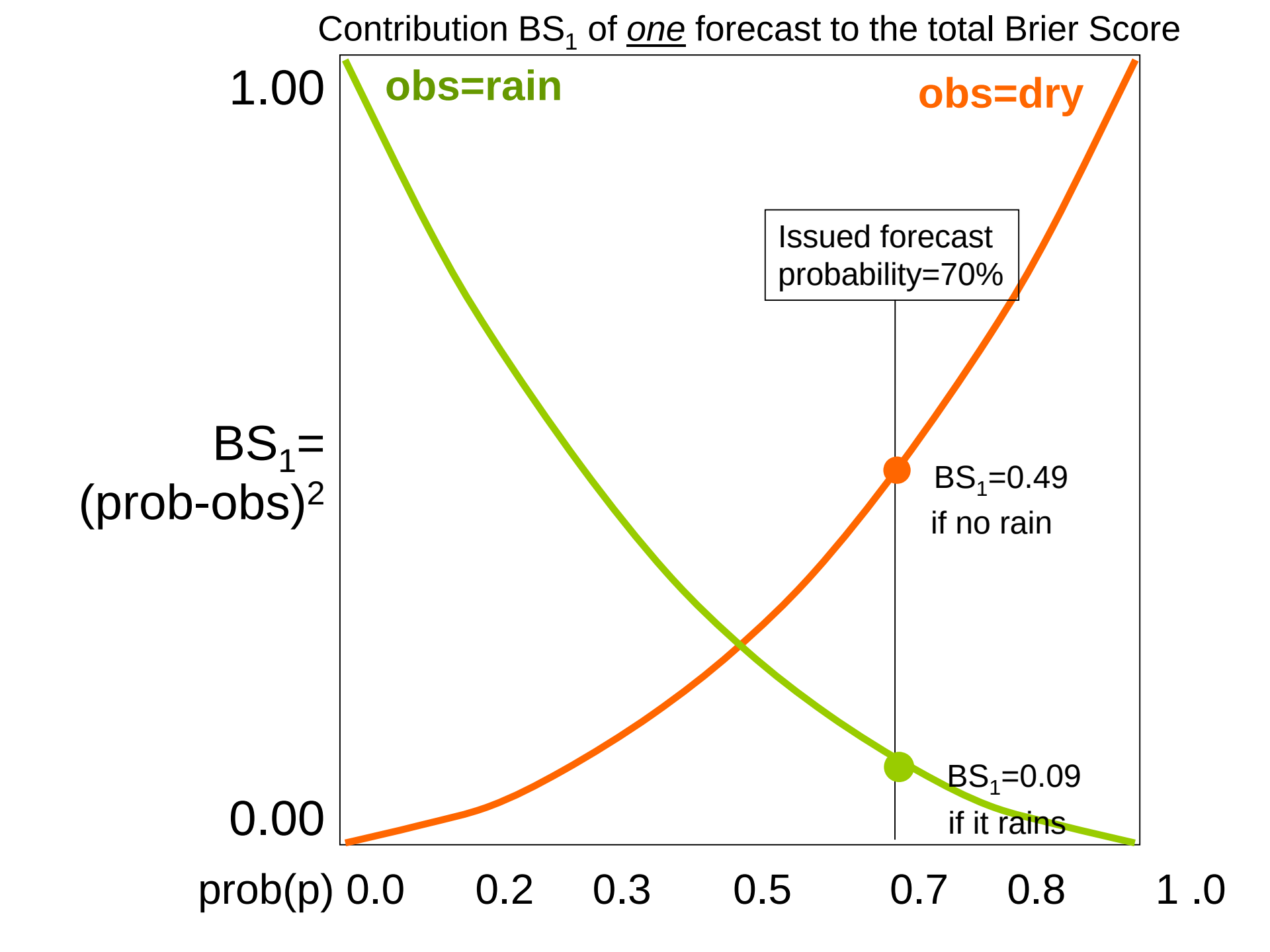
II. Frequentist probabilities

II.3 Forecast system validation

II.3.1 The “strictly proper” Brier Score


Contribution BS_1 of *one* forecast to the total Brier Score
1.00
obs=rain
obs=dry
Issued forecast
probability=70%
BS_1=
$(prob-obs)^2$
BS_1=0.49
if no rain
BS_1=0.09
if it rains
0.00
prob(p) 0.0 0.2 0.3 0.5 0.7 0.8 1 .0

The Brier Score will "punish" you if you, a reliable probability forecaster, put a probability you do not really believe in.

1. You believe the probability is 50% but think people will misunderstand and therefore put 40% or 60%
2. You believe the probability for very severe weather is 40% but in order to make people stay at home you issue a 80% warning
3. You believe (wrongly) that it is tactical to nudge towards the climatological probability

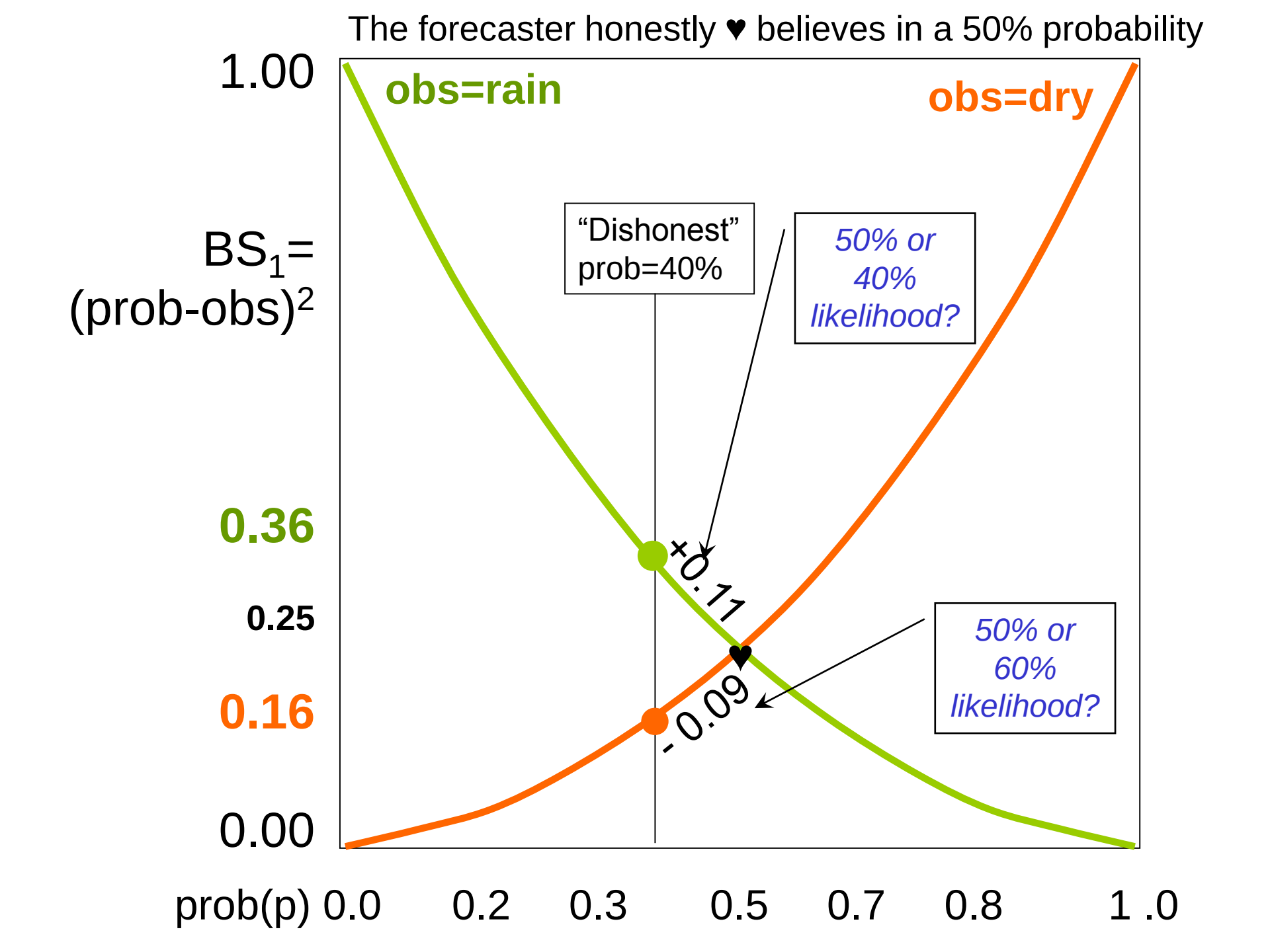
-How does the “proper” Brier score (BS) “know” my true opinion?

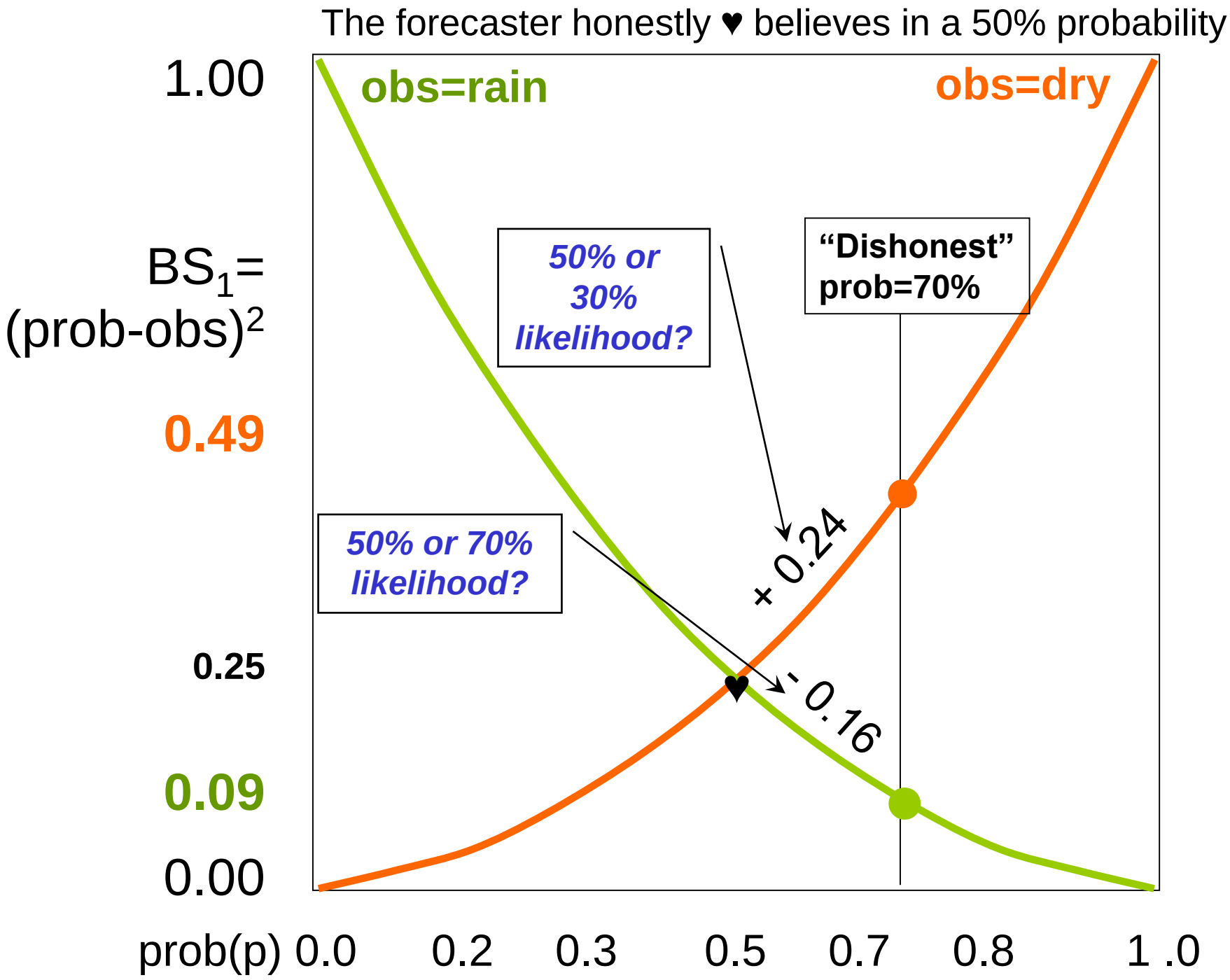
$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

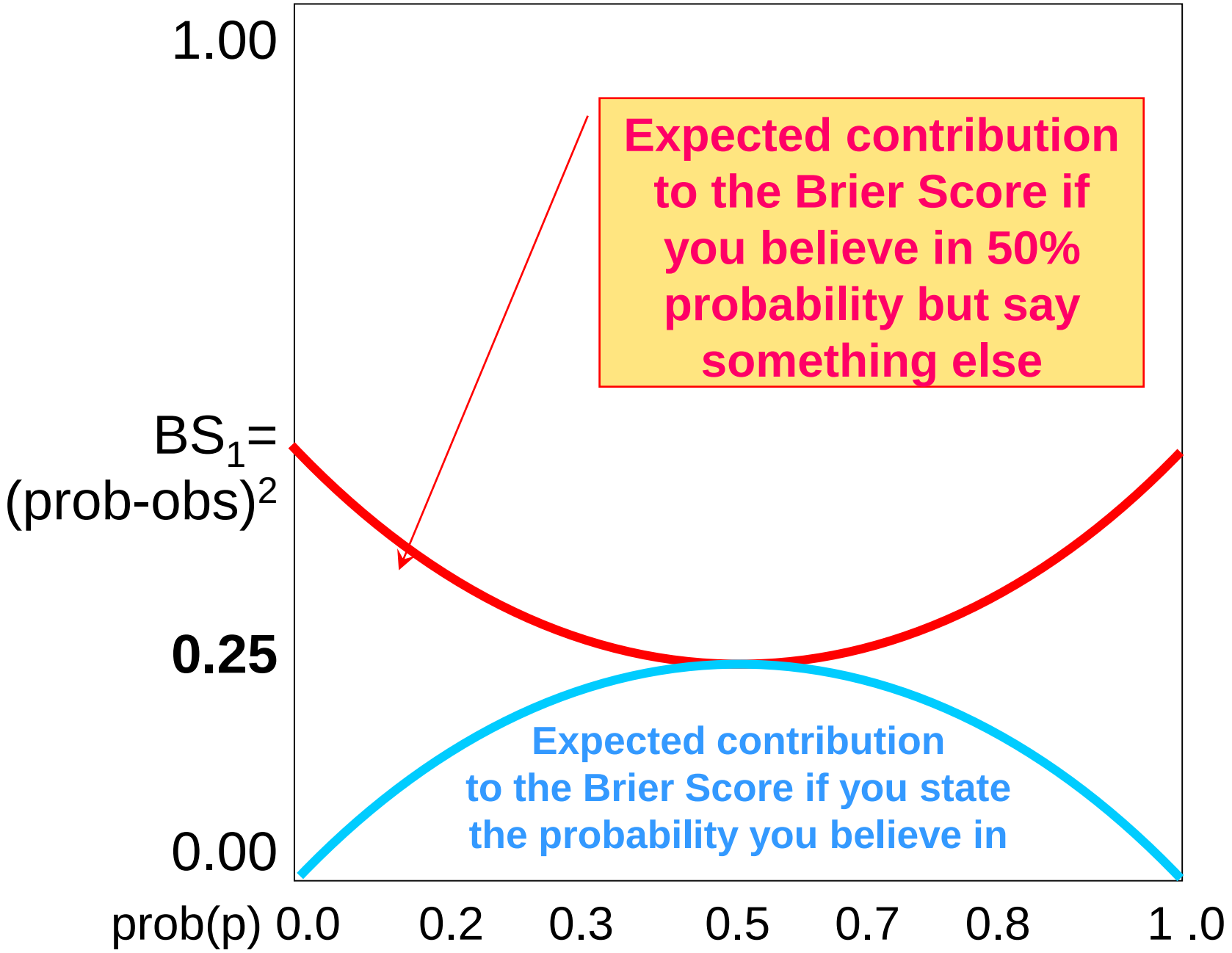
Over N days

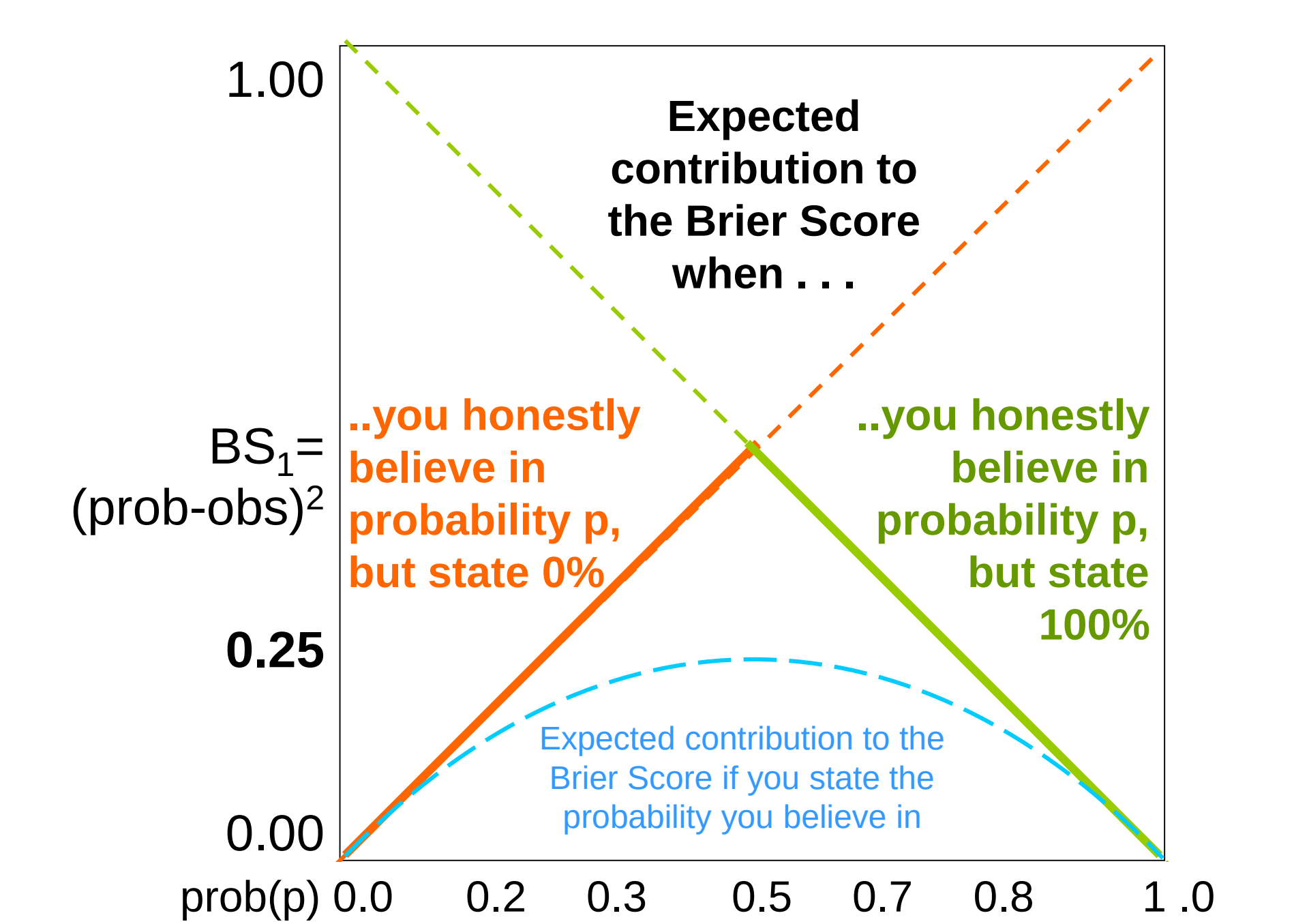
Forecast probability

Observed event (0 or 1)


The forecaster honestly ♥ believes in a 50% probability
1.00
BS1=
(prob-obs)2
0.36
0.25
0.16
0.00
obs=rain
obs=dry
"Dishonest"
prob=40%
50% or
40%
likelihood?
+0.11
- 0.09
50% or
60%
likelihood?
prob(p) 0.0
0.2
0.3
0.5
0.7
0.8
1 .0


The forecaster honestly ♥ believes in a 50% probability
1.00
BS1=
(prob-obs)2
0.49
0.25
0.09
0.00
obs=rain
obs=dry
50% or
30%
likelihood?
"Dishonest"
prob=70%
50% or 70%
likelihood?
+ 0.24
- 0.16
♥
prob(p) 0.0
0.2
0.3
0.5
0.7
0.8
1 .0


1.00
BS_1= (prob-obs)^2
0.25
0.00
Expected contribution to the Brier Score if you believe in 50% probability but say something else
Expected contribution to the Brier Score if you state the probability you believe in
prob(p) 0.0 0.2 0.3 0.5 0.7 0.8 1 .0


1.00
Expected contribution to the Brier Score when . . .
BS1= (prob-obs)2
..you honestly believe in probability p, but state 0%
..you honestly believe in probability p, but state 100%
0.25
Expected contribution to the Brier Score if you state the probability you believe in
0.00
prob(p) 0.0 0.2 0.3 0.5 0.7 0.8 1 .0

II.3.2. The **R**elative **O**perations **C**haracteristics

The **R**elative **O**perating **C**haracteristic curve

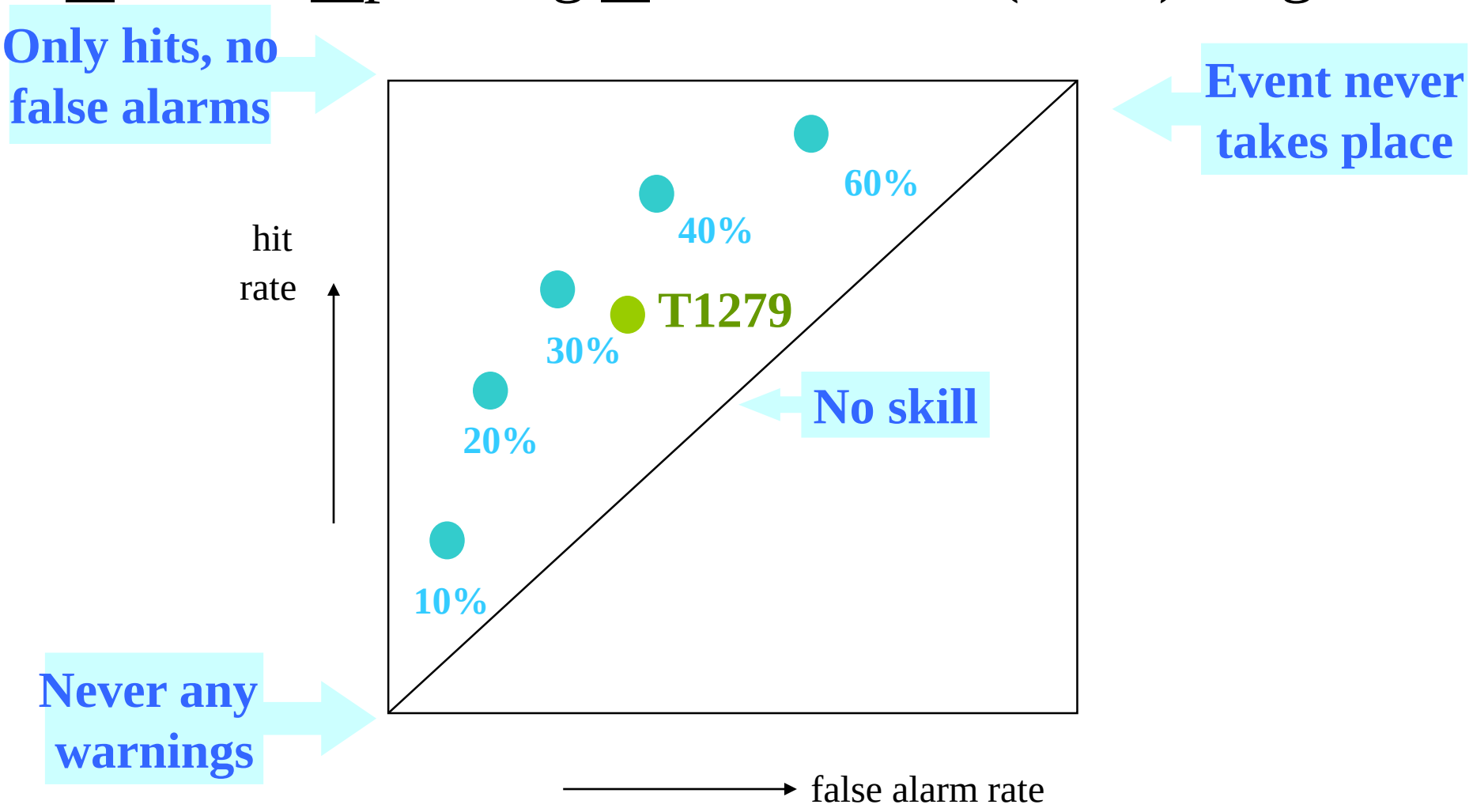
1. **Is able to compares the skill of deterministic and probabilistic forecasts**
2. **For any probability threshold yes/no deterministic forecasts can be obtained**
3. **For each threshold the proportions of "hits" and "false alarms" define the x- and y-axis coordinates**

	Event occurs	Even does not occur
Warning	Hit (**H**)	False alarm (**F**)
No warning	Missed event (**M**)	Correct negative (**N**)

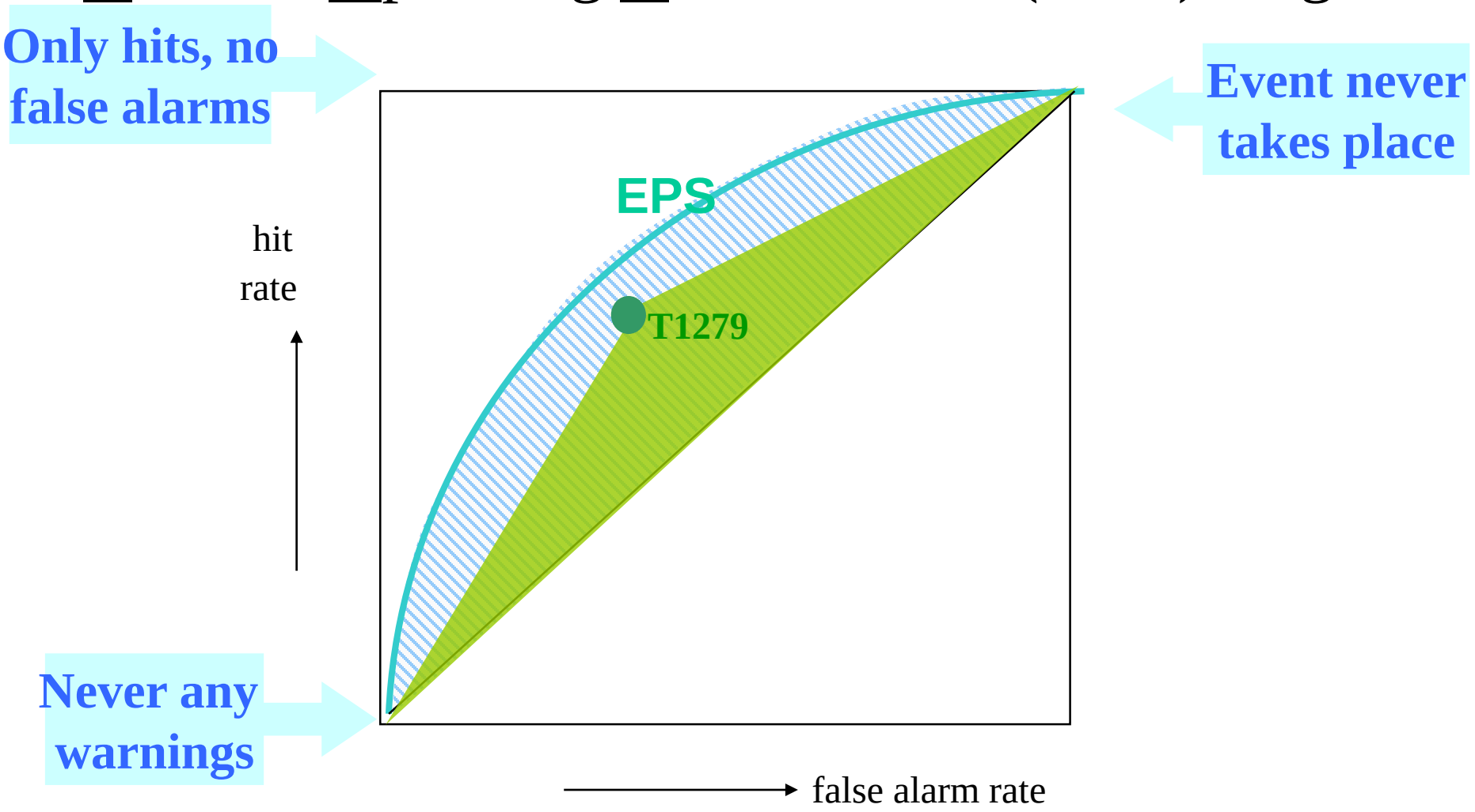
Hit rate (**HR**) = **H**/(**H**+**M**)

False alarm rate (**FAR**) = **F**/(**F**+**N**) *not* F/(F+H)

Relative **O**perating **C**haracteristic (ROC) diagram



Relative **O**perating **C**haracteristic (ROC) diagram



Some draw backs with ROC diagrams

- They do not expose biases/mean errors
- They do not reflect over- or under-confidence
- They are independent of calibration
- They reflect potential skill, like the ACC

II.3.3 The Talagrand diagram

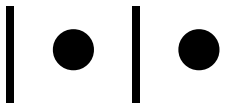
The principle of the Talagrand diagram

With only one ensemble member (|)
all (100%) observations (•) will fall "outside

The principle of the Talagrand diagram

With only one ensemble member (|)
all (100%) observations (•) will fall "outside

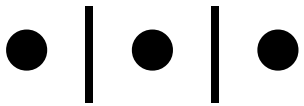
With two ensemble members, two out of three observations (2/3=67%) will fall outside



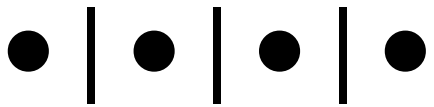
The principle of the Talagrand diagram

With only one ensemble member (|)
all (100%) observations (•) will fall "outside

With two ensemble members, two out of three observations (2/3=67%) will fall outside



With three ensemble members two out of four observations (2/4=50%) will fall outside

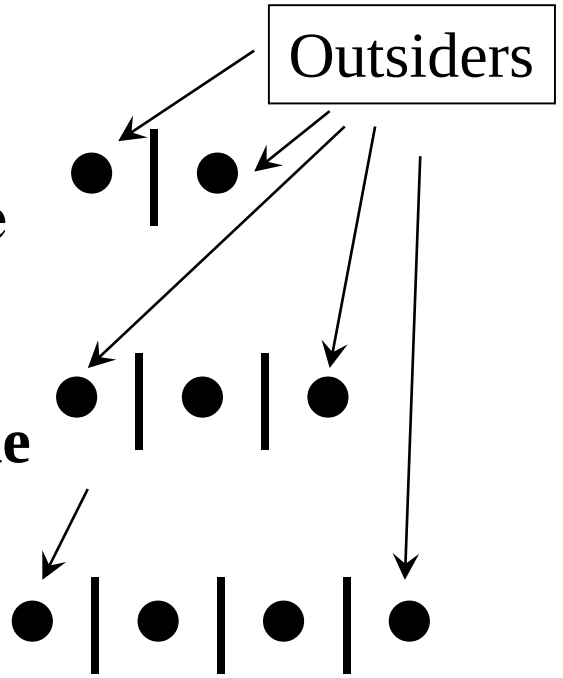


The principle of the Talagrand diagram

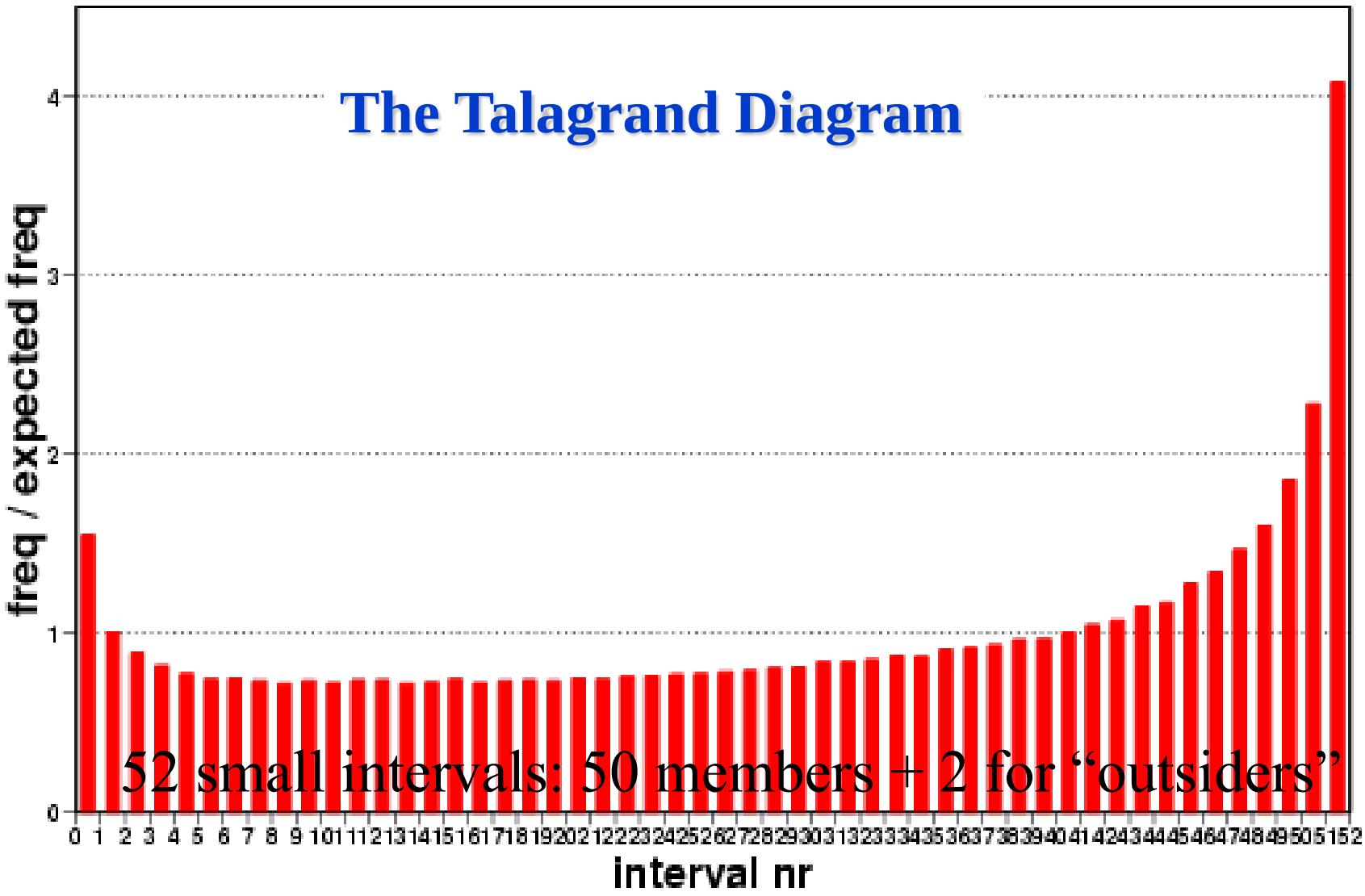
With only one ensemble member (|)
all (100%) observations (•) will fall "outside

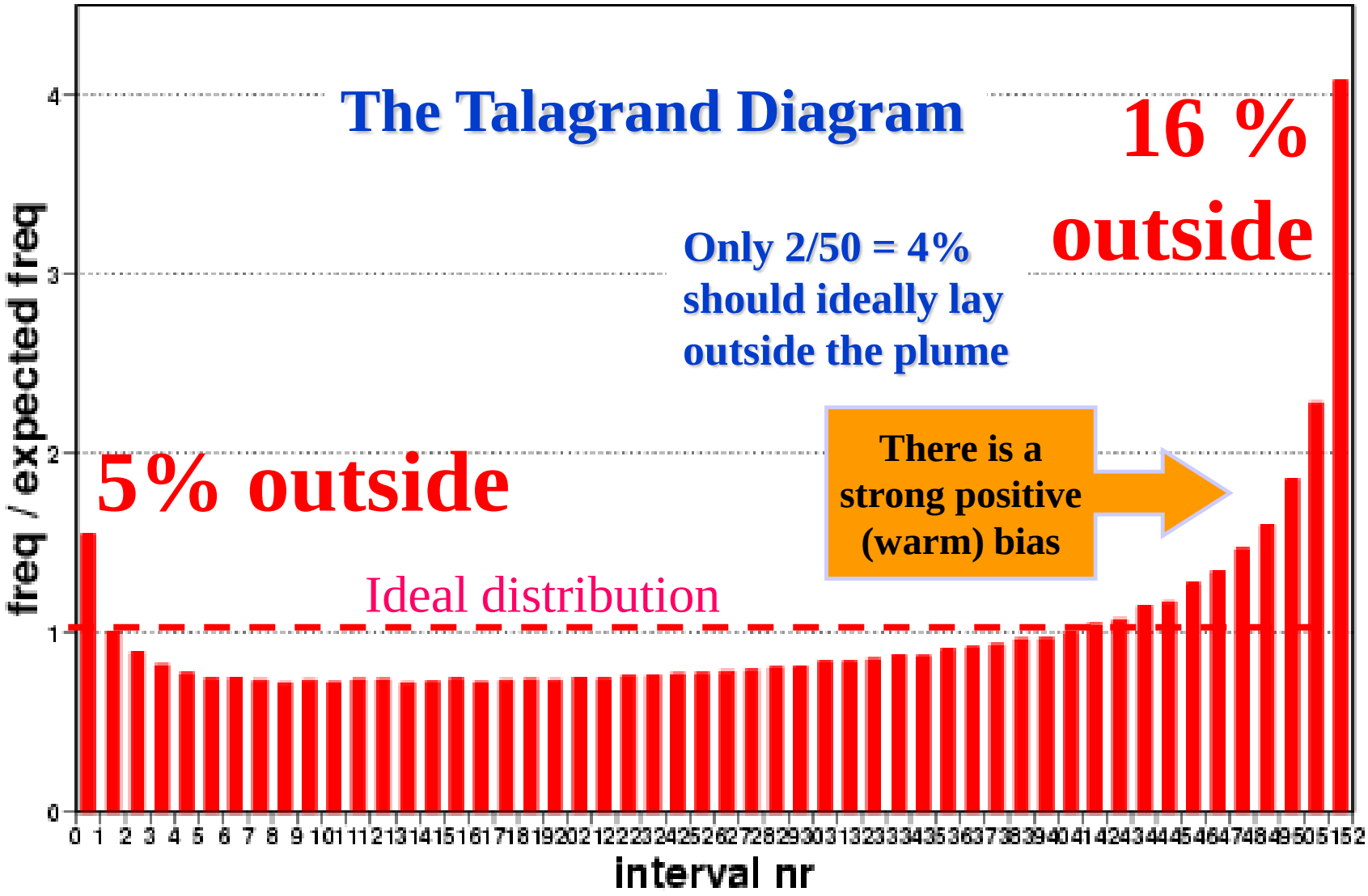
With two ensemble members two out of three observations (2/3=67%) will fall outside

With three ensemble members two out of four observations (2/4=50%) will fall outside



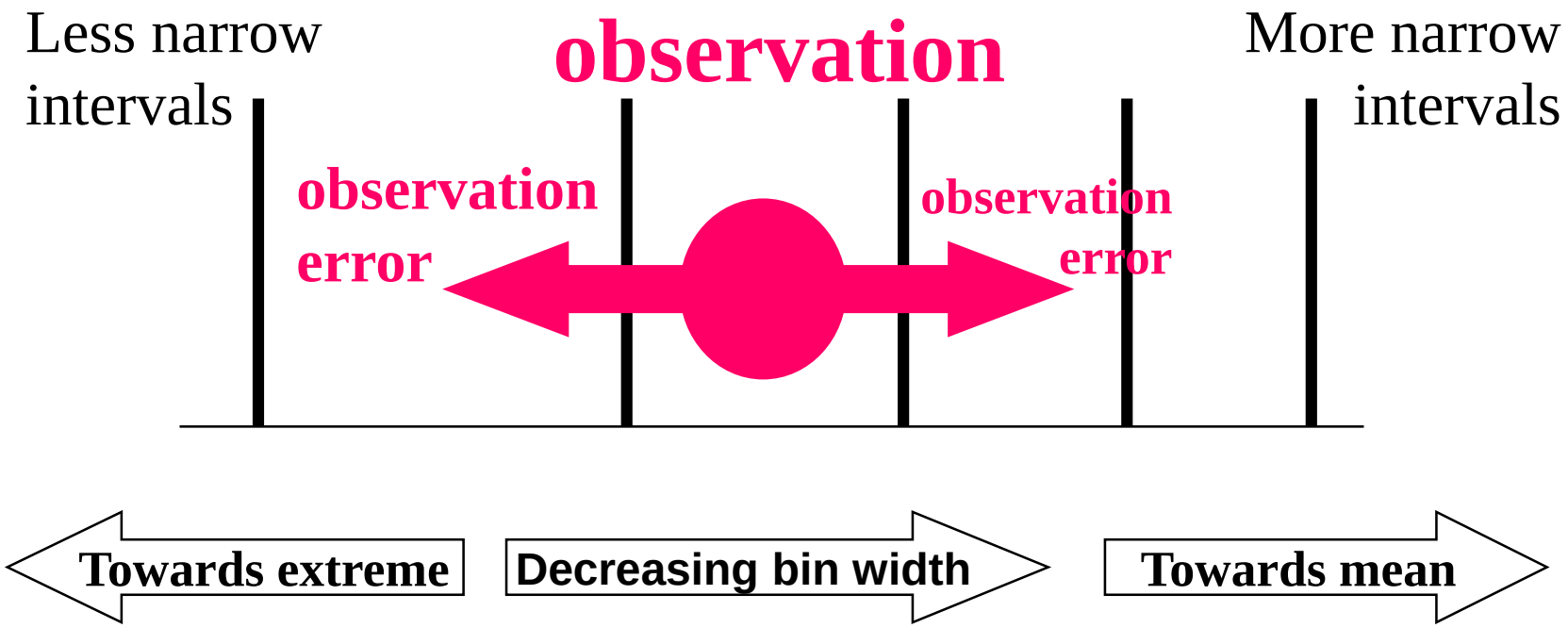
Two observations out of N will always fall outside yielding a proportion of 2/N outside


T(850) anomaly [K] 19991201-20000229 STEP 144
The Talagrand Diagram
freq / expected freq
0
1
2
3
4
52 small intervals: 50 members + 2 for "outsiders"
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
interval nr

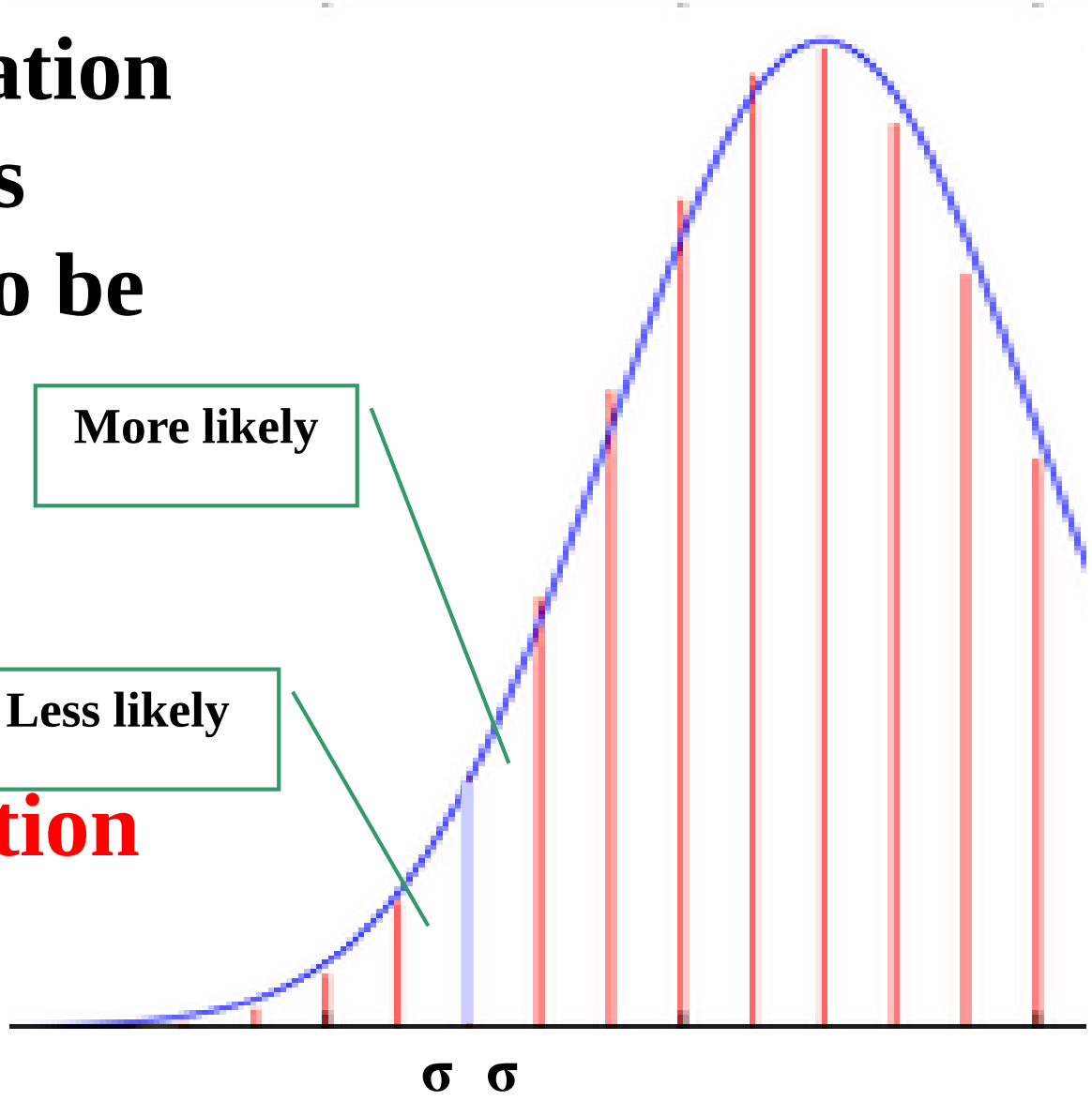

T(850) anomaly [K] 19991201-20000229 STEP 144
The Talagrand Diagram
16 % outside
Only 2/50 = 4% should ideally lay outside the plume
5% outside
There is a strong positive (warm) bias
Ideal distribution
freq / expected freq
interval nr

If the observation error is taken into account, the observation is more likely to occupy a bin away from the norm than closer

But the observation error is not necessarily symmetric….

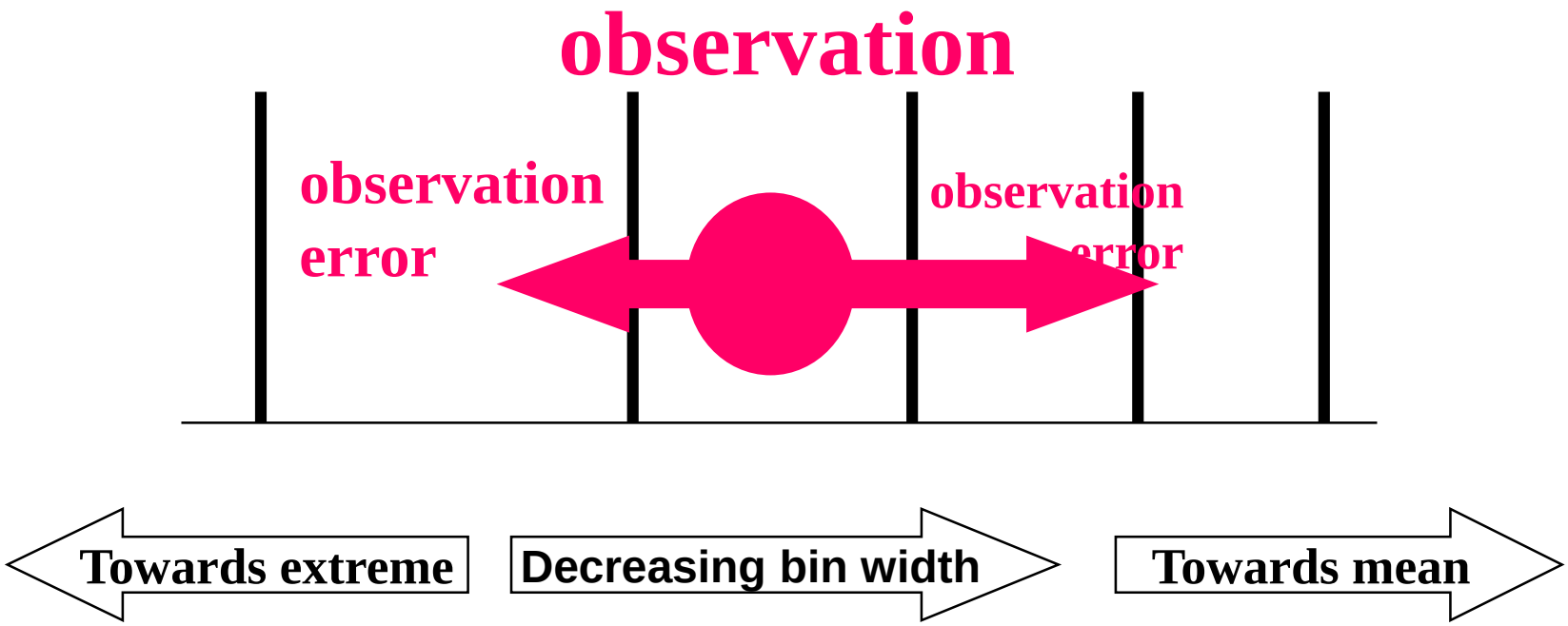


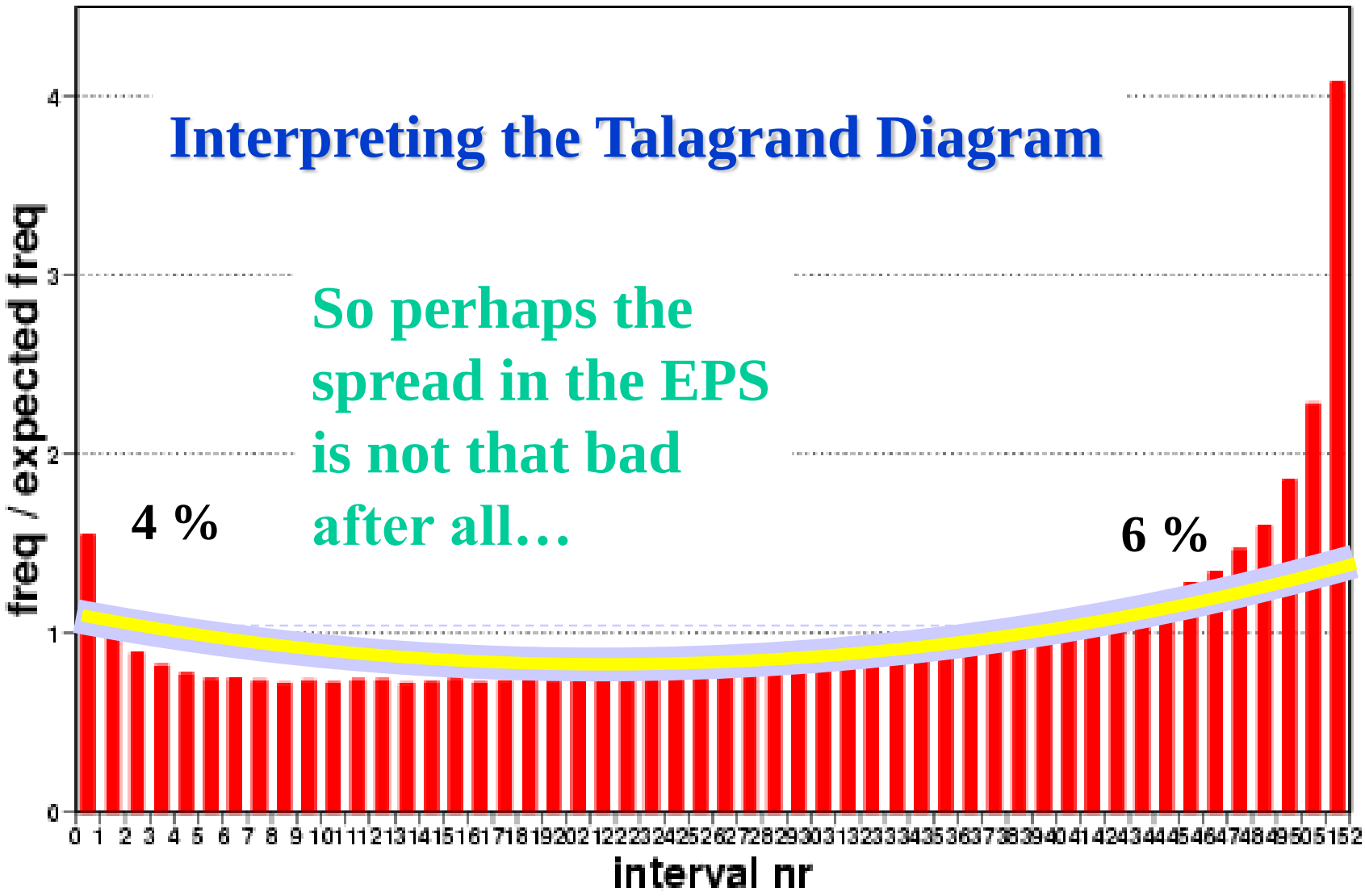
If an observation is wrong, it is most likely to be towards the normal



The observation error is not symmetric

And if the observation error is not symmetric, this might partly compensate….




T(850) anomaly [K] 19991201-20000229 STEP 144
Interpreting the Talagrand Diagram
So perhaps the spread in the EPS is not that bad after all…
4 %
6 %
freq / expected freq
interval nr

The Talagrand diagrams works better for fewer number of members

This avoids that the width of an interval is smaller than the average observation error!

END