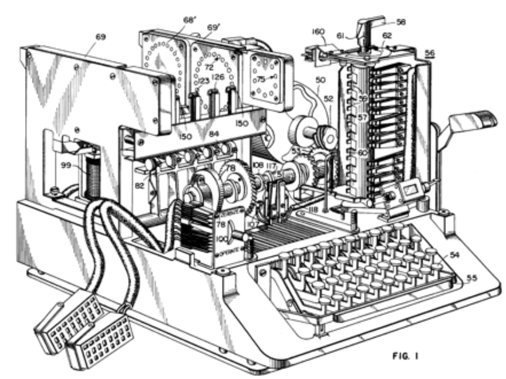
Minds and machines

https://djnavarro.github.io/minds-machines-4103/



Danielle Navarro

d.navarro@unsw.edu.au

Structure to the class

- Content
  - See homepage!
- Assessment
  - See homepage!
- Tips on presenting & participating?
  - Oh gosh… um… see homepage! ☺

Structure for today

- Introductory comments
  - Let's talk… does psychology have theory?
  - Background to computational modelling

- Getting started with R
  - Download and install
  - Work through the introductory sections of the notes at https://psyr.org

A temporary measure

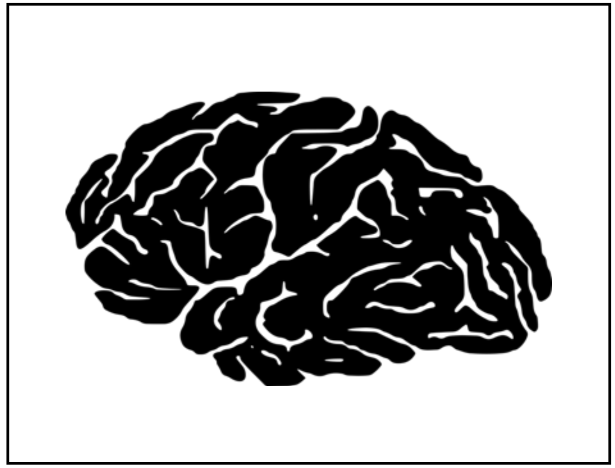
- This is for my other R class, and I'm currently putting together a clean one for this class but the other one will be sufficient for today
- Type this in RStudio:

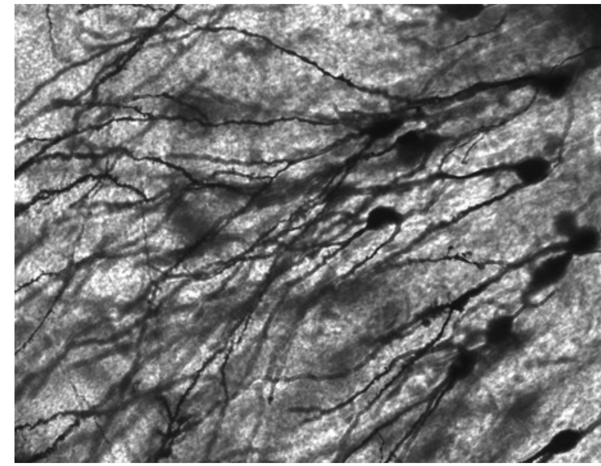
```
install.packages("usethis")
usethis::use_course("psyr.org/part1_core.zip")
```


DECADES OF INCREMENTAL COMPUTATIONAL MODEL DEVELOPMENT
WAAAY TOO MANY PEOPLE
IS THIS THEORETICAL AMNESIA?
imgflip.com

How should we think about cognition?

The view from "below"





Cognition is performed by the brain, and our theories of cognition should informed by the biology of the brain

The view from “above”

Cognition is a feature of intelligent agents, and our theories of cognition should be informed by understanding what intelligent agents do

There are commonalities

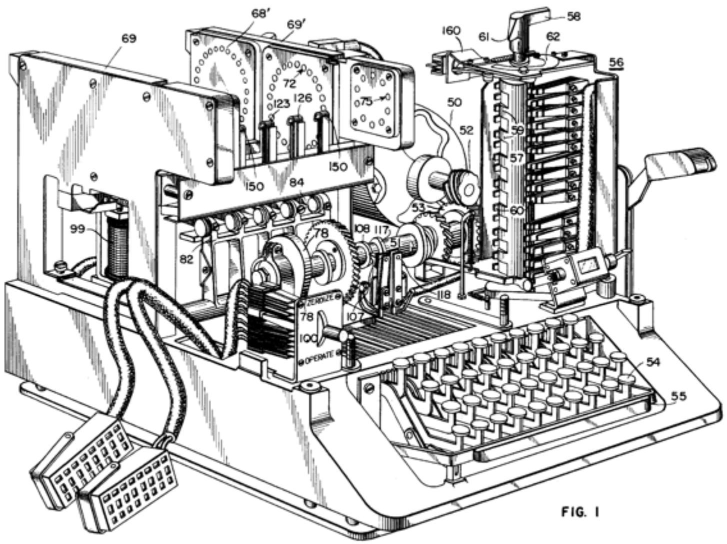
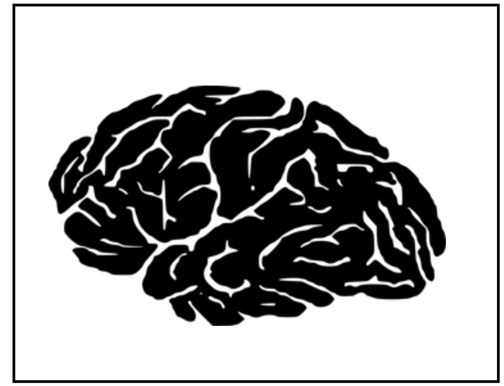
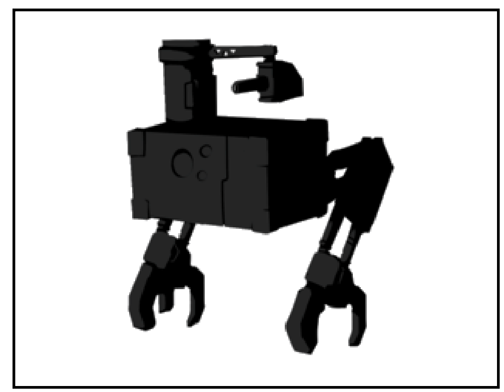


FIG. 1

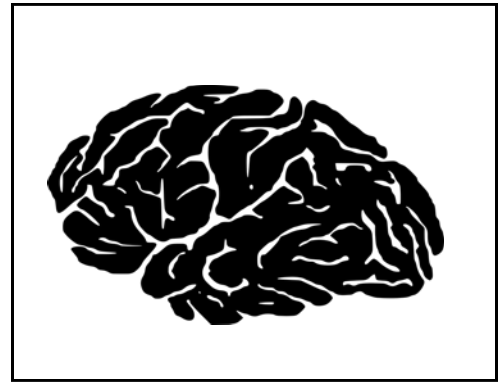




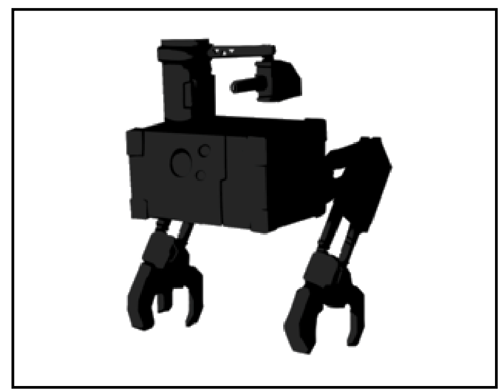
At an abstract level, cognition is a form of computation, and the brain does information processing

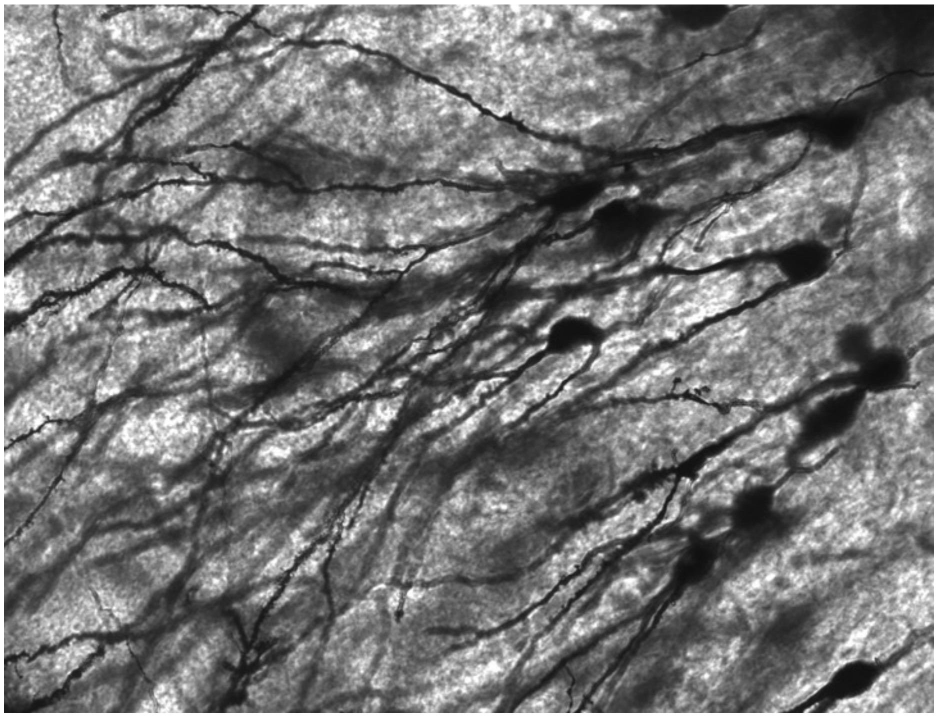
But there are differences

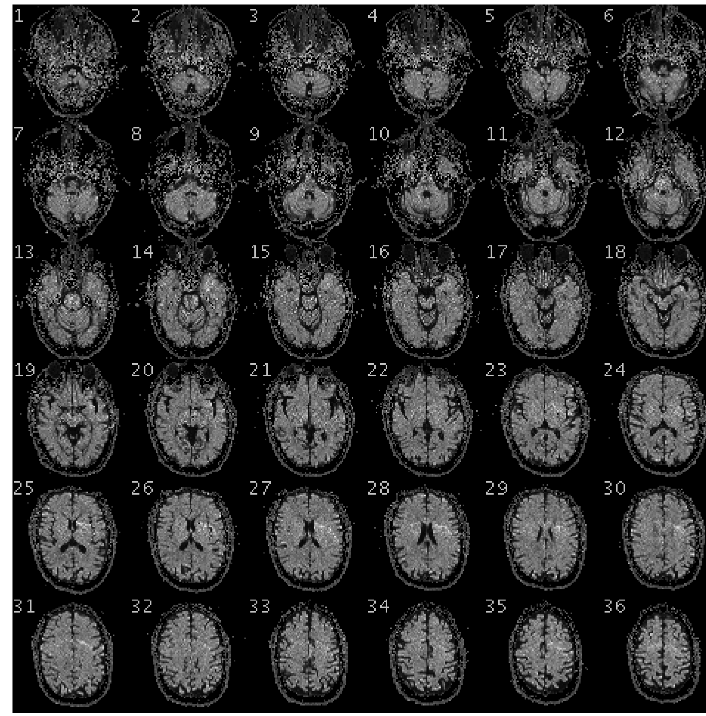
What mechanisms does the brain use to perform computations?



What computational problems does the mind address?

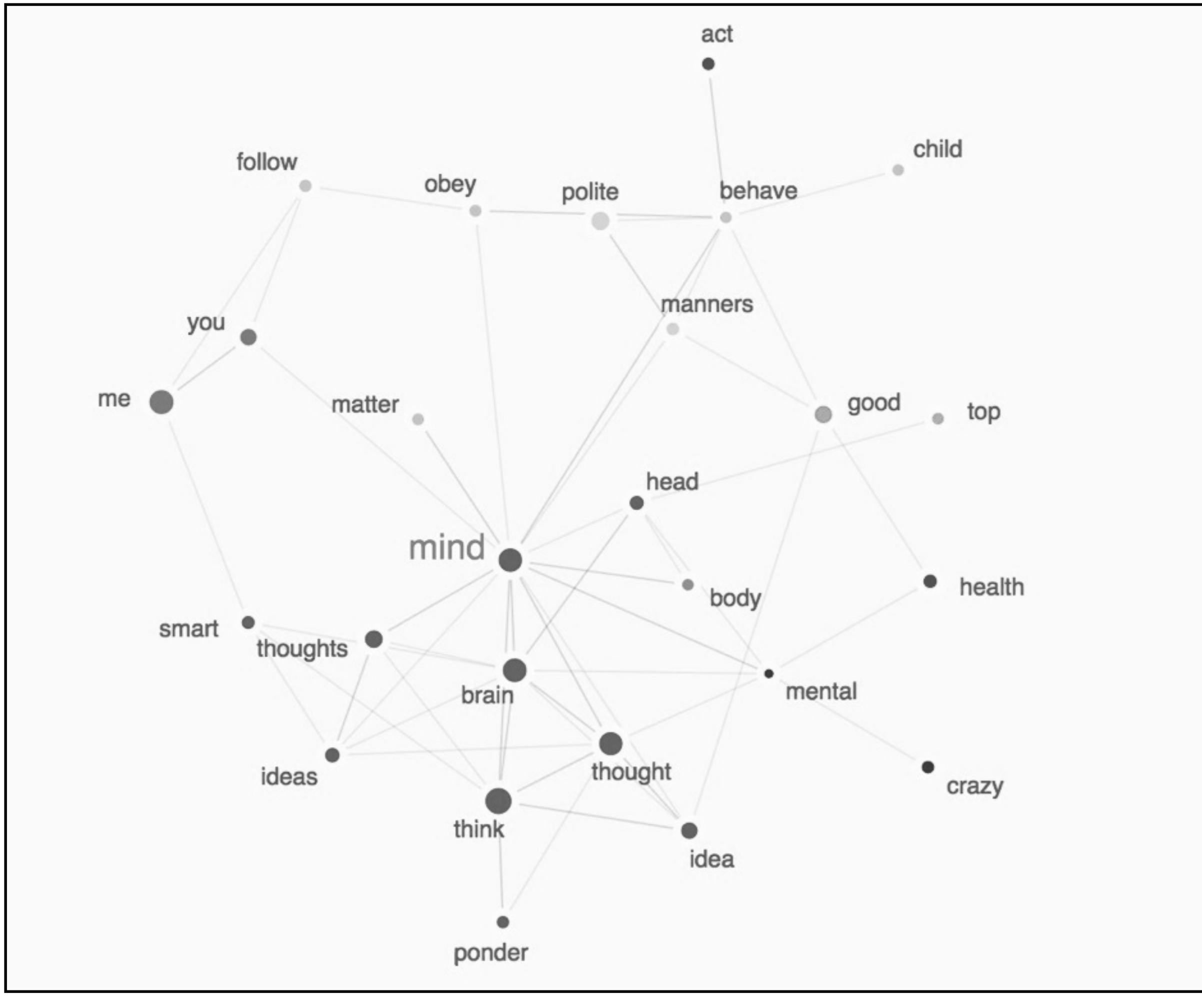


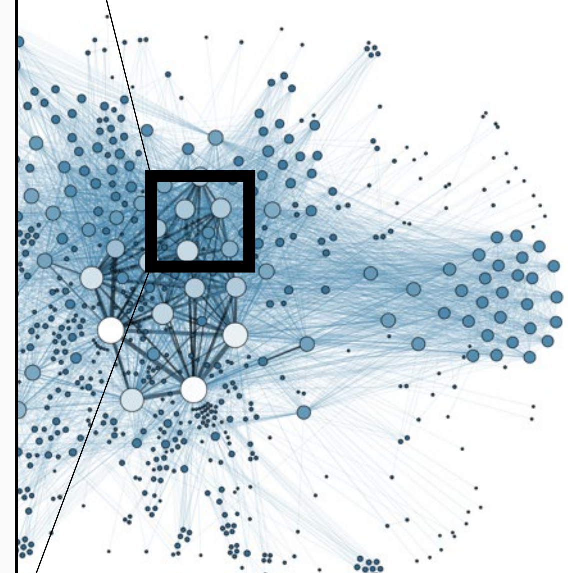


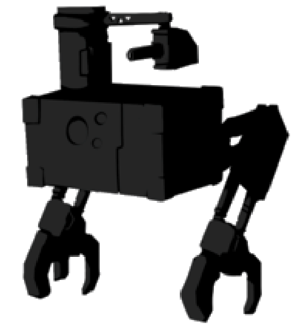


Brains perform computations
using a network of neurons

http://www.smallworldofwords.com/new/visualize





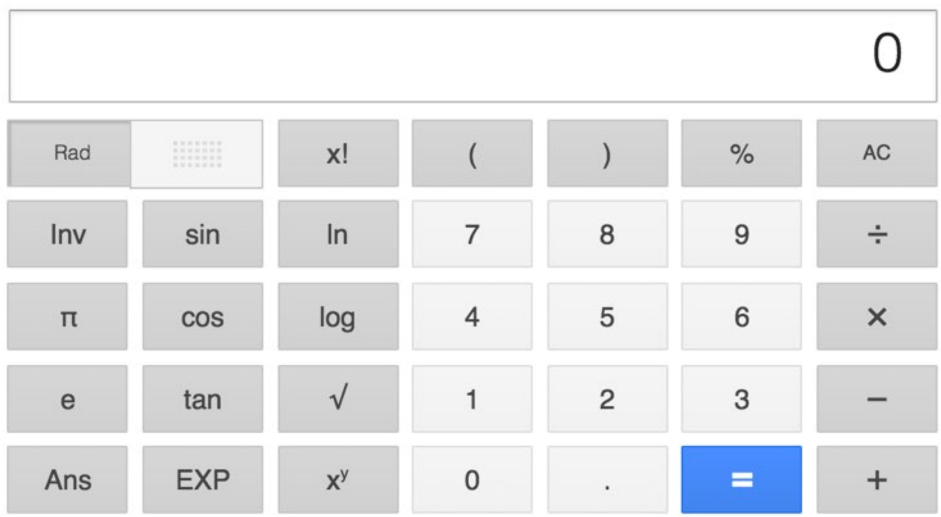


Minds encode meanings using a system of interrelated concepts

Levels of analysis

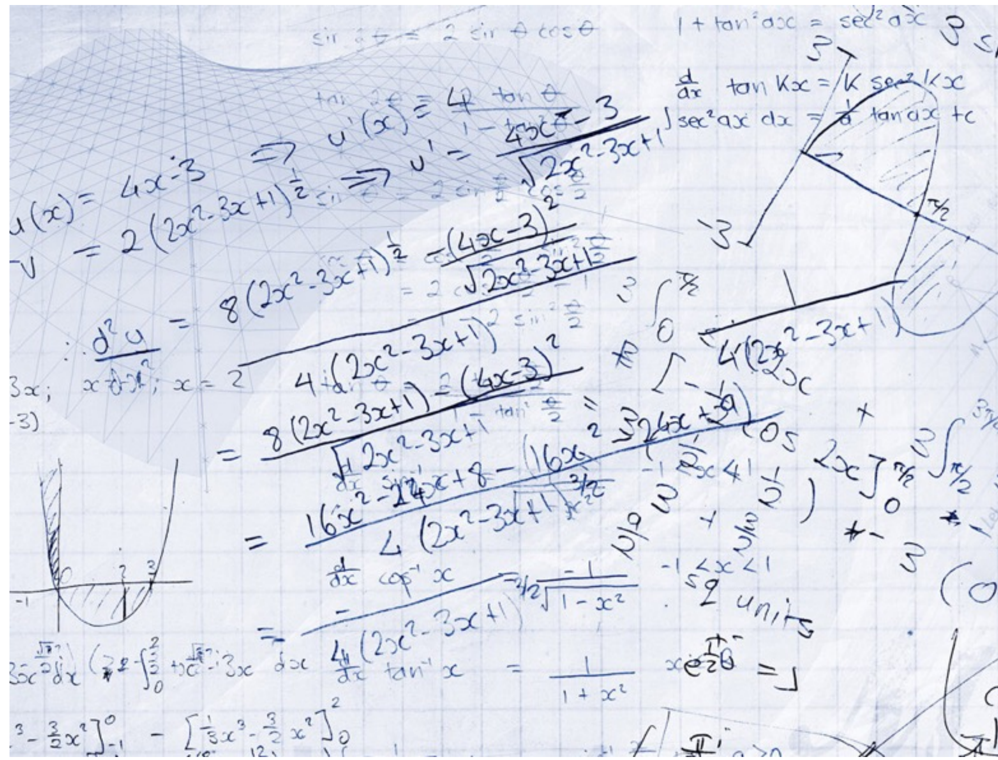
(Marr 1980)

- *Abstract computation:* What problem does a system solve?
- *Algorithm:* What processing steps does it follow to do so?
- *Implementation:* How is this instantiated as a physical entity?



The computational level

"The function of a calculator is to solve arithmetic problems"

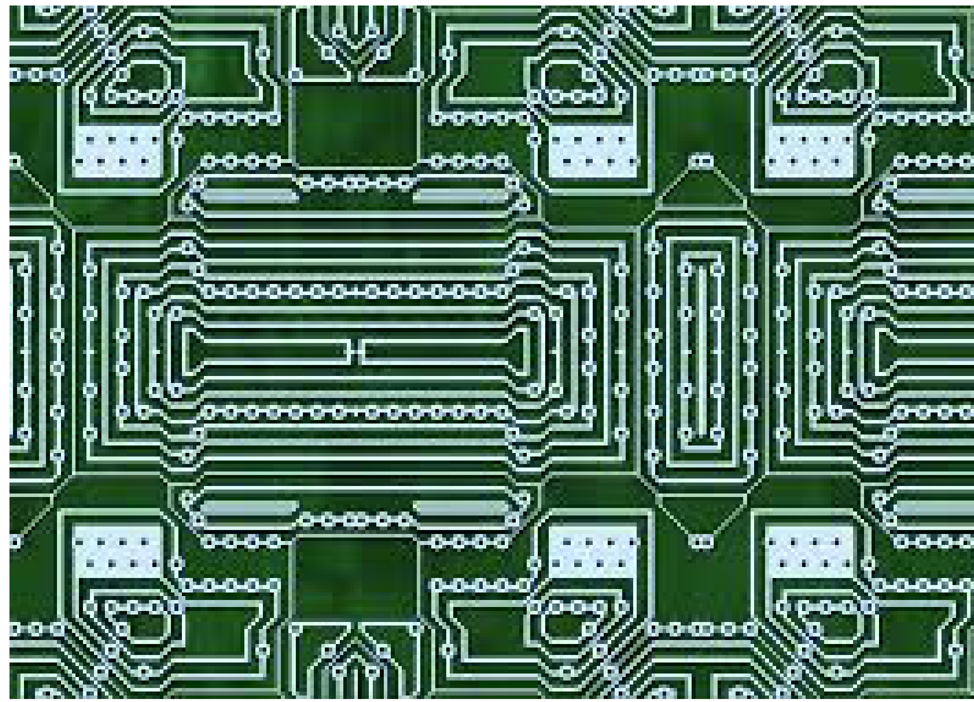


The algorithmic level

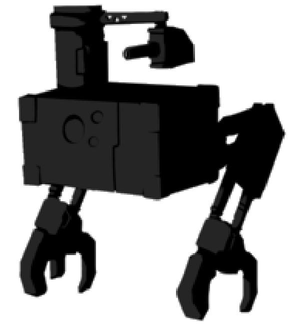
“Addition can be described using a computer program”

The implementation level

"A calculator uses circuitry to do calculations"

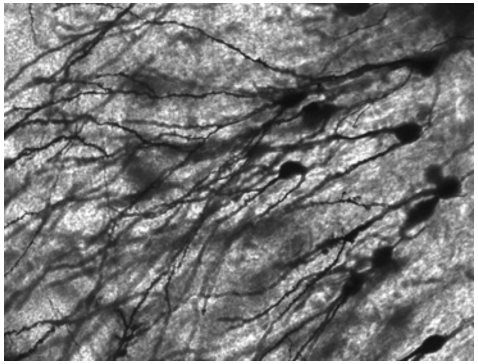


Computation



Algorithm

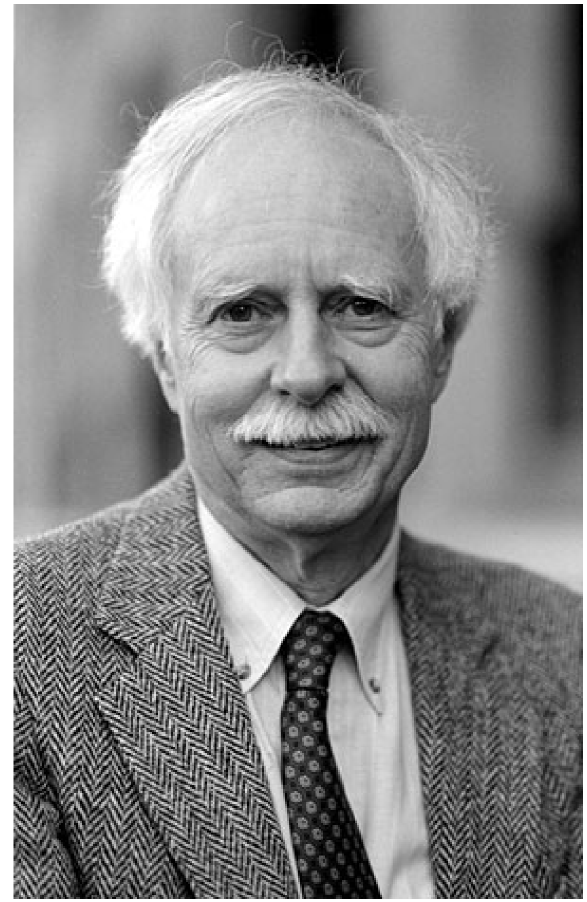
Implementation



The computationalist's dream

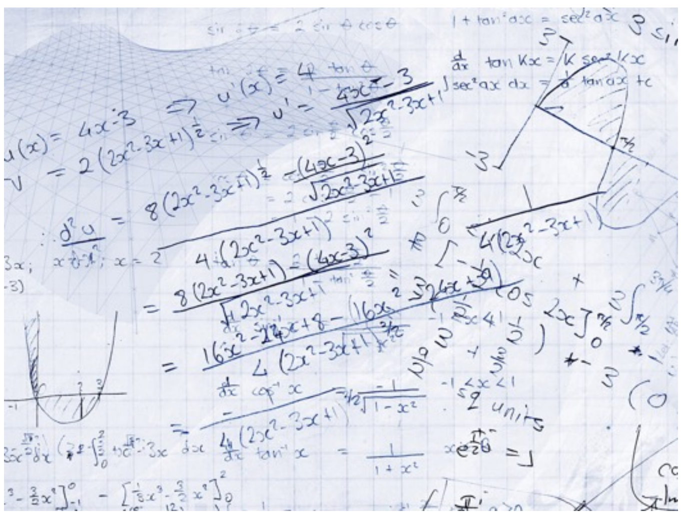
"*Because these regularities reflect universal principles … natural selection may favor their increasingly close approximation in sentient organisms wherever they evolve ... Possibly, behind the diverse behaviors of humans and animals, as behind the various motions of planets and stars, we may discern the operation of universal laws*"

- Roger Shepard, 1987

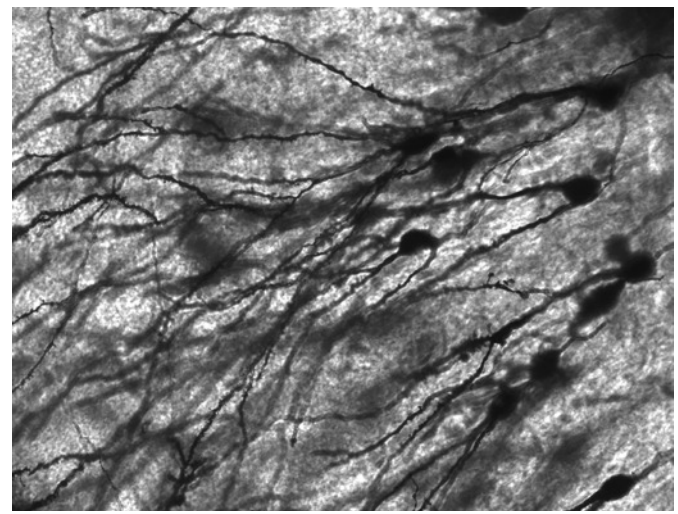


It's a remarkable claim

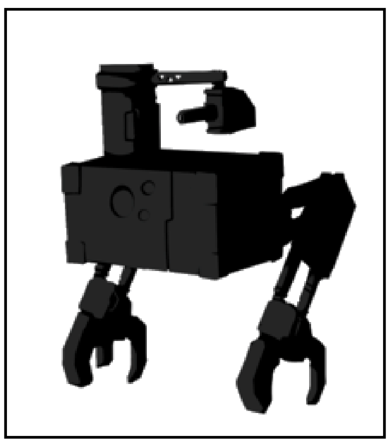
The structure of the learning problem…

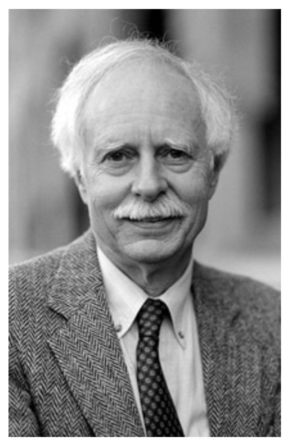


… shapes the structure of the mind that solves it?

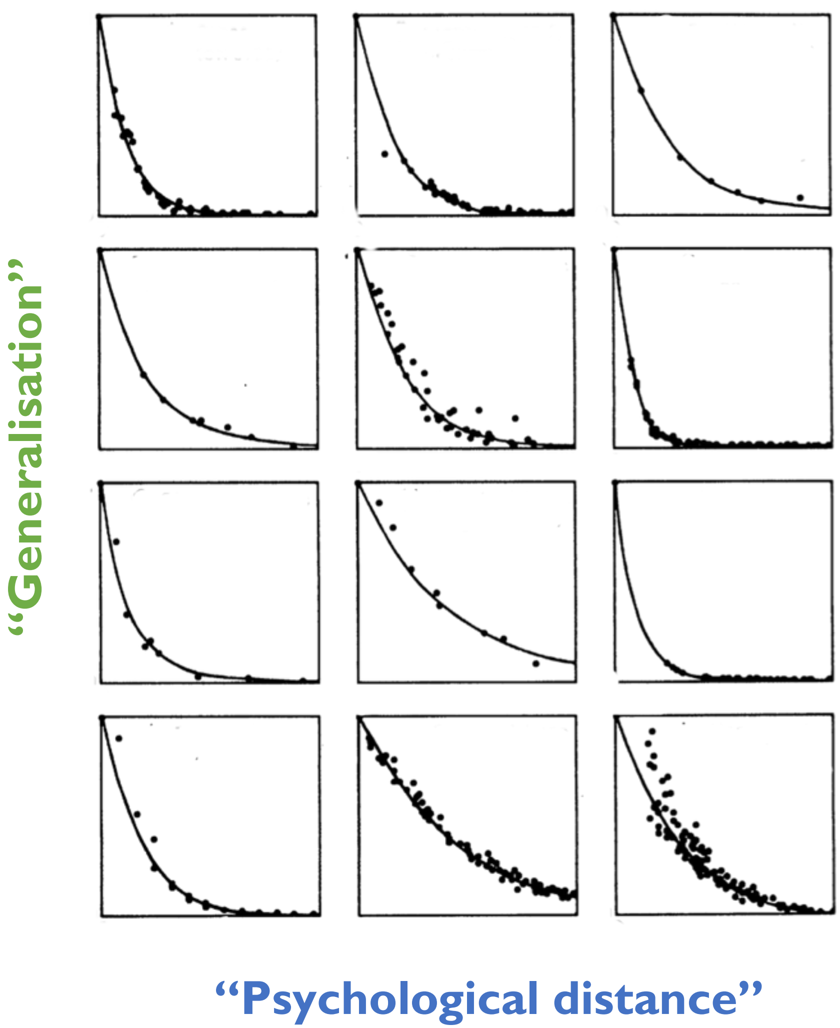


Case study: Shepard's "universal law of generalisation"

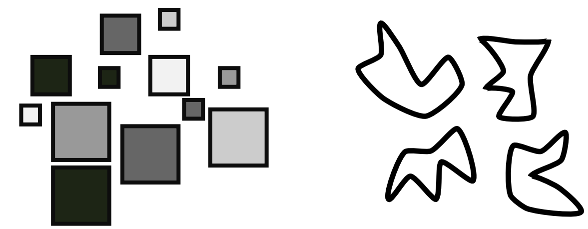




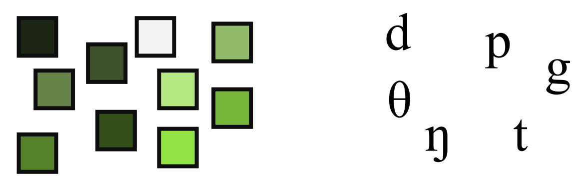
An invariance?



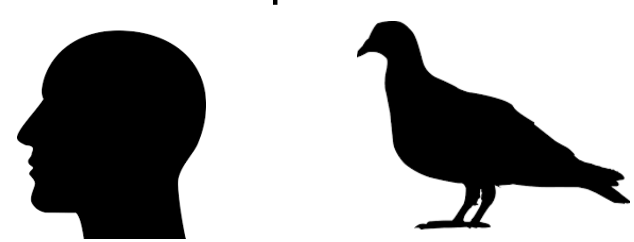
Invariance across stimulus types



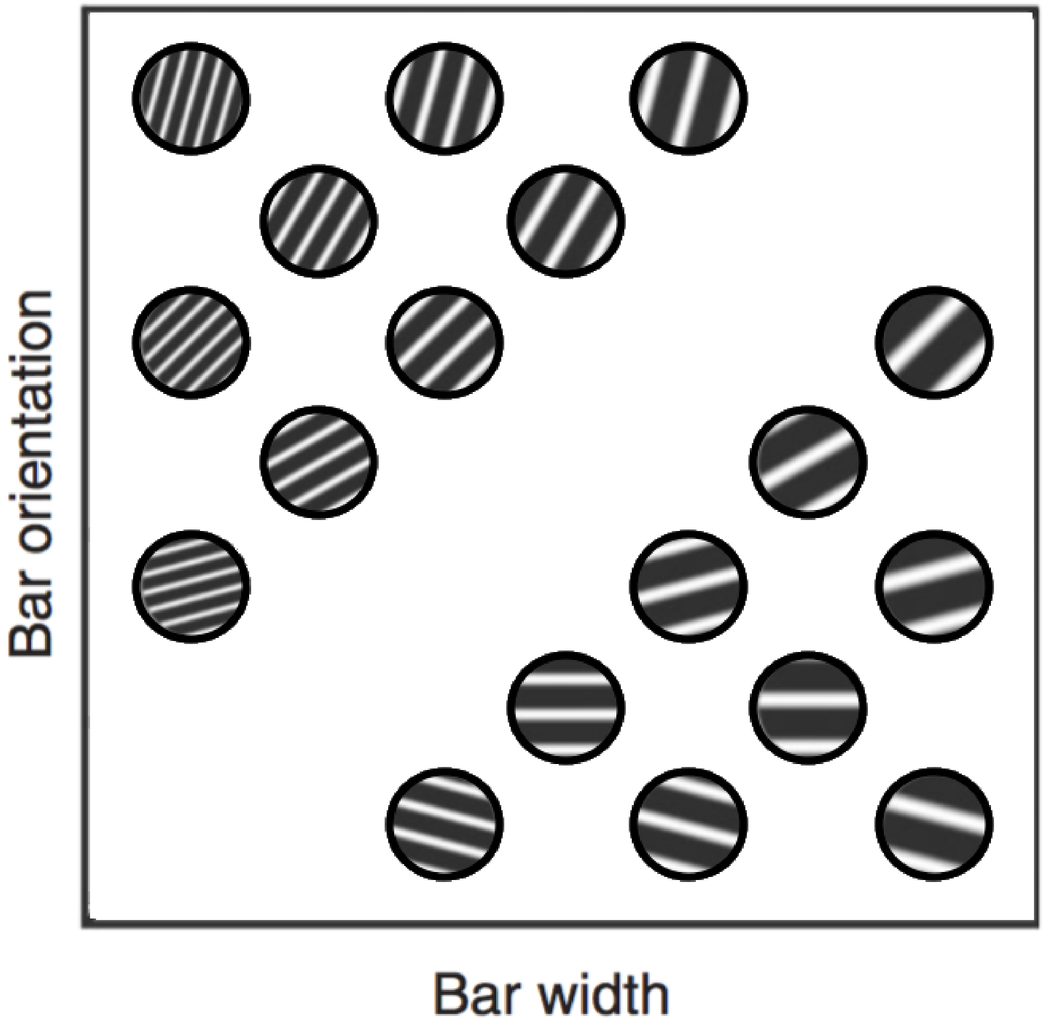
Invariance across sensory modalities



Invariance across species

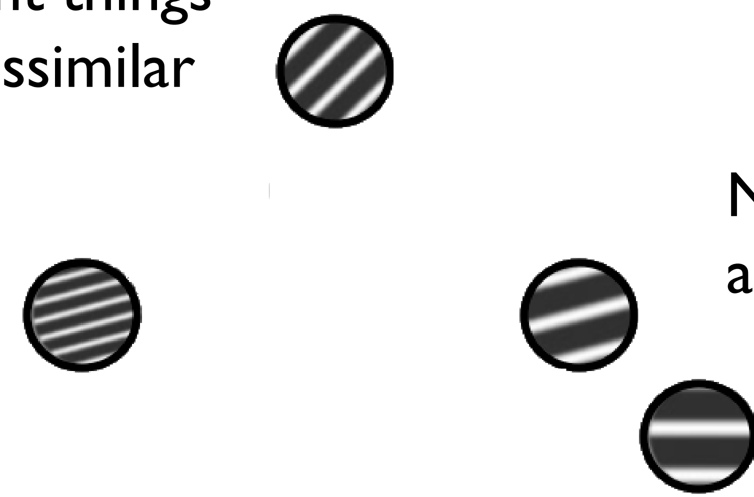


Psychological distance

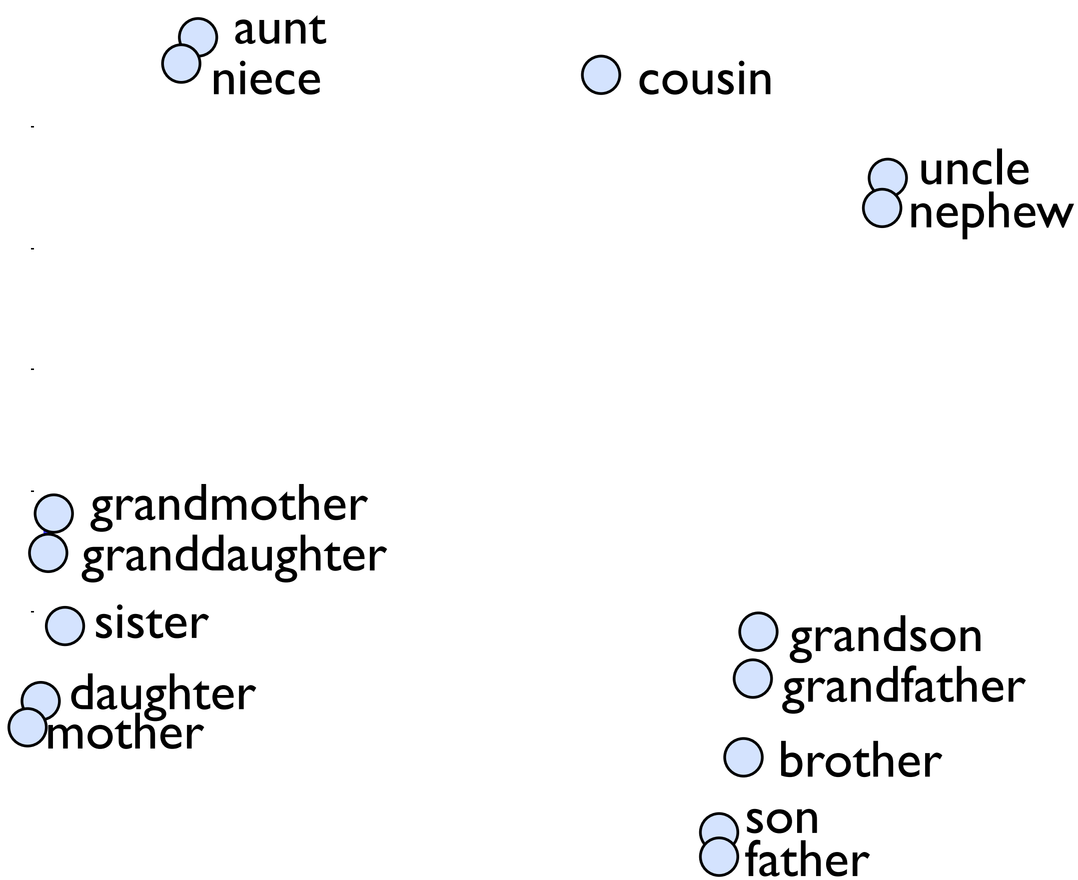


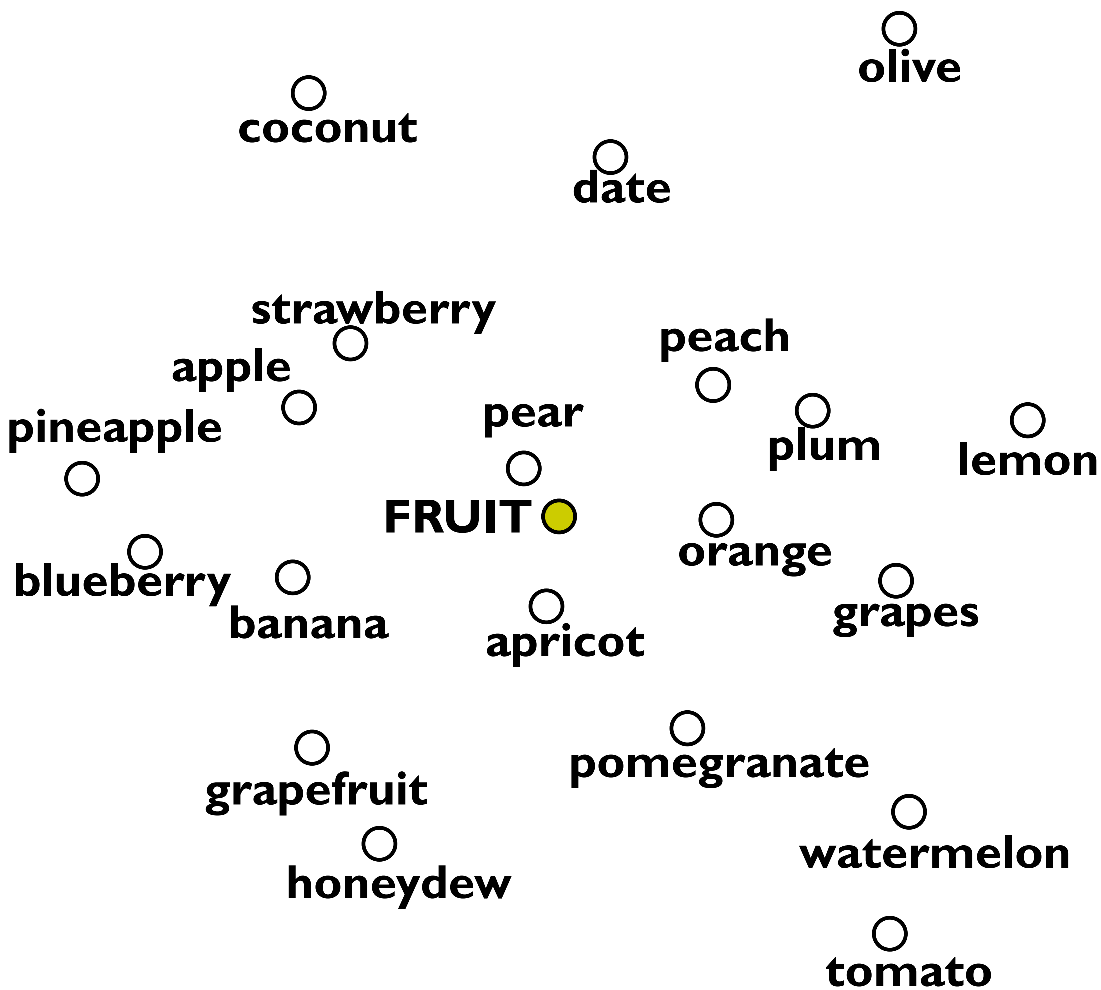
Stimuli differing on one or more perceptual dimensions

Distant things
are dissimilar



Nearby things
are similar


aunt
niece
cousin
uncle
nephew
grandmother
granddaughter
sister
daughter
mother
grandson
grandfather
brother
son
father


olive
coconut
date
strawberry
peach
apple
pear
pineapple
plum
lemon
FRUIT
orange
blueberry
banana
grapes
apricot
pomegranate
grapefruit
watermelon
honeydew
tomato

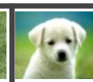
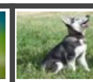
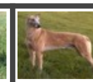
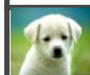
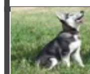
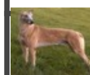
Measurement methods?

- Confusability: probability of mistaking A for B
- Reaction time: time taken to distinguish A from B
- Forced choice: is X more like A or more like B?
- Likert scales: how similar is A to B?
- Sorting tasks: arrange objects into groups

Empirical data…

Distance matrix Δ where $\delta_{i,j}$ is the distance between objects i and j

$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,I} \\ \vdots & \vdots & & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix}.$$

						
	0	1	2	4	5	6
	1	0	1	3	4	6
	2	1	0	1	3	4
	4	3	1	0	2	3
	5	4	3	2	0	1
	6	6	4	3	1	0

… blah blah blah …

- Take this data and use it to work out where items are?
- Use a technique called *Multidimensional Scaling (MDS)*

Basically uses numerical optimisation to find the points in a k-dimensional space that preserves the distances as well as possible - i.e., that minimises a function like the following

$$\min_{x_1,\ldots,x_I} \sum_{i<j} (\|x_i - x_j\| - \delta_{i,j})^2$$

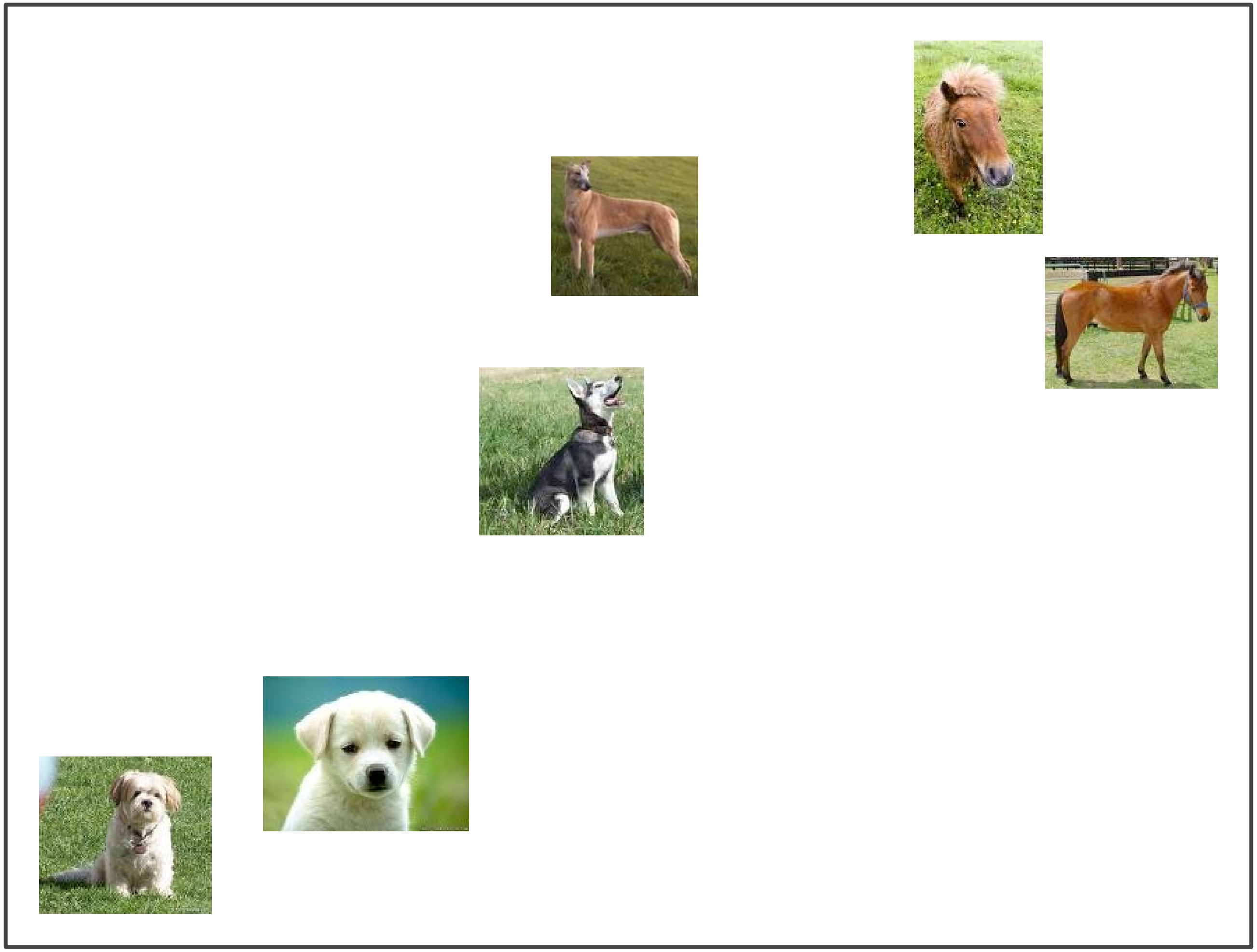
```
> deltas
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    1    2    4    5    6
[2,]    1    0    1    3    4    6
[3,]    2    1    0    1    3    4
[4,]    4    3    1    0    2    3
[5,]    5    4    3    2    0    1
[6,]    6    6    4    3    1    0
> mds <- cmdscale(d=deltas,eig=TRUE)
```

In R: `cmdscale(d,eigs=TRUE)`

distance matrix (d)

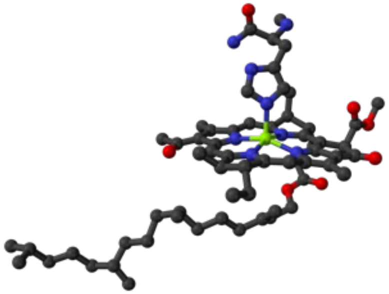
returns eigenvalues for each dimension (eigs=TRUE)

… a psychological space capturing people's intuitions about *similarity*



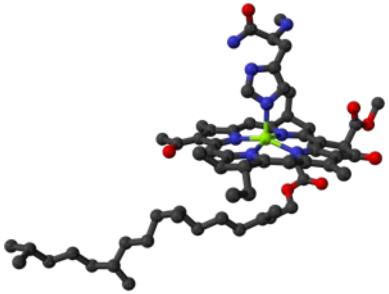
Inductive generalisation

This animal produces TH4 enzyme



Inductive generalisation

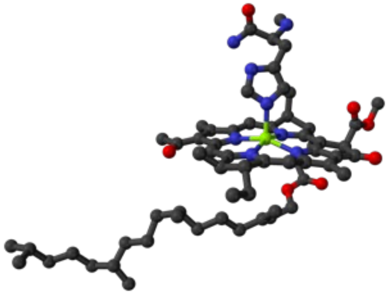
This animal produces TH4 enzyme



Does this one?

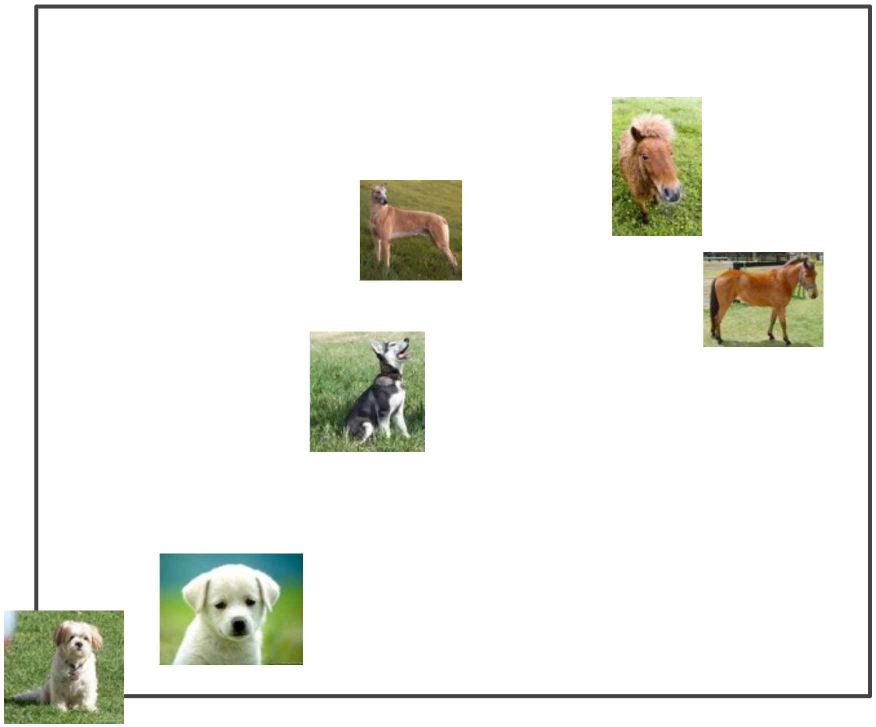
Inductive generalisation

This animal produces TH4 enzyme

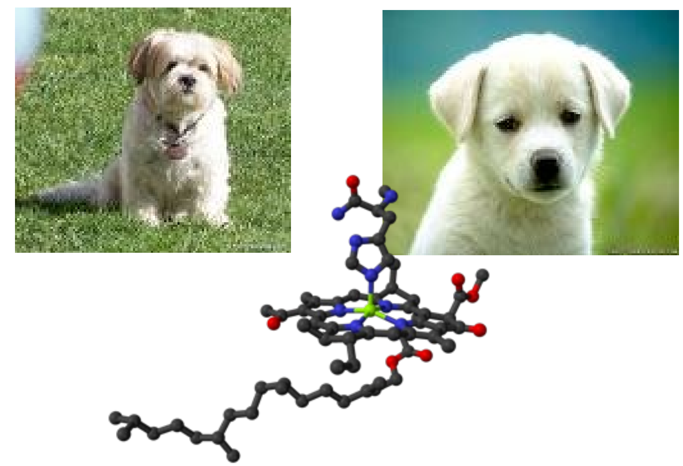


What about this one?

More **similarity**…

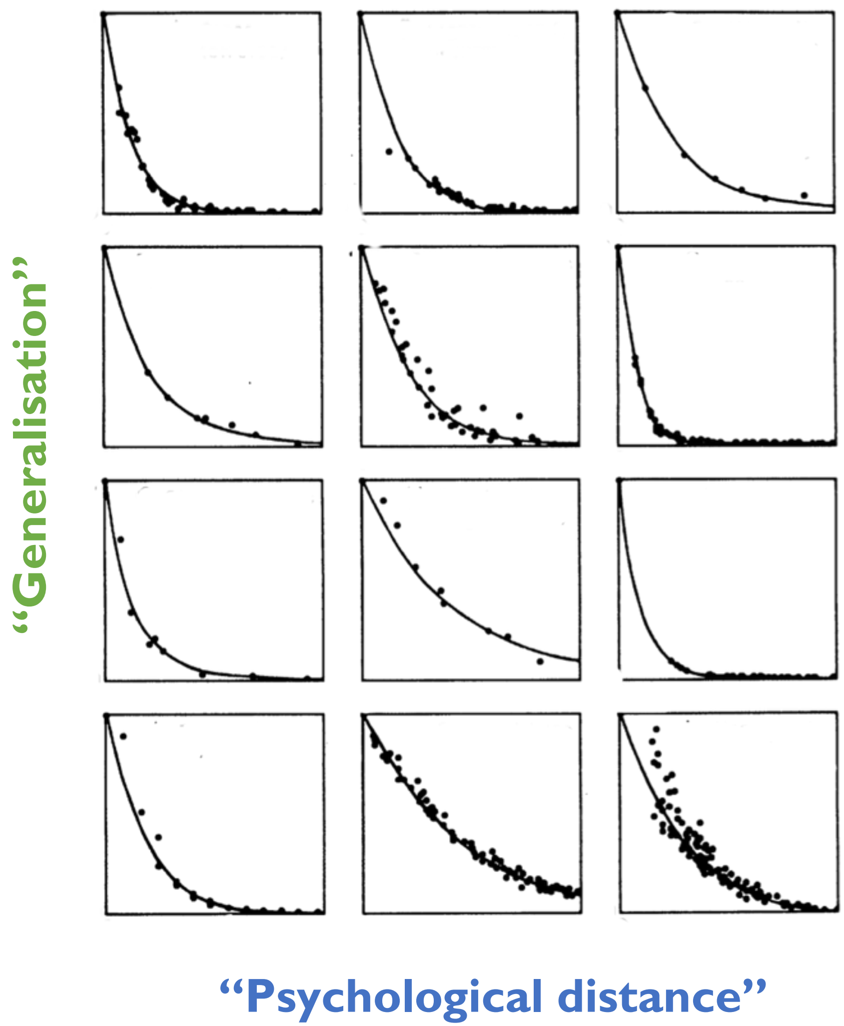


… produces more **generalisation**.

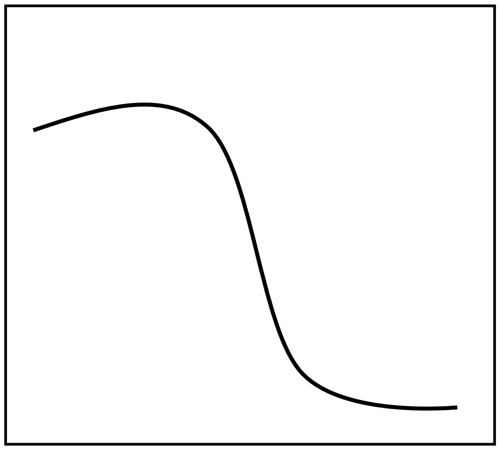


Obviously. But...

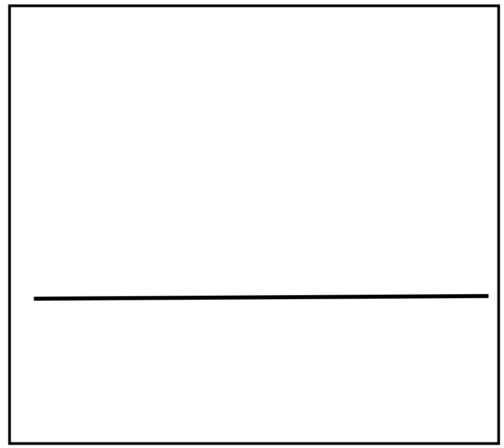
The shape is always the same.
Why is the shape always the same?



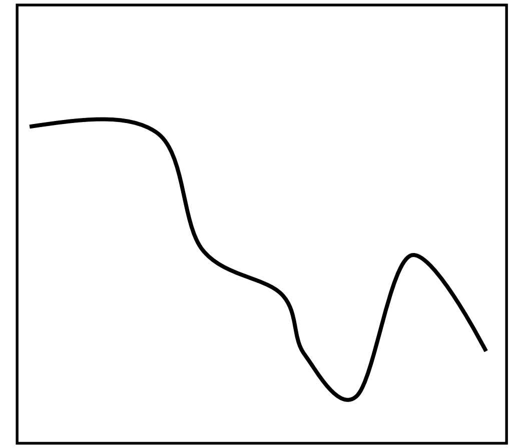
It’s not an inevitability?



This is what you’d expect if people were setting a “hard and fast” boundary



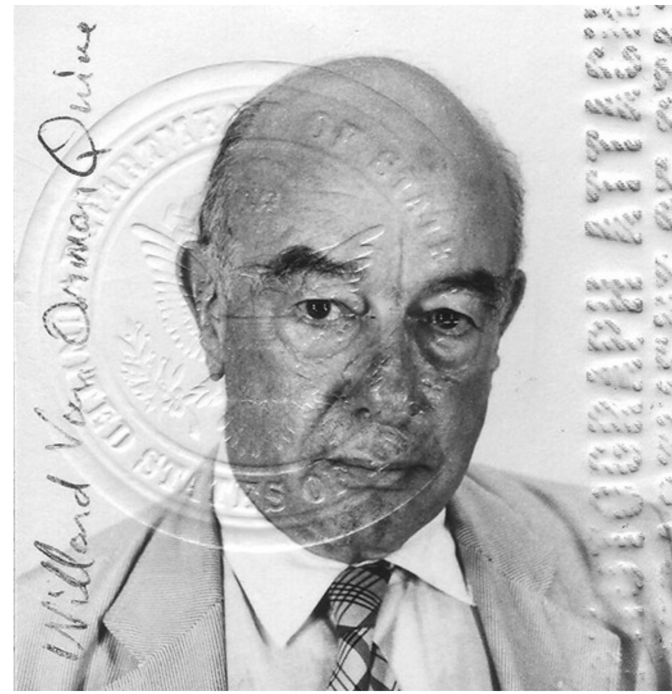
This is what you’d expect on average if people were responding randomly



This is pretty weird, but it you can get this if people are doing something sensible that is only loosely related to similarity

Why similarity matters

"Similarity, is fundamental for learning, knowledge and thought, for only our sense of similarity allows us to order things into kinds so that these can function as stimulus meanings. Reasonable expectation depends on the similarity of circumstances and on our tendency to expect that similar causes will have similar effects"

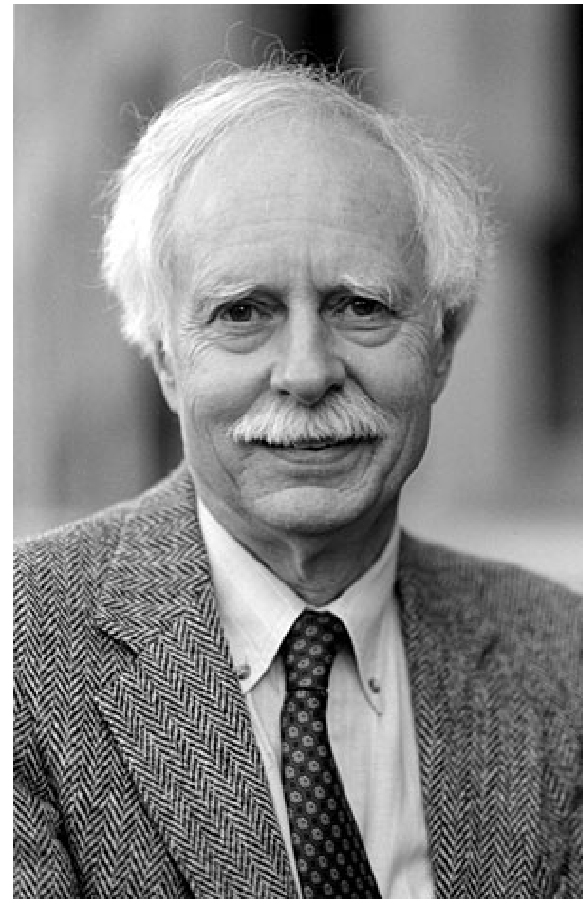


- Willard Van Orman Quine, 1969

Why similarity matters

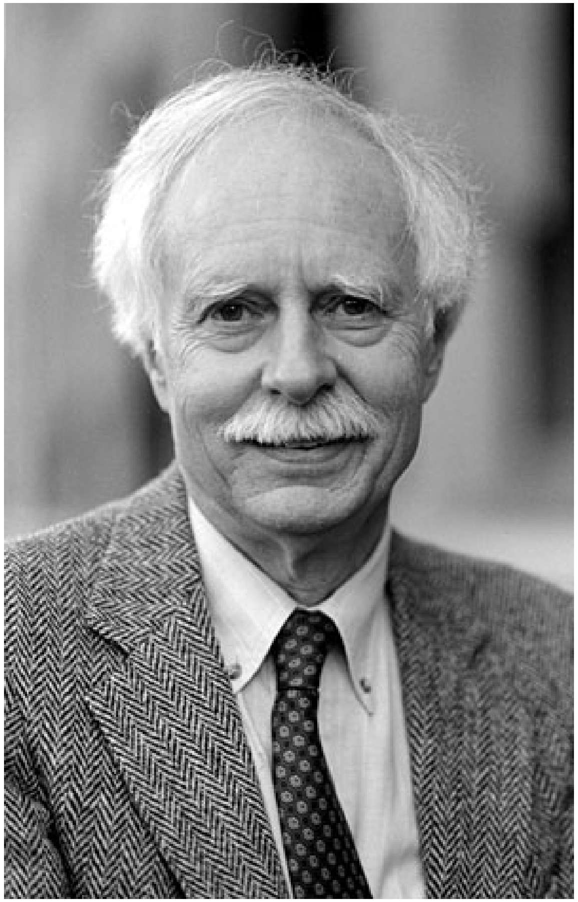
"We generalize from one situation to another not because we cannot tell the difference between the two situations but because we judge that they are likely to belong to a set of situations having the same consequence"

- Roger Shepard, 1987



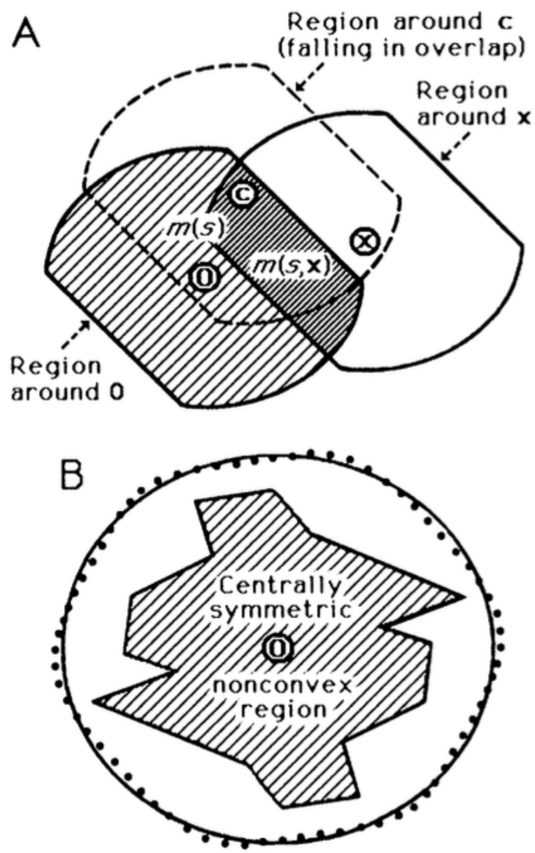
Why similarity matters

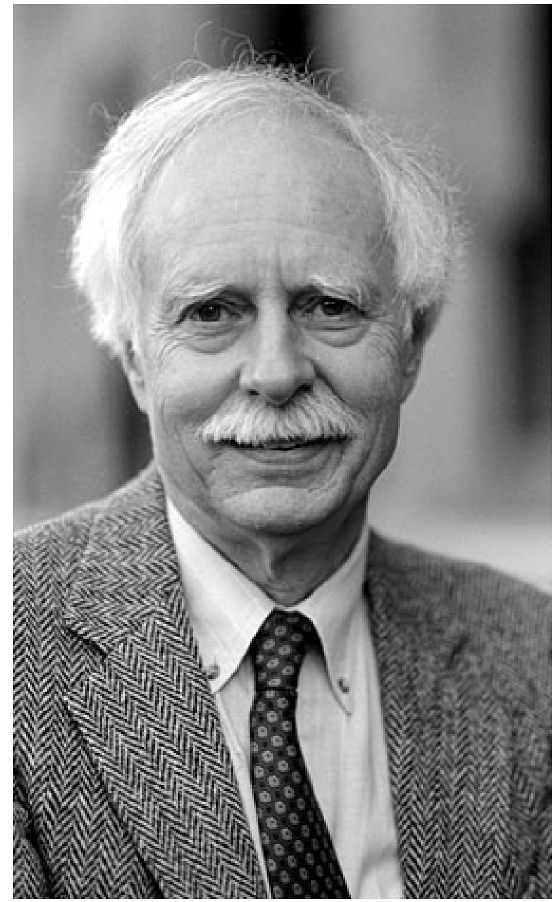
"An object that is significant for an individual … is always a member of a particular class ... Such a class corresponds to some region in the individual's psychological space, which I call a consequential region. I suggest that the psychophysical function that maps physical parameter space into a species psychological space has been shaped ... so that consequential regions are not consistently elongated or flattened in particular dimensions"



Why similarity matters

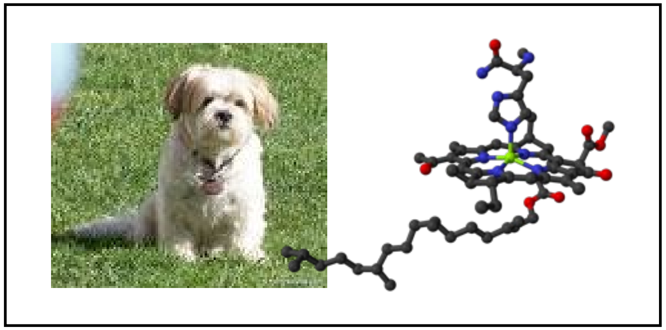
Fig. 2. (**A**) A centrally symmetric convex region shown as centered on **0**, as centered on **x**, and as having a center, **c**, falling within the intersection of the regions centered on **0** and on **x**. (**B**) For an illustrative nonconvex region centered on **0**, the locus of centers, **c**, of similarly shaped regions having a constant (approximately 20%) overlap with the region centered on **0** (dotted curve); and an ellipse corresponding to the Euclidean metric (smooth curve).



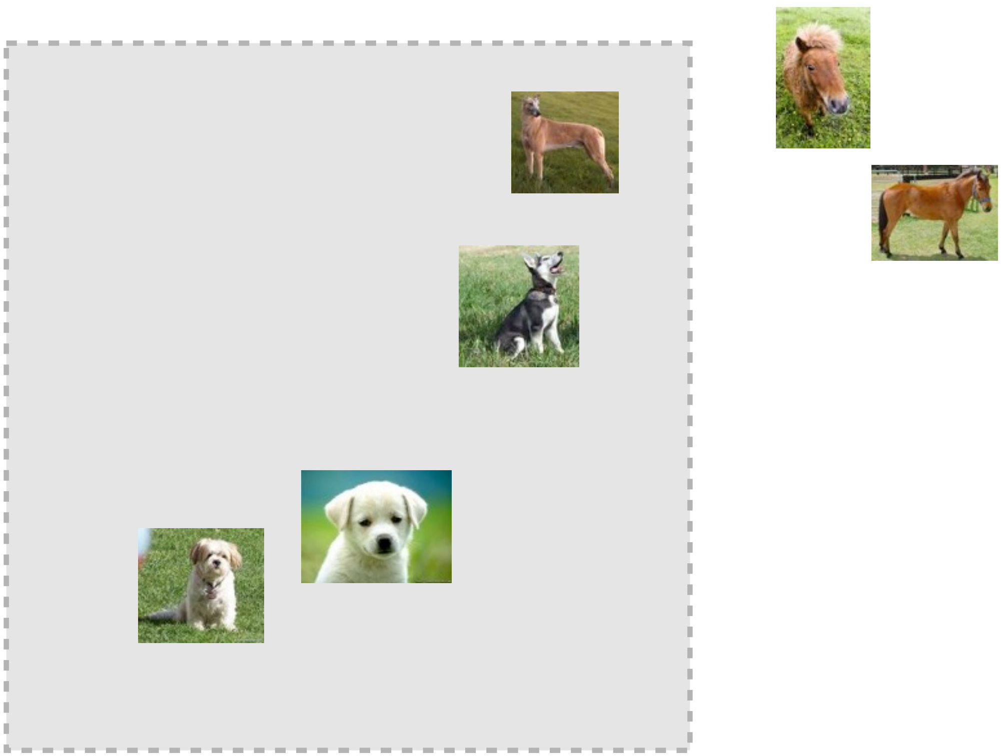


Let's follow Shepard's reasoning…

(Maths warning – this part has equations.
Ignore them and focus on the ideas)

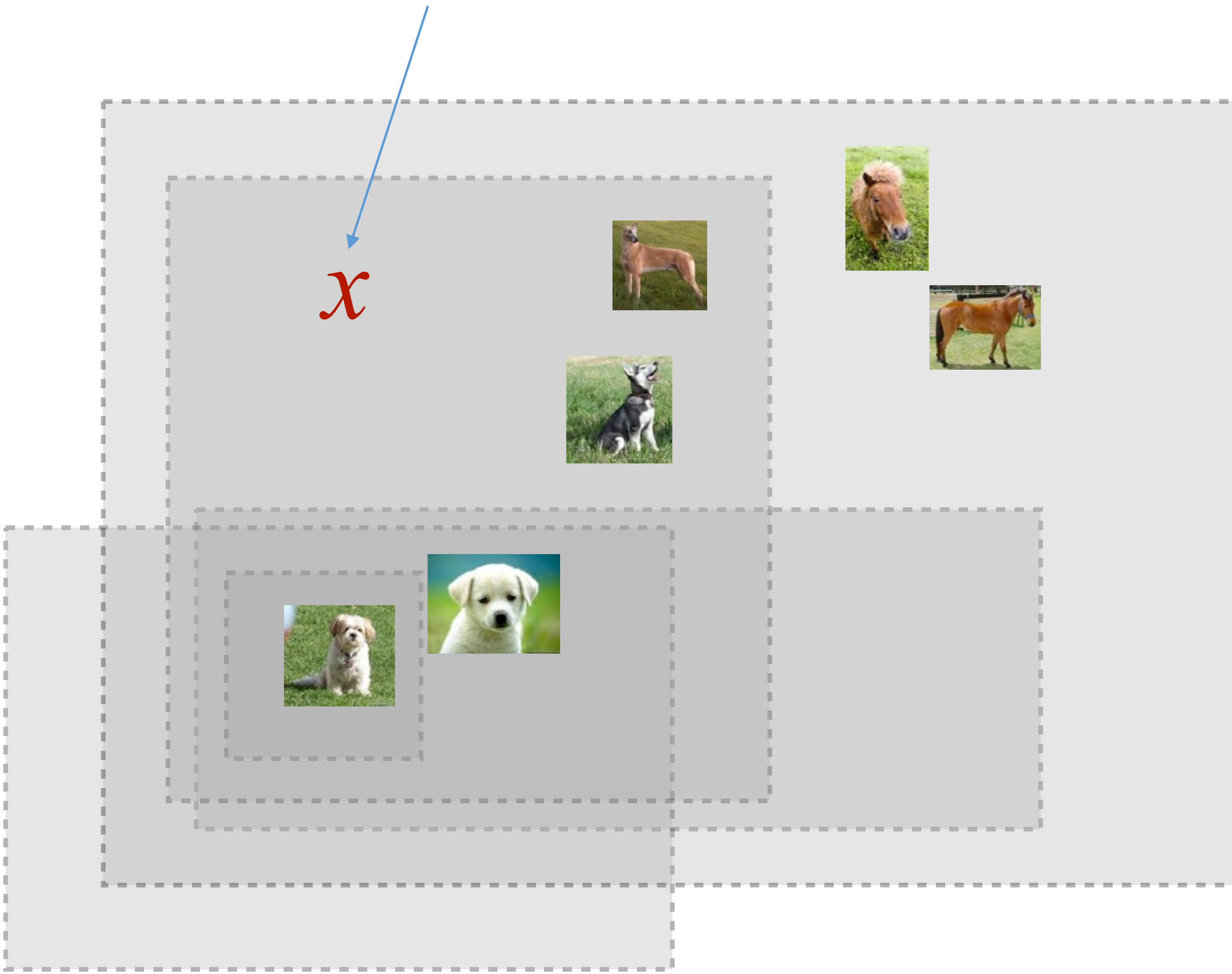


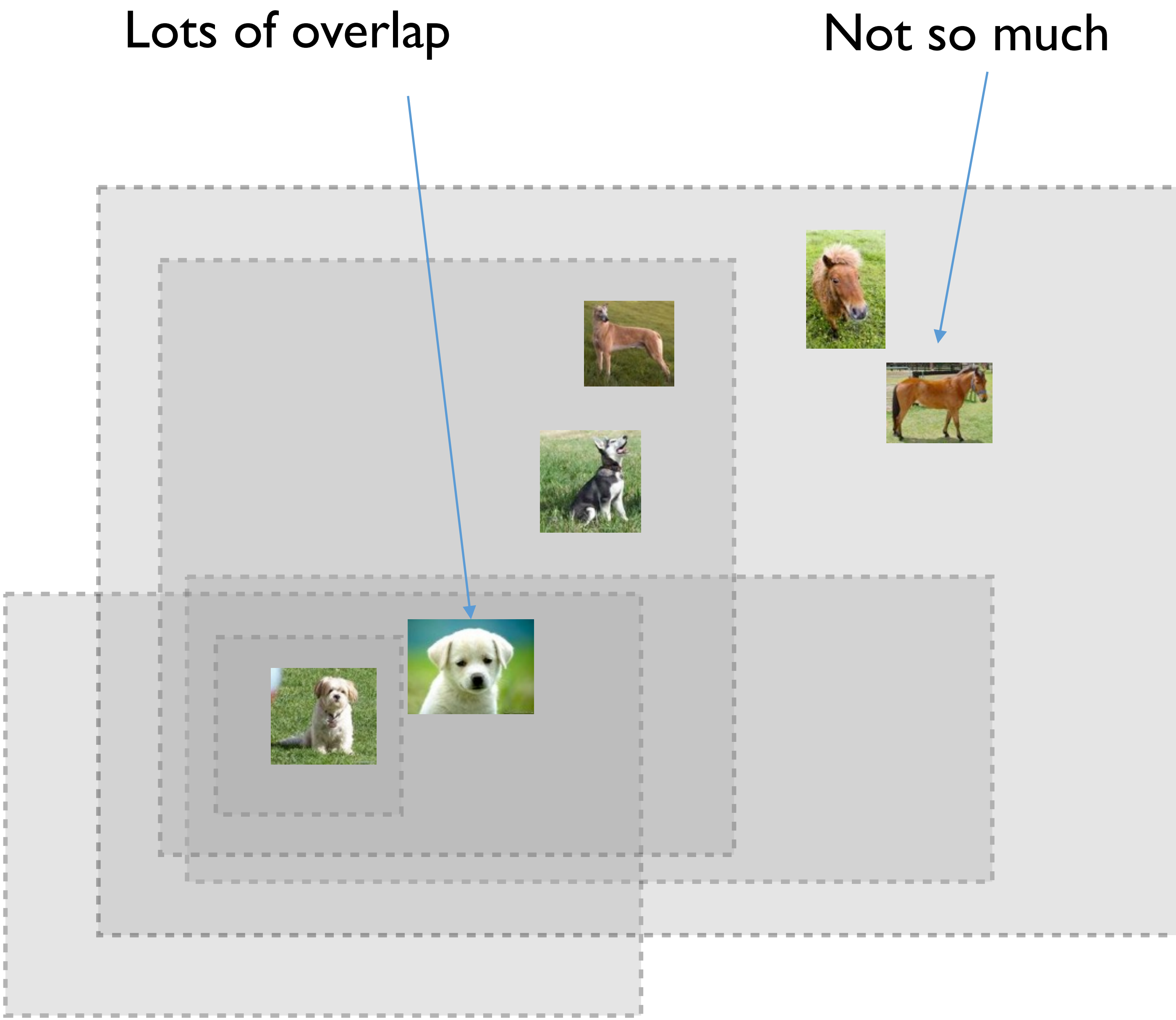
One possible consequential region consistent with the data



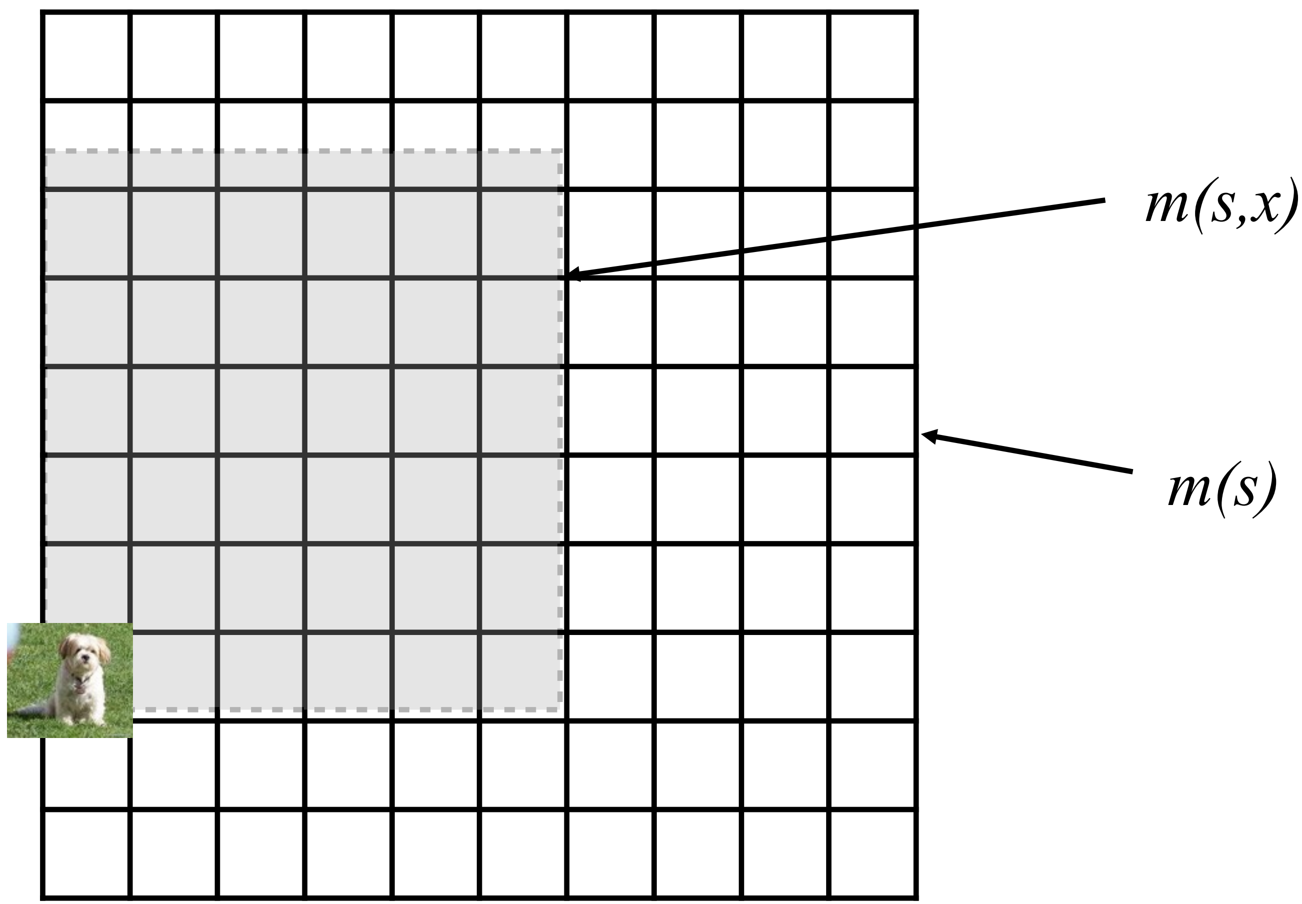
But there are many!

The probability that a new item *x* shares the property depends on the probability that a consequential region includes that point




Lots of overlap
Not so much

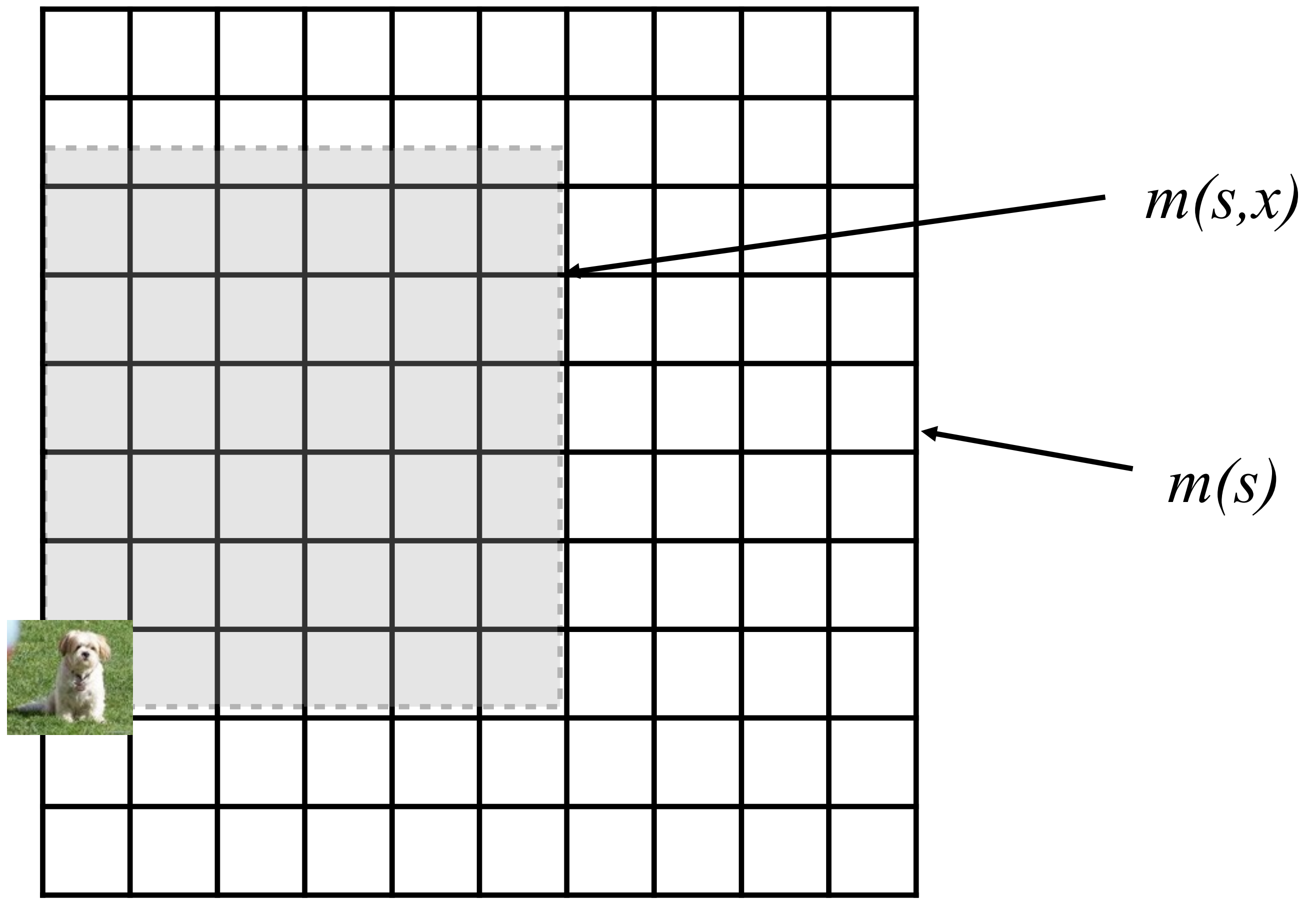
For some *single* consequential region of size s, the conditional probability that x is contained in the region is just the ratio $m(s,x)/m(s)$ of the measure of the overlap to the measure of the whole such region



We calculate the probability that x is contained in any arbitrary region by integrating over all possible regions

$$g(\mathbf{x}) = \int_0^\infty p(s)\,\frac{m(s,\mathbf{x})}{m(s)}\,ds$$

g(x) = probability that a response learned to stimulus 0 will generalise to *x*



We calculate the probability that x is contained in any arbitrary region by integrating over all possible regions

$$g(\mathbf{x}) = \int_0^\infty p(s)\,\frac{m(s,\mathbf{x})}{m(s)}\,ds$$

g(x) = probability that a response learned to stimulus 0 will generalise to *x*

Then we find that generalisation is a function of the distance *d* in the psychological space that…

$$g(d) = \int_d^\infty p(s)\,ds - d\int_d^\infty \frac{p(s)}{s}\,ds$$

Has unit value at $d=0$

$$g'(d) = -\int_d^\infty \frac{p(s)}{s}\,ds$$

Monotonically decreases with increasing *d*

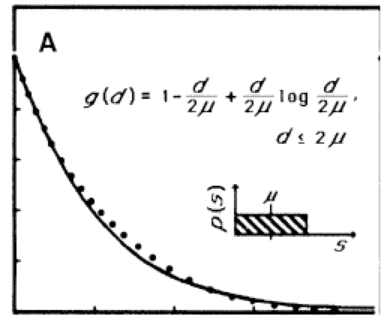
$$g''(d) = \frac{p(d)}{d}$$

Is concave upward (mostly)

Shape of the generalisation gradient?

$$g(\mathbf{x}) = \int_0^\infty p(s)\,\frac{m(s,\mathbf{x})}{m(s)}\,ds$$

That depends on the nature of *p(s)* ...

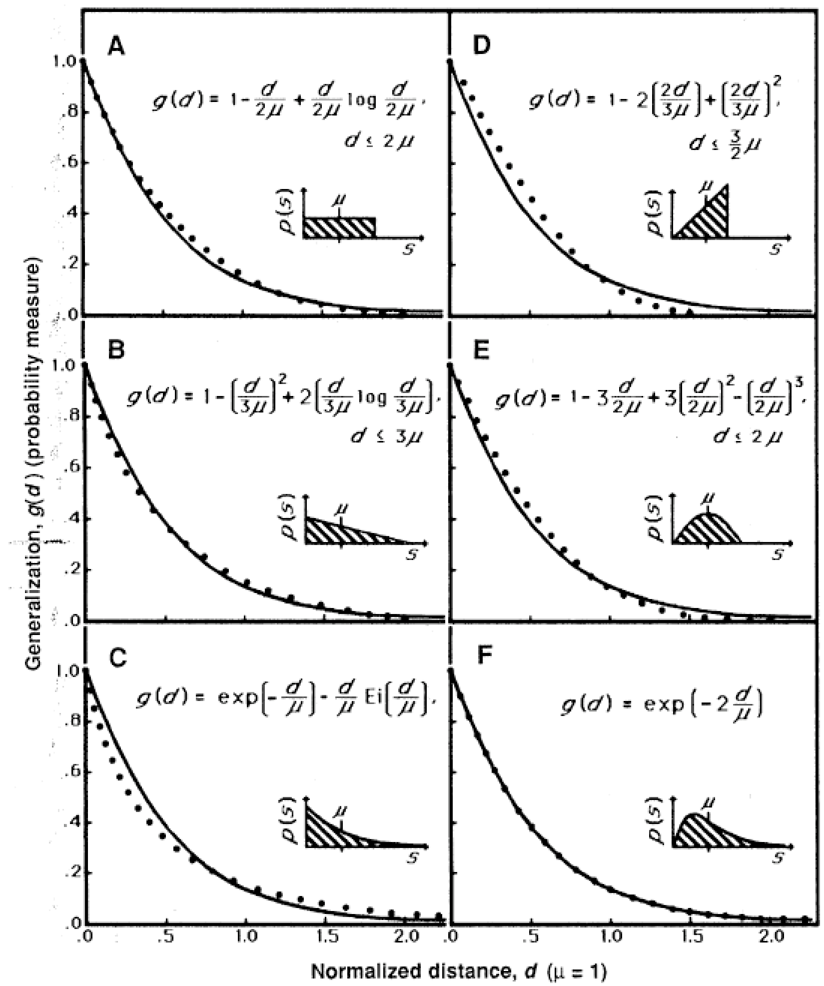


Shape of the generalisation gradient?

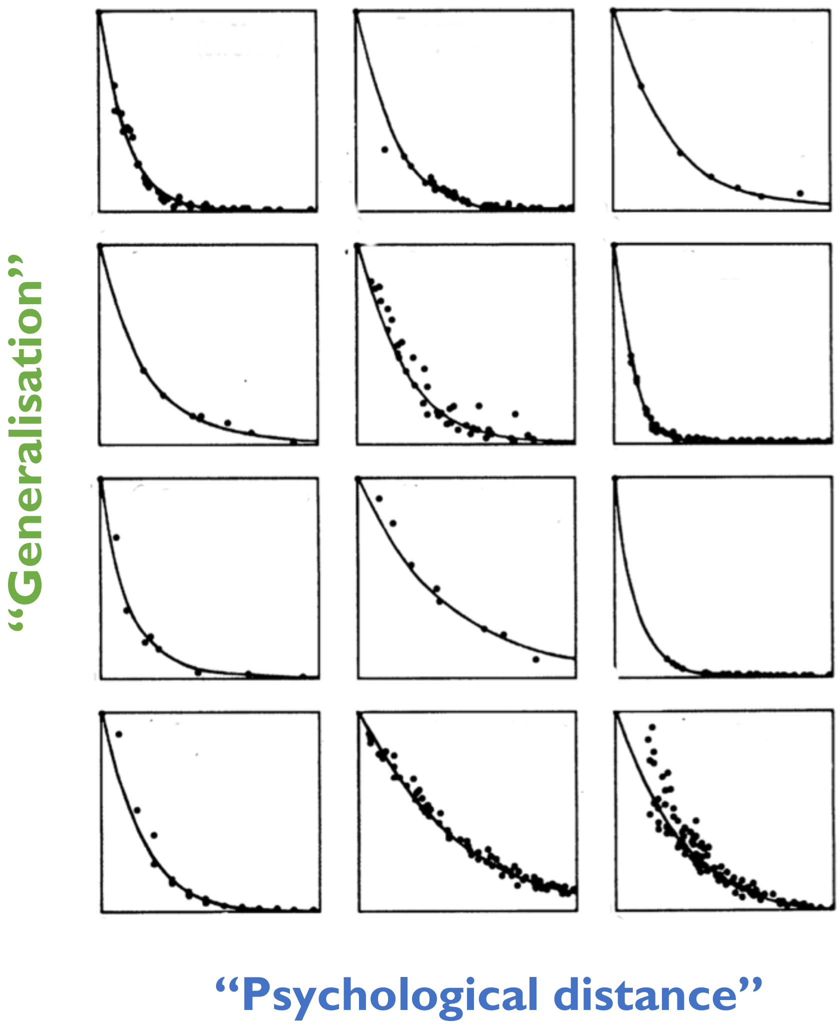
$$g(\mathbf{x}) = \int_0^\infty p(s)\, \frac{m(s,\mathbf{x})}{m(s)}\, ds$$

That depends on the nature of *p(s)* …

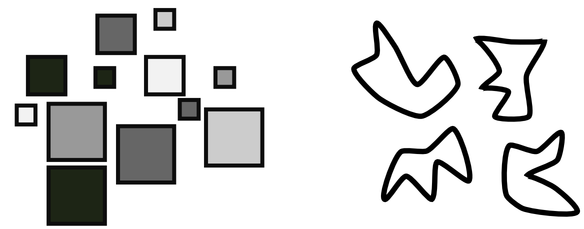
… but not very much



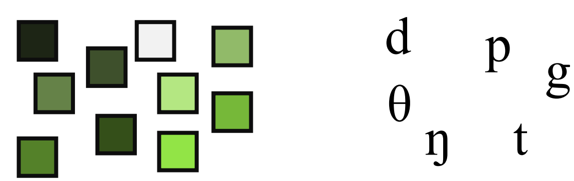
So why do we see this invariance?



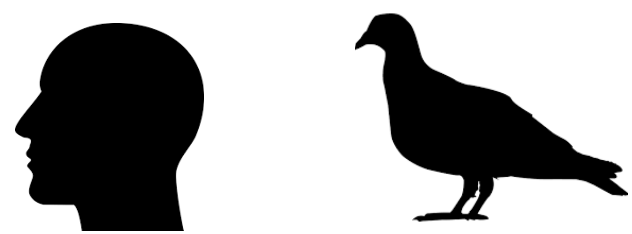
Invariance across stimulus types



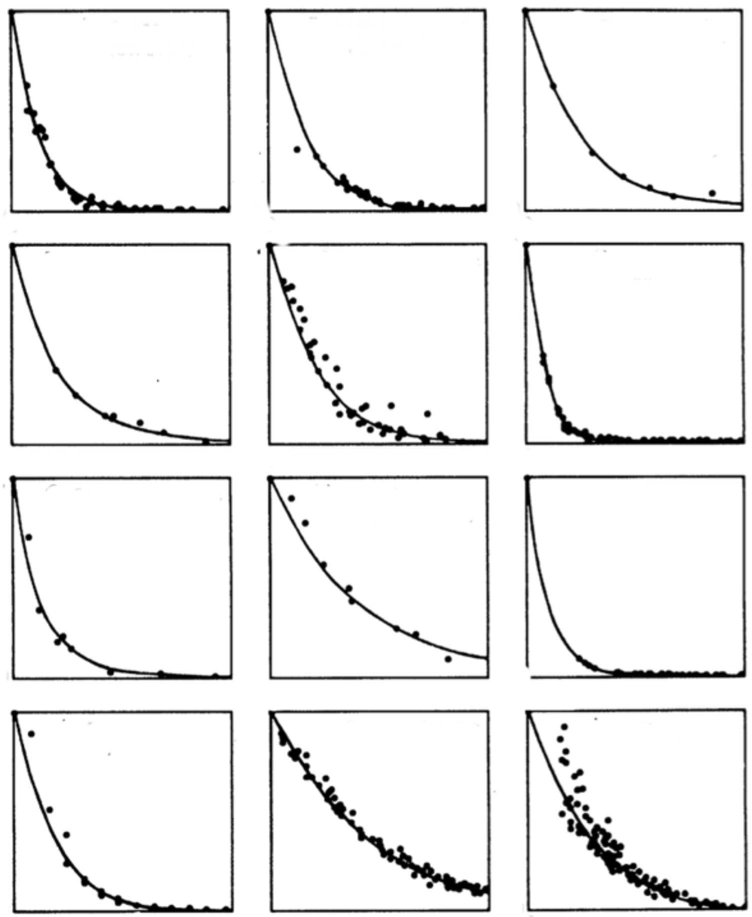
Invariance across sensory modalities



Invariance across species

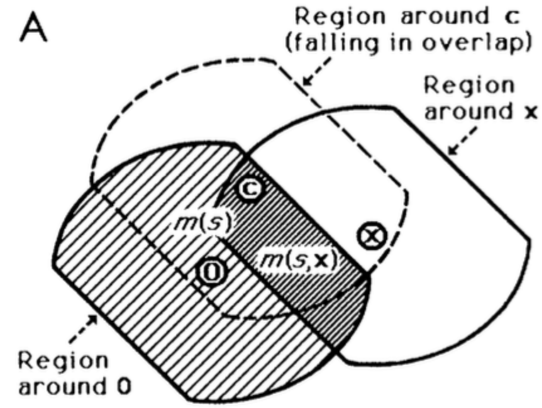


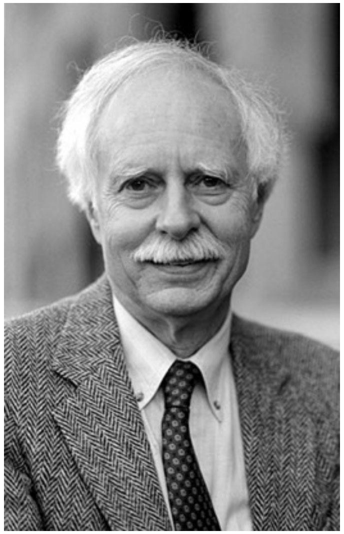
Structure of the problem!



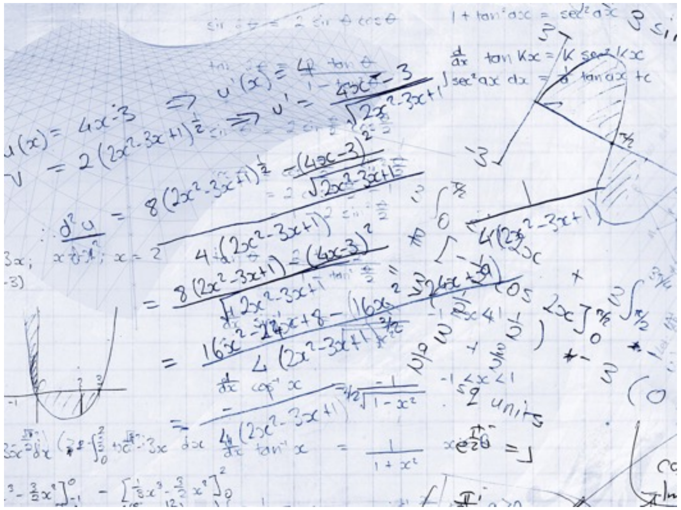
“Generalisation”

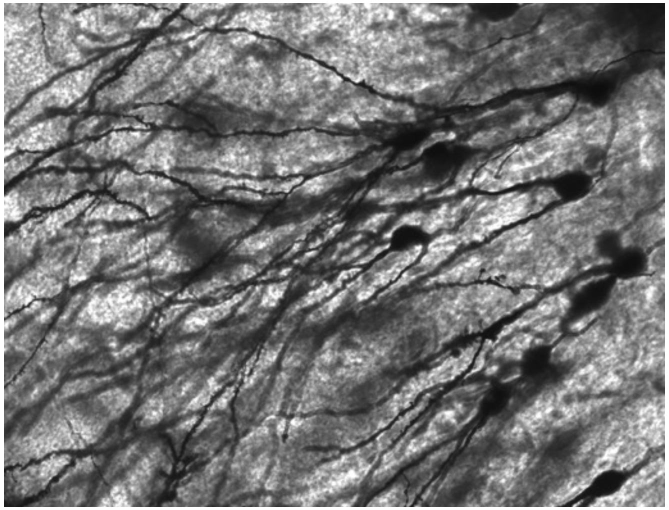
“Psychological distance”



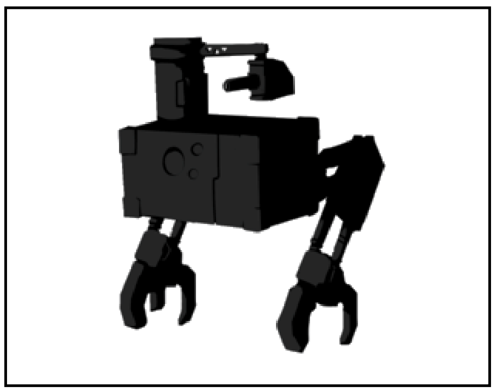


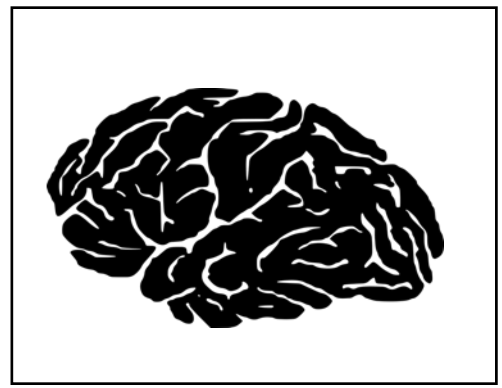
Structure of the problem!



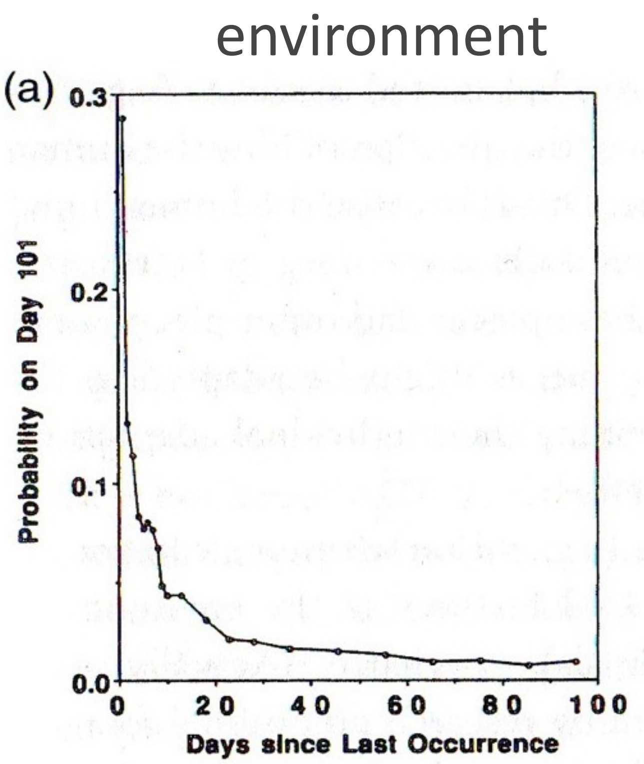


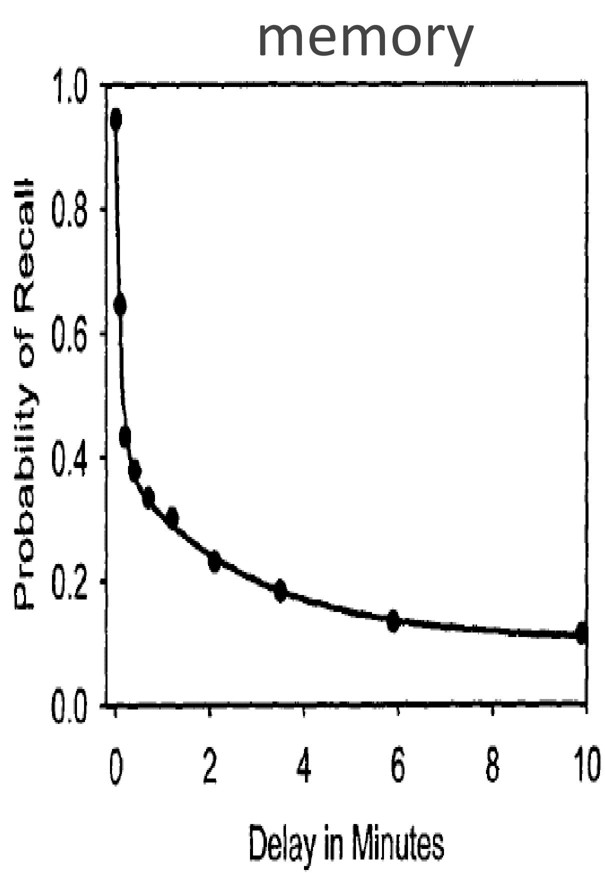
Are there other examples?





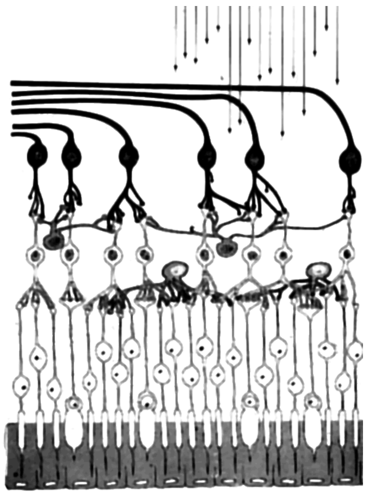
Memory?

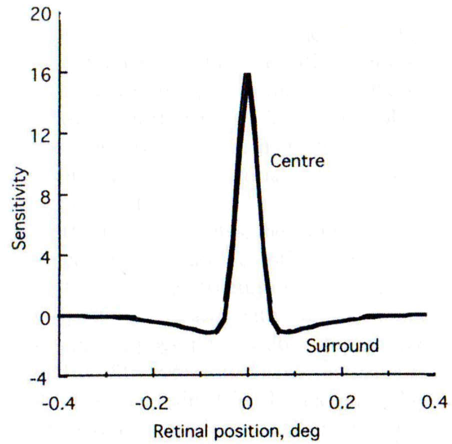




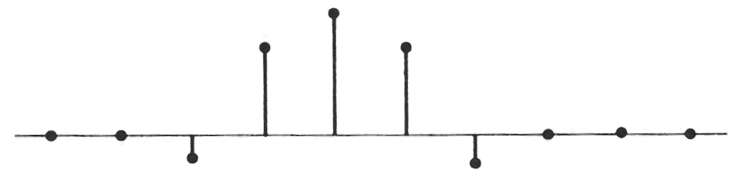
The probability that people recall a particular piece of information is closely matched to the probability that it is needed by the person. (Anderson & Schooler, 1991)

Vision?

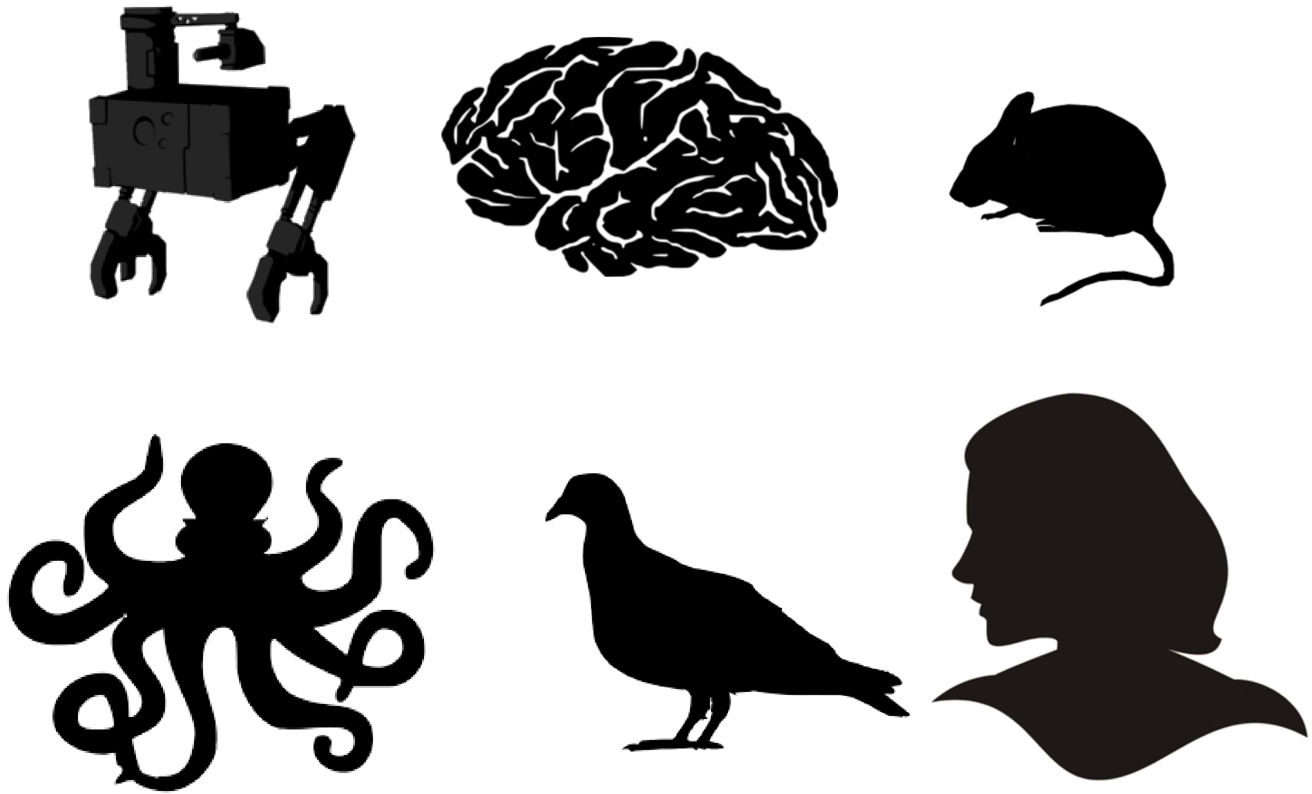




Centre-surround receptive fields have a mirror in optimal signal processing…



… because high pass filters are good for edge detection?



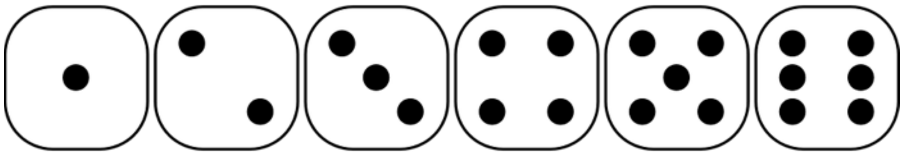
“Bonus lecture”
Introduction to Bayes

Structure

- What does the word “probability” mean?
- What are the rules that probabilities obey?
- Discovering Bayes’ rule
- What is Bayes’ rule used for?
- An example of Bayesian reasoning

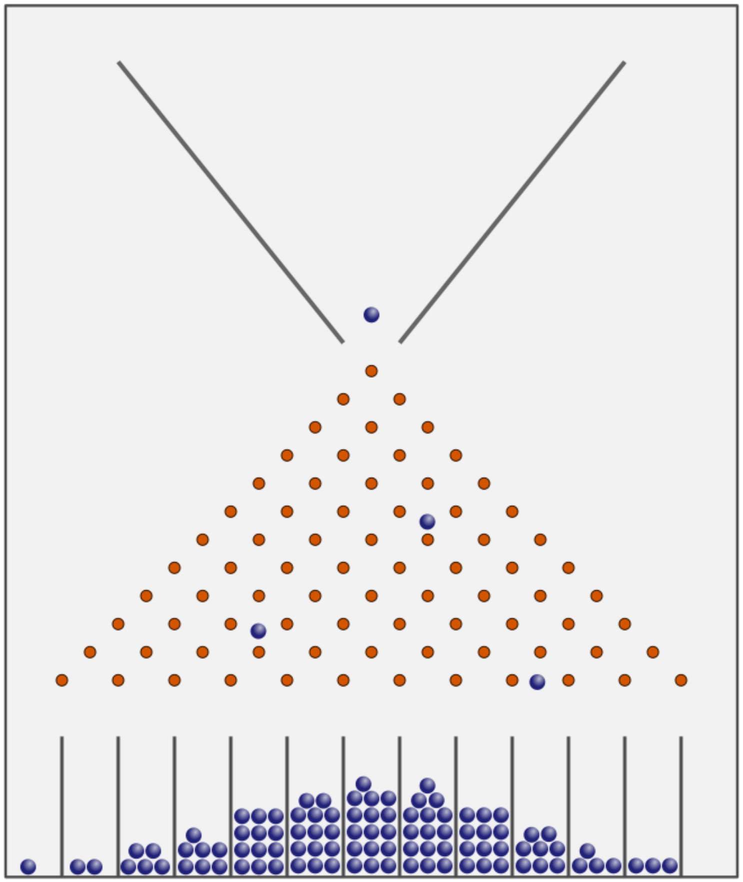
What does the word “probability” mean?

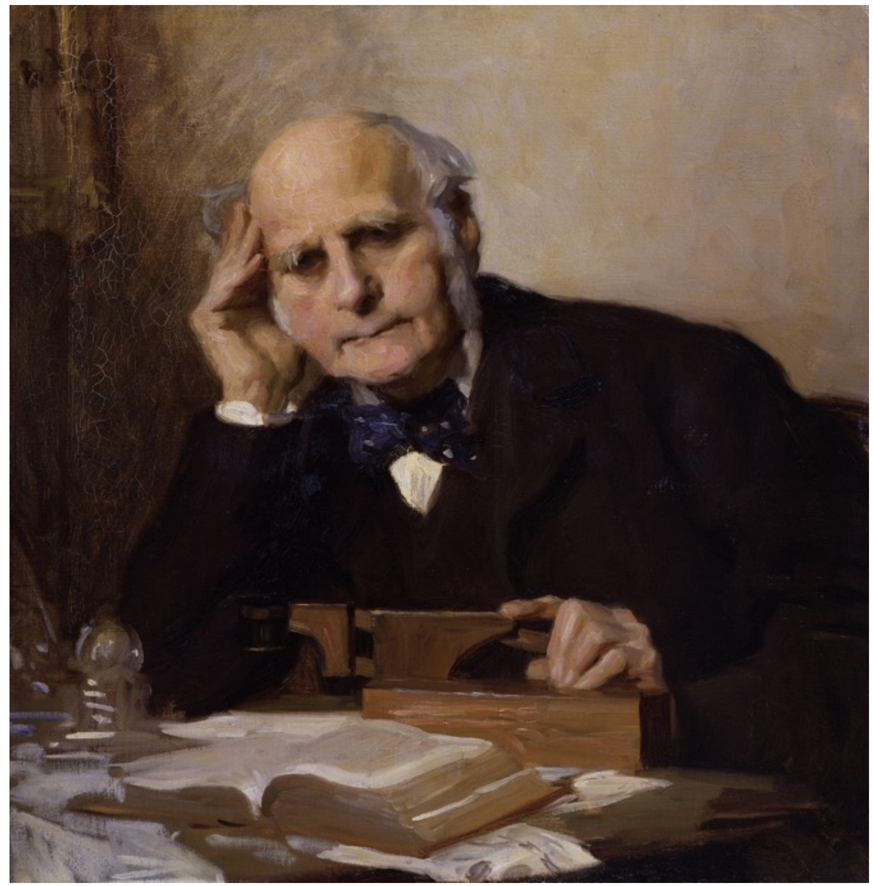
“Aleatory” processes

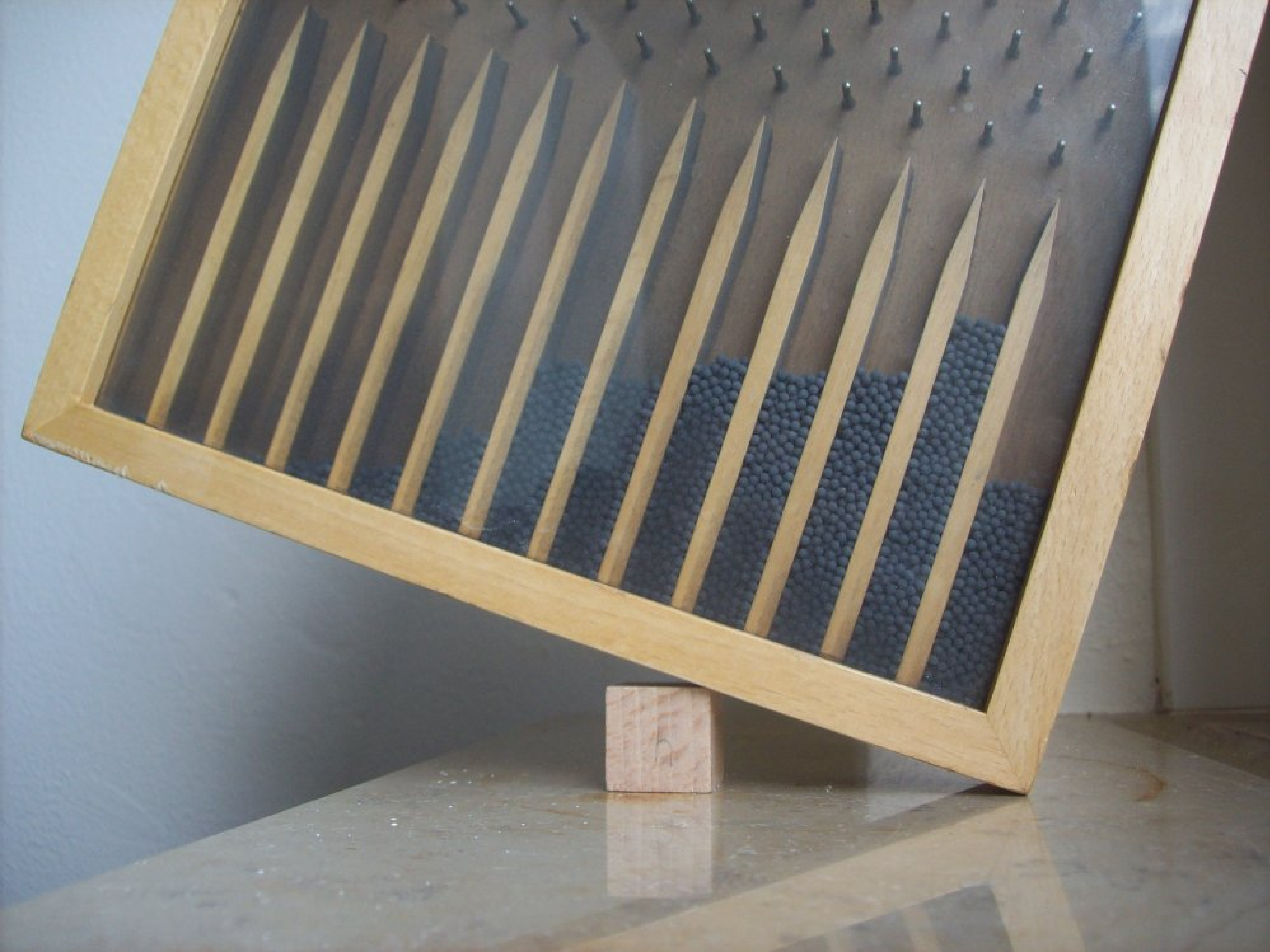


Probability is an objective characteristic associated with physical processes, defined by counting the relative frequencies of different kinds of events when that process is invoked

“Aleatory” processes

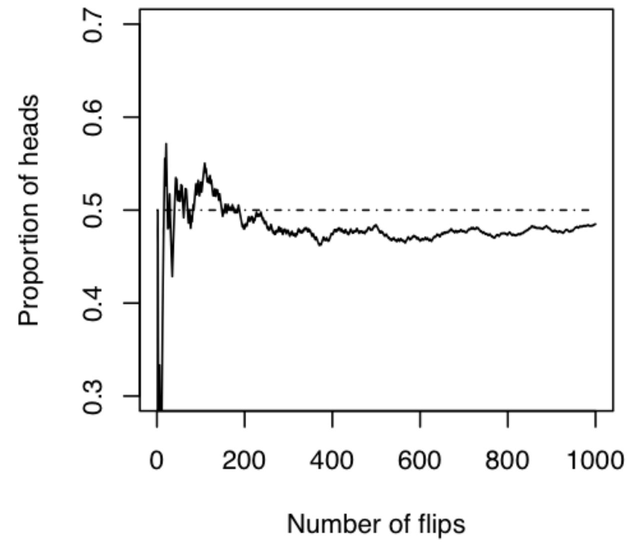






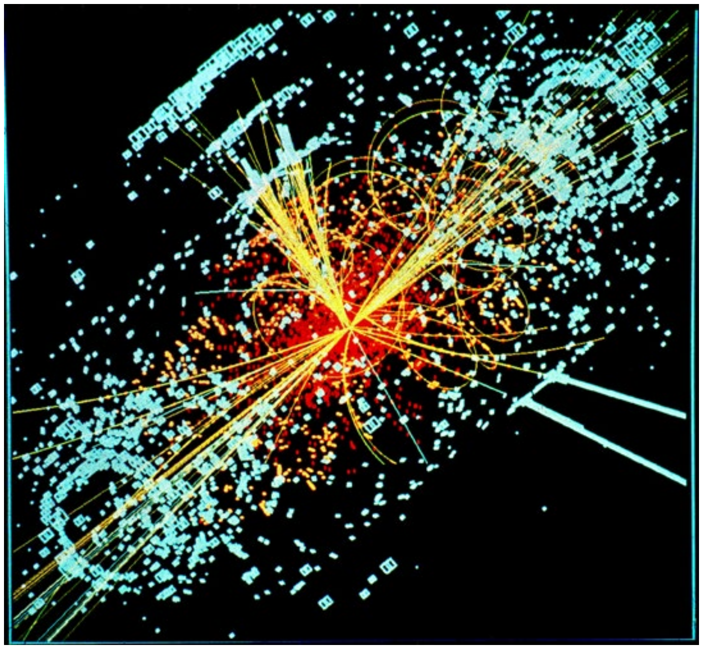
Frequentist statistics

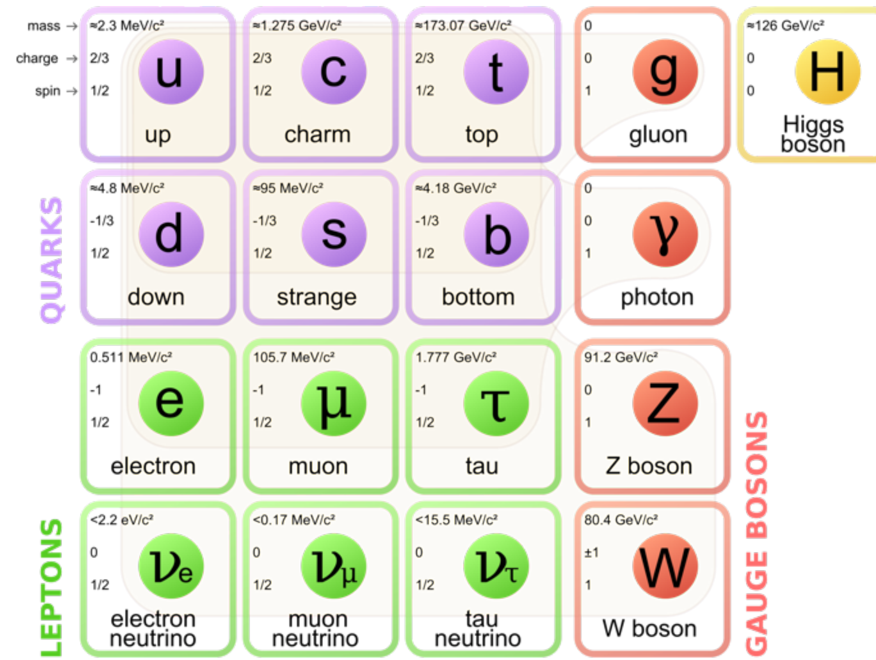
Coin flipping is an aleatory process, and can be repeated as many times as you like



The probability of a head is defined as the long-run frequency

Frequentist statistics



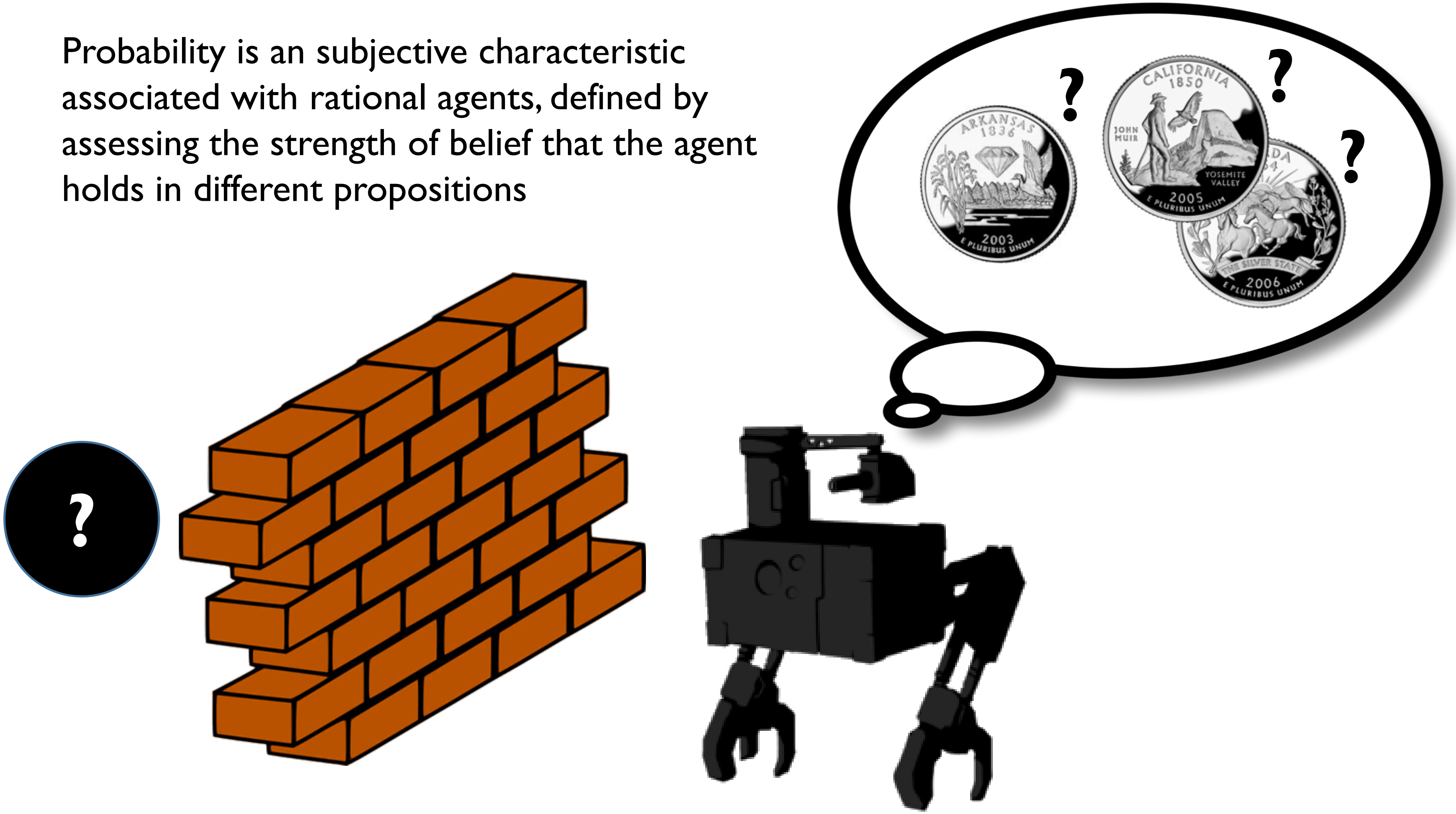


A particle physics experiment is a repeatable procedure, and thus a frequentist probability can be constructed to describe its outcomes

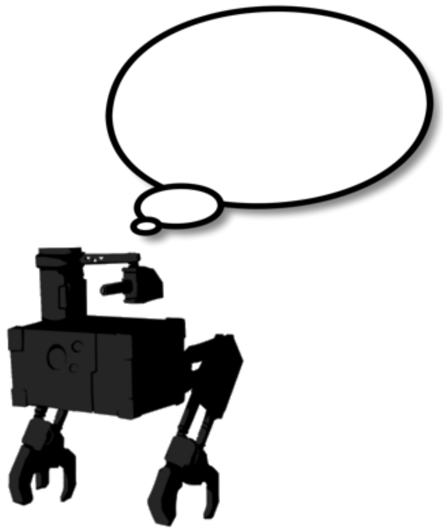
A scientific theory is <u>not</u> a repeatable procedure, and cannot be assigned a probability: there is no such thing as "the probability that my theory is true"

Epistemic uncertainty

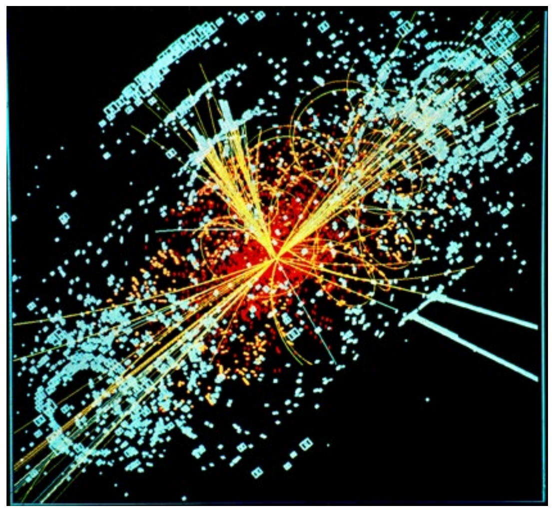
Probability is an subjective characteristic associated with rational agents, defined by assessing the strength of belief that the agent holds in different propositions



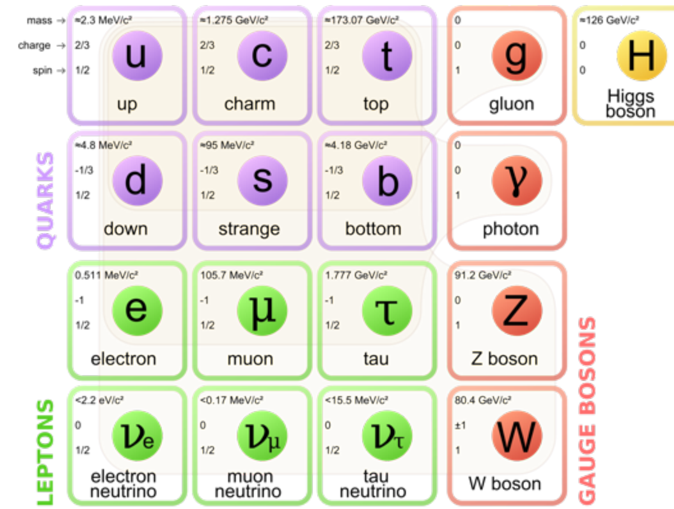
“Bayesian” statistics



Probabilities can be attached to any proposition that an agent can believe

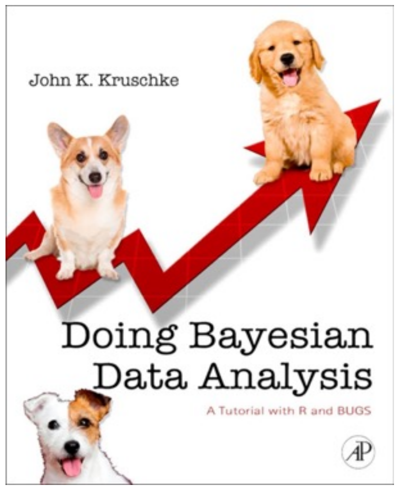


A particle physics experiment generates observable events about which a rational agent might hold beliefs



A scientific theory contains a set of propositions about which a rational agent might hold beliefs

Two different ways to use this idea

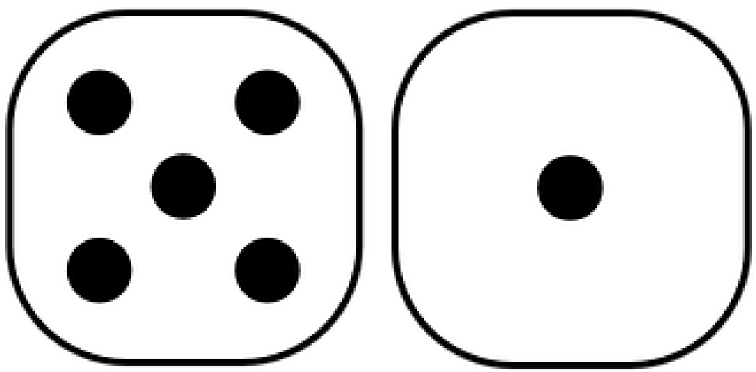


"Psychologists should use Bayesian statistics to analyse data"

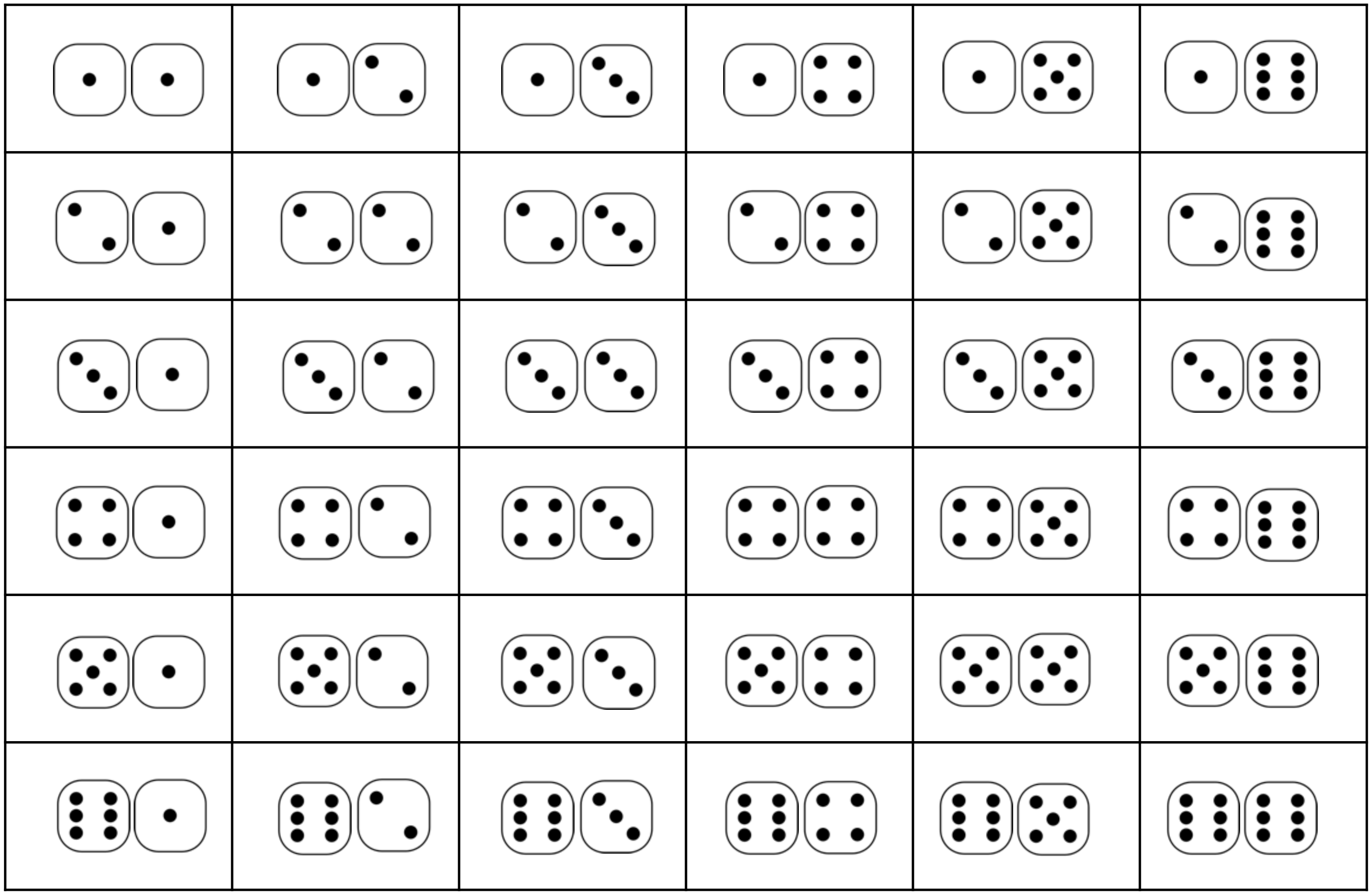
"Human reasoning can be described as a form of Bayesian inference"

How probabilities <u>behave</u>

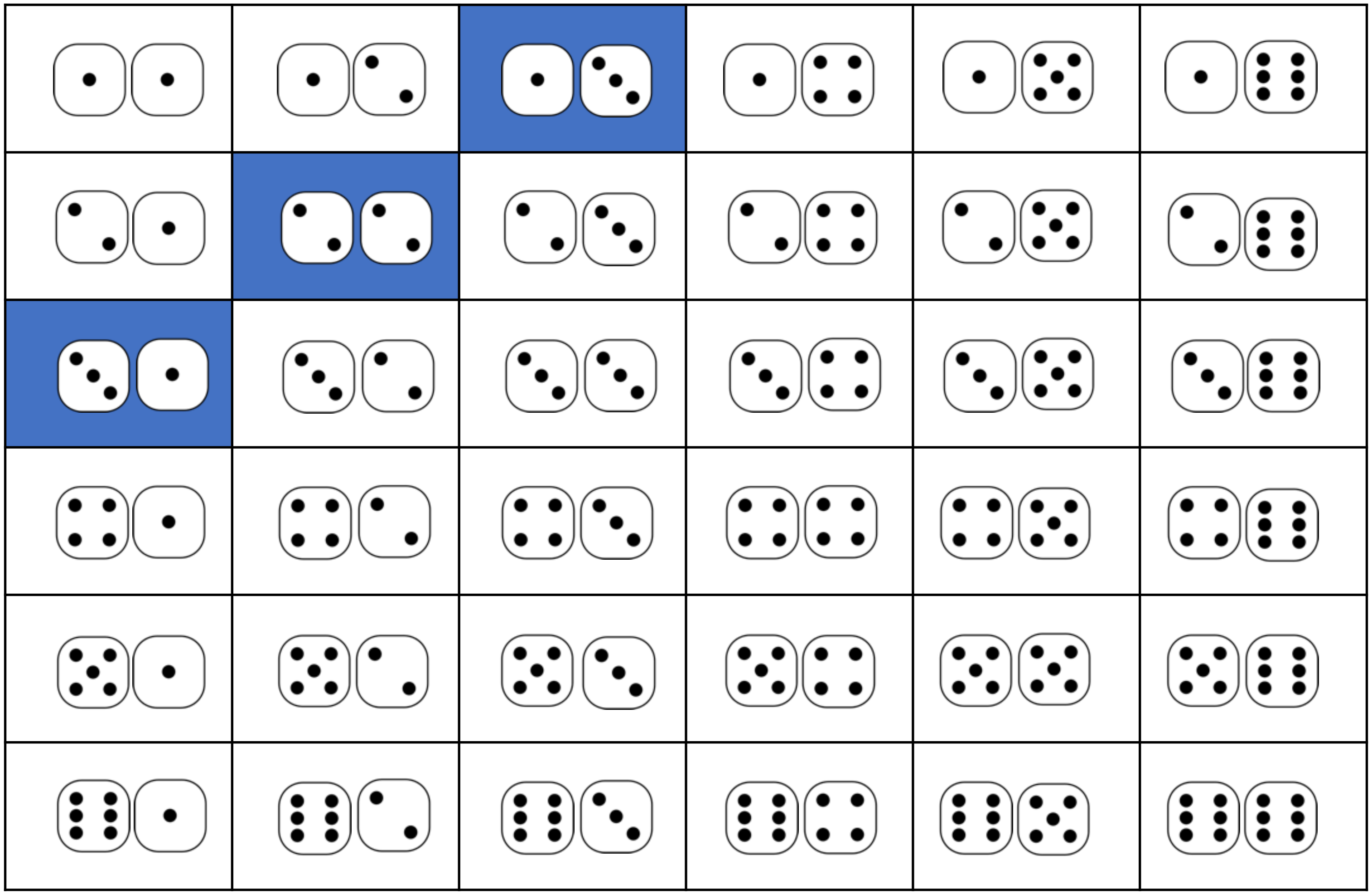
(note: Bayesians and frequentists agree on these rules)



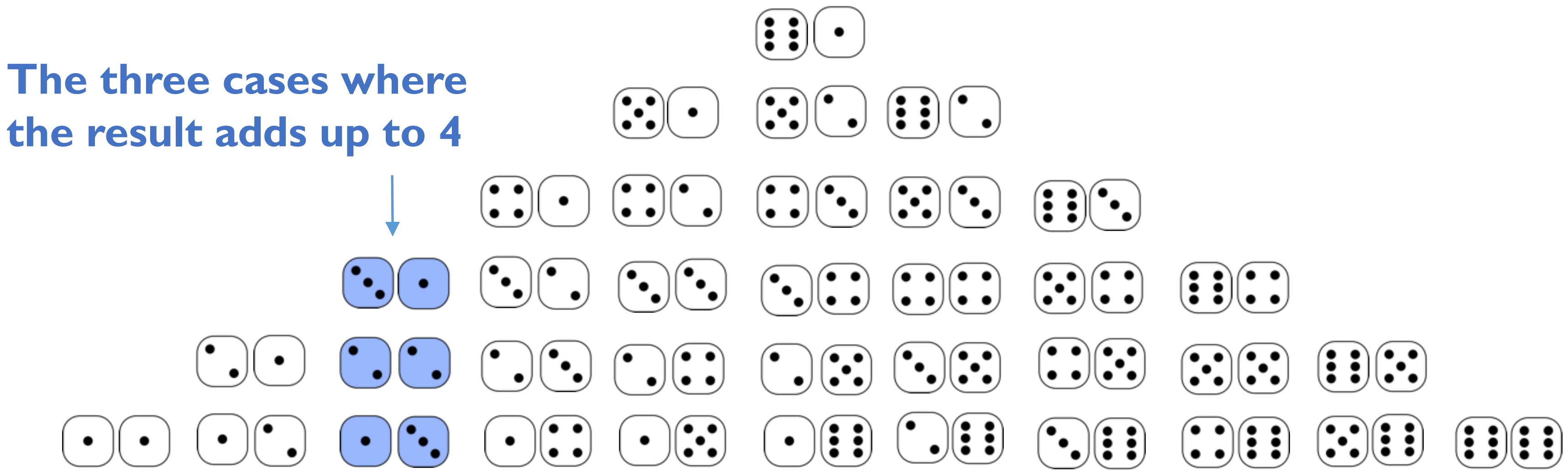
Thirty six cases in total

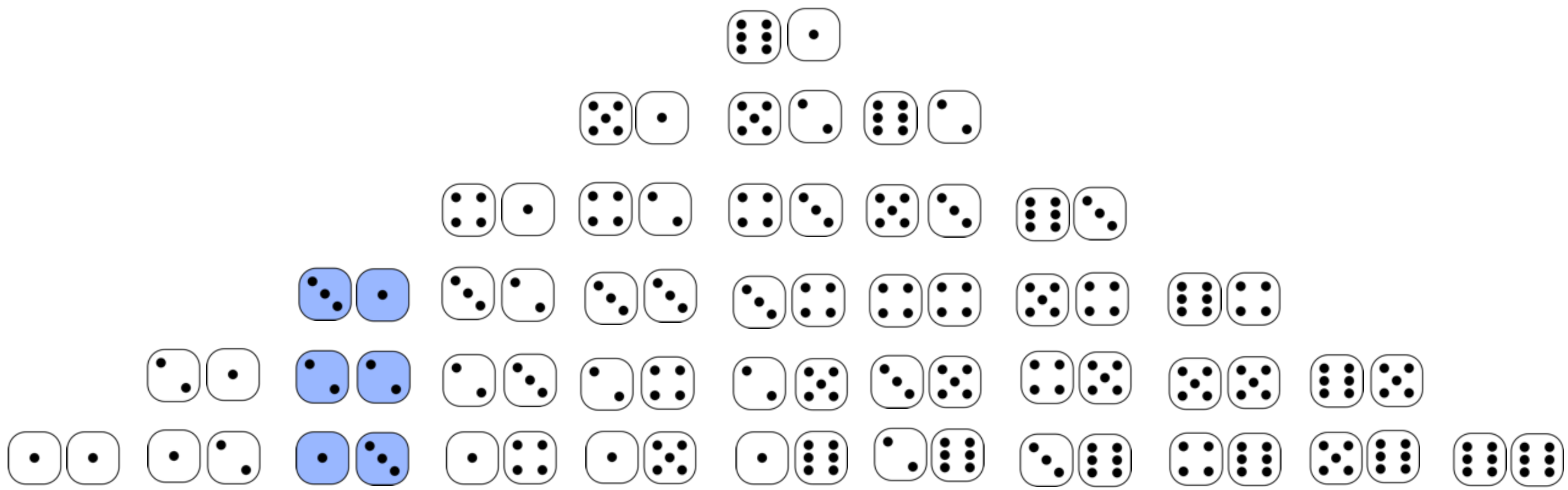


Three cases where
the dice add up to 4



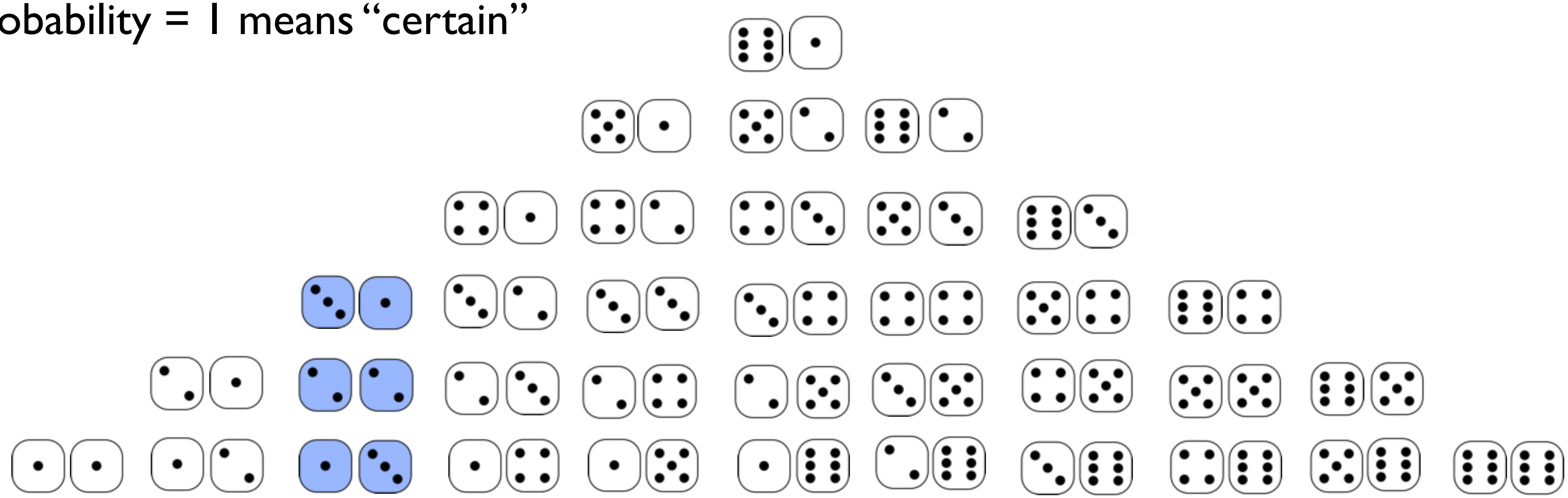
All 36 cases organised by outcome





Roll	2	3	4	5	6	7	8	9	10	11	12
N	1	2	3	4	5	6	5	4	3	2	1

- Probabilities are numbers between 0 and 1
- Probability = 0 means "impossible"
- Probability = 1 means "certain"

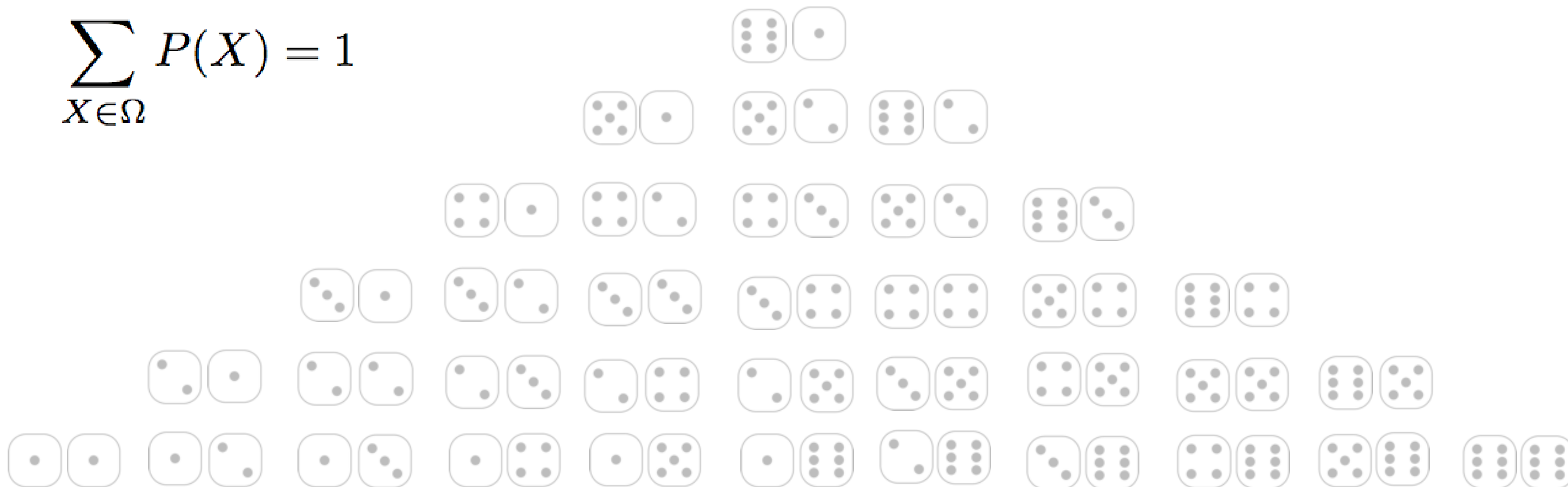


Roll	2	3	4	5	6	7	8	9	10	11	12
N	1	2	3	4	5	6	5	4	3	2	1
Prob	.028	.056	.083	.111	.139	.167	.139	.111	.083	.056	.028

Probability = 3/36 = .083

- Probabilities sum to 1
- "Law of total probability"
- "Conservation of belief"

$$\sum_{X \in \Omega} P(X) = 1$$

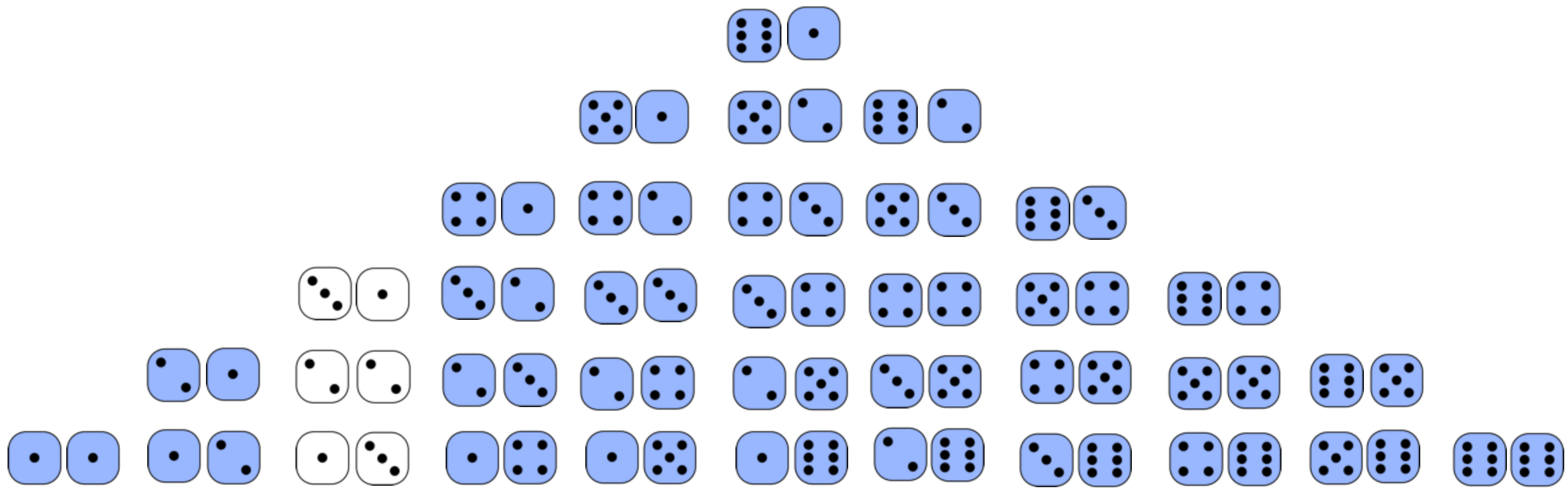


Roll	2	3	4	5	6	7	8	9	10	11	12
N	1	2	3	4	5	6	5	4	3	2	1
Prob	**.028**	**.056**	**.083**	**.111**	**.139**	**.167**	**.139**	**.111**	**.083**	**.056**	**.028**

.028 + .056 + .083 + .111 + .139 + .167
+ .139 + .111 + .083 + .056 + .028 = 1

$\neg A$ “not A” means “something other than A happens”

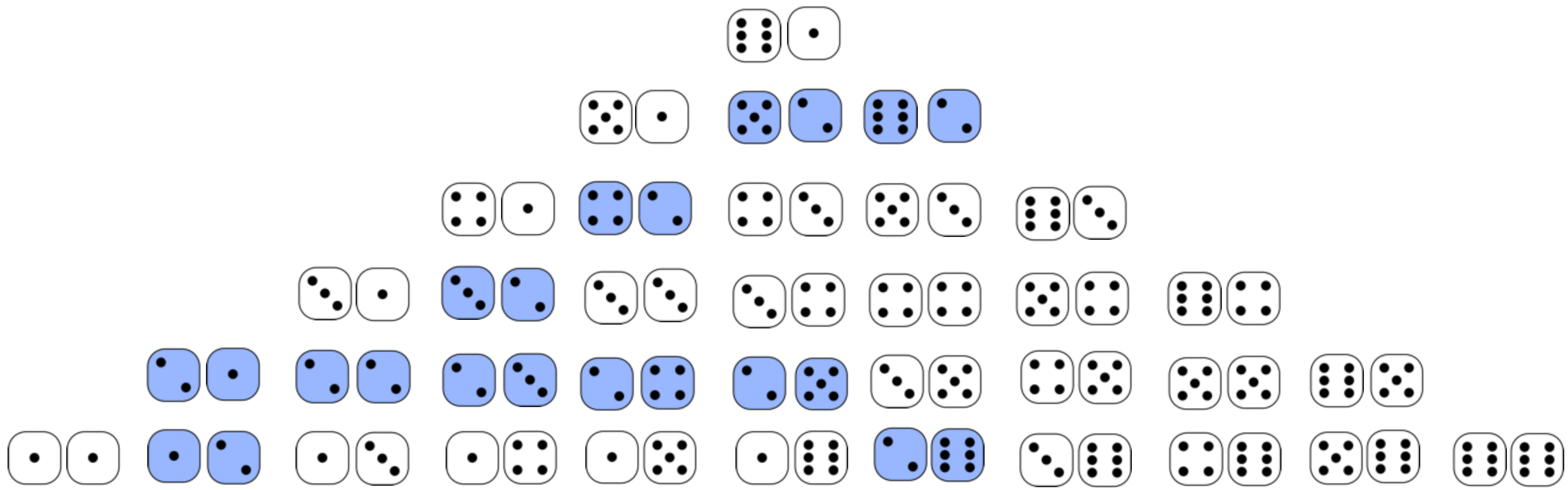
NOT



The probability of “not rolling a four” equals one minus the probability of rolling a four

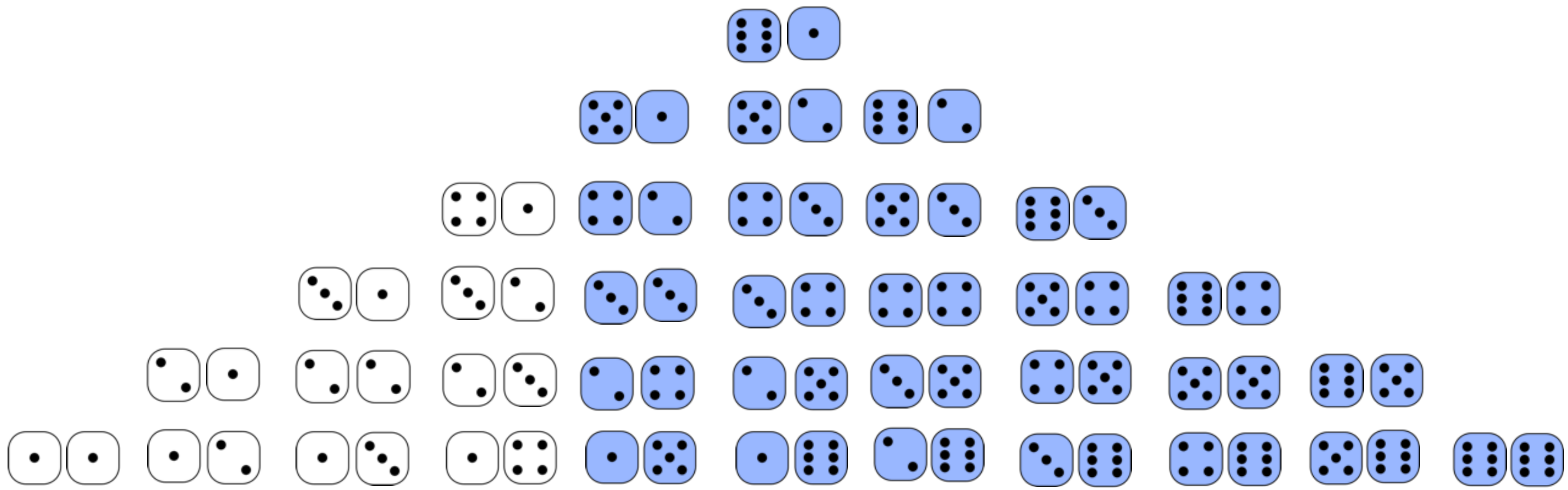
$$P(\neg A) = 1 - P(A)$$

A = “at least one die has a value of 2”



$$P(A) = \frac{11}{36} = .31$$

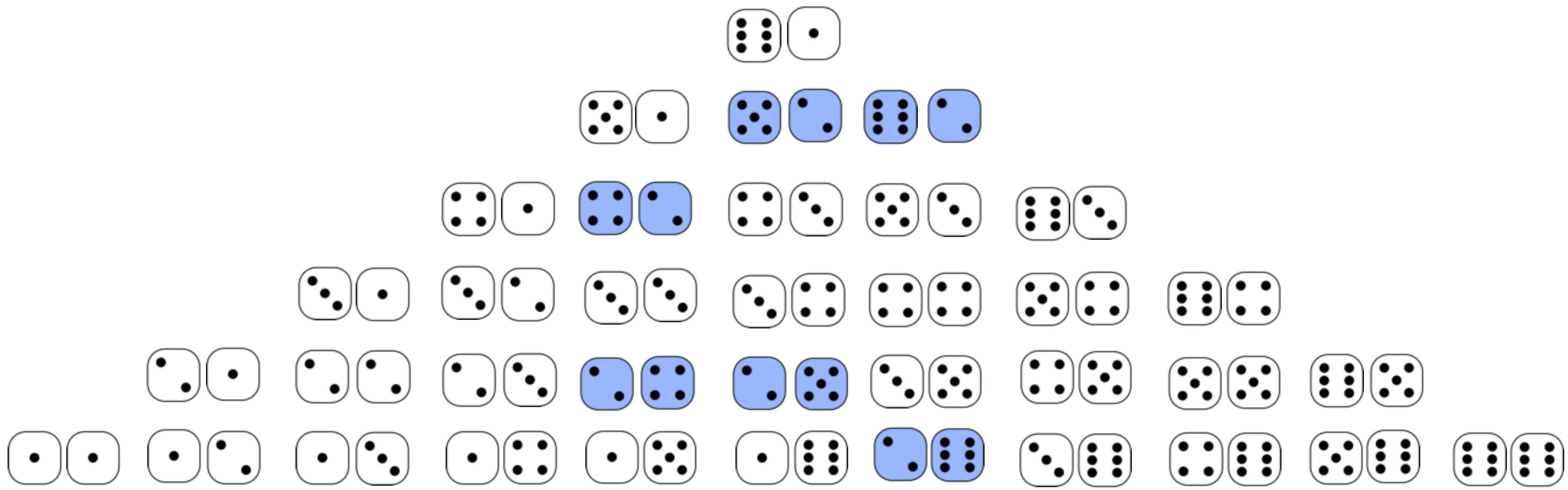
B = “the total is at least six”



$$P(B) = \frac{26}{36} = .72$$

AND

A = “at least one die has value 2”
B = “the total is at least six”



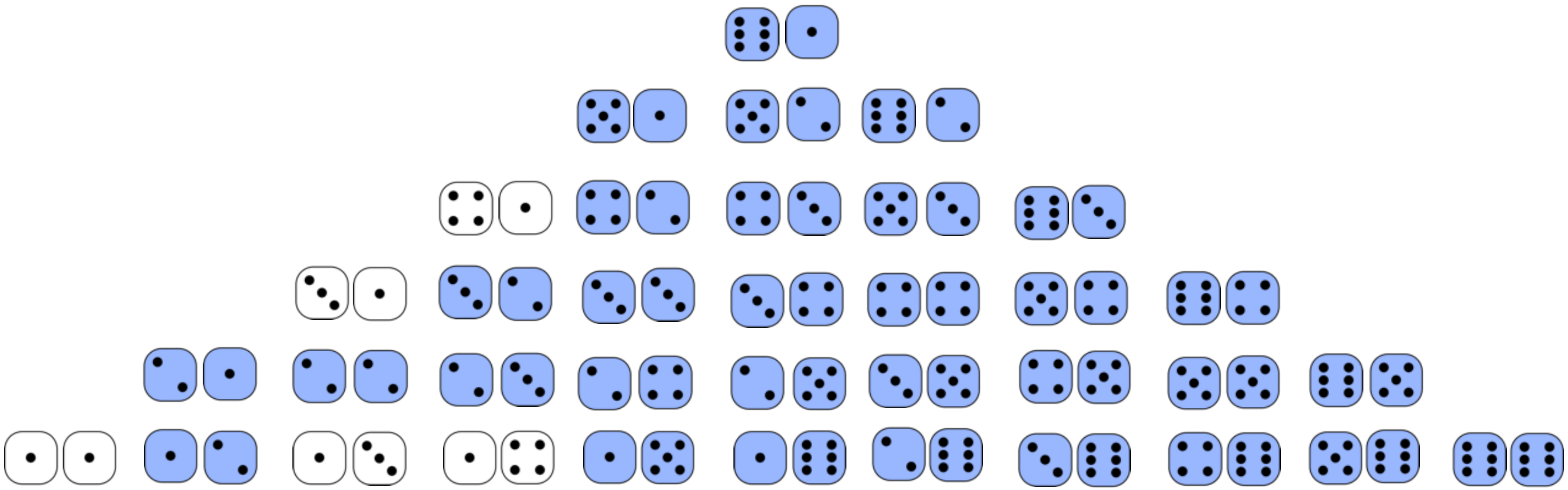
$$P(A, B) = \frac{6}{36} = .17$$

The *joint probability* that “A and B” both occur

$$P(A \cap B)$$

A = “at least one die has value 2”
B = “the total is at least six”

OR

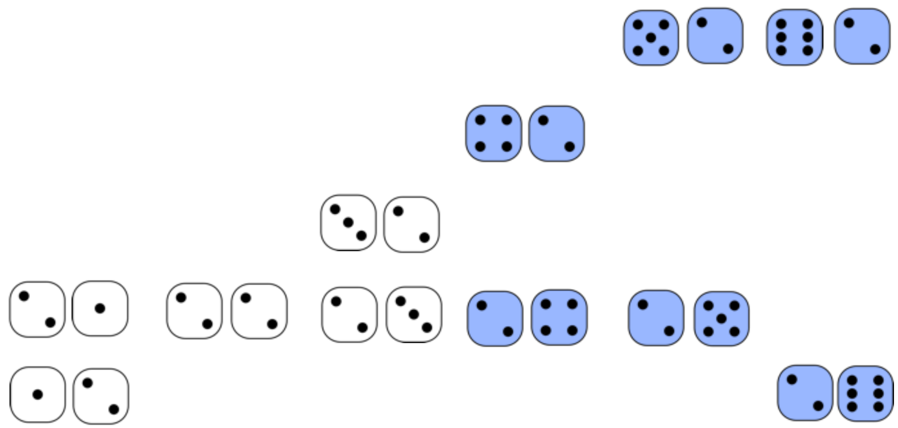


$$P(A \cup B) = \frac{31}{36} = .86$$

The probability that **at least** one of A or B occurs

A = “at least one die has value 2”
B = “the total is at least six”

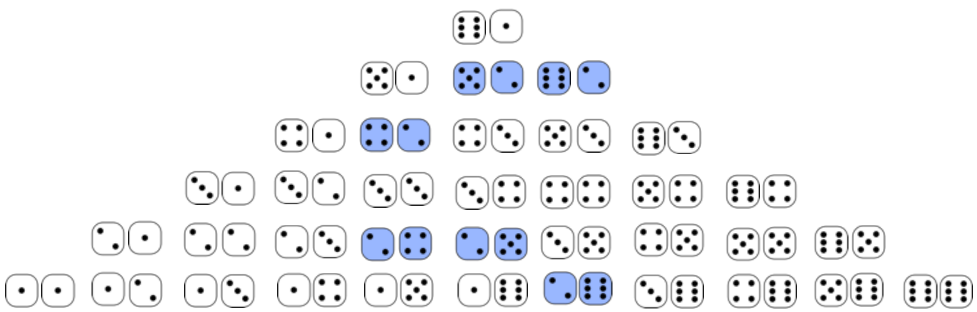
IF



$$P(B|A) = \frac{6}{11} = .55$$

The *conditional probability* that B occurs **given** that A occurs

AND **IF**

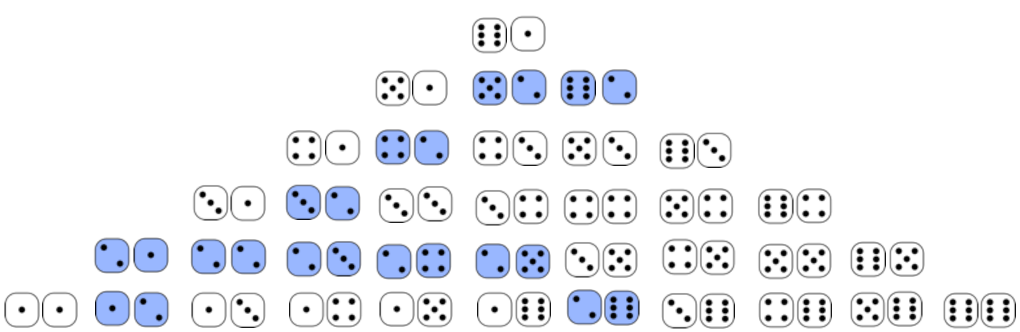


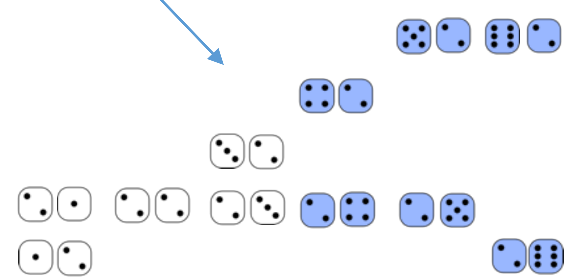
6/36

$$P(A, B) = P(A) \times P(B|A)$$

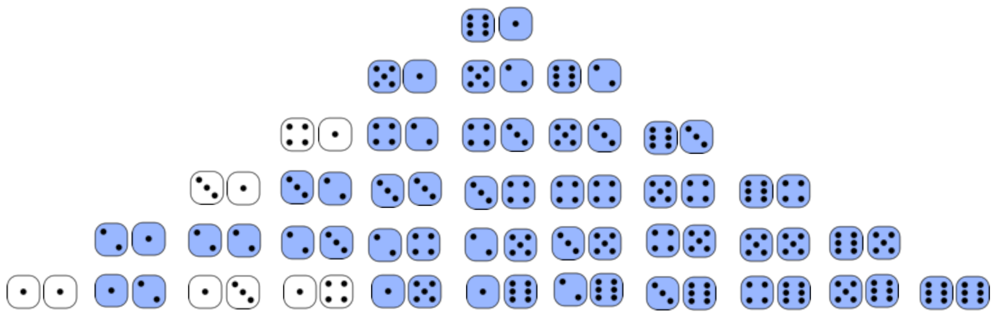
11/36

6/11

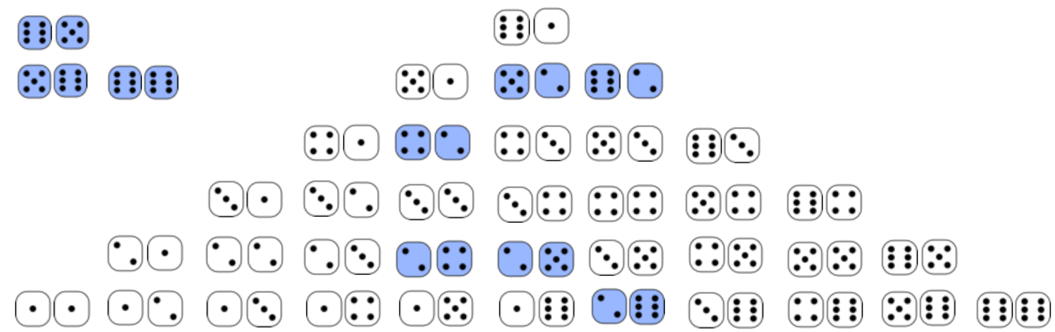




OR **AND**



31/36

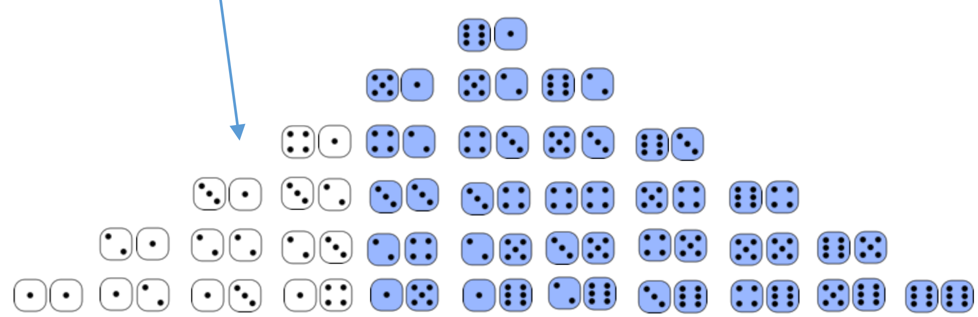


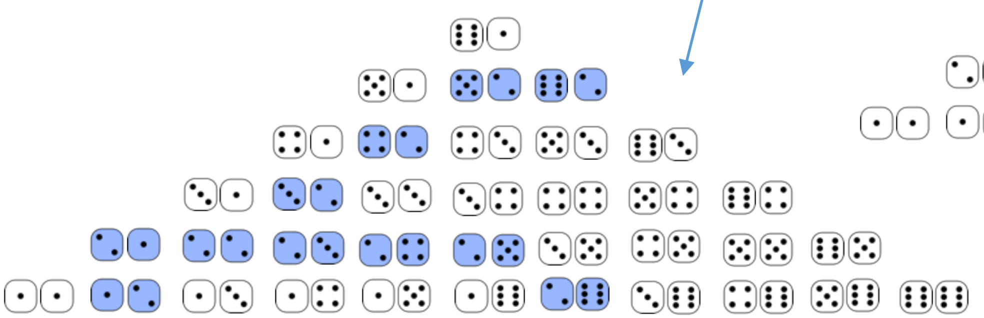
6/36

$$P(A \cup B) = P(A) + P(B) - P(B \cap A)$$

11/36

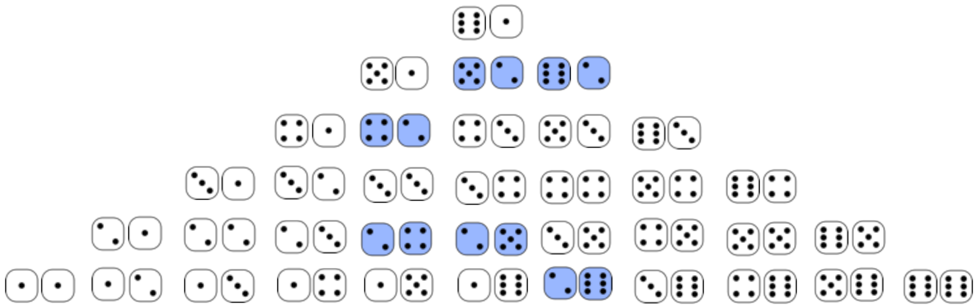
26/36





Discovering Bayes' rule

Exchangeability – when the order doesn't matter

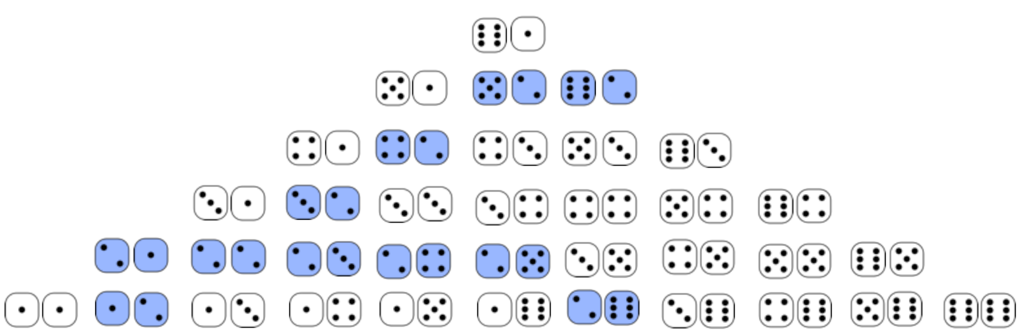


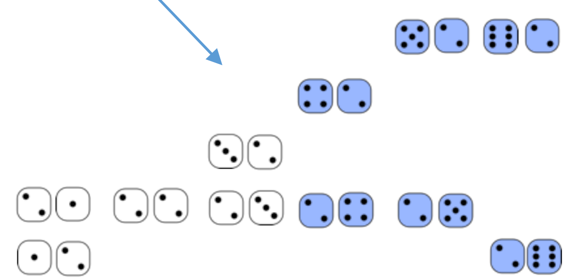
6/36

$$P(A, B) = P(A) \times P(B|A)$$

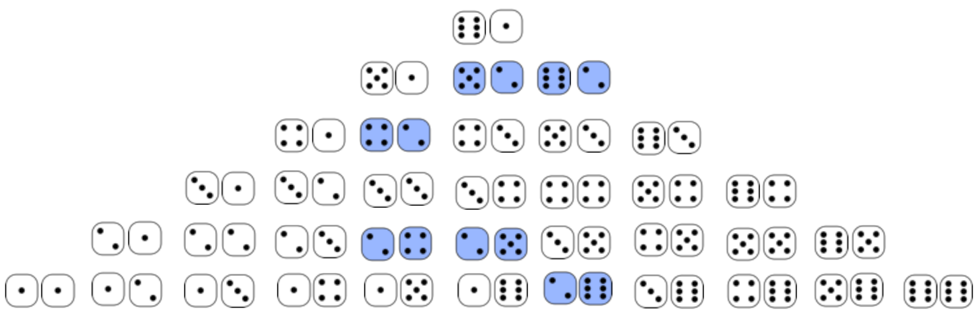
11/36

6/11





Exchangeability – when the order doesn't matter

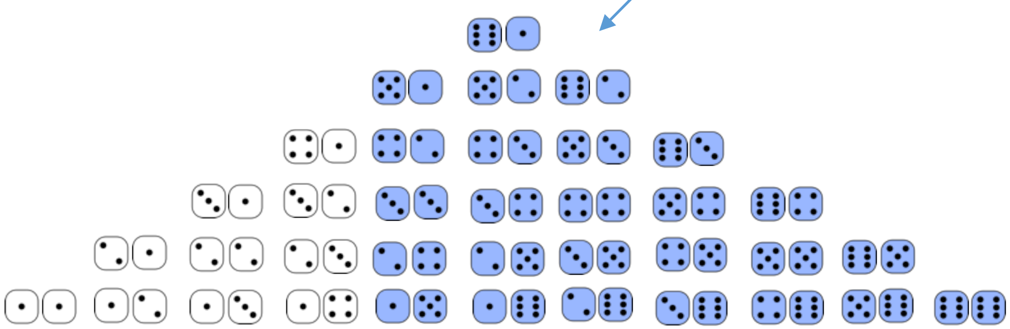


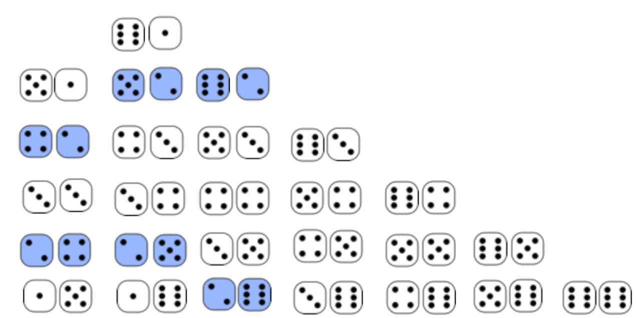
6/36

$$P(A,B) = P(B) \times P(A|B)$$

26/36

6/26





So if... $P(A, B) = P(A) \times P(B|A)$

… and … $P(A, B) = P(B) \times P(A|B)$

Then … $P(A) \times P(B|A) = P(B) \times P(A|B)$

So if... $$P(A,B) = P(A) \times P(B|A)$$

… and ... $$P(A,B) = P(B) \times P(A|B)$$

Then ... $$P(A) \times P(B|A) = P(B) \times P(A|B)$$

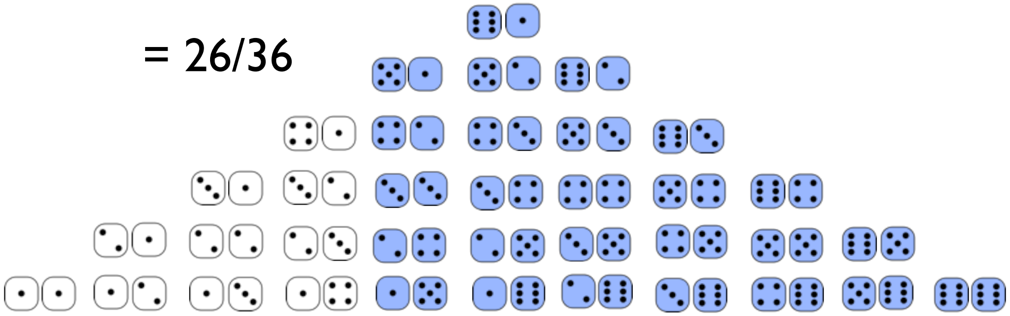
And if ... $$P(A) \times P(B|A) = P(B) \times P(A|B)$$

Then ... $$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

Bayes' rule

$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

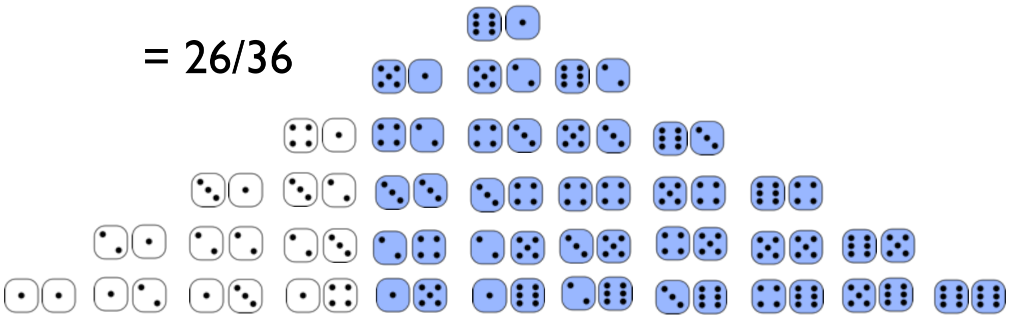
= 26/36



Probability that the total is at least 6

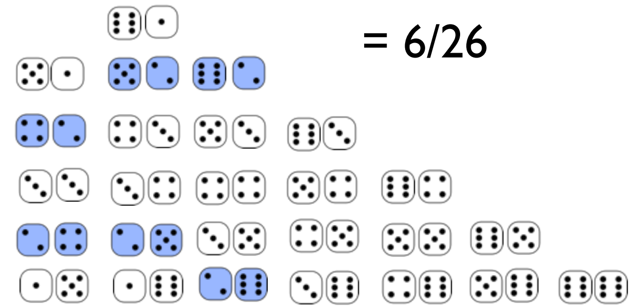
$$\frac{P(B) \times P(A|B)}{P(A)}$$

= 26/36



Probability that the total is at least 6

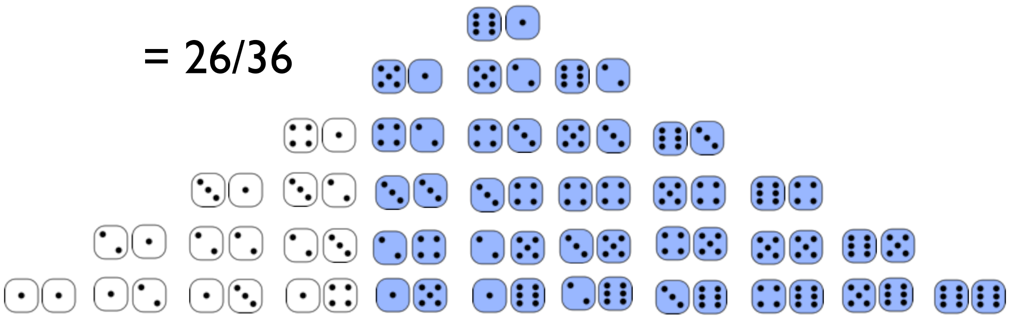
= 6/26



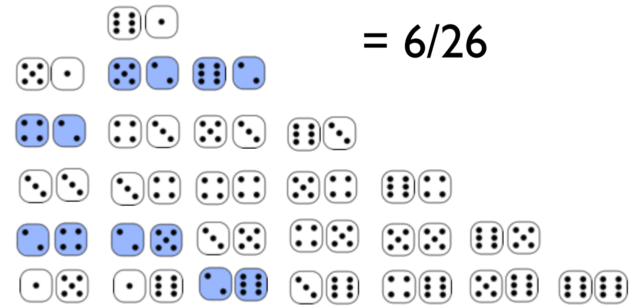
Probability that at least one die has a 2 given that the total is at least 6

$$\frac{P(B) \times P(A|B)}{P(A)}$$

= 26/36



Probability that the total is at least 6

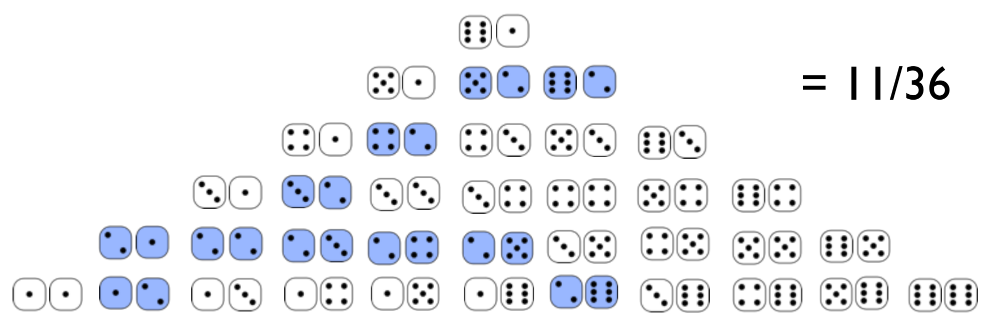


= 6/26

Probability that at least one die has a 2 given that the total is at least 6

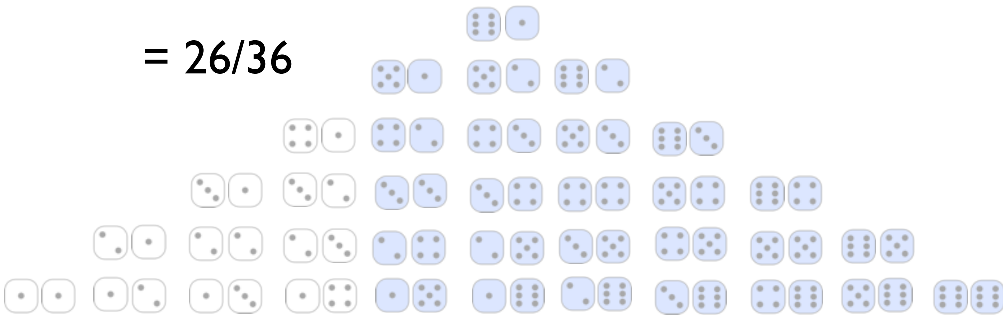
$$\frac{P(B) \times P(A|B)}{P(A)}$$

Probability that at least one die has a 2



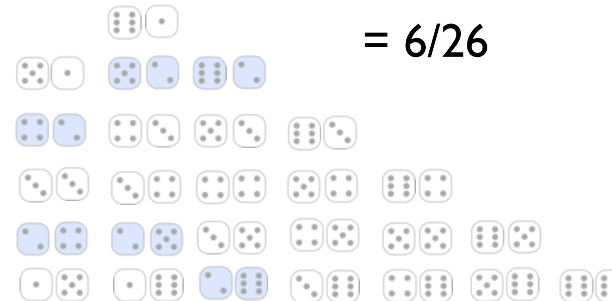
= 11/36

= 26/36



Probability that the total is at least 6

= 6/26

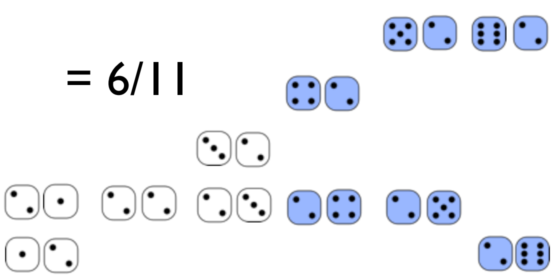


Probability that at least one die has a 2 given that the total is at least 6

$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

Probability that the total is at least 6 given that at least one die has a 2

= 6/11



Probability that at least one die has a 2

= 11/36

What is Bayes' rule used for?

Bayes' rule is a mathematical fact that probabilities must obey

26/36

6/26

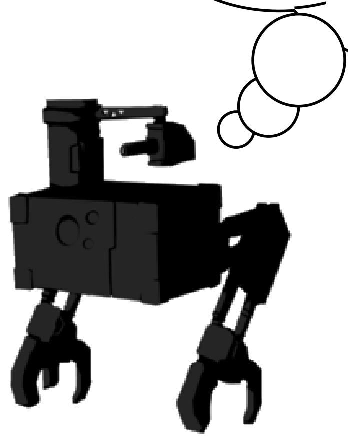
$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

6/11

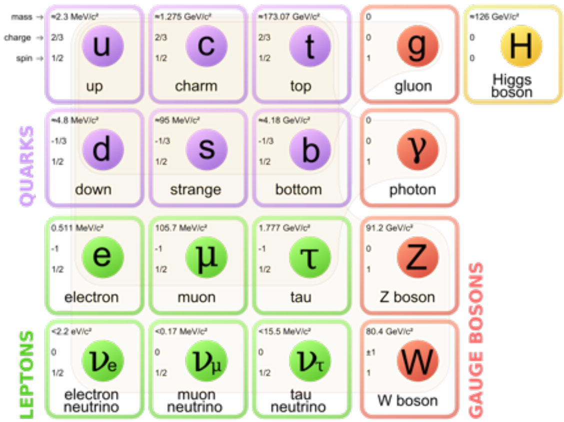
11/36

Bayesian reasoning happens when we combine this mathematical rule with epistemic probability

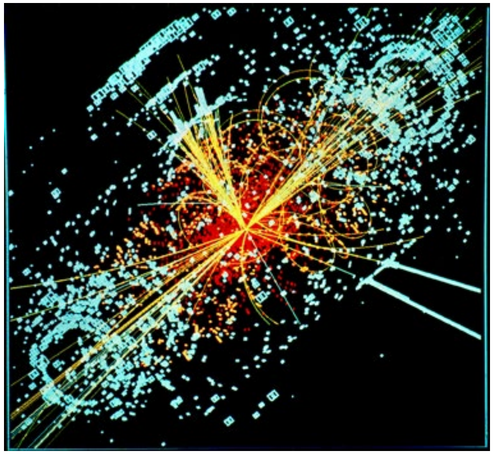
$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$



What do we use this for??



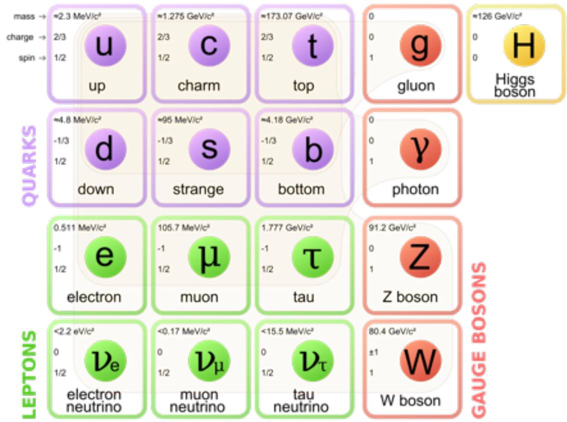
h = A hypothesis about the world

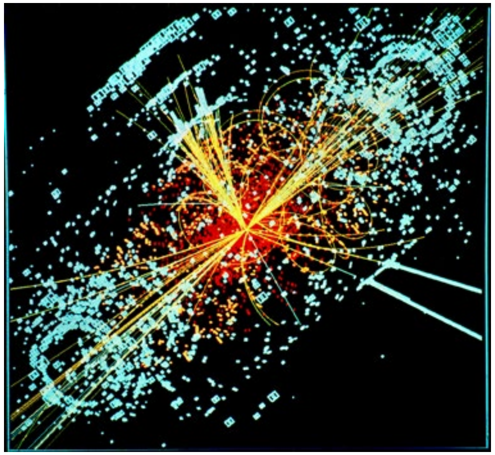


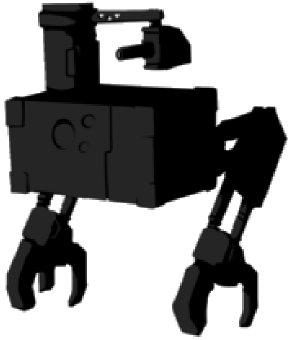
d = Some observable data

How strongly should I believe in this hypothesis…

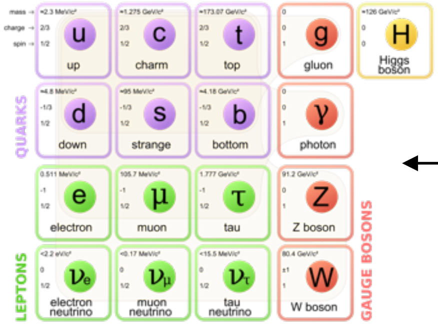
… given that I have observed these data?



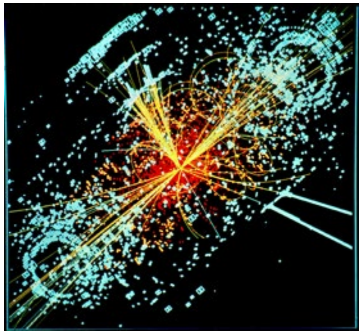




h|d



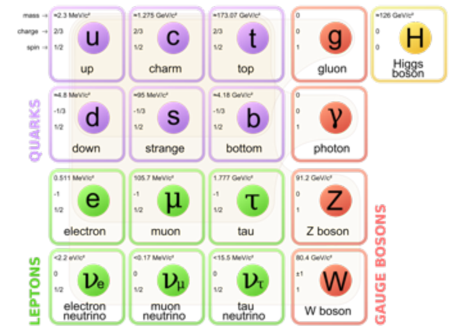
←



The **posterior probability** that my hypothesis is true given that I have observed these data…

$$P(h|d) = \frac{P(d|h) \times P(h)}{P(d)}$$

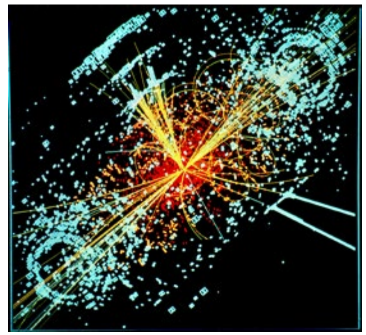
h

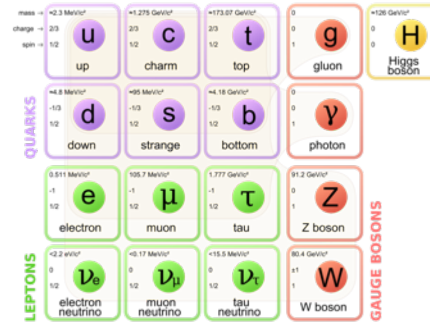


The **<u>prior probability</u>** that I assigned to this hypothesis before observing the data

$$P(h|d) = \frac{P(d|h) \times P(h)}{P(d)}$$

d|h



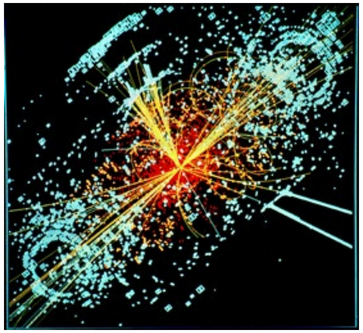


The **likelihood** that I would have observed these data if the hypothesis is true

$$P(h|d) = \frac{P(d|h) \times P(h)}{P(d)}$$

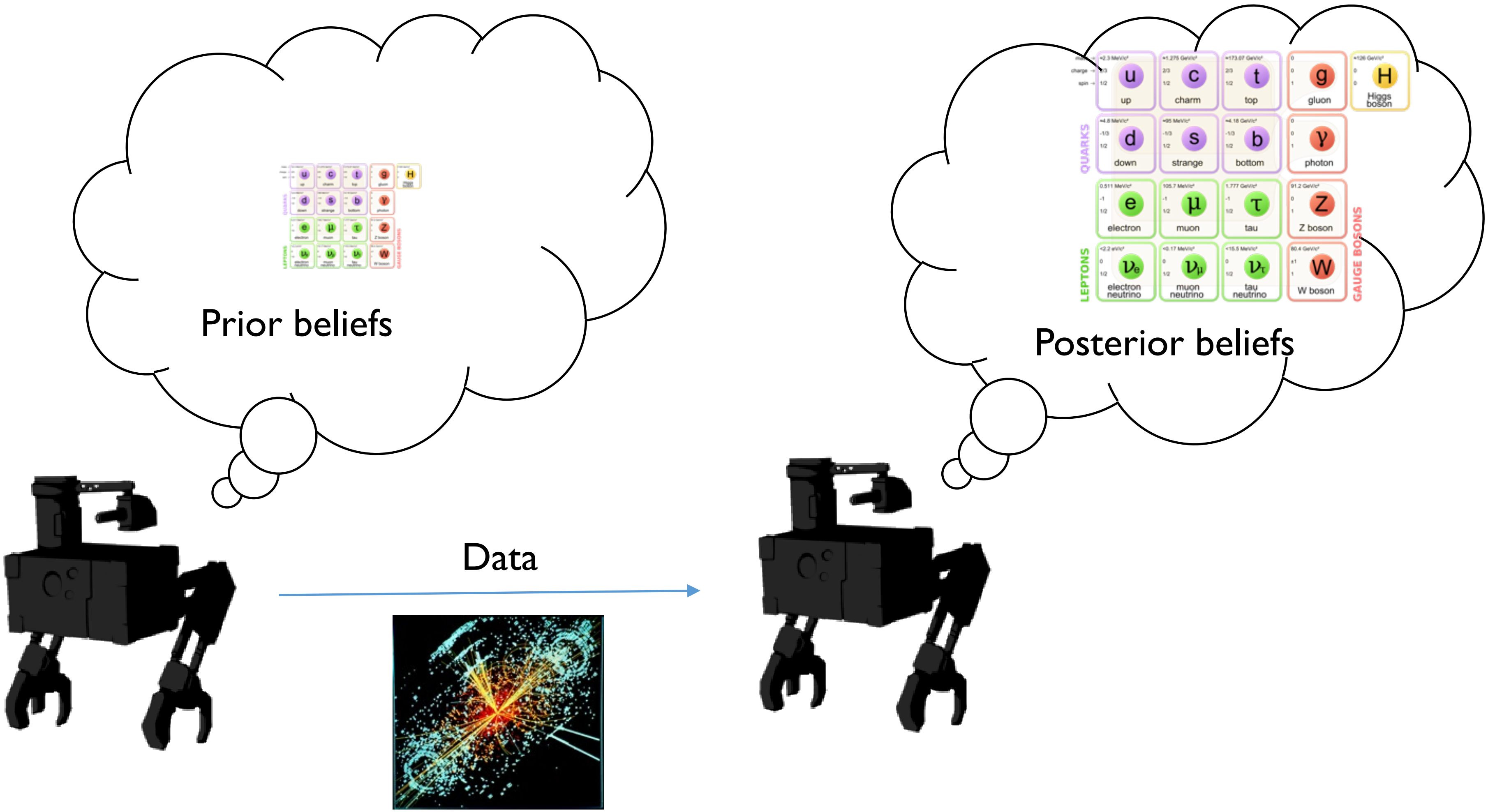
$$P(h|d) = \frac{P(d|h) \times P(h)}{P(d)}$$

d



The "marginal" probability of observing these particular data (more on this shortly)

Belief revision!



P(d|h) : the likelihood of observing d if h is true

P(h) : the prior probability that h is true

P(h|d) : the posterior probability that h is true

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

P(d) : discussed later

What happened here?
An example of Bayesian reasoning

There are many possibilities

dropped a wine glass

broke a window

psychic explosion

earthquake

a wizard did it

etc…

Let’s just think about two of them

I dropped a wine glass

Kids broke the window

"Prior odds"

$$\frac{P(h_1)}{P(h_2)} = \frac{\text{[image 1]}}{\text{[image 2]}} = 0.1$$

Before learning anything else I think "wine glass dropping" is 10 times more plausible than "broken window"

Some data

There is a cricket ball
next to the broken glass

Likelihood of the data

When I drop a wine glass…

… It's very unlikely that I just happen to do so right next to a cricket ball

$P(d|h) = 0.001$

Likelihood of the data

When the kids break a window…

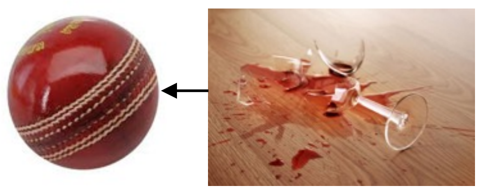
… It's not at all uncommon for a cricket ball to end up near the glass

$P(d|h) = 0.15$

“Likelihood ratio”

(a.k.a. Bayes factor)

$$\frac{P(d|h_1)}{P(d|h_2)} =$$



$$= \frac{0.15}{0.001} = 150$$

I think it is 150 times more likely that I would find a cricket ball when a window breaks than when a wine glass is broken

Posterior odds

$$\frac{P(h_1|d)}{P(h_2|d)} = \frac{P(d|h_1)}{P(d|h_2)} \times \frac{P(h_1)}{P(h_2)}$$

Posterior odds = 15

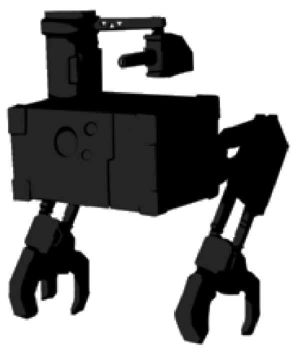
Likelihood ratio = 150

Prior odds = .1

In light of the evidence, I now think the window-breaking hypothesis is 15 times more likely than the wine-glass hypothesis

The learner has a hypothesis space

…



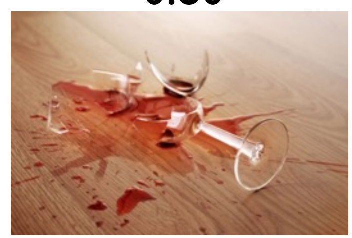
Priors assign probabilities P(h)

0.01

0.01

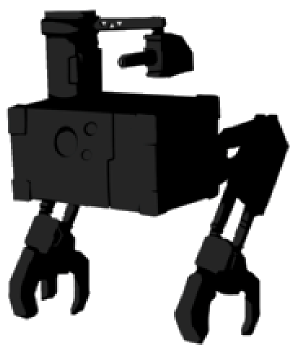
0.01

0.80

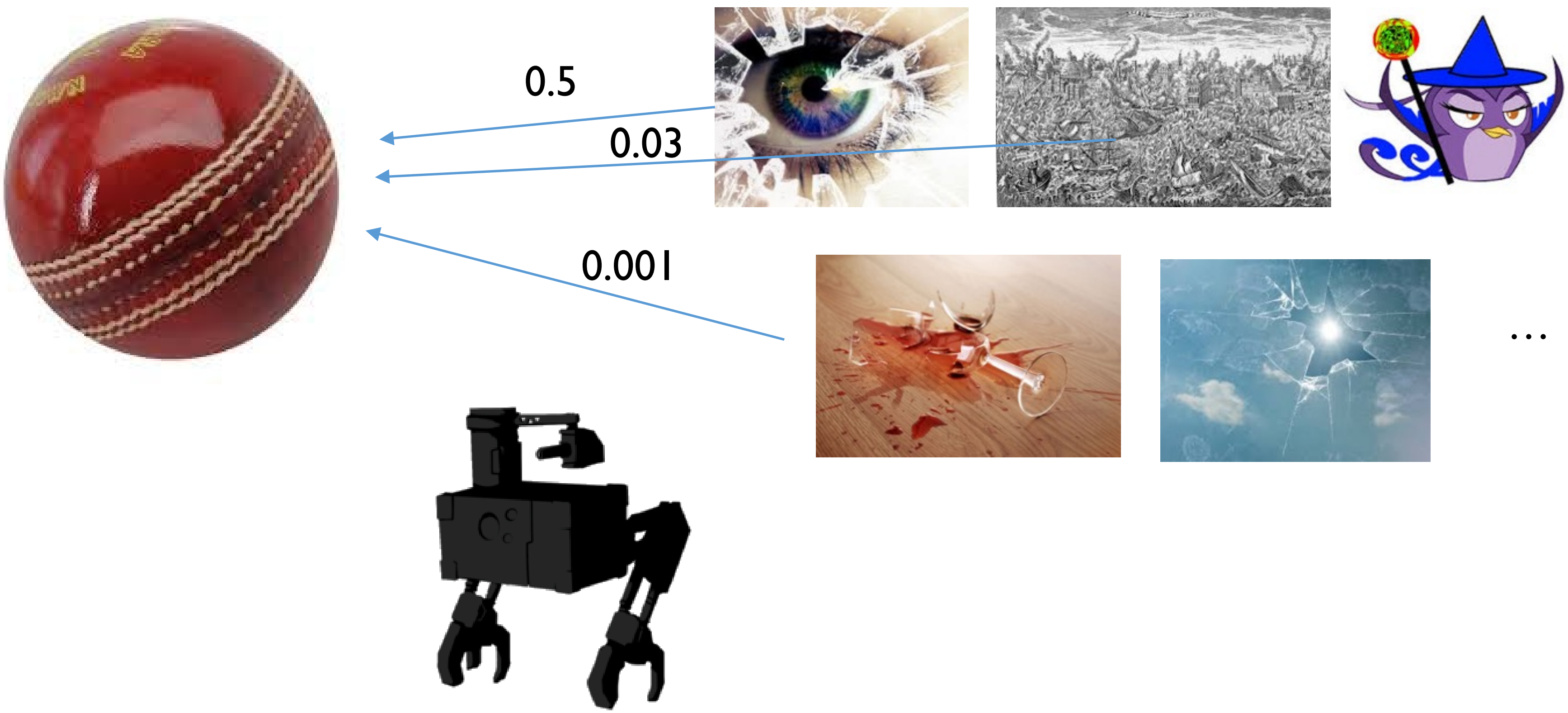


0.08

...



Each hypothesis provides a likelihood P(d|h)

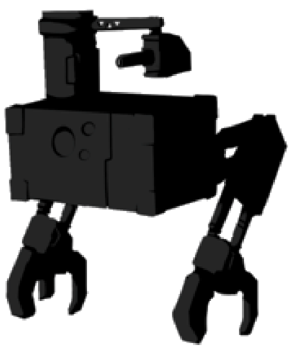


Posteriors computed by Bayes' rule



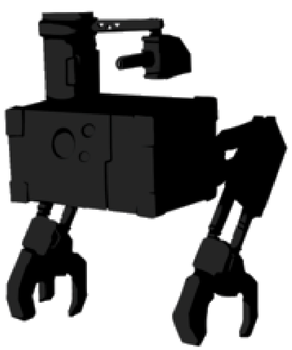






$$P(h|d) = \frac{P(d|h) \times P(h)}{P(d)}$$

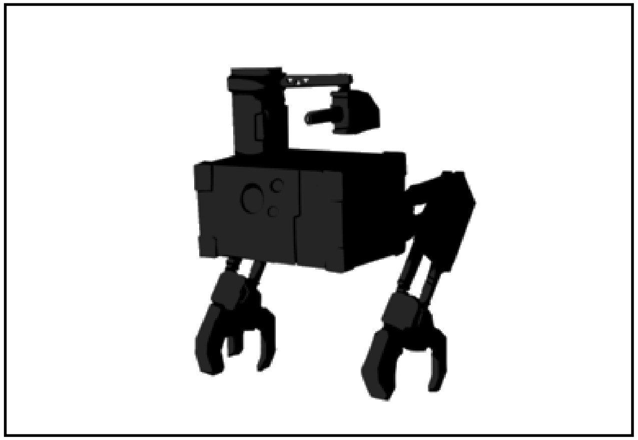
Posteriors computed by Bayes' rule



Sum taken across all possible hypotheses

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P(d|h')P(h')}$$

The question…



Bayesian reasoning is a powerful tool for building intelligent robots…

… but is it a useful tool for helping us to understand human cognition?