Returns to Precision

Barbara Mellers
University of Pennsylvania

IARPA hosted 4 forecasting tournaments from 2011 to 2015

Will Venezuela experience a domestic political crisis between December 2013 and April 2014?

Before April 2015, will foreign national military forces launch a ground offensive against the Houthi rebels in Yemen?

Will the six-party talks with North Korea resume before 1 May 2014?

The Brier Scoring Rule was the Accuracy Metric

0

Perfect theory of deterministic system

.5

Just Guessing

2

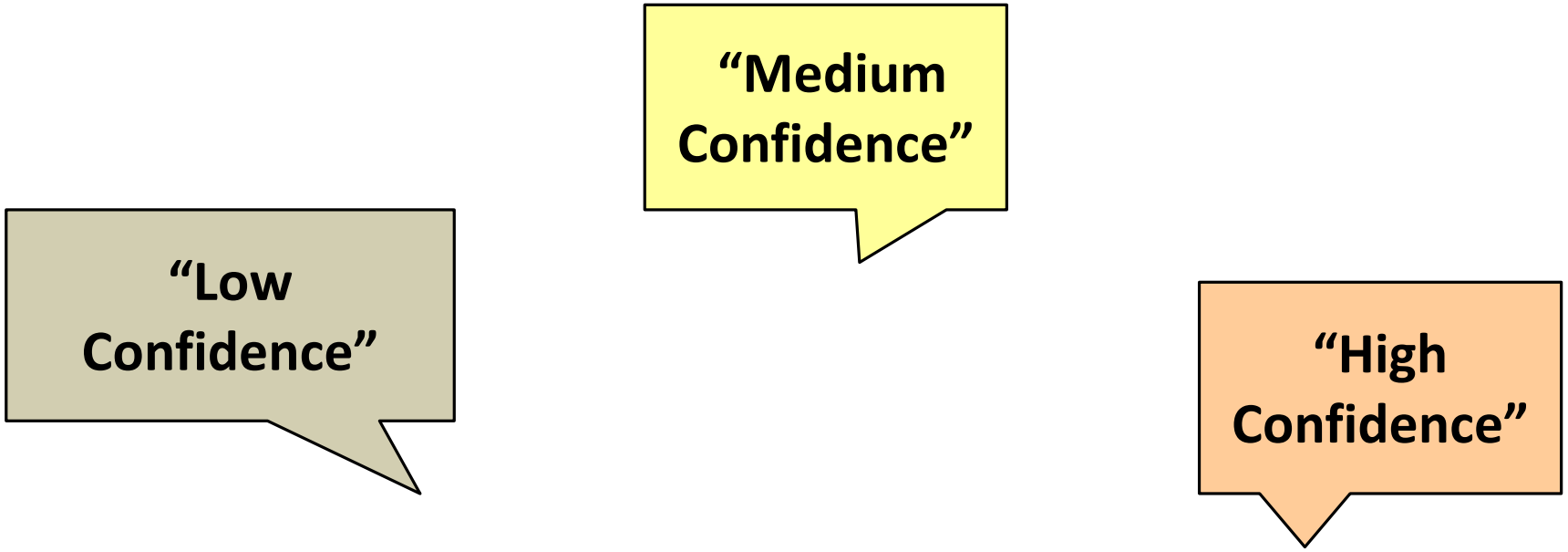
Reverse Clairvoyance

Brier Scores for a Meteorologist

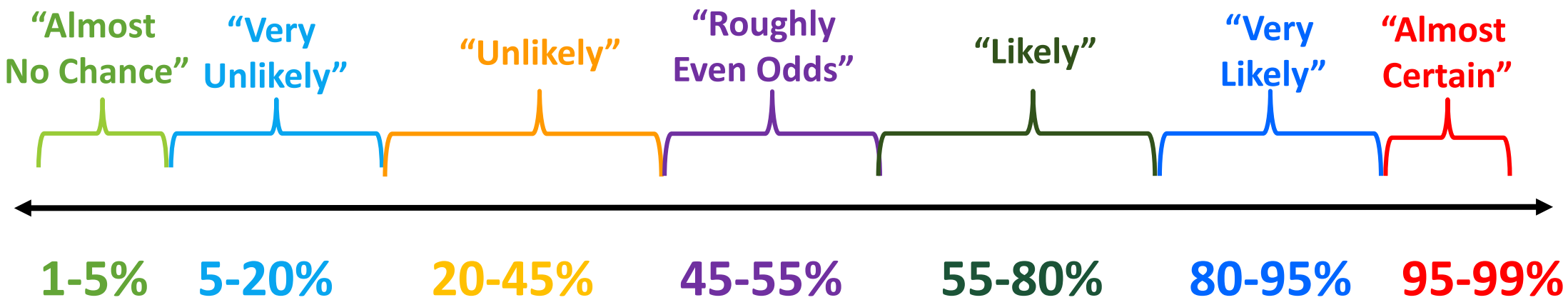
Day	"Rain"	Rain	Brier Score
1	90%	1	$(1-.9)^2 + (0-.1)^2 = 0.02$
2	40%	0	$(0-.6)^2 + (1-.4)^2 = 0.72$
3	90%	0	$(0-.9)^2 + (1-.1)^2 = 1.62$
	73%	33%	0.75

How Do Analysts Make Forecasts?

How Do Analysts Make Forecasts?



Current Analytic Standards



1. Verbal uncertainty phrases mean different things to different people.

Plausible Ranges of Analytic Standards Given by GJ Forecasters

2. You can’t score verbal uncertainty phrases for accuracy. Without an accuracy metric, it is impossible to learn how to do better.

Steam Engines

Harnessing steam power required many innovations, among the most important was a new way to measure the energy output of engines.

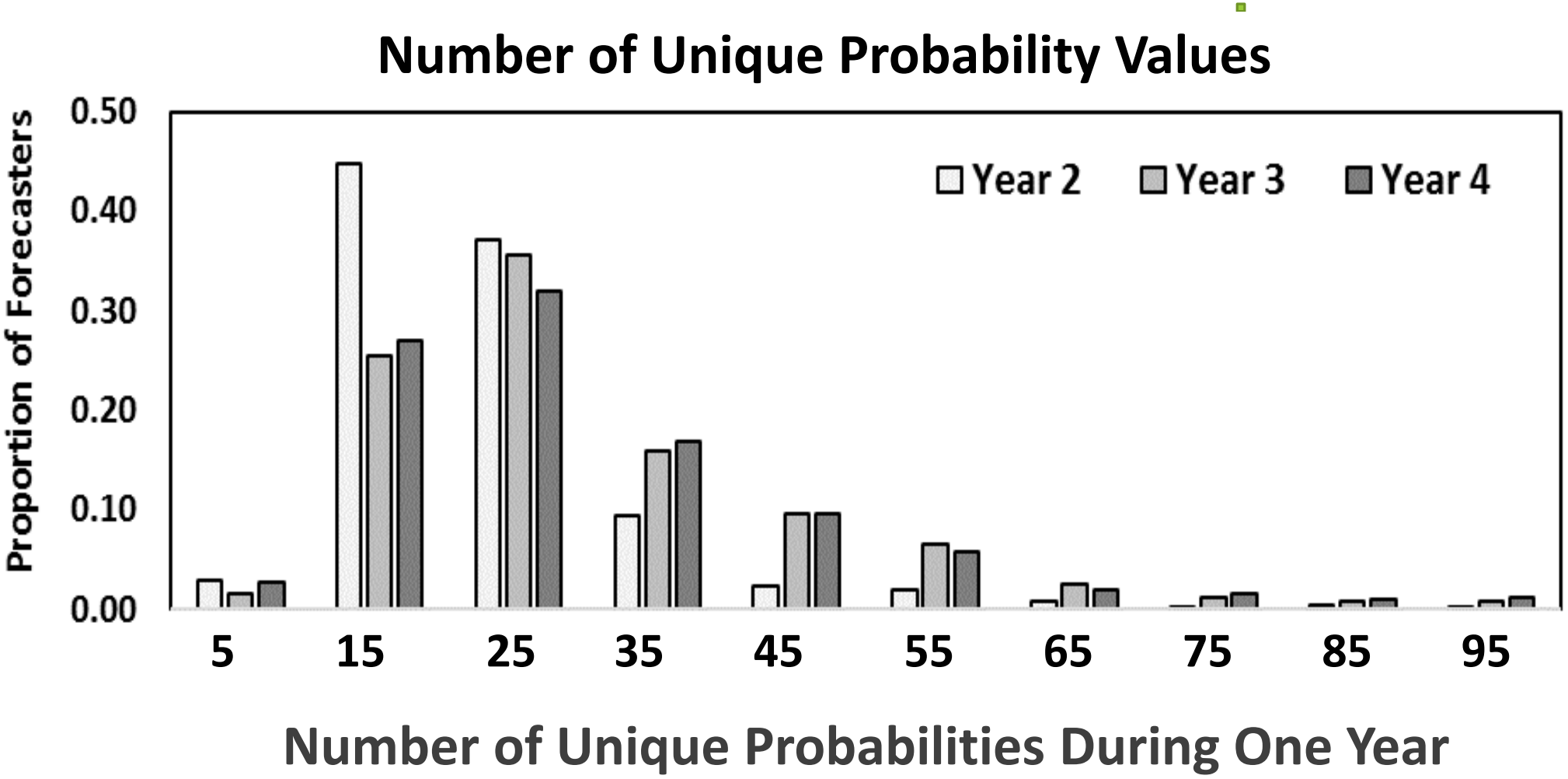
Measurements were done with a tool called a micrometer (aka the “Lord Chancellor”) that measured tiny distances and allowed innovators to evaluate their ideas.

“I have been struck by how important measurement is to improving the human condition. You can achieve incredible progress if you set a clear goal and find a measure that will drive progress toward that goal…”

Do the current Analytic Standards provide enough gradations for analysts to convey their beliefs?

How many distinctions did GJ Forecasters make when stating their beliefs?

How many distinctions do GJ Forecasters need to state their beliefs?


Number of Unique Probability Values
Proportion of Forecasters
0.50
0.40
0.30
0.20
0.10
0.00
Year 2
Year 3
Year 4
5
15
25
35
45
55
65
75
85
95
Number of Unique Probabilities During One Year

Effects of Rounding to 3 and 7 Categories

Forecasters	Original Brier Scores	Confidence Ratings (3)	Verbal Unequal Intervals (7)	Phrases Equal Intervals (7)
All Forecasters	0.153	11.8%***	0.7%***	1.9%***
Super Forecasters	0.093	236.1%***	6.1%***	40.4%***

Note: *** $p < 0.001$.

Thanks