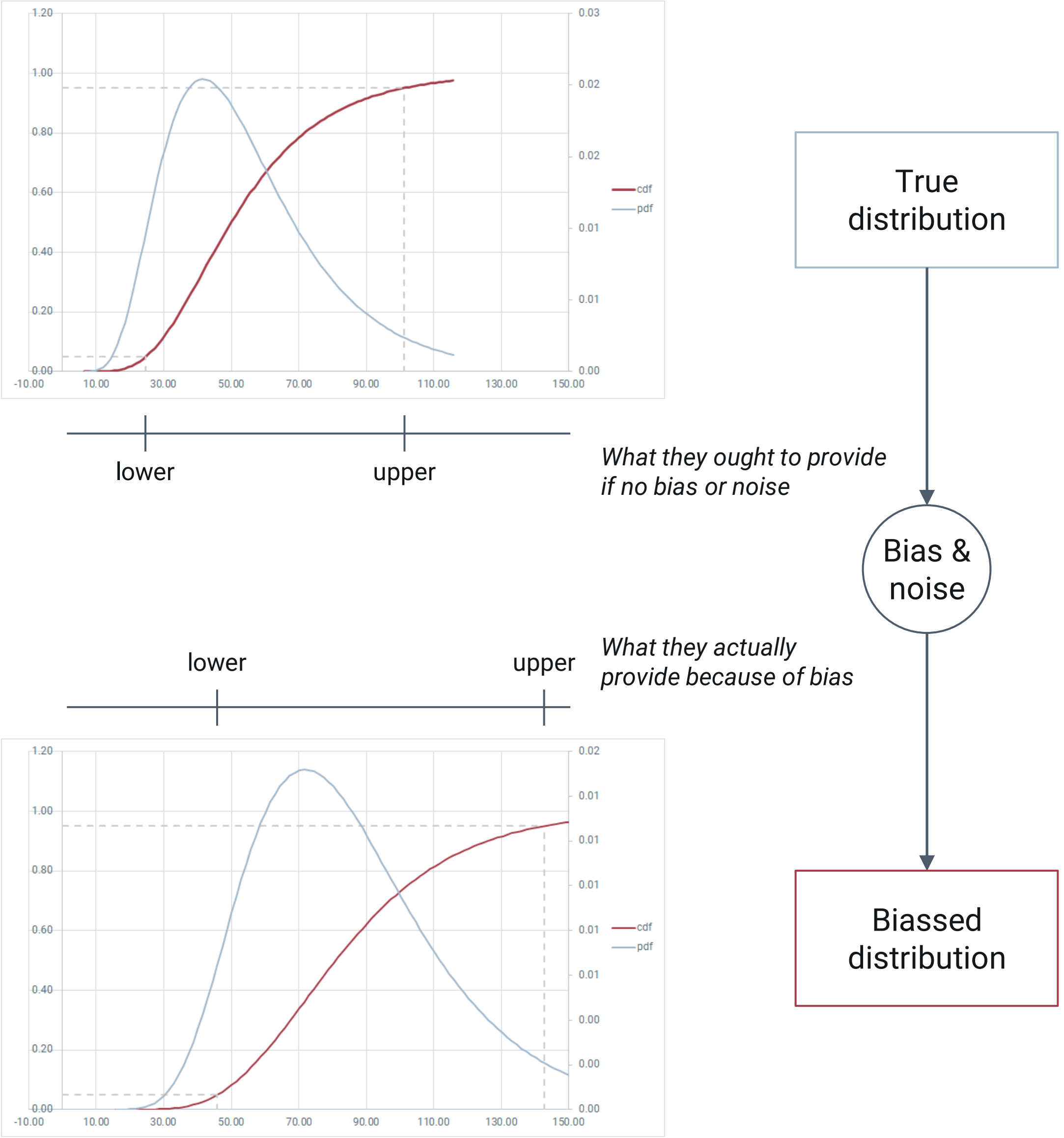
What they ought to provide
if no bias or noise

True distribution

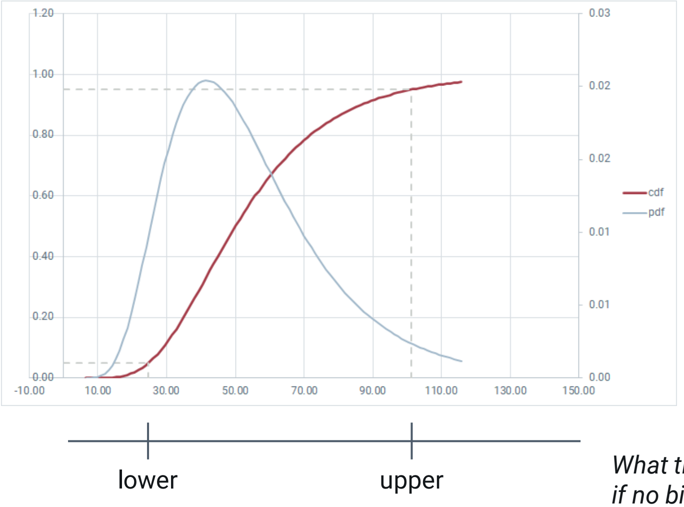
Postulate that each SME has a “true” or faithful uncertainty distribution corresponding to the data they have to hand

But that bias, ineptitude and noise distort the SMEs perception of that uncertainty

And the range they give corresponds to the distorted perception of the uncertainty



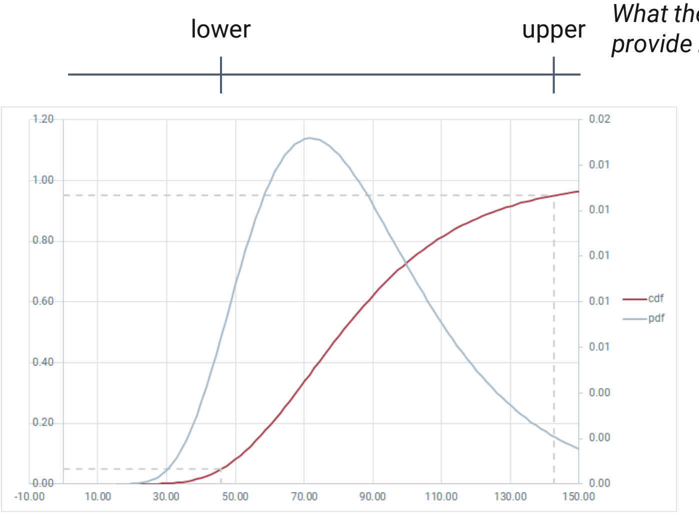
Postulate that each SME has a “true” or faithful uncertainty distribution corresponding to the data they have to hand



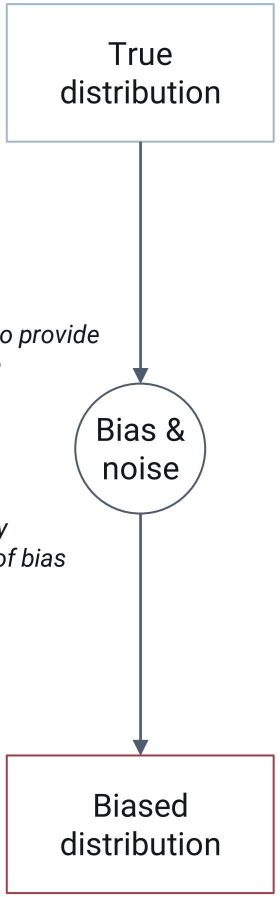
What they ought to provide if no bias or noise

But that bias, ineptitude and noise distort the SMEs perception of that uncertainty

What they actually provide because of bias

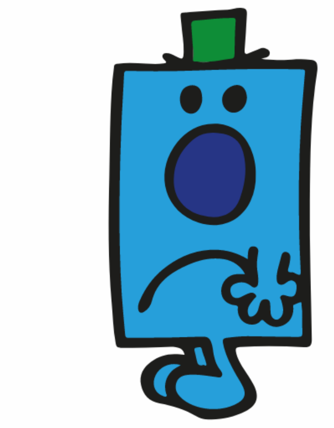


And the range they give corresponds to the distorted perception of the uncertainty



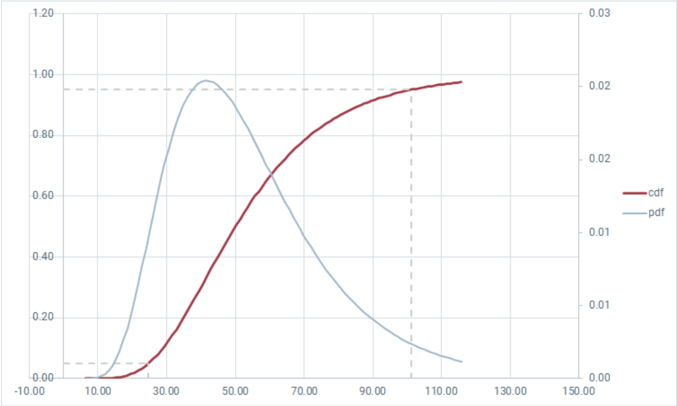
Describe the bias with two numbers

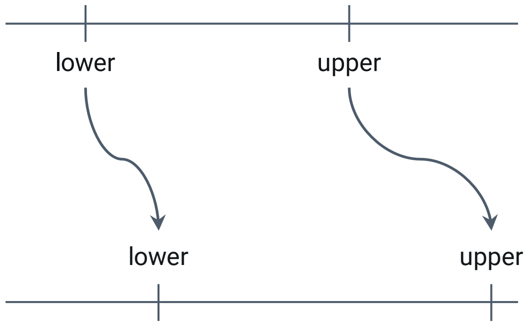
1. A number describing how much the middle of the distribution moves
   - Optimism / pessimism - δ
2. A number describing how much the range is narrowed or broadened
   - Confidence / Vagueness - ε



Mr Grumpy combines

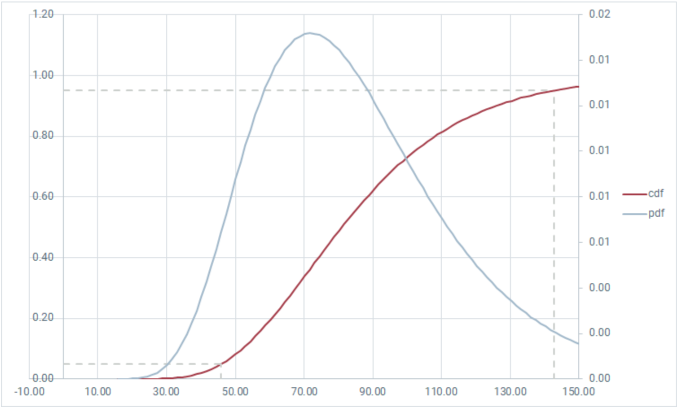
- pessimism
  - middle of the distribution shifts right
- confidence
  - range narrows relative to middle of distribution

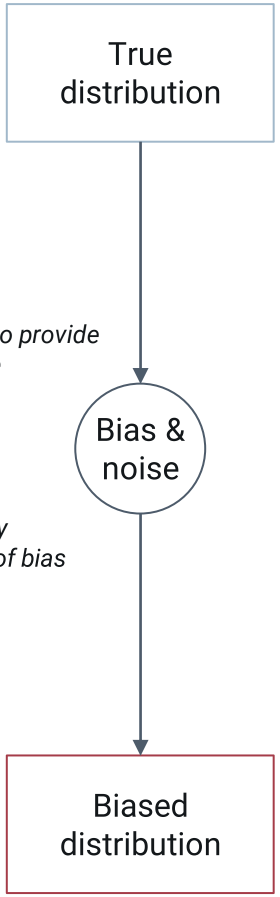




What they ought to provide if no bias or noise

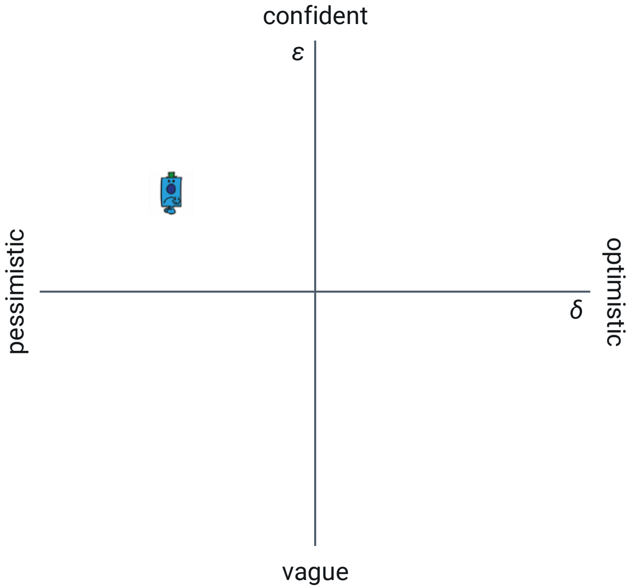
What they actually provide because of bias





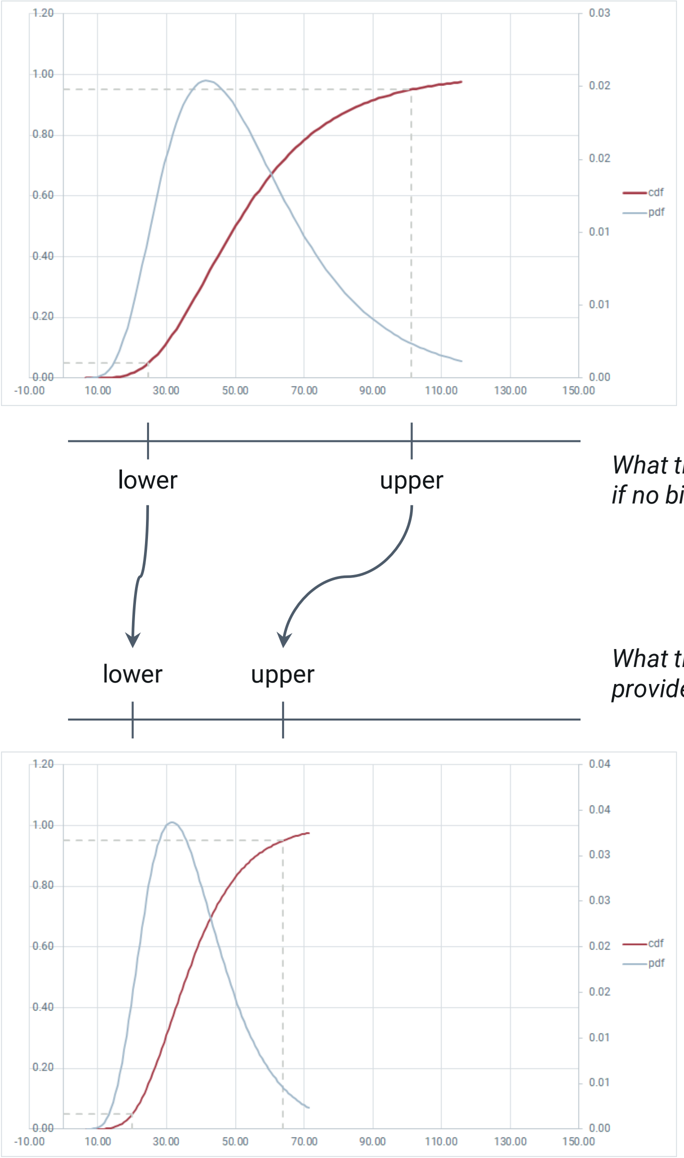
Describe the bias with two numbers

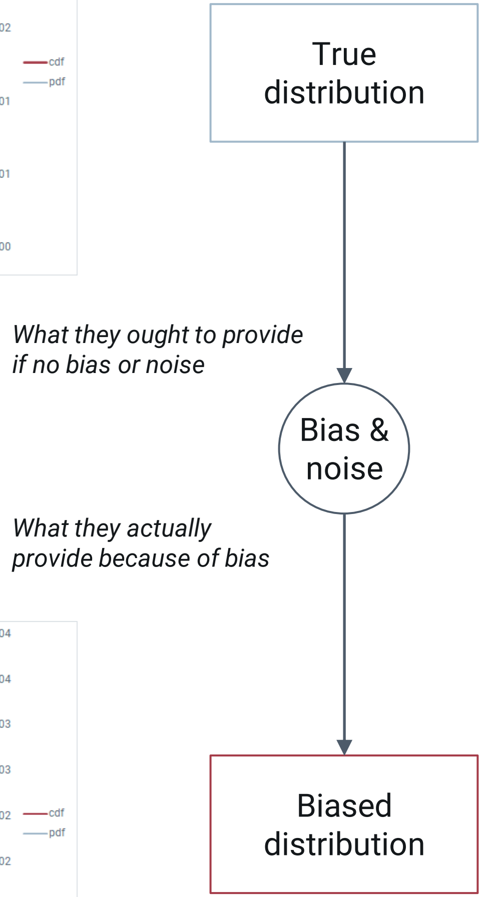
1. Optimism / pessimism - δ
2. Confidence / Vagueness - ε



Mr Happy combines

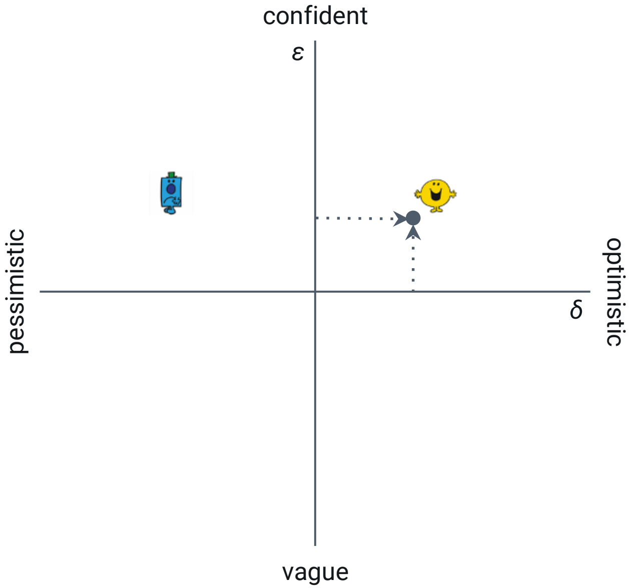
- optimism
  - middle of the distribution shifts left
- confidence
  - range narrows relative to middle of distribution

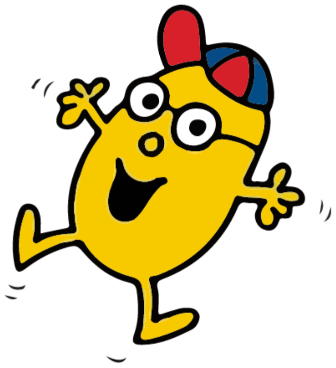




Describe the bias with two numbers

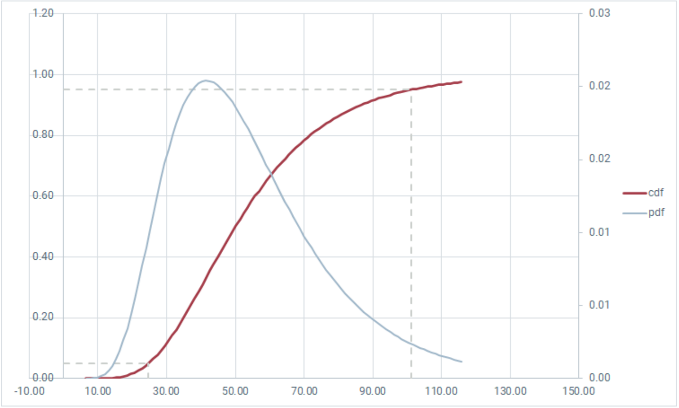
1. Optimism / pessimism - δ
2. Confidence / Vagueness - ε

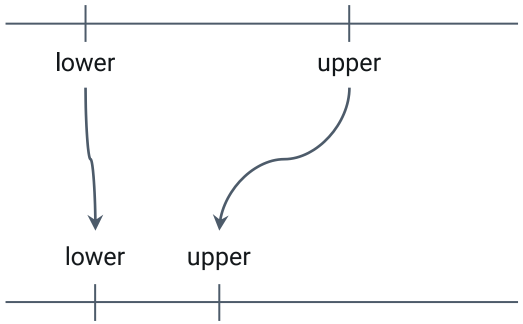


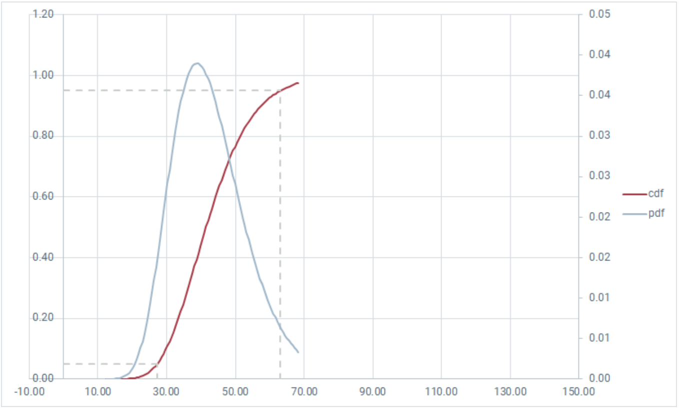


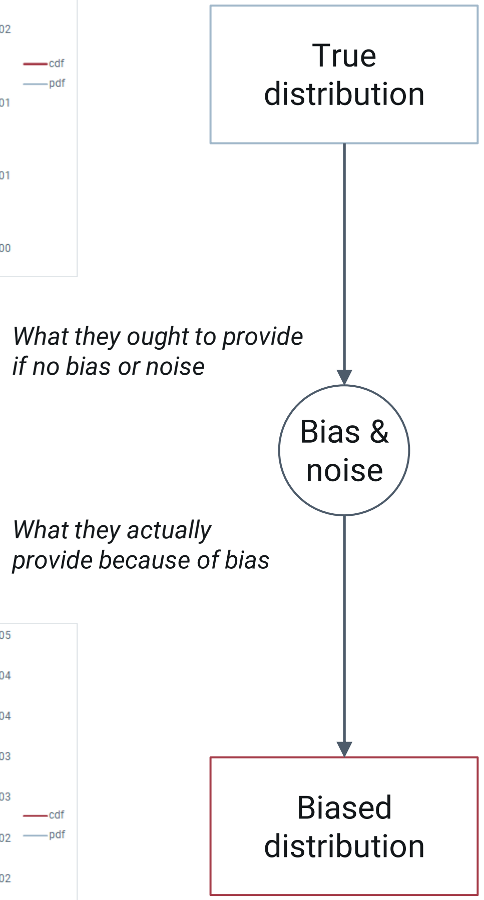
Mr Brave combines

- neutrality
  - middle stays the same (on average)
- strong confidence
  - range narrows relative to middle of distribution



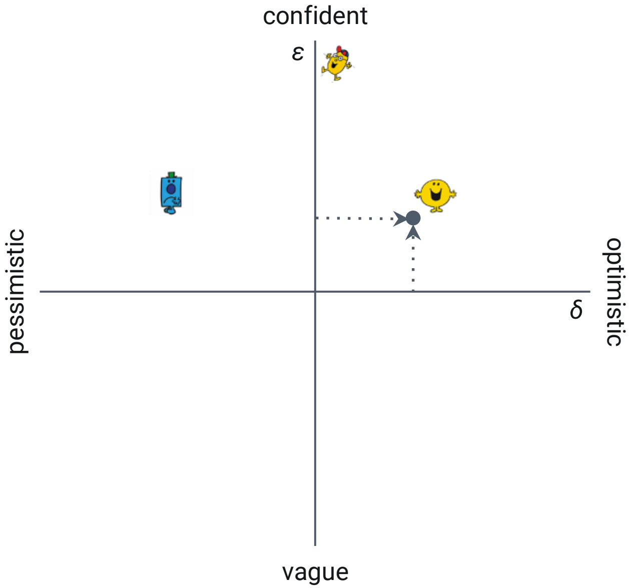




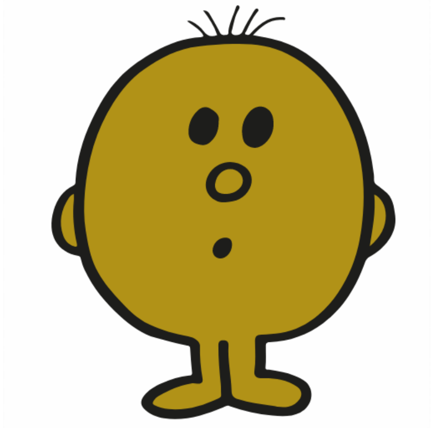


Describe the bias with two numbers

1. Optimism / pessimism - δ
2. Confidence / Vagueness - ε

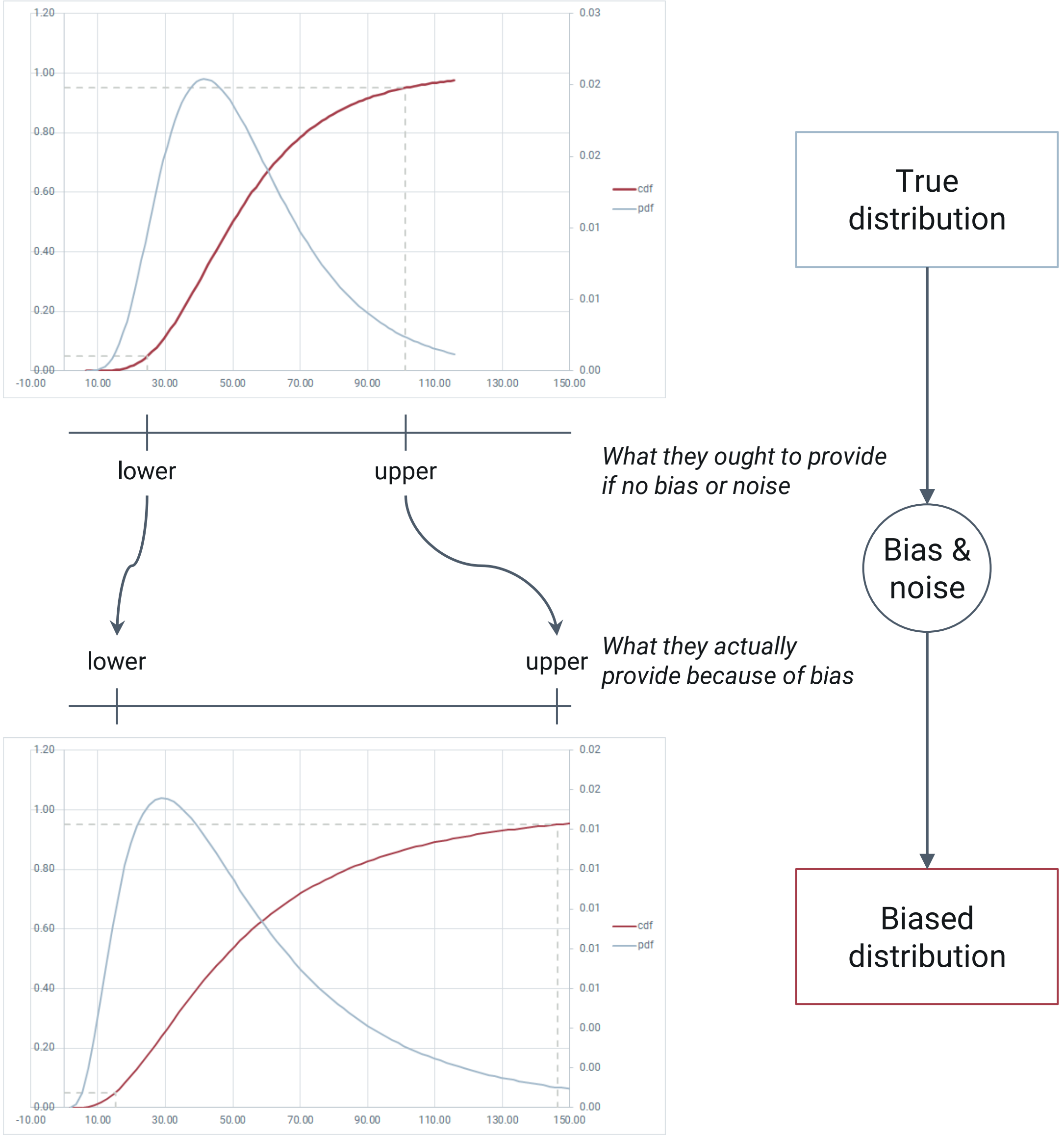


Stochastic ApS
Mathematical modelling for business & the business of mathematical modelling



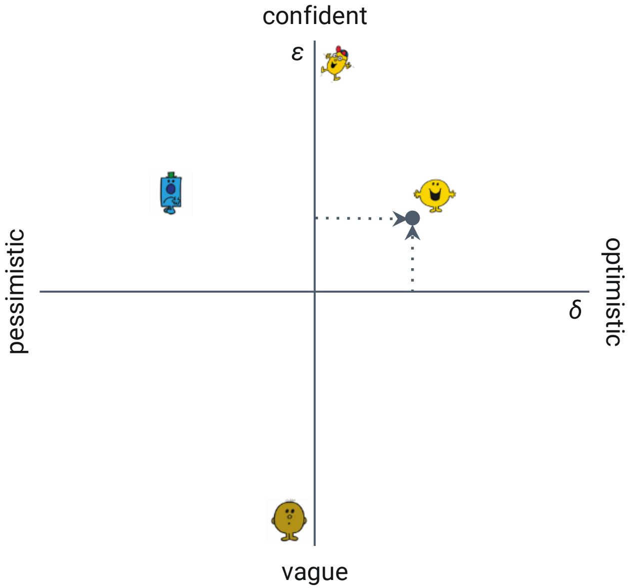
Mr Quiet combines

- neutrality
  - middle stays the same (on average)
- strong vagueness
  - range widens relative to middle of distribution

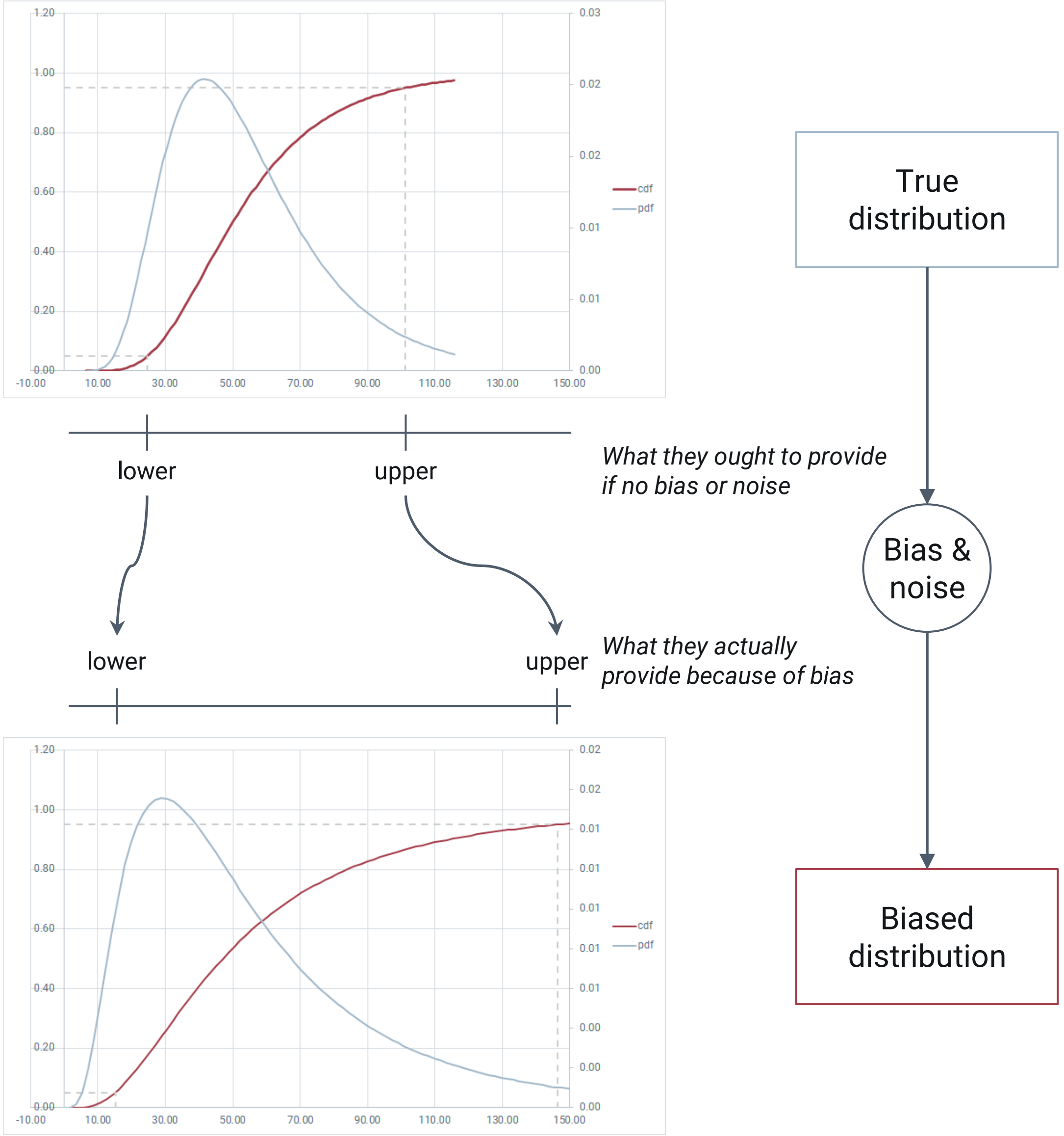


Describe the bias with two numbers

1. Optimism / pessimism - δ
2. Confidence / Vagueness - ε

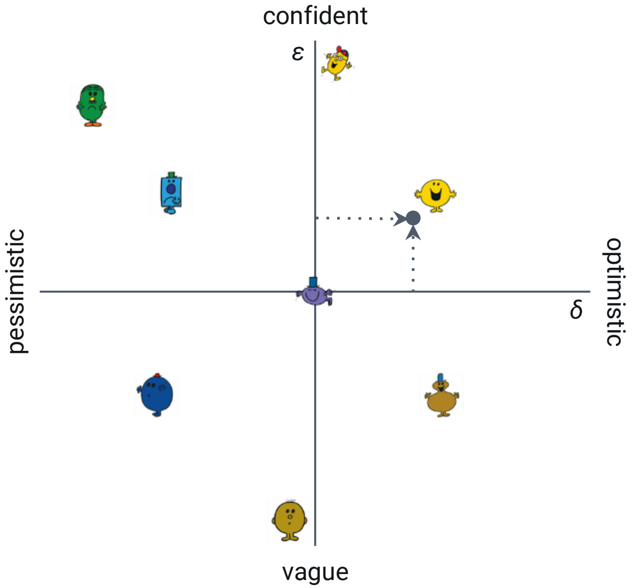


We can characterize all sorts of characters in this way, but a much more interesting question is how to infer an SME’s bias parameters given test results



Describe the bias with two numbers

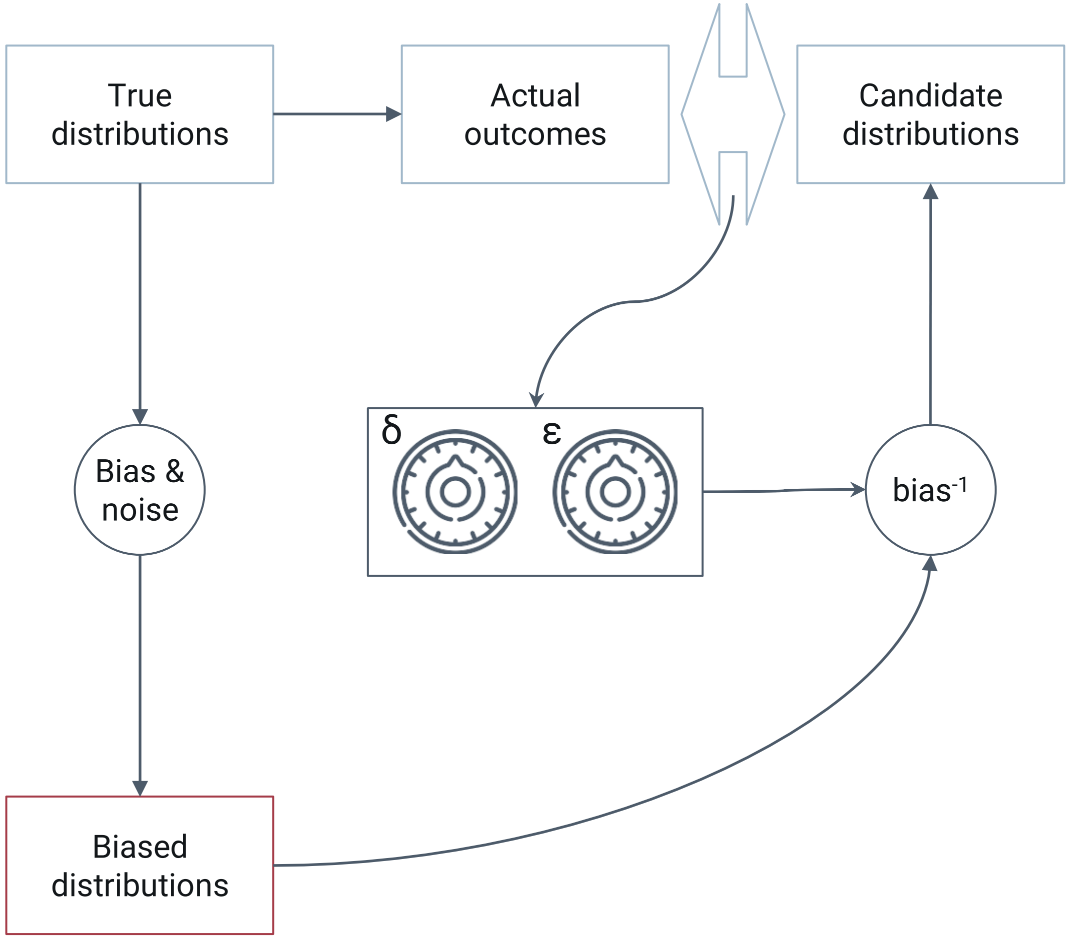
1. Optimism / pessimism - δ
2. Confidence / Vagueness - ε



Bias inference

Given a sequence of upper and lower bound estimates and a corresponding list of actual outcomes

1. Choose a form of the “true” uncertainty distribution to characterize the “true” uncertainty
2. Use this distribution form to convert all the lower and upper bounds (interpreted as percentiles) in the series into the parameters of that distribution.
3. Postulate a form for the bias transformations, including noise terms
4. For a choice of δ and ε, reverse the bias transformation to give candidates for “true” distributions
5. Compare the statistics of the actual outcome series with the statistics predicted by the distributions corrected by the bias transform
6. Find the bias parameters δ and ε that give the distributions that give the best match to the actual outcomes



Bias inference

Given a sequence of upper and lower bound estimates and a corresponding list of actual outcomes

1. Choose a form of the “true” uncertainty distribution to characterize the “true” uncertainty
2. Use this distribution form to convert all the lower and upper bounds (interpreted as percentiles) in the series into the parameters of that distribution.
3. Postulate a form for the bias transformations, including noise terms
4. For a choice of δ and ε, reverse the bias transformation to give candidates for “true” distributions
5. Compare the statistics of the actual outcome series with the statistics predicted by the distributions corrected by the bias transform
6. Find the bias parameters δ and ε that give the distributions that give the best match to the actual outcomes

Notes

1. With two bias parameters, it’s natural to use two-parameter distributions
   - Normal if there is a natural “correct” scale to the answer and a reasonable range is well-removed from zero. Parameterize with mean and variance
   - Log-normal if the question is greater than zero and may range over several orders of magnitude, especially if underlying uncertainties are multiplicative.
     Parameterize with normal mean and normal variance
   - Beta for percentages, probabilities and proportions.
     Parameterize with mean and either variance or samples
2. This is simple for standard distributions
3. Usually a percentage shift in the mean and a multiplicative factor for the variance.
4. This is also simple for simple transformations

Bias inference

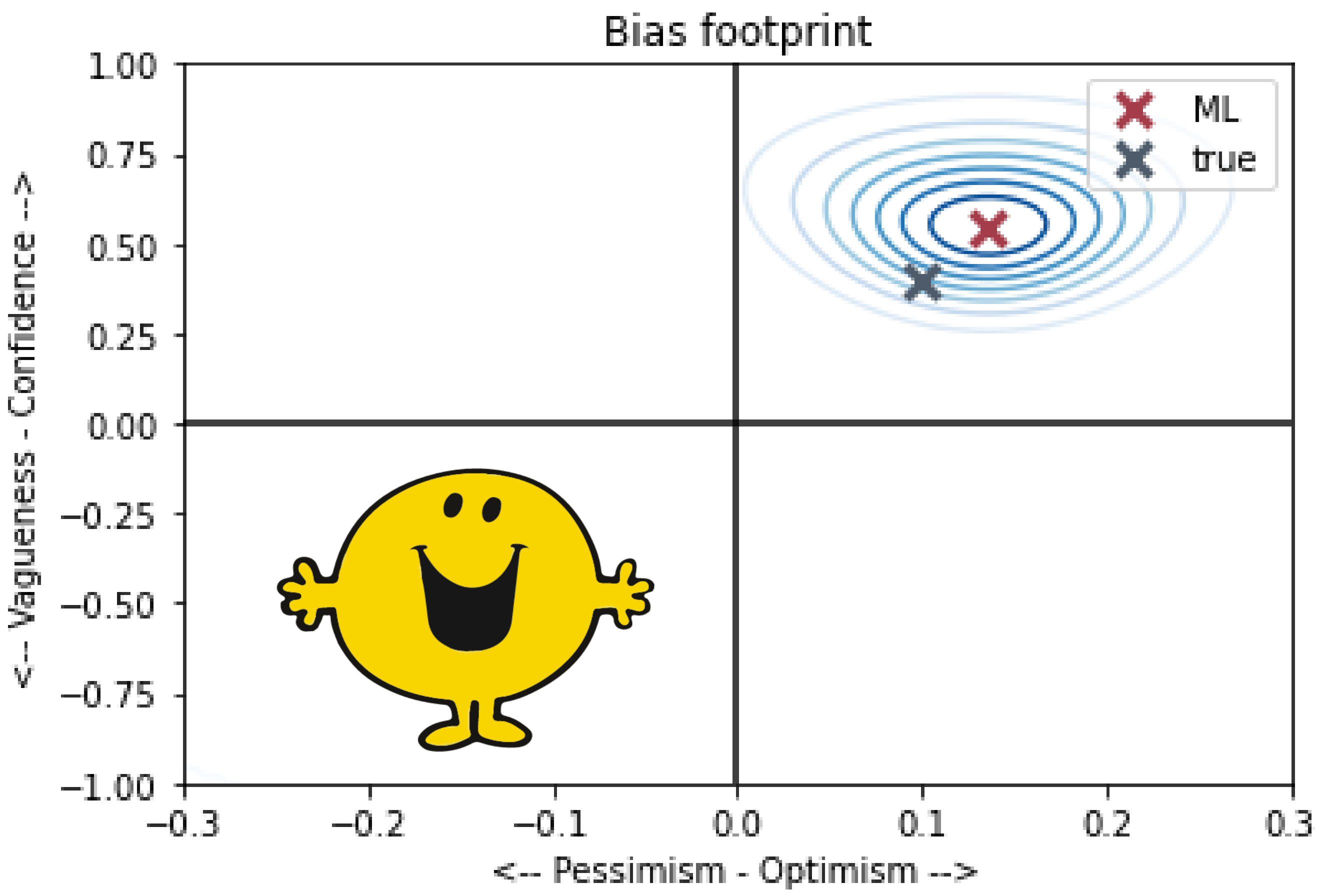
Given a sequence of upper and lower bound estimates and a corresponding list of actual outcomes

1. Choose a form of the “true” uncertainty distribution to characterize the “true” uncertainty
2. Use this distribution form to convert all the lower and upper bounds (interpreted as percentiles) in the series into the parameters of that distribution.
3. Postulate a form for the bias transformations, including noise terms
4. For a choice of δ and ε, reverse the bias transformation to give candidates for “true” distributions
5. Compare the statistics of the actual outcome series with the statistics predicted by the distributions corrected by the bias transform
6. Find the bias parameters δ and ε that give the distributions that give the best match to the actual outcomes

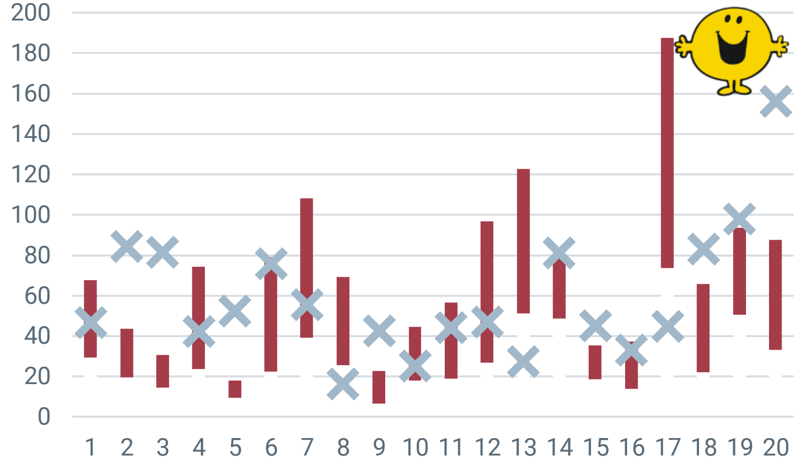
Notes

1. With two bias parameters, it’s natural to use two-parameter distributions
2. This is simple for standard distributions
3. Usually a percentage shift in the mean and a multiplicative factor for the variance.
4. This is also simple for simple transformations
5. There are several methods for doing this.
   Most of the obvious don’t work.
   - Matching predicted mean to observed mean trivially gives optimism / pessimism bias parameter
   - Matching predicted variance to observed variance doesn’t work because variance converges slowly and because we can’t see whether variance is coming from true uncertainty or errors in assessments
   - Least squares minimization also gives optimism, but gives nothing fruitful for confidence
   - A Bayesian approach allows you to marginalize noise and gives a likelihood function, which gives a good sense of the error in the assessment

Bias footprint based on 20 assessments



Plot shows **likelihood function** for Mr. Happy's sequence of assessments and actuals

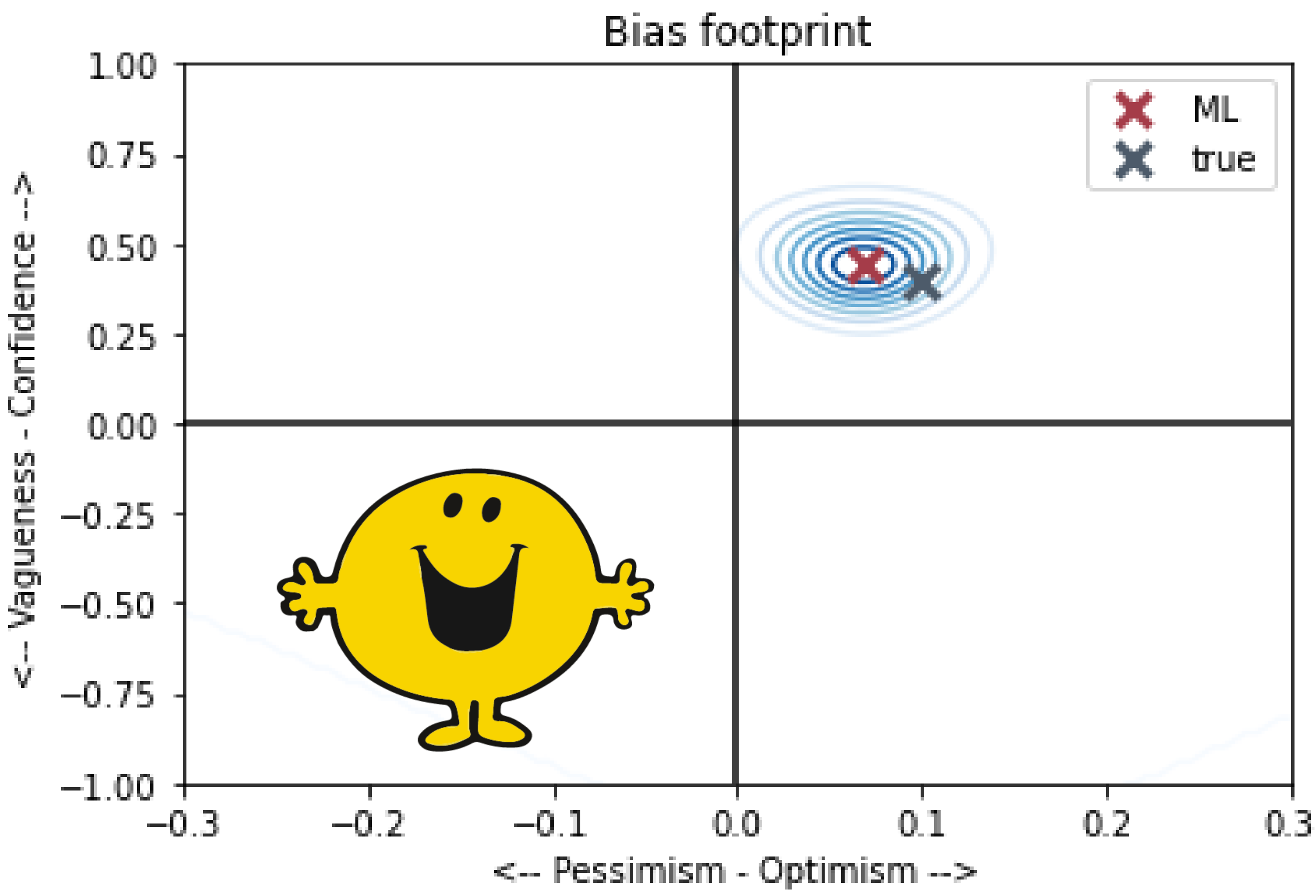


Effectively it shows the probability of seeing the sequence of actuals given Happy's assessments as a function of the bias parameters.

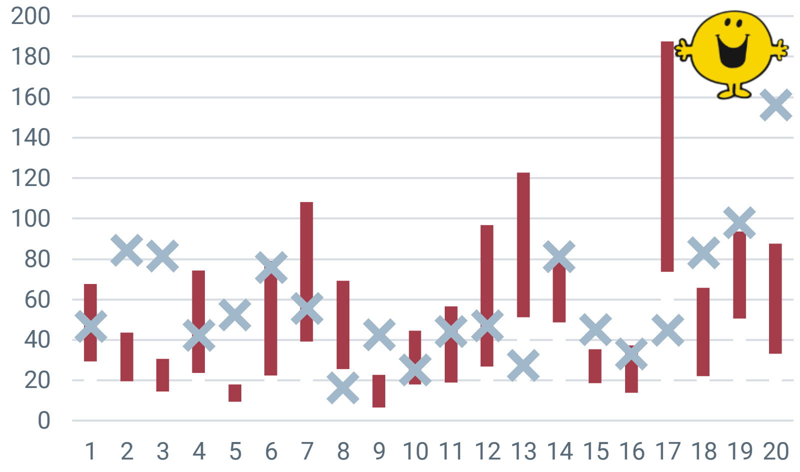
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Happy's sequence

Bias footprint based on 50 assessments



Plot shows **likelihood function** for Mr. Happy's sequence of assessments and actuals

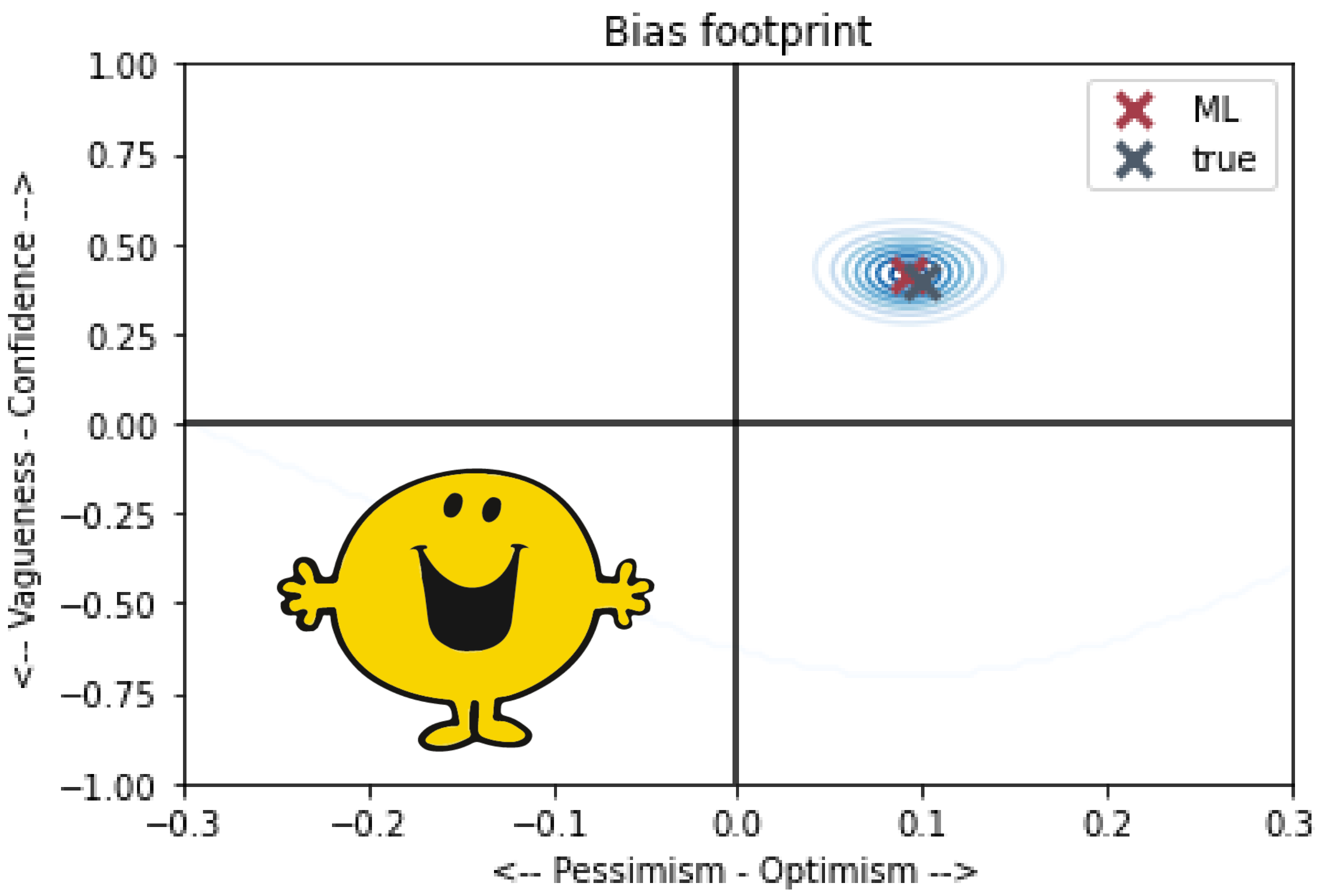


Effectively it shows the probability of seeing the sequence of actuals given Happy's assessments as a function of the bias parameters.

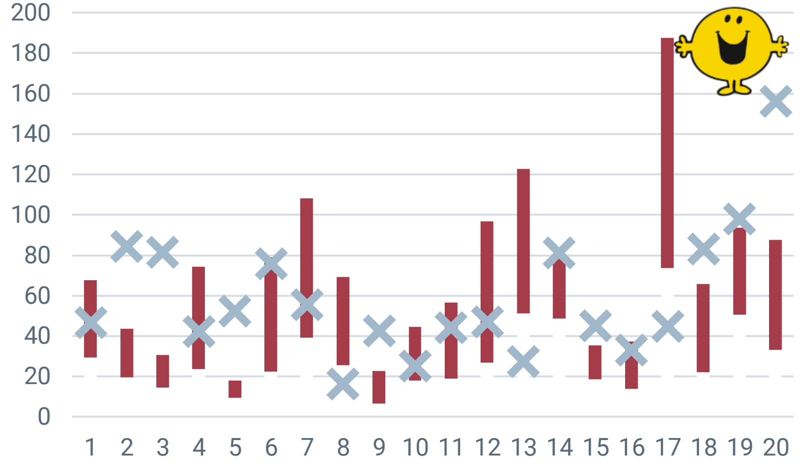
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Happy's sequence

Bias footprint based on 100 assessments



Plot shows **likelihood function** for Mr. Happy's sequence of assessments and actuals

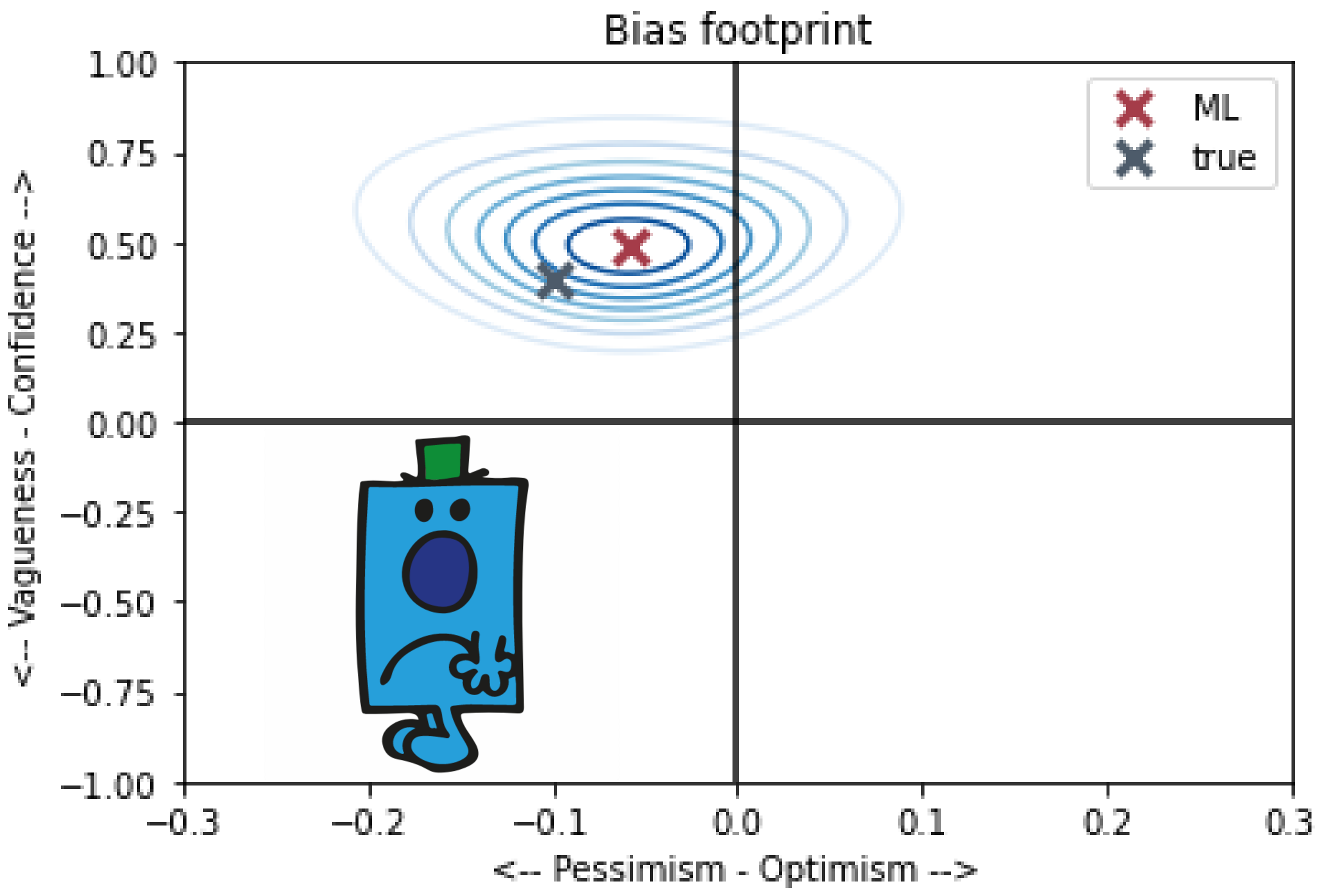


Effectively it shows the probability of seeing the sequence of actuals given Happy's assessments as a function of the bias parameters.

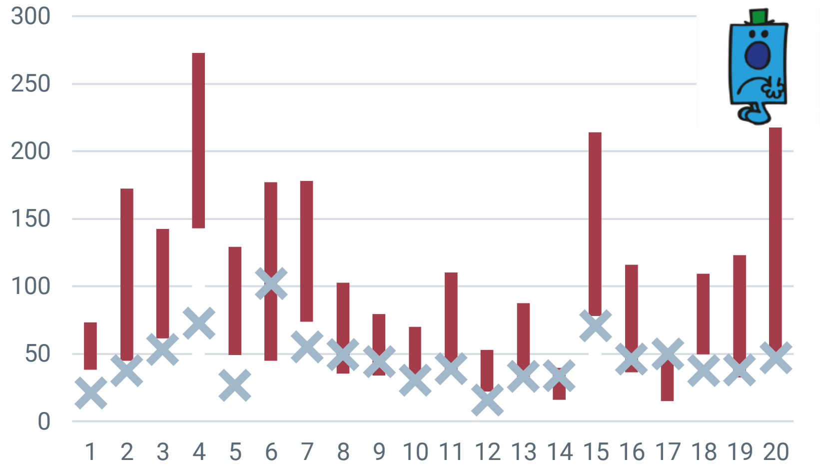
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Happy's sequence

Bias footprint based on 20 assessments



Plot shows **likelihood function** for Mr. Grumpy's sequence of assessments and actuals

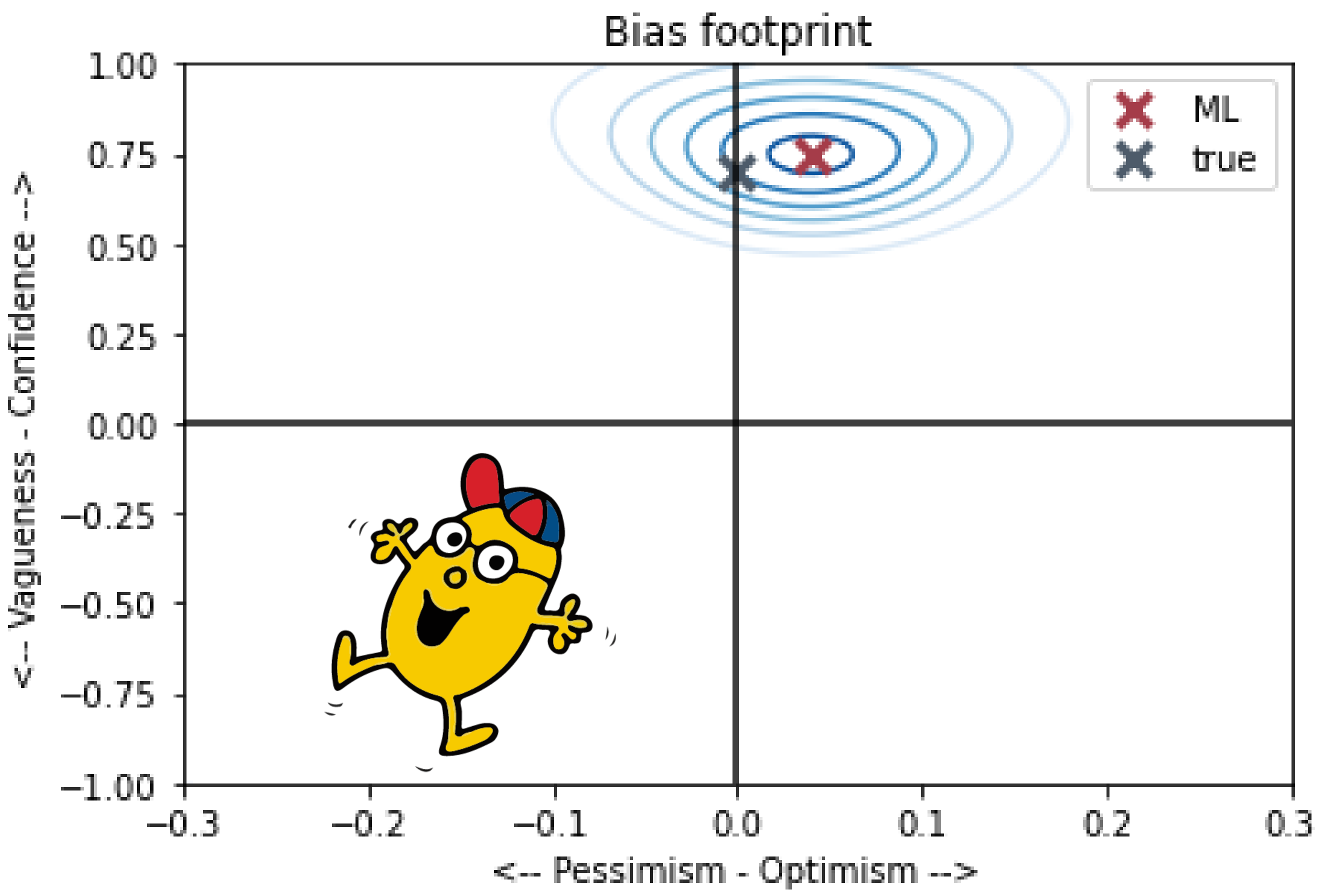


Effectively it shows the probability of seeing the sequence of actuals given Grumpy's assessments as a function of the bias parameters.

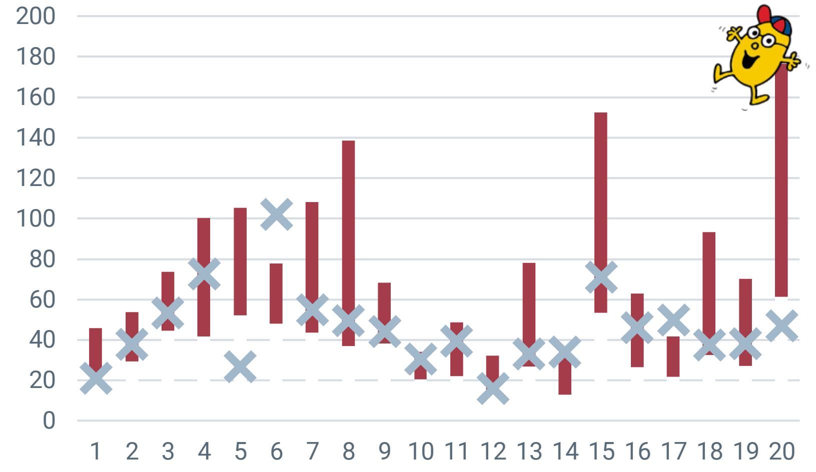
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Grumpy's sequence

Bias footprint based on 20 assessments



Plot shows **likelihood function** for Mr. Brave's sequence of assessments and actuals

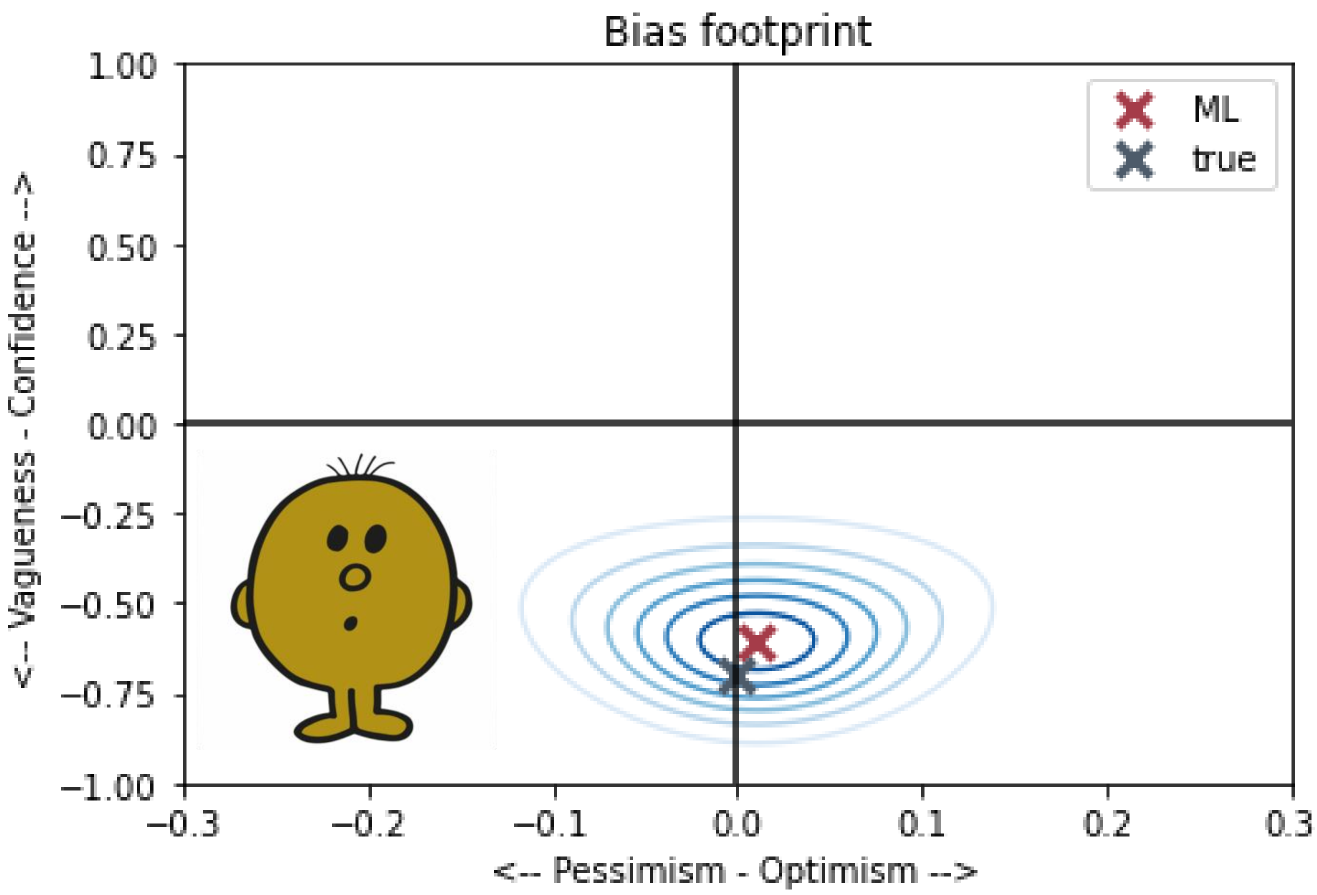


Effectively it shows the probability of seeing the sequence of actuals given Brave's assessments as a function of the bias parameters.

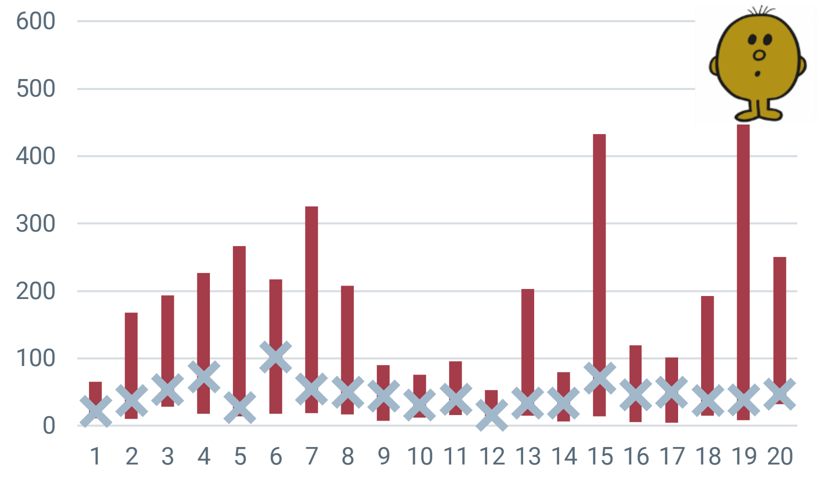
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Brave's sequence

Bias footprint based on 20 assessments



Plot shows **likelihood function** for Mr. Quiet's sequence of assessments and actuals

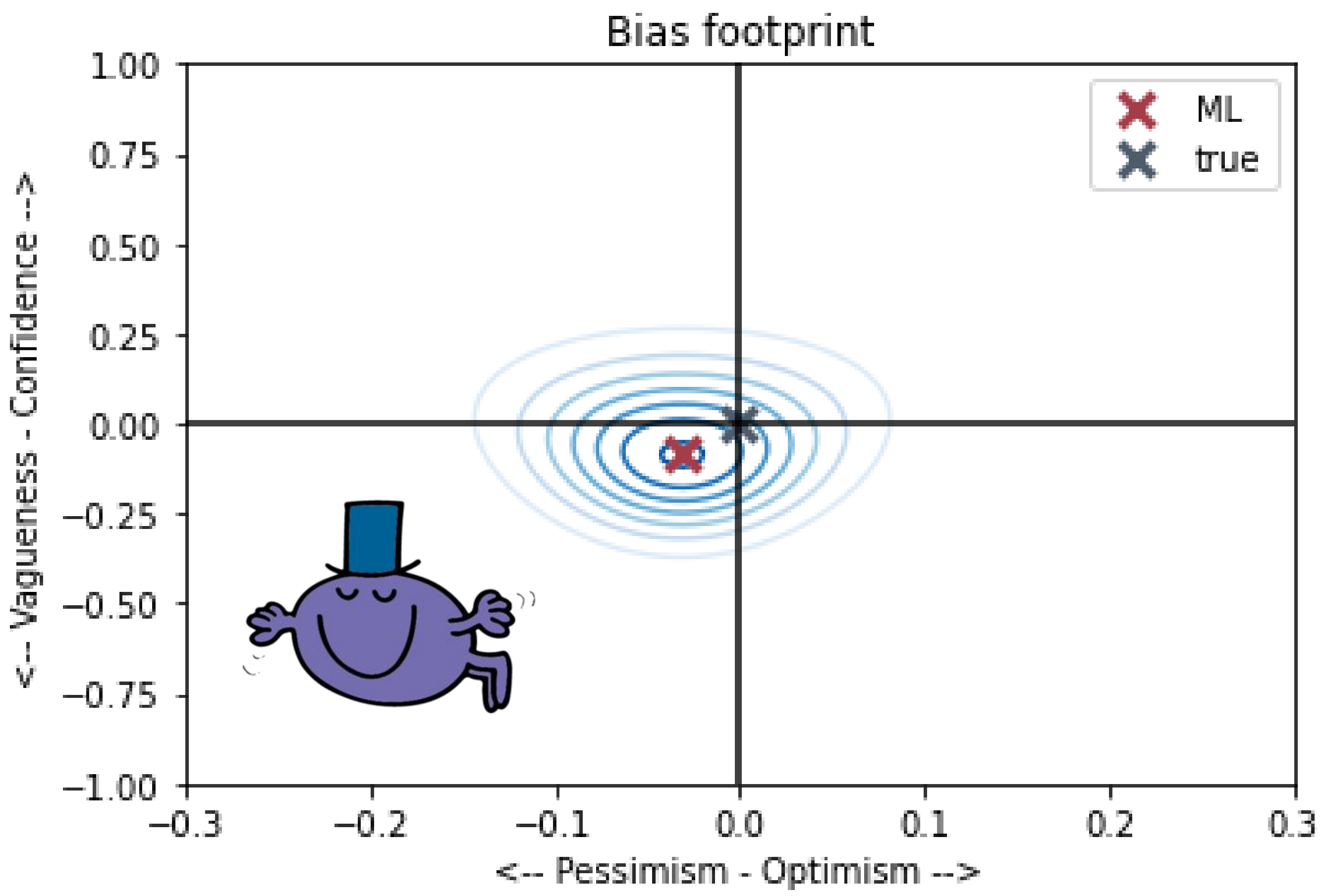


Effectively it shows the probability of seeing the sequence of actuals given Quiet's assessments as a function of the bias parameters.

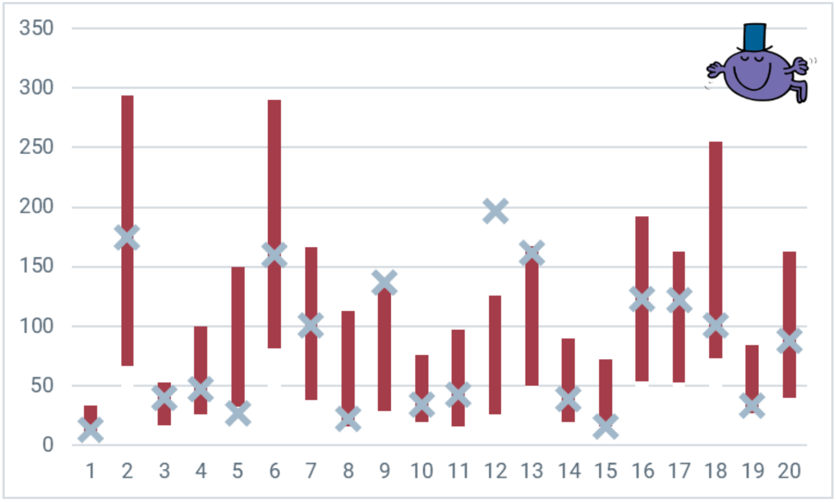
The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Quiet's sequence

Bias footprint based on 20 assessments



Plot shows **likelihood function** for Mr. Impossible's sequence of assessments and actuals



Effectively it shows the probability of seeing the sequence of actuals given Impossible's assessments as a function of the bias parameters.

The red cross is the most likely combination of bias parameters, based on this analysis.

The green cross shows the values of bias parameters used to generate Impossible's sequence

Stochastic ApS
Mathematical modelling for business &
the business of mathematical modelling

Take aways

Bias parameter elicitation

- Quantifies optimism / pessimism
- Quantifies over-confidence / vagueness
- Shows clearly the uncertainty in the assessment

The uncertainty can arise from

- Number of assessments
- Lack of consistency in bias
- (often) both

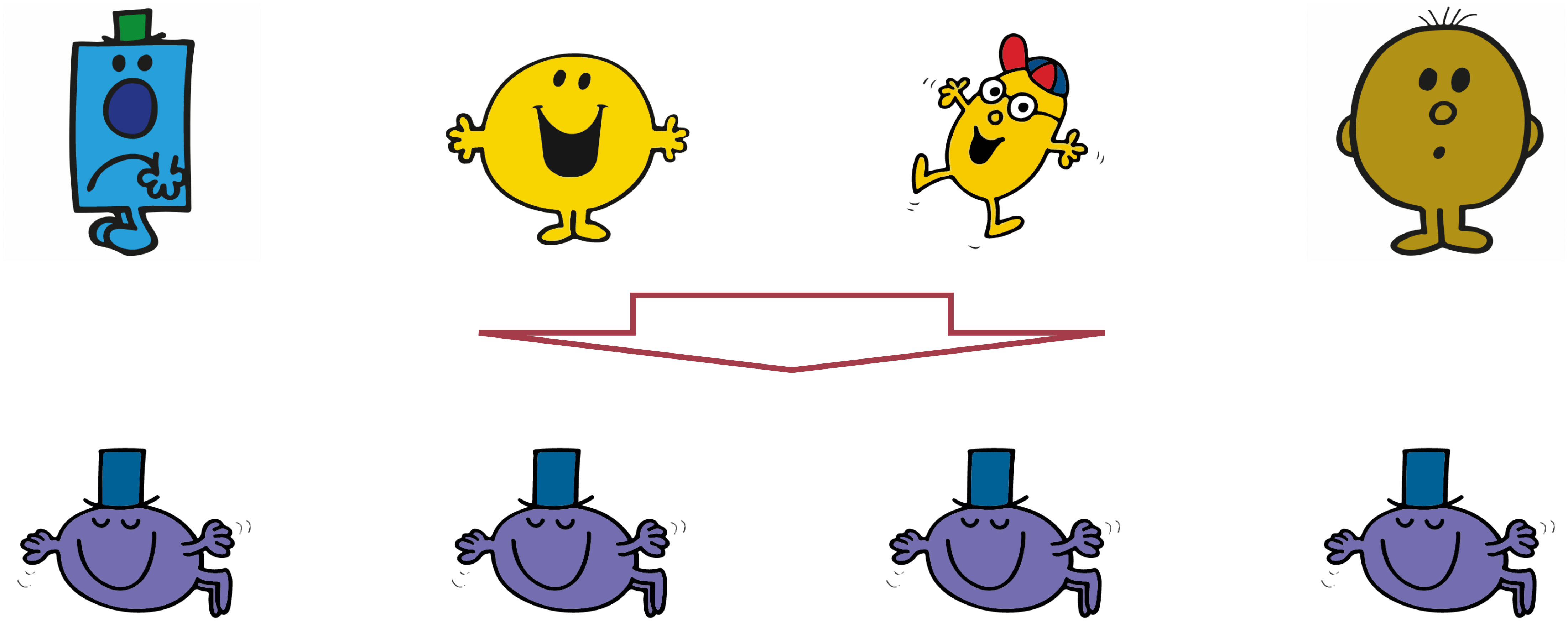
The parameters can be used

- To train SMEs
- To correct SMEs
- As a step in combining SME scores to make better assessments

Aggregation

Bride of FrankenSME

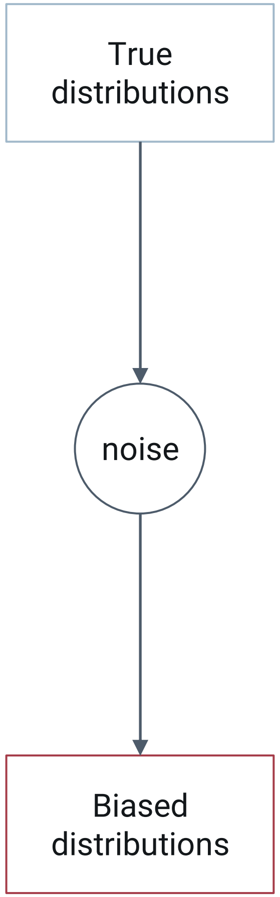
I have now corrected or calibrated my four subject matter experts…



Aggregation

We have removed the bias, but there is still noise

1. Use the same transformation as before
2. Treat the biased distributions as “samples” from the “true” distribution
3. Calculate the likelihood distributions for each of the distribution parameters
    - Effectively calculating the probability of each parameter given that the SME assessments are noisy, but not biased
4. Take the maximum likelihood values as the “true” values
    - Or calculate the expected value of each parameter



FrankenSME and Bride of FrankenSME

Doug Hubbard's FrankenSME method is essentially **empirical**

- Based on and requires a large database of assessments
- Really impressive aggregate performance
- Captures "extra-statistical" effects
- Data gathered for groups of up to seven SMEs

Bride of FrankenSME (this method) is pure statistical

- Requires a single parameter that describes deviation between SMEs (and which can be measured from historical assessments)
- Aggregate performance not as robust
- Makes no account of extra-statistical effects
- Very easy to implement, no solving or seeking
- Extendable to arbitrarily many SMEs