

The following presentation was given at:

# **SCAF Workshop**

## **“Modelling Risk and Uncertainty”**

*Tuesday 20th November 2018*  
*BAWA Centre, Bristol*

Released for distribution by the Author  
[www.scaf.org.uk/library](http://www.scaf.org.uk/library)



# The Good, The Bad & The Ugly

Andy Nolan & Judy Nolan

This information is provided by Rolls-Royce in good faith based upon the latest information available to it; no warranty or representation is given; no contractual or other binding commitment is implied.

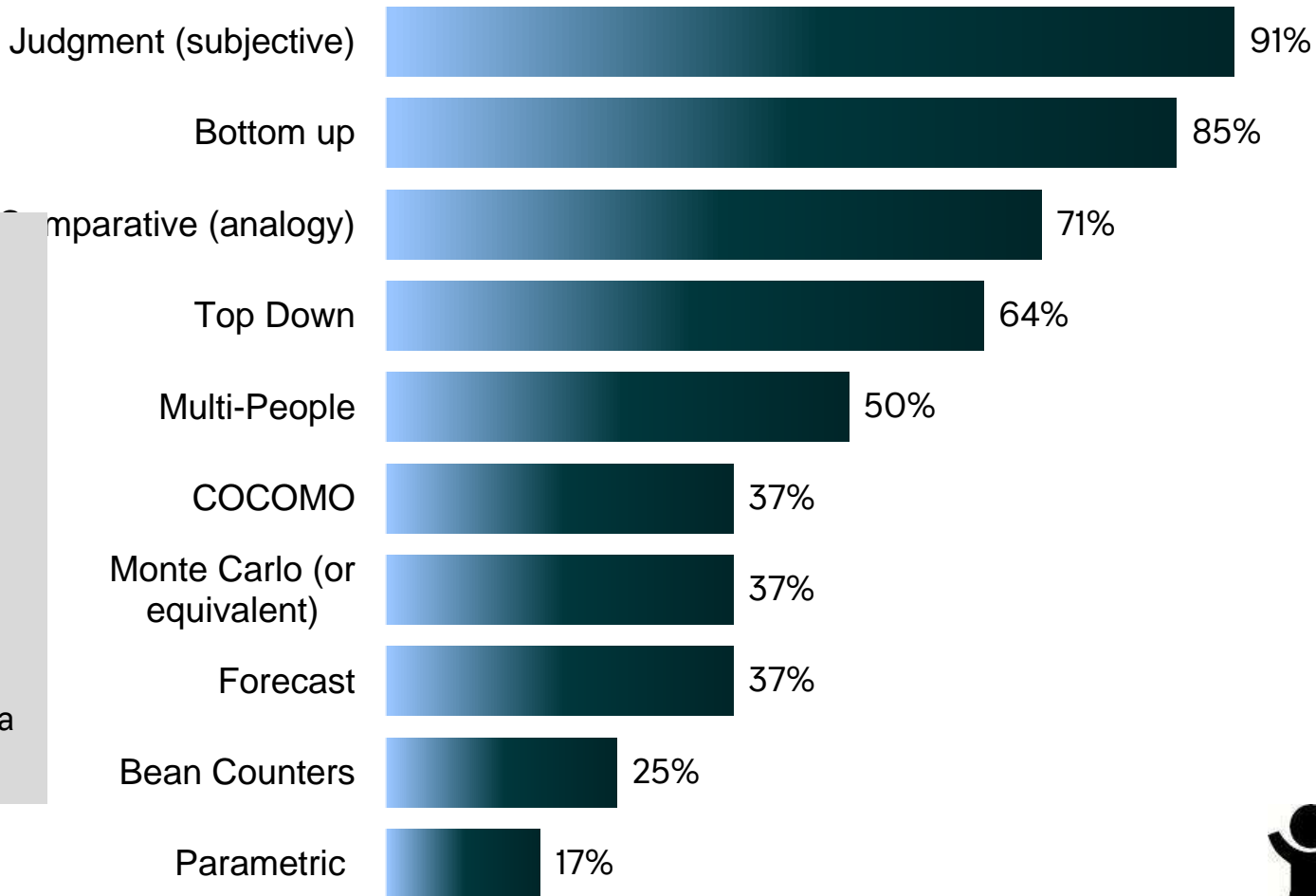


## % Of Population That Used Each Technique

### Who Uses Judgement?

A study in 2014 of over 300 estimates inside Rolls-Royce showed that 91% of us will use some form of judgement when estimating. This is consistent with industry research.

Judgement is not just about guessing, it's also about applying your experience and common sense to know you have a good estimate.

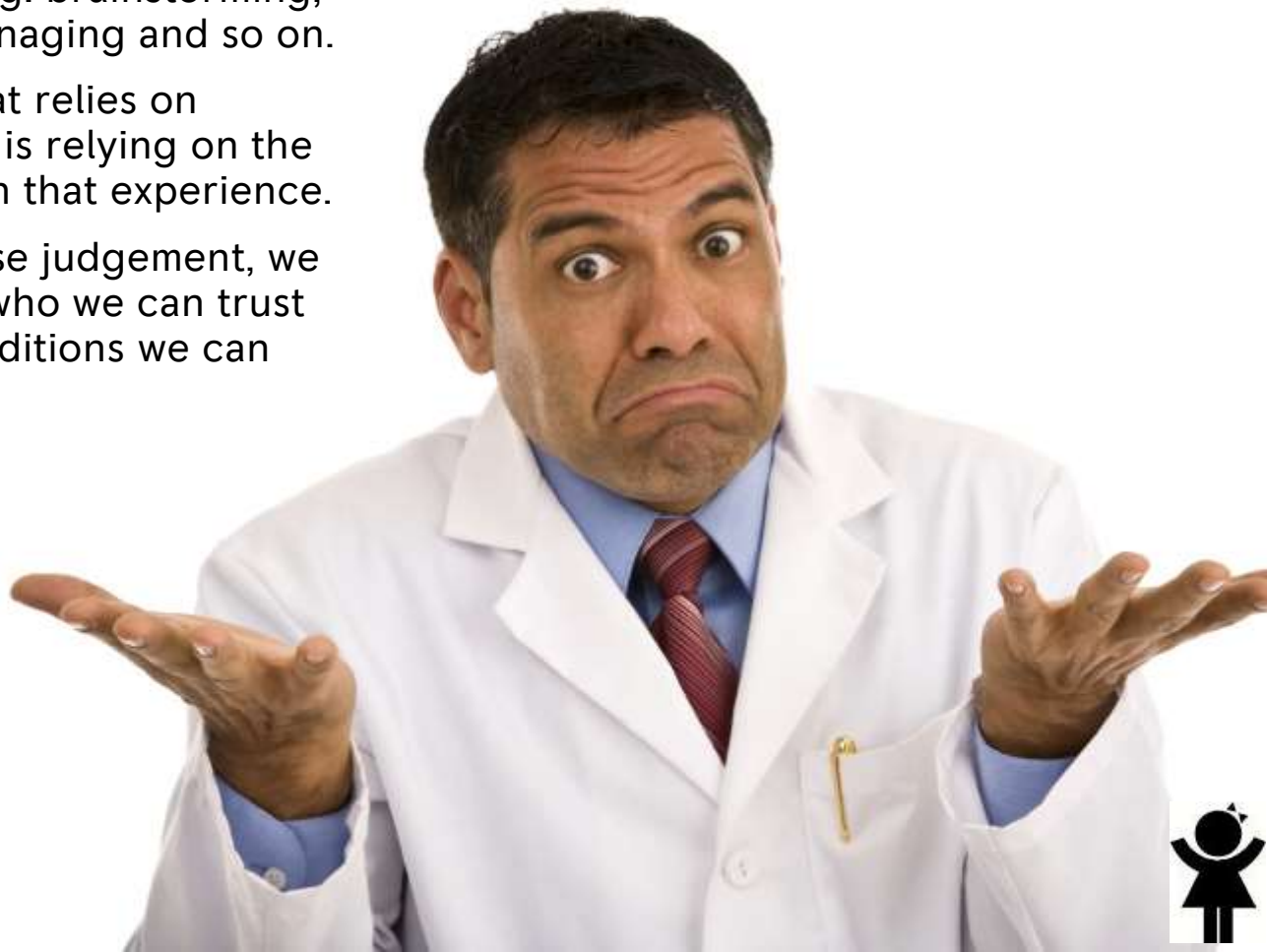




We rely on judgement in many business activities e.g. brainstorming, problem solving, managing and so on.

Any job / activity that relies on experienced people is relying on the Judgement based on that experience.

If we are going to use judgement, we might as well learn who we can trust and under what conditions we can trust them





## Are some people good at it?

In 2016/18, we asked people to estimate things like the height of mountains & buildings, the populations of countries....and got 6168 responses from 578 people.

The table shows the average scores for each person. This enables us to identify a person's consistent judgement style.

Confidence vs % Error		Confidence Level									
		0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100
% Error	-5% - 5%	0	1	0	0	0	0	0	0	0	0
	-10% - 10%	0	0	0	1	1	0	1	1	0	0
	-17% - 20%	1	0	0	3	3	2	3	0	0	0
	-29% - 40%	0	1	2	6	14	18	4	6	3	0
	-45% - 80%	2	3	12	21	47	37	15	3	0	0
	-62% - 160%	1	13	21	50	61	23	9	1	1	0
	-77% - 320%	5	9	27	33	32	14	6	1	2	0
	-87% - 640%	1	9	4	21	8	2	2	2	0	0
	-93% - 1280%	1	2	3	1	0	1	0	0	0	0
	-96% - 2560%	0	0	0	1	0	0	0	0	0	0

For each guess, the estimator had to say how confident they were in their answer where 0 means that had no idea and 100 meant the estimator was certain.

The estimate % error was calculated by comparing the guesses against the correct answers - as defined in Google and Wikipedia

The chart shows the average confidence plotted against the average % error for each person. This enables us to identify a persons consistent level of accuracy.

Consistently Over Confident

Confidence and %Error align

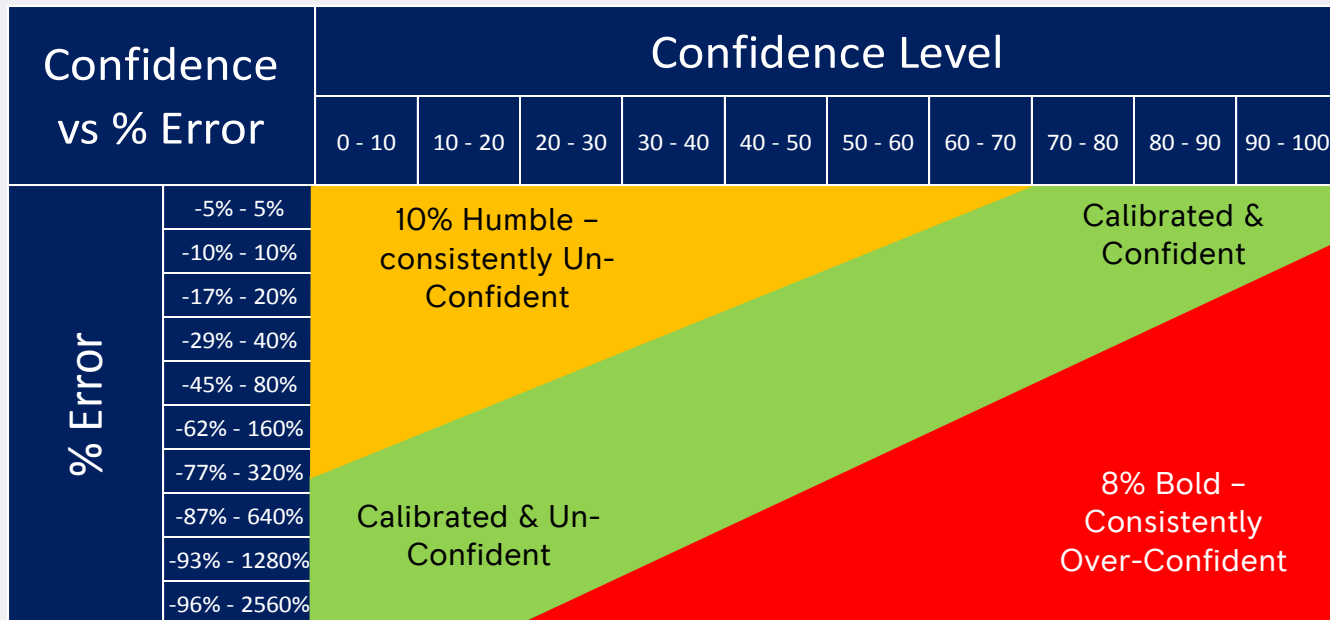
Consistently Under confident





## Are some people good at it?

From the last table, we overlaid 3 types of people, the Calibrated (green), the Humble (amber) and the Bold (red) based on their % error in relation to their confidence



Calibrated (Green): 82% of those who responded were calibrated i.e. their confidence and there % error aligned i.e. when they were confident their error was small and when they were un confident there error was high.

Humble (Amber): 10% of people lay in the Amber zone where their error was less than their confident i.e. they were better then they thought they were.

Bold (Red) 8% of people lay on the red area where there error was higher than there confidence i.e. they thought they knew but did not





## A “Test” for Calibration

We wanted to understand if this relationship could be tested for and quantified, so we could easily identify a person’s level of calibration

To determine a person’s Calibration, we asked each person to complete a 20 question test.

Each question needed a yes/no response and each person had to score their Confidence that they got the right answer.

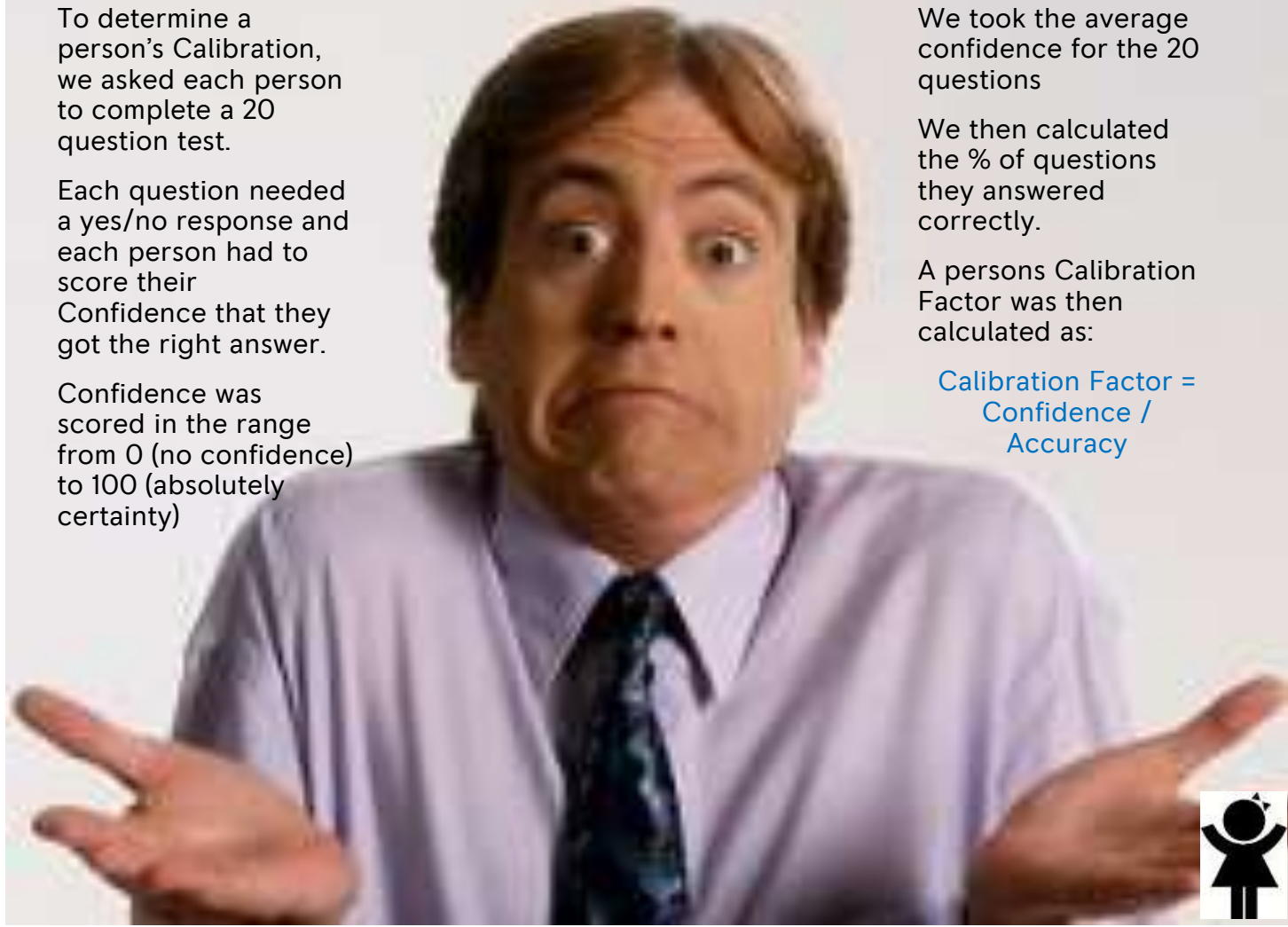
Confidence was scored in the range from 0 (no confidence) to 100 (absolutely certainty)

We took the average confidence for the 20 questions

We then calculated the % of questions they answered correctly.

A persons Calibration Factor was then calculated as:

$$\text{Calibration Factor} = \frac{\text{Confidence}}{\text{Accuracy}}$$



		Y/N	Conf
1	The Eiffel Tower is taller than the Empire State Building?	Yes	6
2	A Tennis ball is bigger in diameter than a Cricket ball	No	3
3	On average cats lives longer than dogs	Yes	7
4	Titanium has a higher melting point then stainless steel	No	5
5	More people live in Germany than Japan		
6	A golf ball is bigger (diameter) than a table tennis ball		
7	Ice is more dense than water		
8	The largest animal in the world is a mammal		
9	A ice hockey puck will fit in a golf hole		
10	The deepest point in any ocean is in the Pacific Ocean		





		Y/N	Conf
1	The Eiffel Tower is taller than the Empire State Building?	No	
2	A Tennis ball is bigger in diameter than a Cricket ball	No	
3	On average cats lives longer than dogs	Yes	
4	Titanium has a higher melting point then stainless steel	Yes	
5	More people live in Germany than Japan	No	
6	A golf ball is bigger (diameter) than a table tennis ball	Yes	
7	Ice is more dense than water	No	
8	The largest animal in the world is a mammal	Yes	
9	A ice hockey puck will fit in a golf hole	Yes	
10	The deepest point in any ocean is in the Pacific Ocean	Yes	



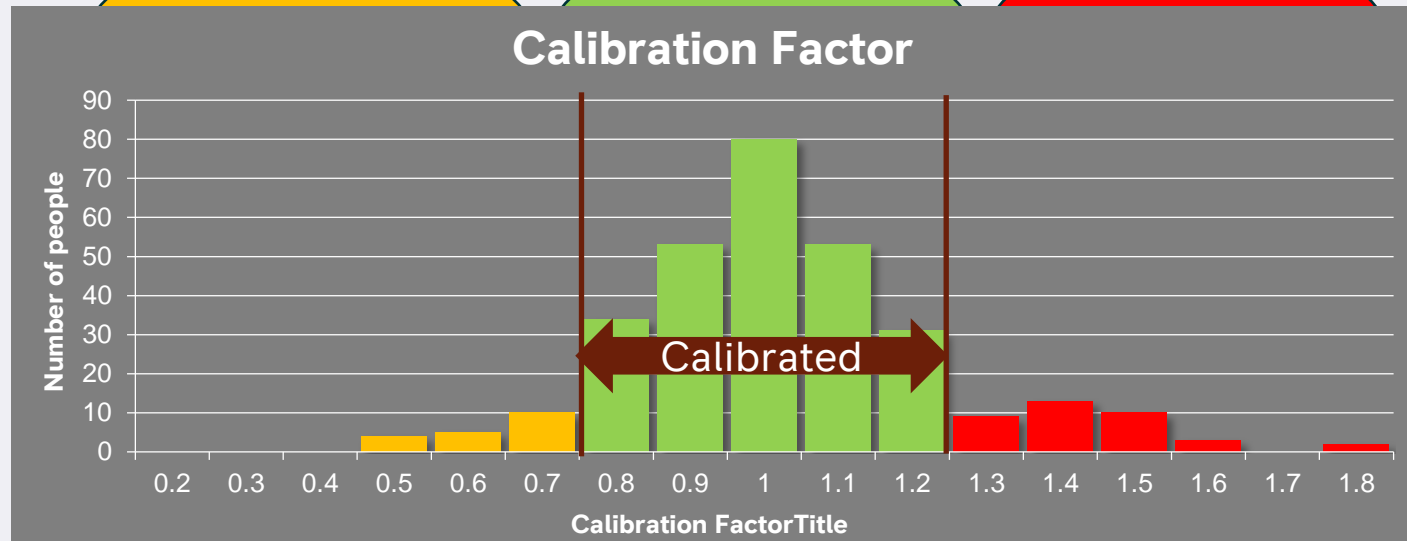
A photograph of several young children sitting at desks in a classroom, focused on writing with pencils. The children are diverse in appearance. The text is overlaid in a large, yellow, sans-serif font.

Calibration Factor =  
Confidence / Accuracy





## Calibration Factor = Confidence / Accuracy



### Humble

Calibration Factor < 0.8. A person whose Confidence is lower than their Accuracy and this means a person got more answers right than they believed they should. For example, if they had a Confidence of 25% but got 50% of the questions right, they would have a Calibration Factor of 0.5.

### Calibrated

Calibration Factor between 0.8 and 1.2. A person whose Confidence matches their Accuracy. If a person had an average Confidence of 50% and they got 50% of the answers right, they would have a Calibration Factor of 1. So would a person who was 30% Confident and 30% Accurate. Also for a person who was 80% Confidence and 80% Accurate.

### Bold

Calibration Factor > 1.2. A person whose Confidence is higher than their Accuracy and this means a person got less answers right than they believed they should. For example, if they had a Confidence of 80% but got 40% of the questions right, they would have a Calibration Factor of 2.





## So how does calibration affect judgement accuracy



We then asked people to guess at general knowledge questions

Each person had to score their confidence in the answers they gave.

The research was to see if there was a relationship between Calibration Factor and estimate accuracy and as you will see, there is a good correlation...

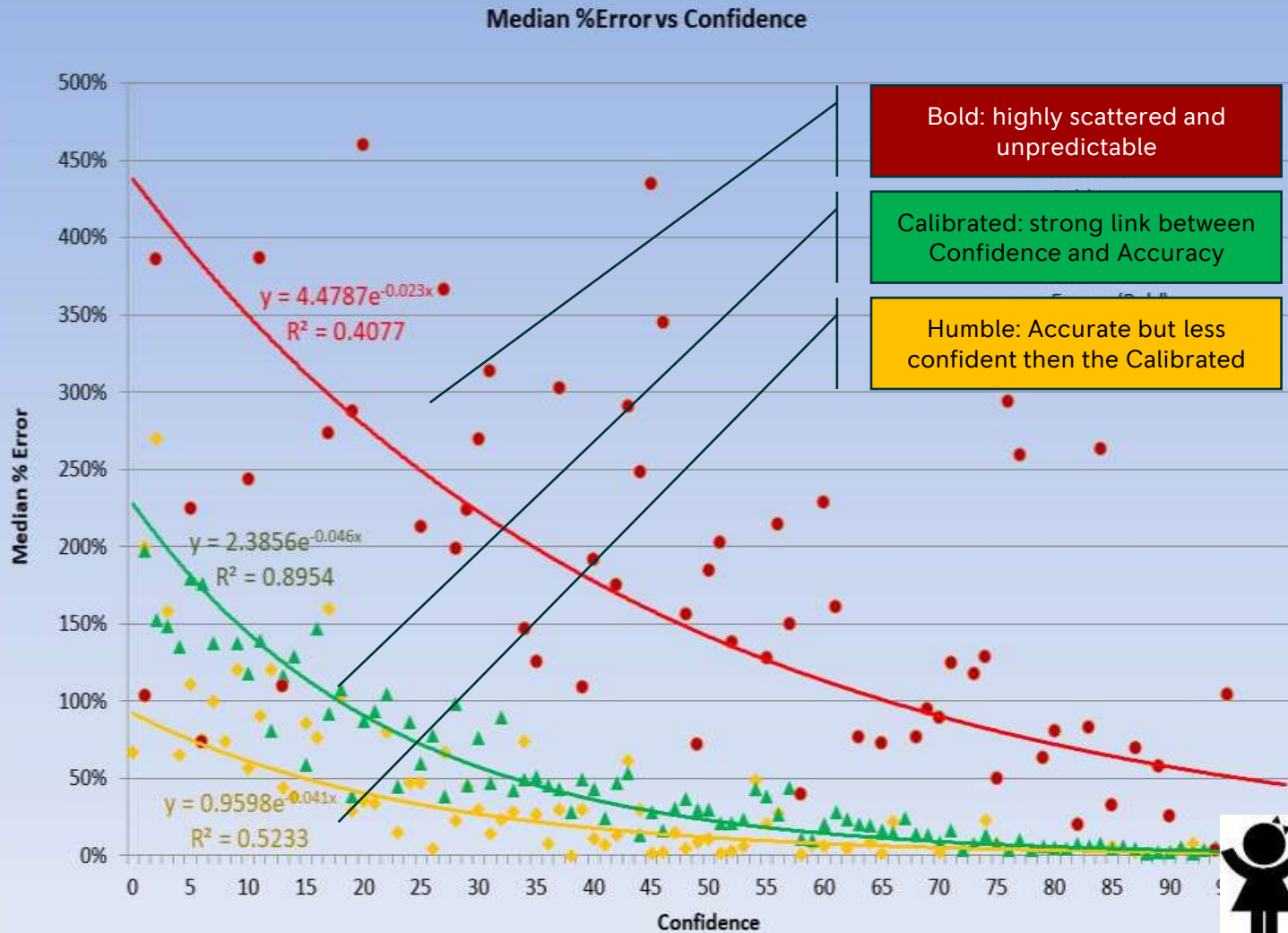




## The Bold, the Humble and the Calibrated

The chart shows the guesses from 205 people, split into the 3 groups of **Bold**, **Humble** and **Calibrated**.

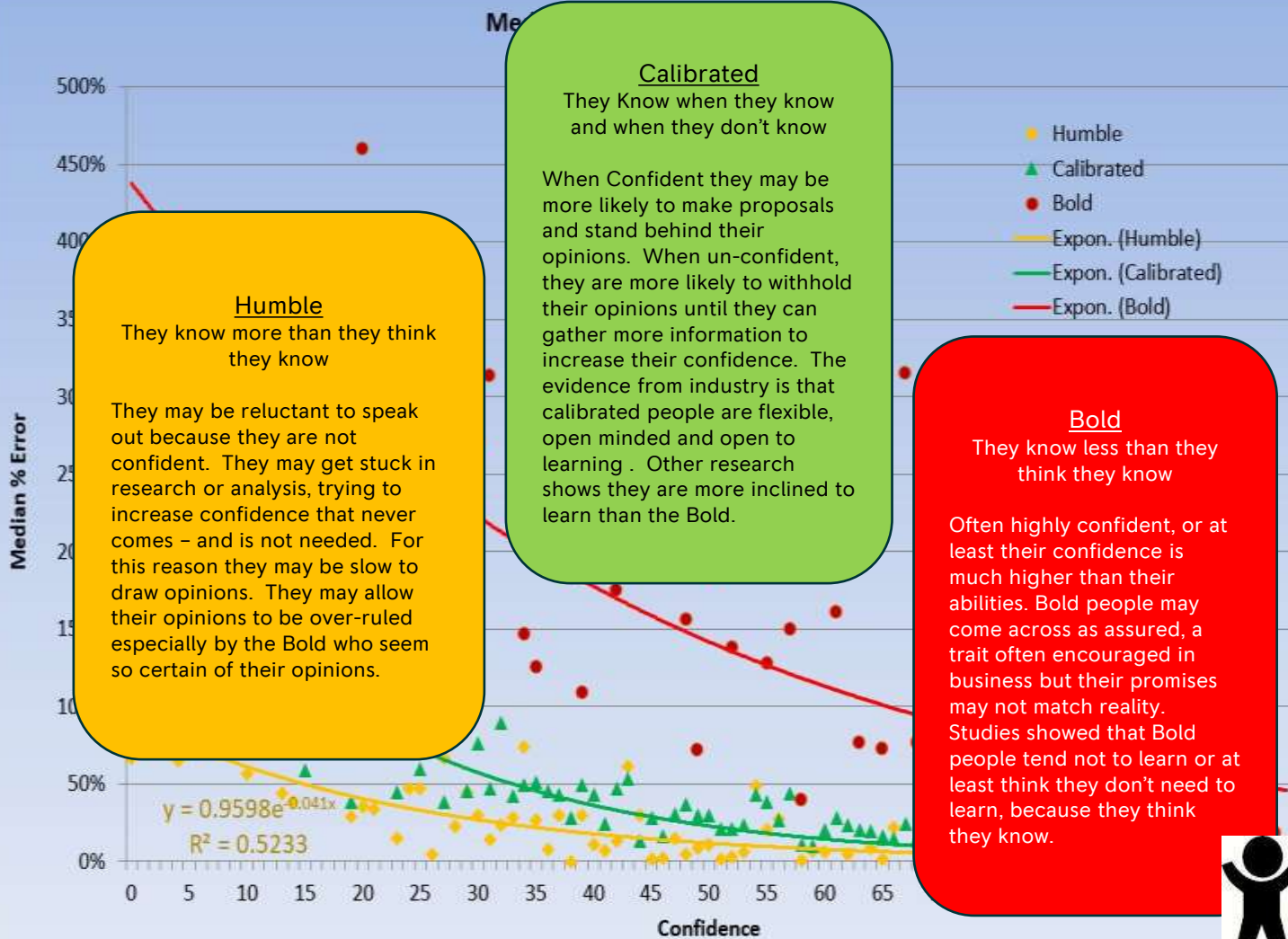
The chart shows horizontally the Confidence of each persons guess and vertically the % error of their guesses. Each data point represents the median value of people with the same confidence.

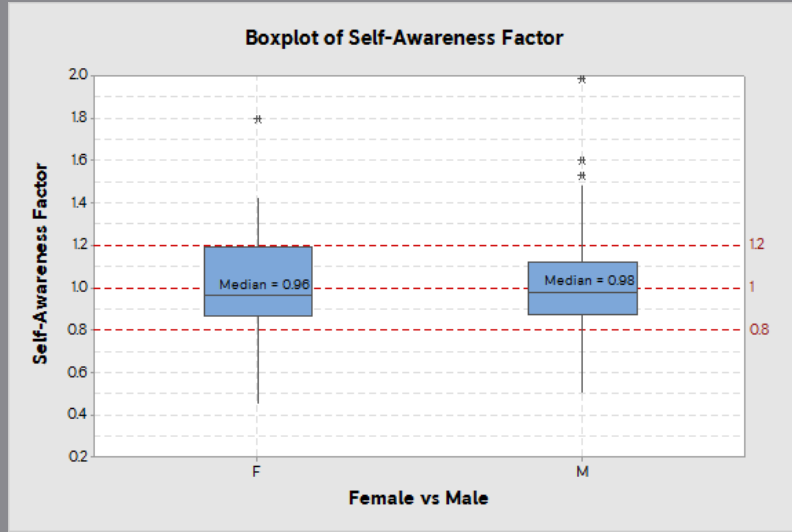




## What does this mean?

What does this research and other research from industry tell us about your judgement skills?

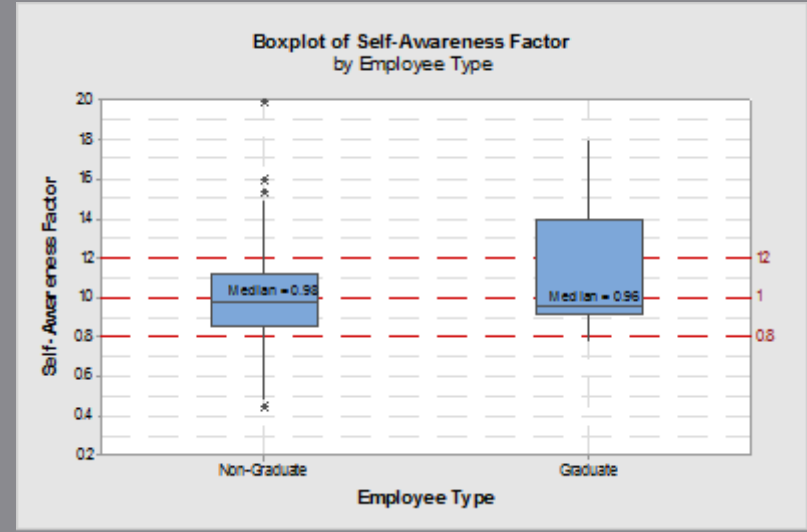




## Female vs Male

There is no statistical difference between Female and Male Self-Awareness factor – neither group is more self-aware

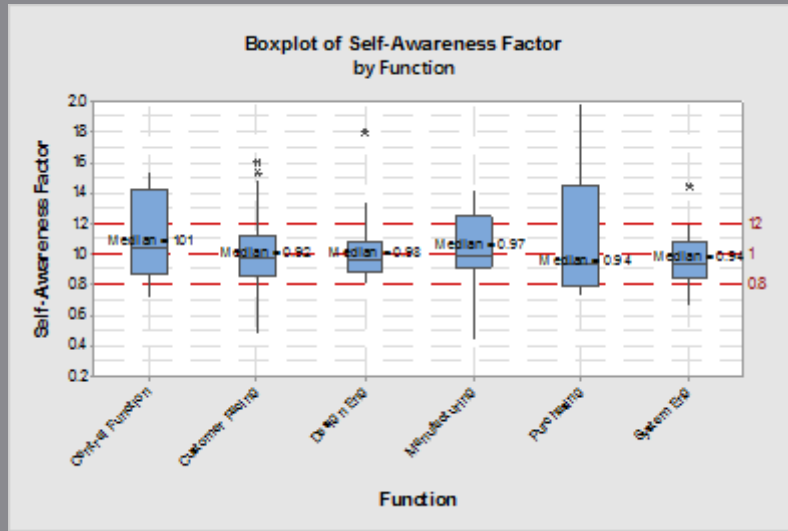
Neither group has a higher % of bold



## Post Grad Scheme vs Grad scheme

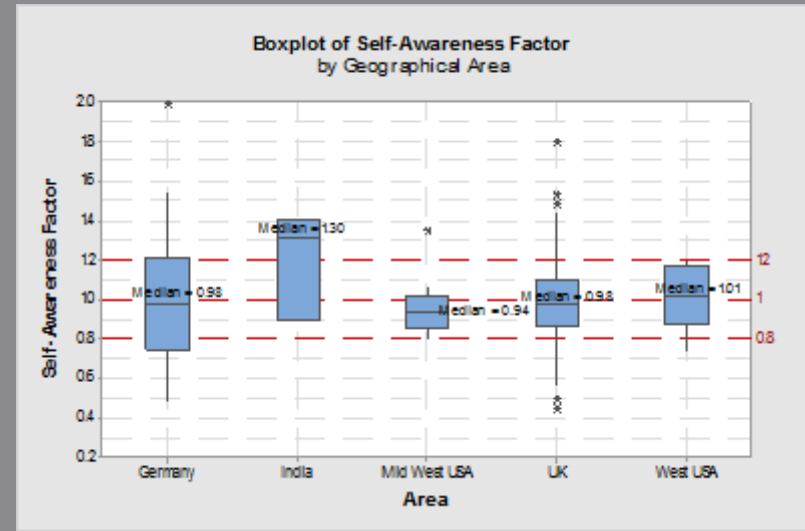
There is not statistical difference, but people on still the graduate scheme seem bolder





## Functional

Central functions and Purchasing seem to be biased to above 1, so are bolder in nature




## Geographical location

There are some differences between geographical locations. These are not statistically significant, but should be taken into account when considering an estimate





A background image of an archer's hands holding a yellow bow, with a target visible in the blurred background.

# Accuracy = Confidence × Calibration

Confidence alone does not mean a person will be accurate, they need to be calibrated as well.

You also may think its better to be Humble but it is better to be Calibrated than Humble.



Like an archer that learns from each shot, it is possible to improve your Calibration by leaning from your results and taking another test.

Of the 31 volunteers who used the latest calibration tests, we saw an overall improvement in Calibration scores from repeated testing.

**Test 1**  
Take test 1, look at your results and then re-calibrate yourself

**Test 2**  
Take test 2, look at your results and then re-calibrate yourself

**Test 3**  
Take test 3, look at your results and then re-calibrate yourself

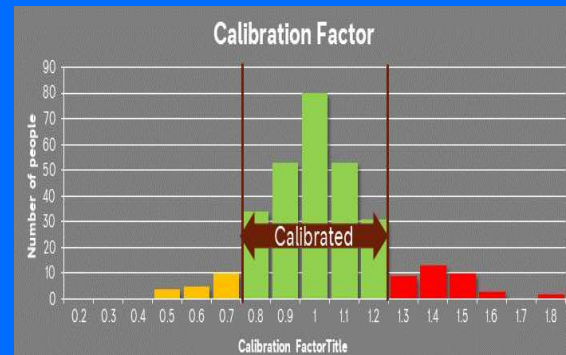


# Conclusions

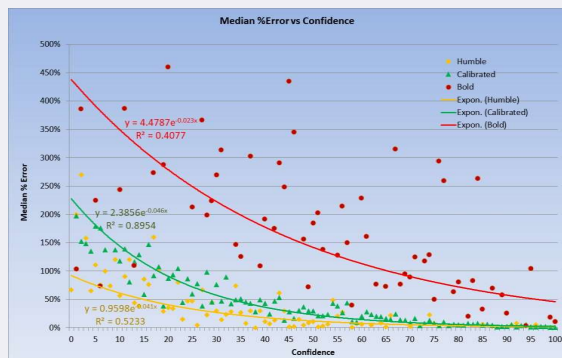
## 1 Not everyone is good at judgement



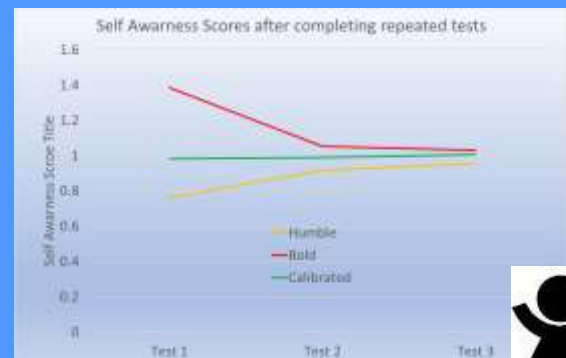
## 2 Calibration = Confidence / Accuracy



## 3 Accuracy = Confidence \* Calibration



## 4 Calibration can be learned





# Any questions?

