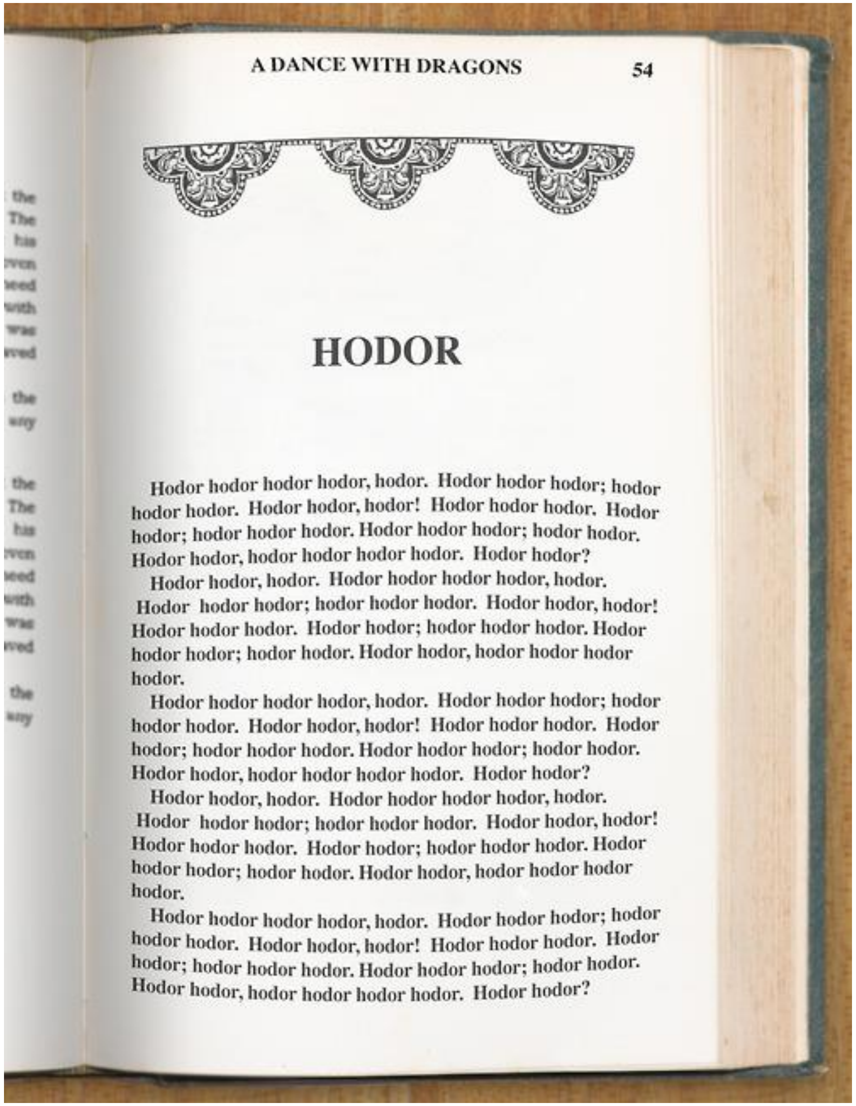
What if … we could predict *Game of Thrones?*

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

- *A song of Ice and Fire* – series of novels by George R. R. Martin, 1996-
- *Game of Thrones,* TV series, 2011-
- Five books so far; two more anticipated
- Characters frequently killed off; subject of intense speculation.

Each chapter from point of view of one character



Q. *Can we predict which characters will show up next?*

A. *No, not really.*

Subject of this talk:

- Prediction in general
- Why is it so hard?
- Why are some things nevertheless predictable?

Types of prediction (1)

Treasury, 19 August 2014:

real production gross domestic product (GDP) is forecast to grow by 2.8% on average over the four years to March 2018

Point prediction. Examples:

- Growth will be exactly 2.8%.
- Die roll will be a 3.
- Tyrion will have exactly 9 POV chapters.

Types of prediction (2)

Australian budget 2013:

real GDP growth in 2013-14 is expected to be 2½ per cent, with the 70 per cent confidence interval ranging from 1½ to 3¼ per cent. In other words, if forecast errors are similar to those in the past 15 years, there is a 70 per cent probability that the growth rate will lie in this range.

Interval prediction. Examples:

- Growth will be probably be around 2.8%.
- Die roll will be a 1, 2 or 4.
- Tyrion will have between 5 and 15 POV chapters.

How to be right

- Make sure your audience wants your prediction to come true (e.g. Malachy/Spellman 1958; eow)
- Manipulate events so that your predictions come true (e.g. Lavoisin and Lavigoreux 1679)
- Make a lot of predictions, but only report the correct ones. (e.g. Paul the Octopus 2010?)
- Be vague (e.g. "You will meet a tall dark stranger") – trade-off between accuracy and uncertainty.

How can we measure uncertainty?

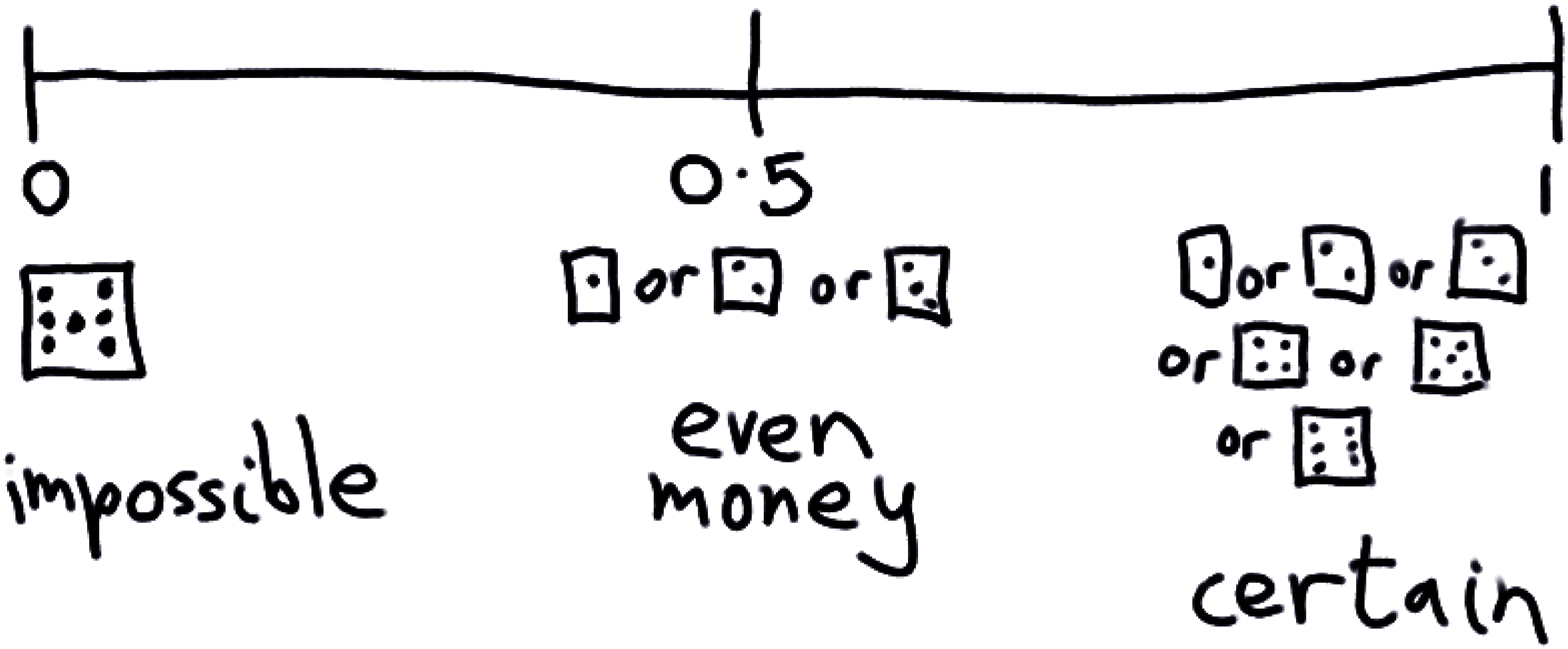
Using **probability**. The probability of an event is a number between 0 and 1 which describes how likely it is to happen.

0 = impossible (or *false*)

1 = certain (or *true*)

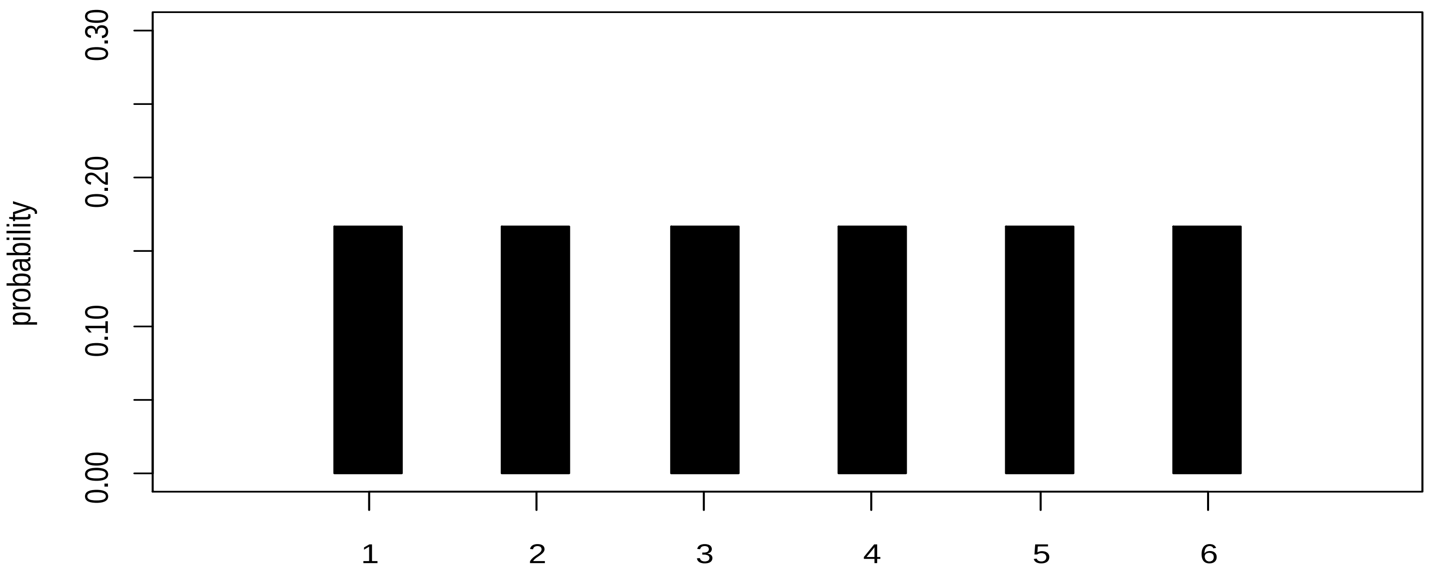
other values = differing degrees of certainty

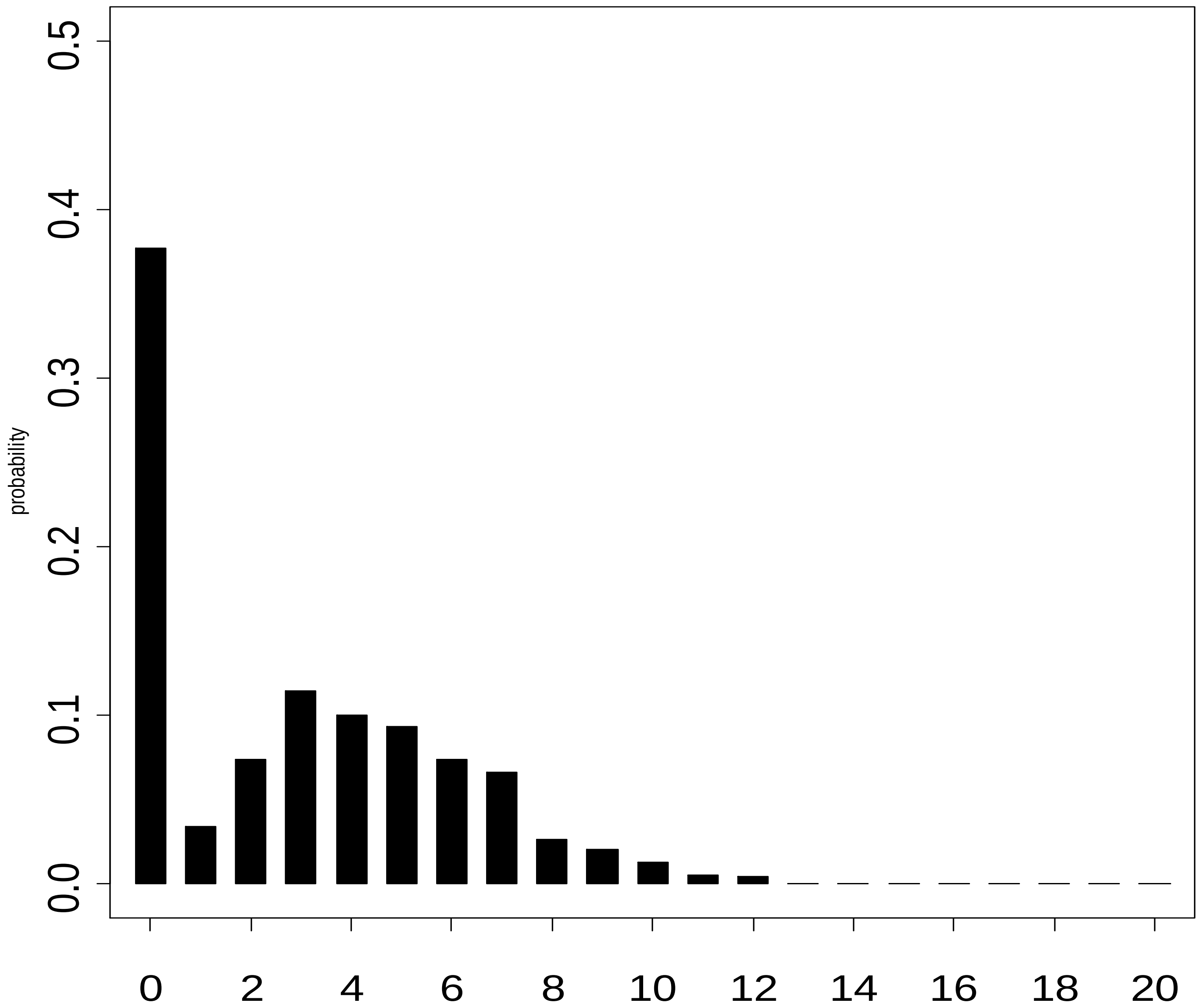
How can we measure uncertainty?



Instead of predicting “We will roll a 3”, predict “The probability of rolling a 3 is 1/6.”

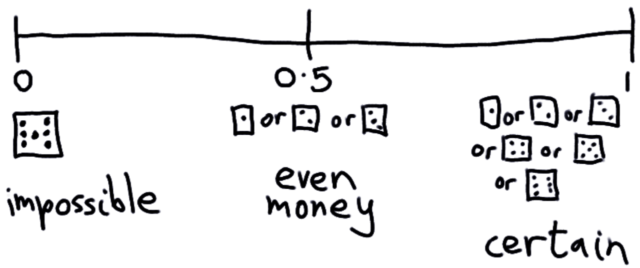
A **probability distribution** gives the probability of every possible outcome

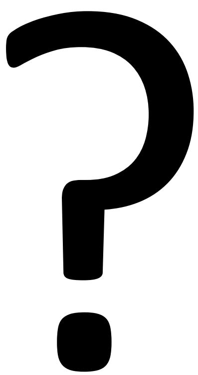


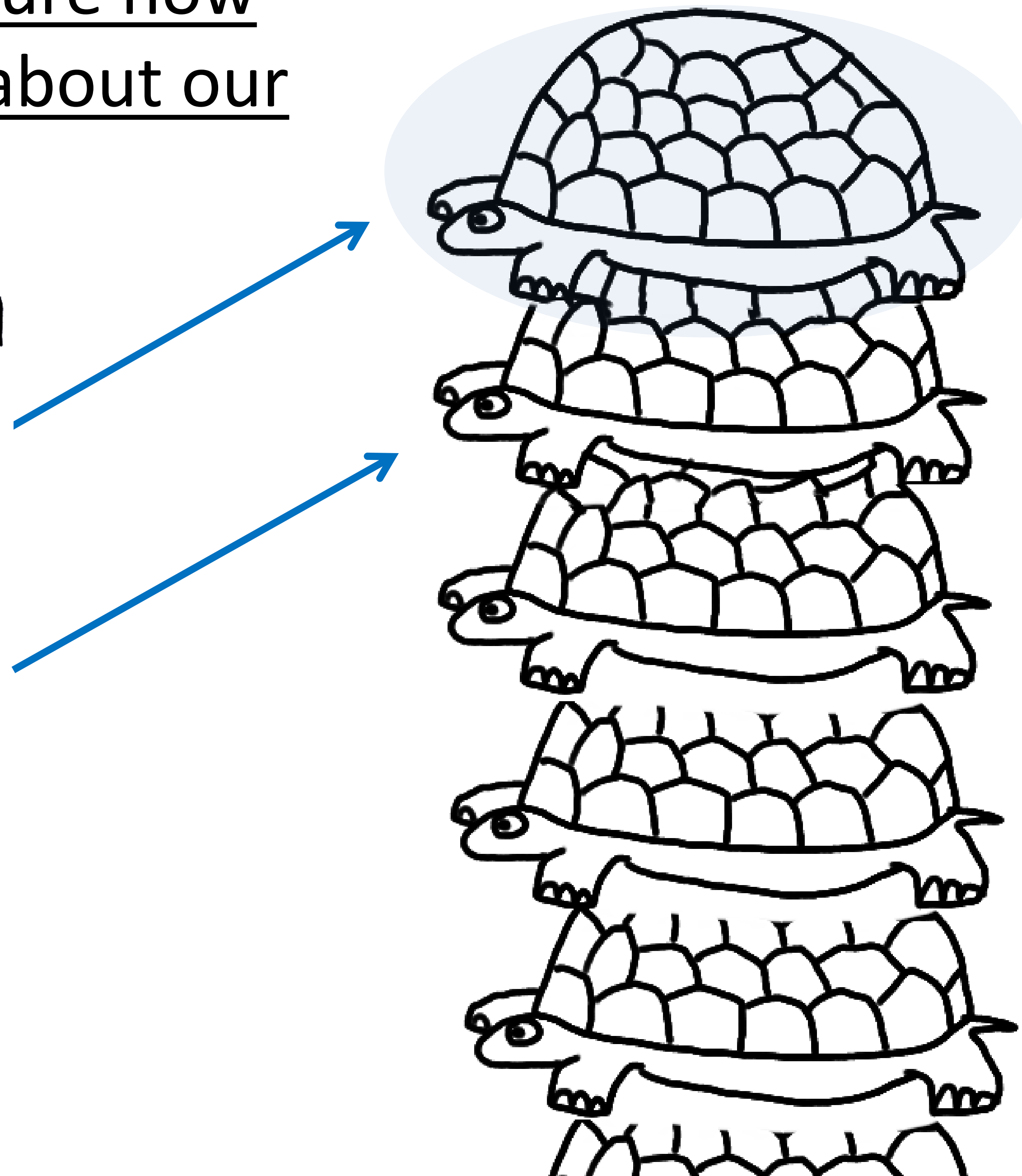


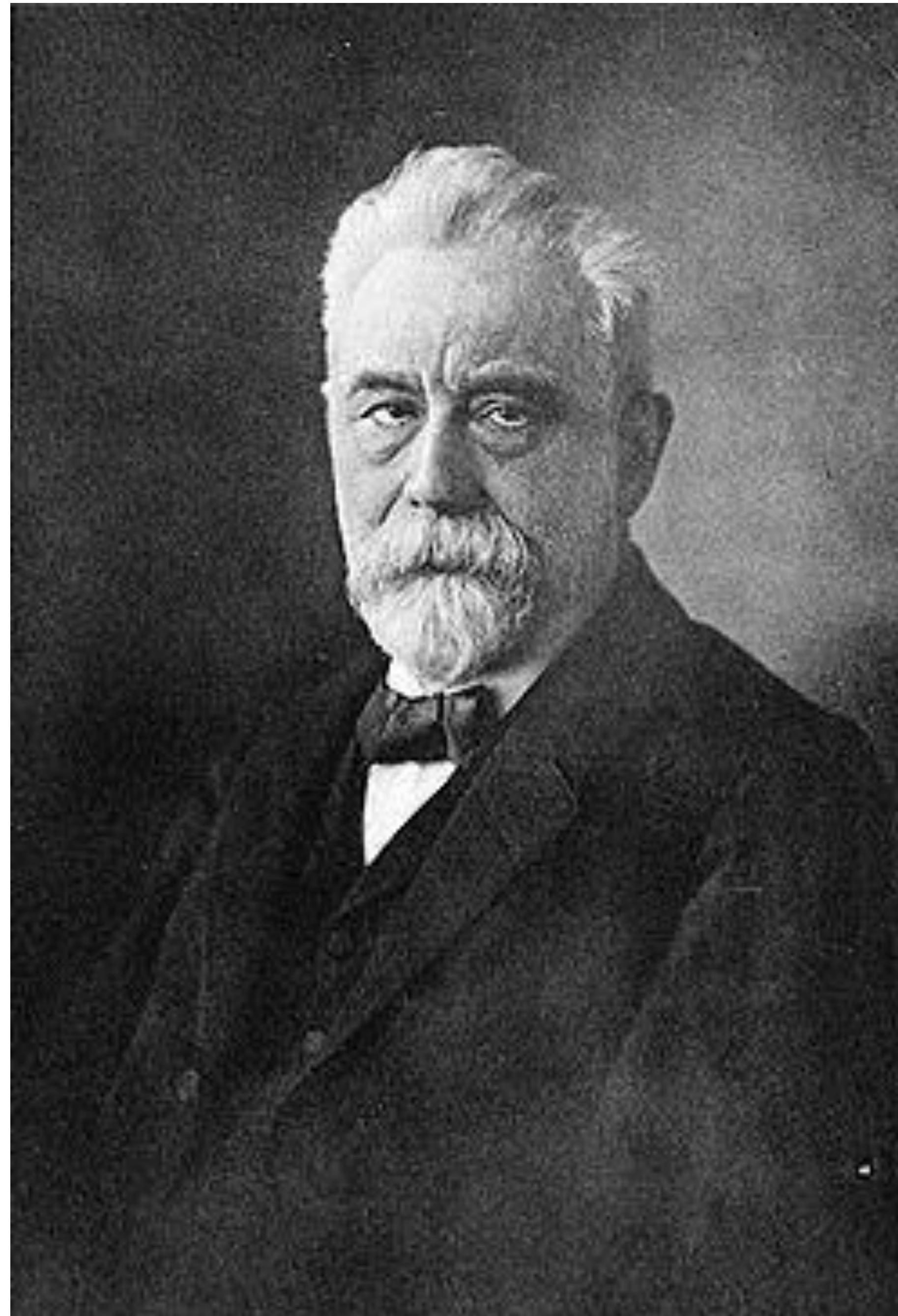
probability
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
12
14
16
18
20

How can we measure how uncertain we are about our uncertainty?







Extrapolation

We can’t predict something that hasn’t happened before.

e.g. *Grue* = green before 2024 and blue afterwards (slight variant of Goodman’s *New Riddle of Induction*)

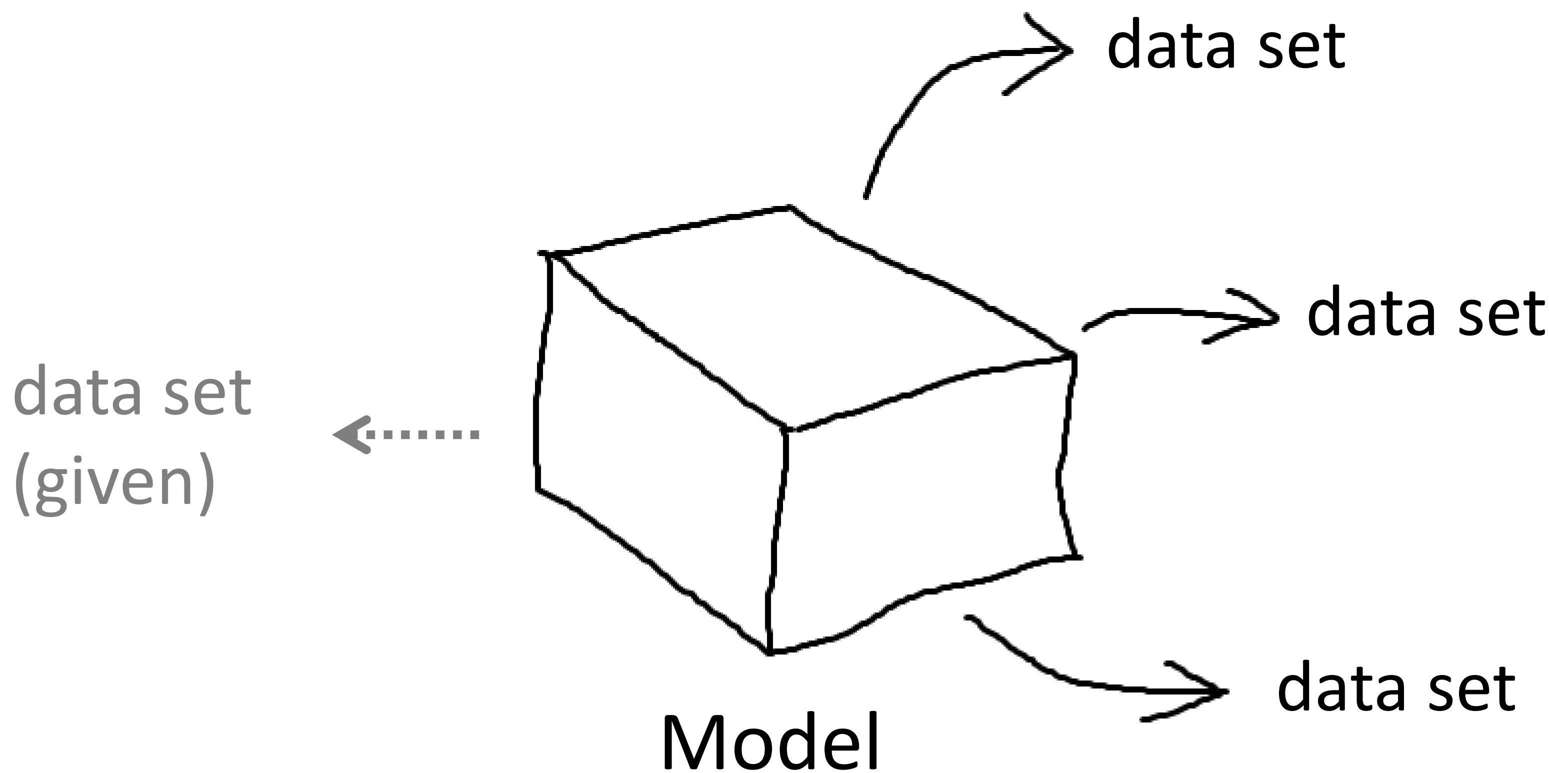
Deduction = Reasoning from causes to effects

Induction = Reasoning from effects to causes

Problem of Induction:

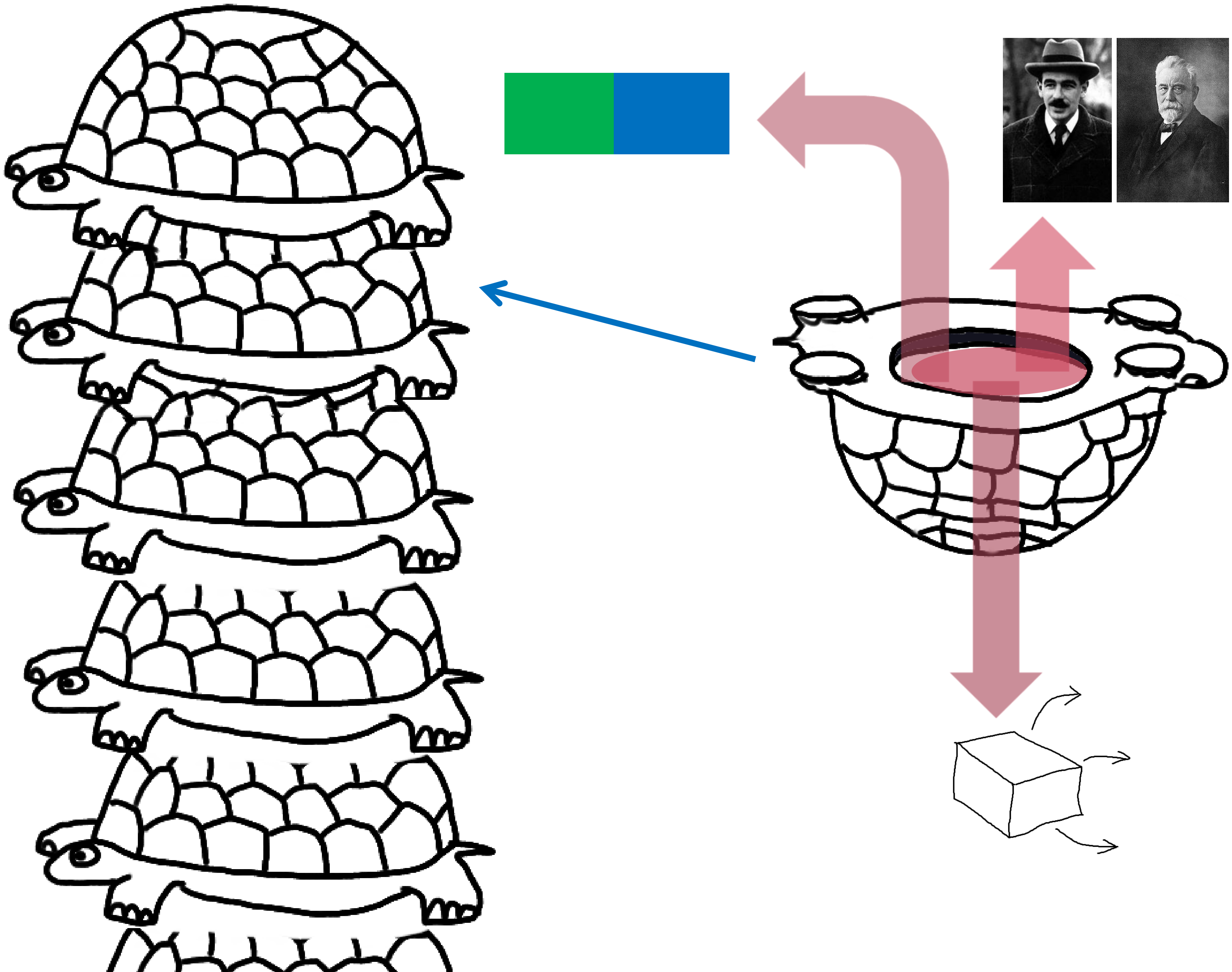
Can induction lead to knowledge at all?

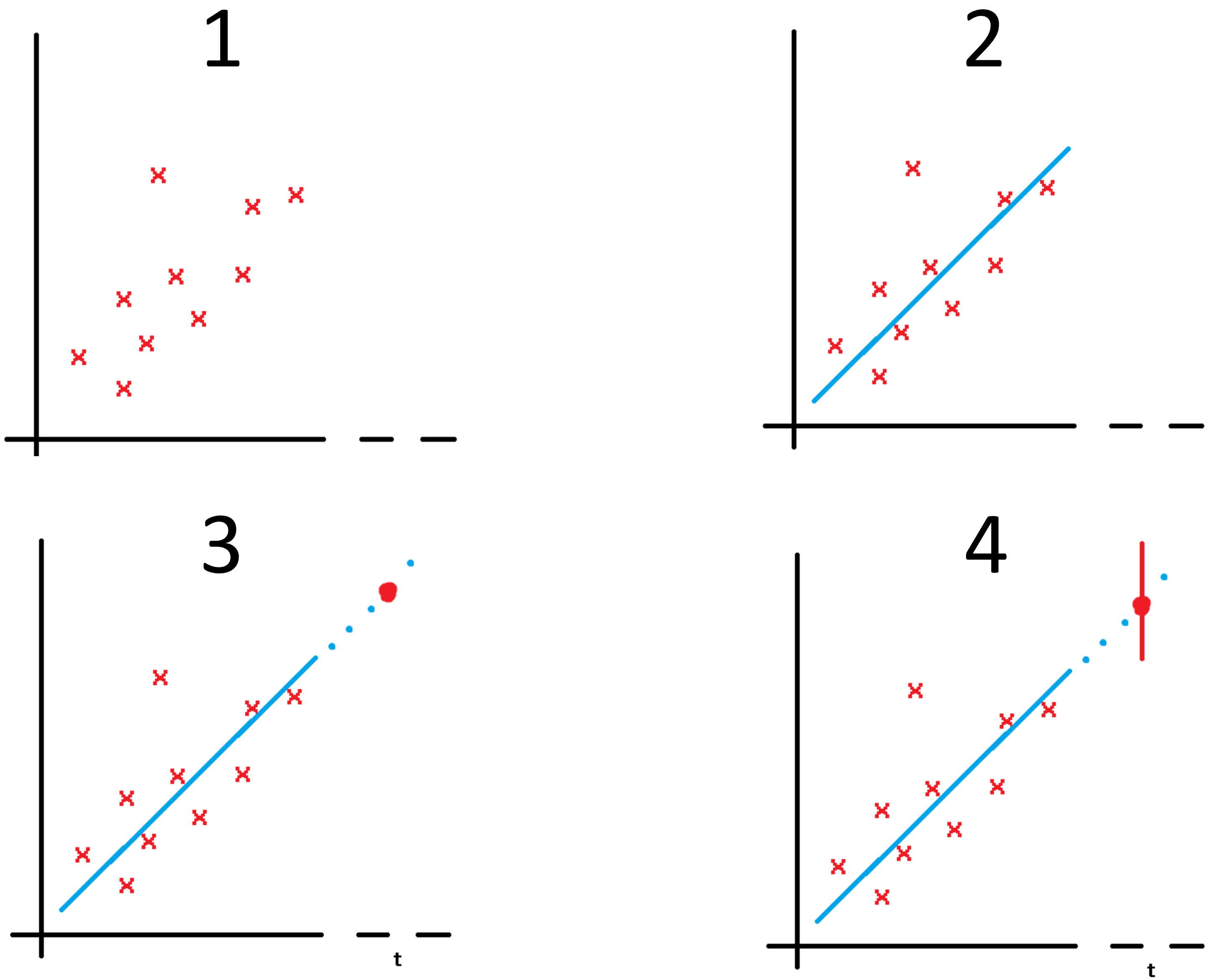
Statistics is a subject which attempts to use probability in order to do inductive reasoning



- Start with a data set.
- Find a model which could plausibly have generated your data set.
- Measure your uncertainty by looking at what other kinds of data set your model *could have* generated.

- Q. How do you know whether you have the correct model?




1
2
3
t
4
t

1

character	AGOT	ACOK	ASOS	AFFC	ADWD
Eddard	15	0	0	0	0
Catelyn	11	7	7	0	0
Sansa	6	8	7	3	0
Arya	5	10	13	3	2
Bran	7	7	4	0	3
Jon Snow	9	8	12	0	13
Daenerys	10	5	6	0	10
Tyrion	9	15	11	0	12
Theon	0	6	0	0	7
Davos	0	3	6	0	4
Samwell	0	0	5	5	0
Jaime	0	0	9	7	1
Cersei	0	0	0	10	2
Brienne	0	0	0	8	0
Areo	0	0	0	1	1
Arys	0	0	0	1	0
Arianne	0	0	0	2	0
Asha	0	0	0	1	3
Aeron	0	0	0	2	0
Victarion	0	0	0	2	2
Quentyn	0	0	0	0	4
Jon Connington	0	0	0	0	2
Melisandre	0	0	0	0	1
Barristan	0	0	0	0	4

2

2.5. Model.

$$X_{it} \sim \begin{cases} \text{Pois}(\lambda_i) & \text{if } |t-\beta_i| < \tau_i \\ 0 & \text{otherwise.} \end{cases}$$

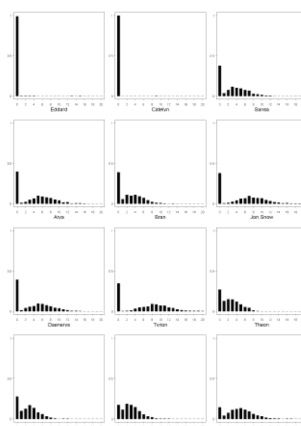
for $1 \leq i \leq N$, and $t \in \{1,2,3,4,5,6,7\}$, with

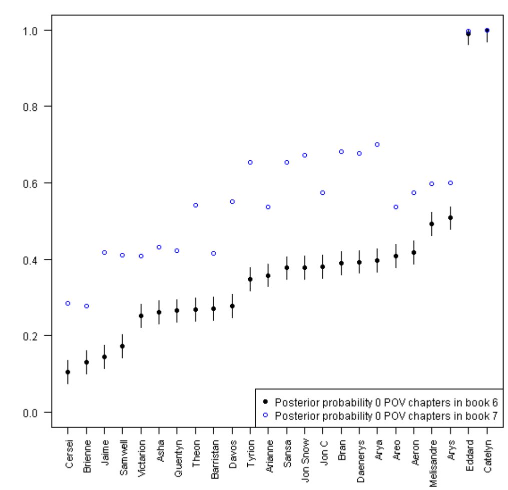
$$\log(\lambda_i) \sim N(\mu_\lambda, \sigma_\lambda^2)$$
$$\tau_i \sim N(\mu_\tau, \sigma_\tau^2) \text{ truncated to } [0,7]$$
$$\beta_i \sim N(\mu_\beta, \sigma_\beta^2) \text{ truncated to } [0,7]$$

where $\sigma_\lambda, \sigma_\tau, \sigma_\beta > 0$ and $\mu_\lambda, \mu_\tau, \mu_\beta \in \mathbb{R}$. For fixed i, the X_{it} are assumed to be conditionally independent given λ_i, τ_i and β_i. For fixed t and $i \neq j$, the X_{it} and X_{jt} are assumed to be conditionally independent given the values of $\lambda_i, \tau_i, \beta_i$ and λ_j, τ_j and β_j.

3

4





- Is the model correct?

  No (domain knowledge.)

- How can we evaluate the predictions? Note: not like rolling a die; it only happens once (Sickmiller story) What does probability even mean for a one-time event?

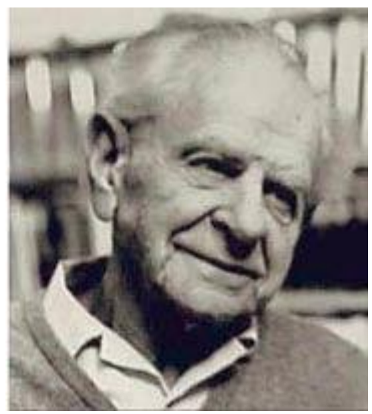
- Is there a better approach?

Probably! Via textual analysis (e.g. Rasmussen, Lambert and Bernth project) but it's very complicated


Statistical Models
BIG DATA

Why predict things anyway?

How else can we know whether a theory is **in**correct? (Popper)



e. g.

- LHC €7.5 bn because physicists care whether their theories are correct.
- 

 2008: GFC?