Advanced Psychophysics

John Greenwood
Department of Experimental Psychology

NEUR3001

Contact: john.greenwood@ucl.ac.uk

Remember

- Understanding the brain requires approaches with different scales of analysis
- Psychophysics characterises the relationship between physical (e.g. visual) stimuli & the psychological experience of them via behaviour
- Gives us a range of approaches to measure performance
- From this we can derive insights about the visual system
  - e.g. that discrimination thresholds tend to be a constant proportion of the reference intensity (Weber's Law)

Today

- More advanced things we can do with psychophysics
- Adaptive Procedures
- Signal Detection Theory
- Is there a sensory threshold?
  - Signal Detection Theory says no
  - Remains a useful concept regardless
    - Illustrated with a comparison between behavioural measurements and the responses of a single neuron

From behaviour to function

- We need more methods!

Stimulus
- Luminance patch
- Gabor
- etc.

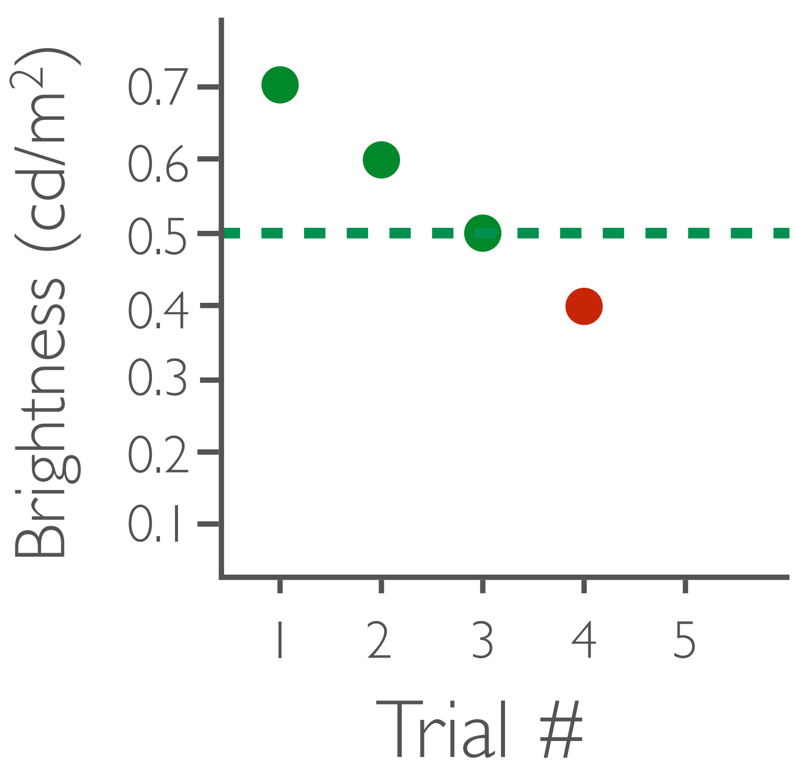
Task
- detection
  - yes/no
  - forced choice
- discrimination
  - yes/no
  - forced choice

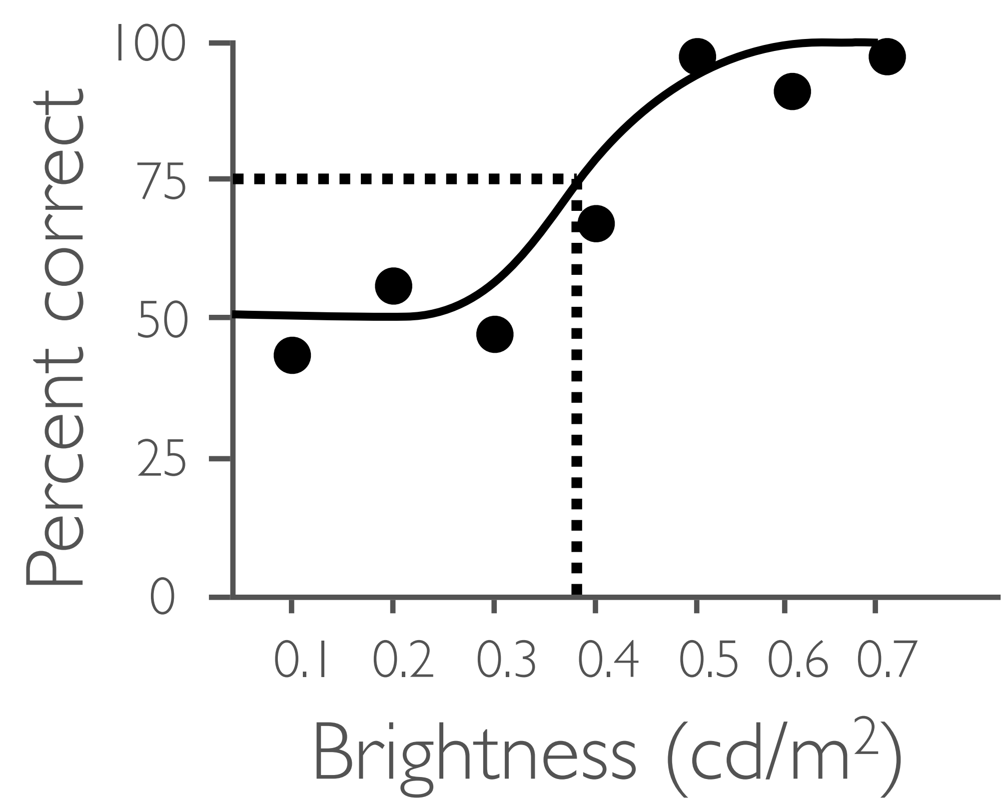
Method
- Limits
- Adjustment
- Constant stimuli
- Adaptive procedures
- SDT

Outcome
- performance
  - Reaction times
  - Percent correct
  - Thresholds
  - d’ and c
- appearance
  - PSEs
  - Matching
  - Scaling

Previously

- Method of limits gives rapid threshold estimates
  - But suffers from errors of habituation/expectation
- Method of Constant Stimuli avoids these errors
  - But it's slow and needs pre-defined intensity levels
- Is there another way?



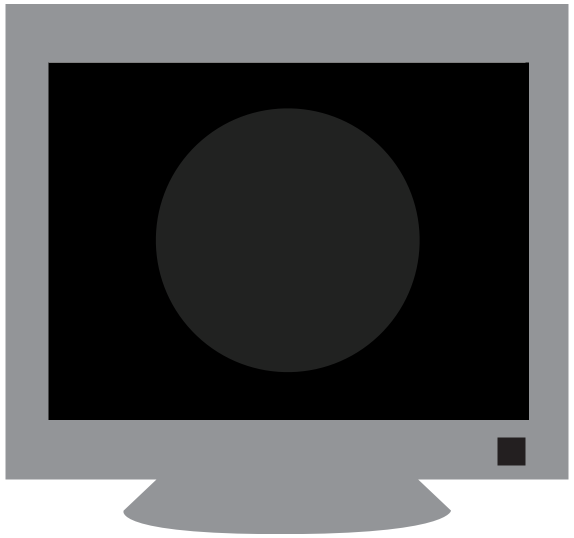


Adaptive Procedures

- Use our knowledge of perception to increase efficiency
- Select the stimulus intensity to present on each trial based on the responses to prior trials
- Tend to be quicker than classical methods as a result
- There are many approaches in this domain - we're going to look at two of them
  - Staircase procedures
  - QUEST

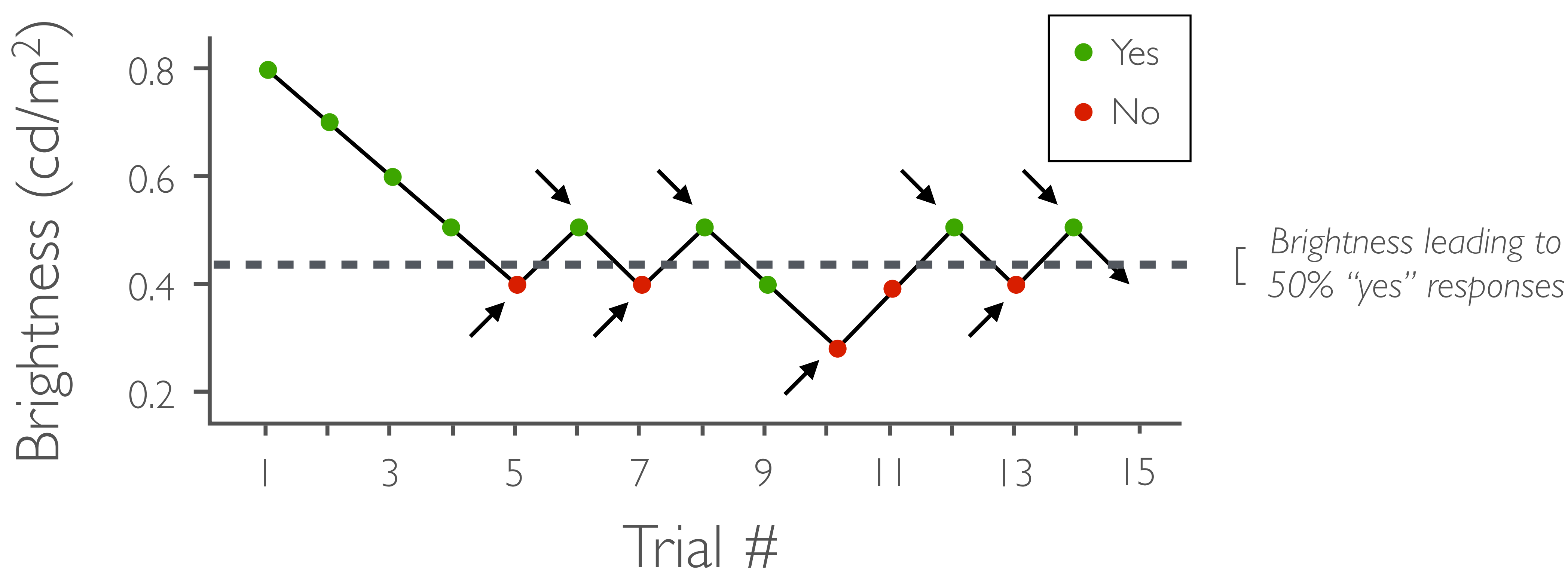
Staircase Procedures

- Derives from the Up/Down Method
  - Dixon & Mood (1948) sought to find the threshold height from which dropped weights would make gunpowder mixtures explode
  - Refined by Cornsweet (1962)
- Basic approach resembles the Method of Limits
- Results in a single intensity value where a desired performance level is reached (e.g. 50% ‘yes’ responses)
- Simplest example: detection threshold task for luminance patch with a yes/no procedure



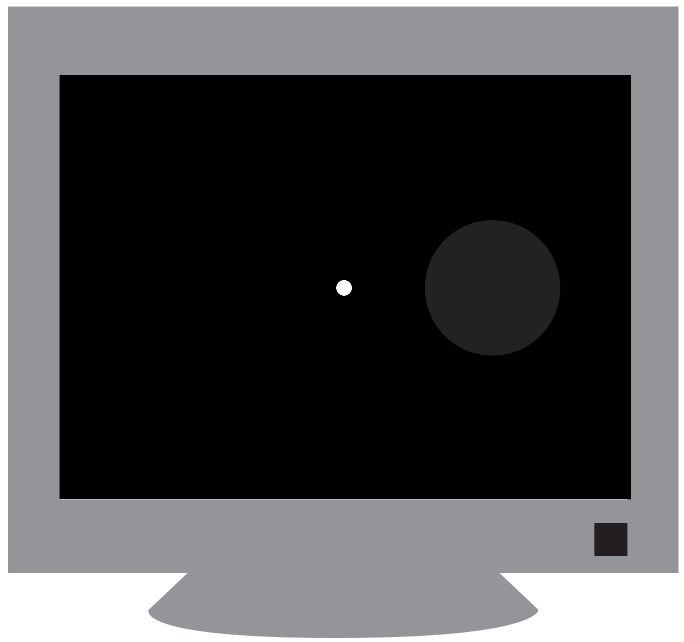
One down/one up staircase

- Start at (e.g.) high intensity and reduce intensity with each ‘yes’ until response changes
- Now raise intensity with each ‘no’ until it changes to ‘yes’, and so on
- One down/one up: one ‘yes’ gives a decrease, one ‘no’ an increase
- Converges on 50% - take the average of the reversal points for threshold



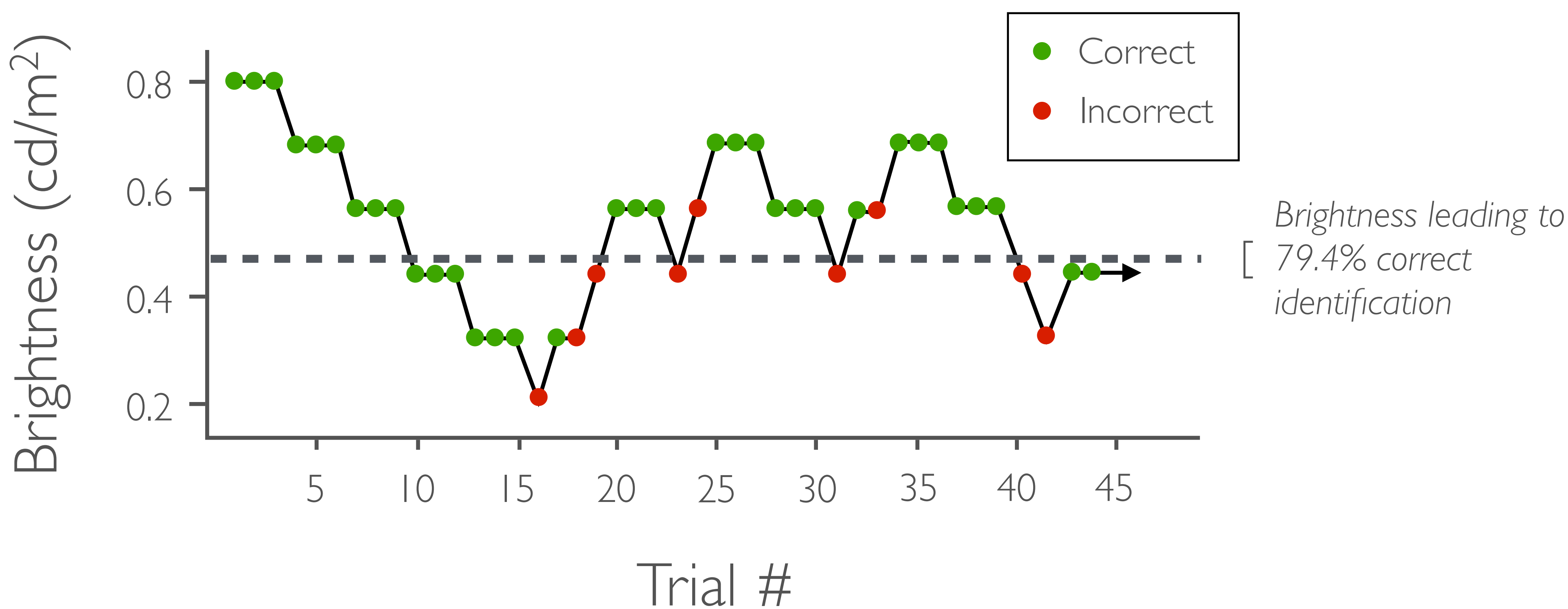
Transformed staircases

- But what if we want a different performance level?
  - e.g. the 75% correct threshold for a 2AFC task
- If the criterion to decrease intensity is more strict, then we can target a higher performance level
  - e.g. if two responses are required in a row to move down, performance will be above chance
- Convergence point is $\sqrt[n]{0.5}$ for **n** down/one up staircase
  - Two down/one up: 70.7%
  - Three down/one up: 79.4%



Three down/one up staircase

- Three correct responses *in a row* will decrease intensity
- A single incorrect response with increase intensity
- Converges on 79.4% correct: average the reversal points to estimate threshold (typically 8, often ignore first 2)

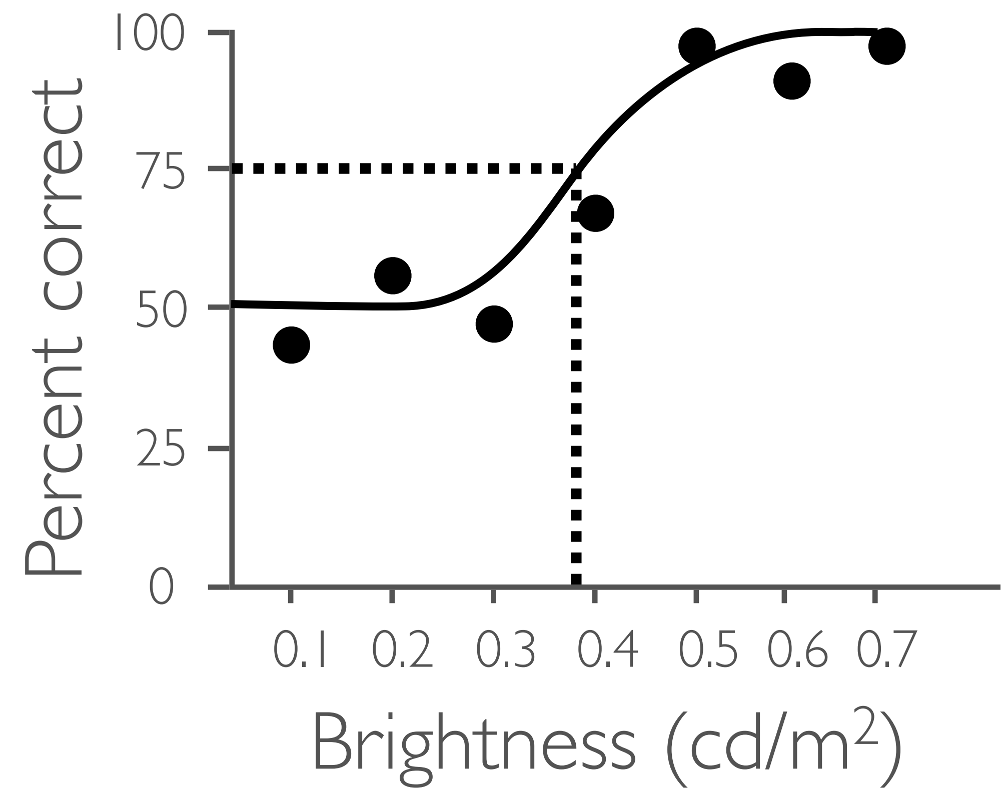


Staircases

- Quicker than the Method of Constant Stimuli
- But still requires quite a few trials to reach threshold, particularly with interleaved staircases
- Can we do better?

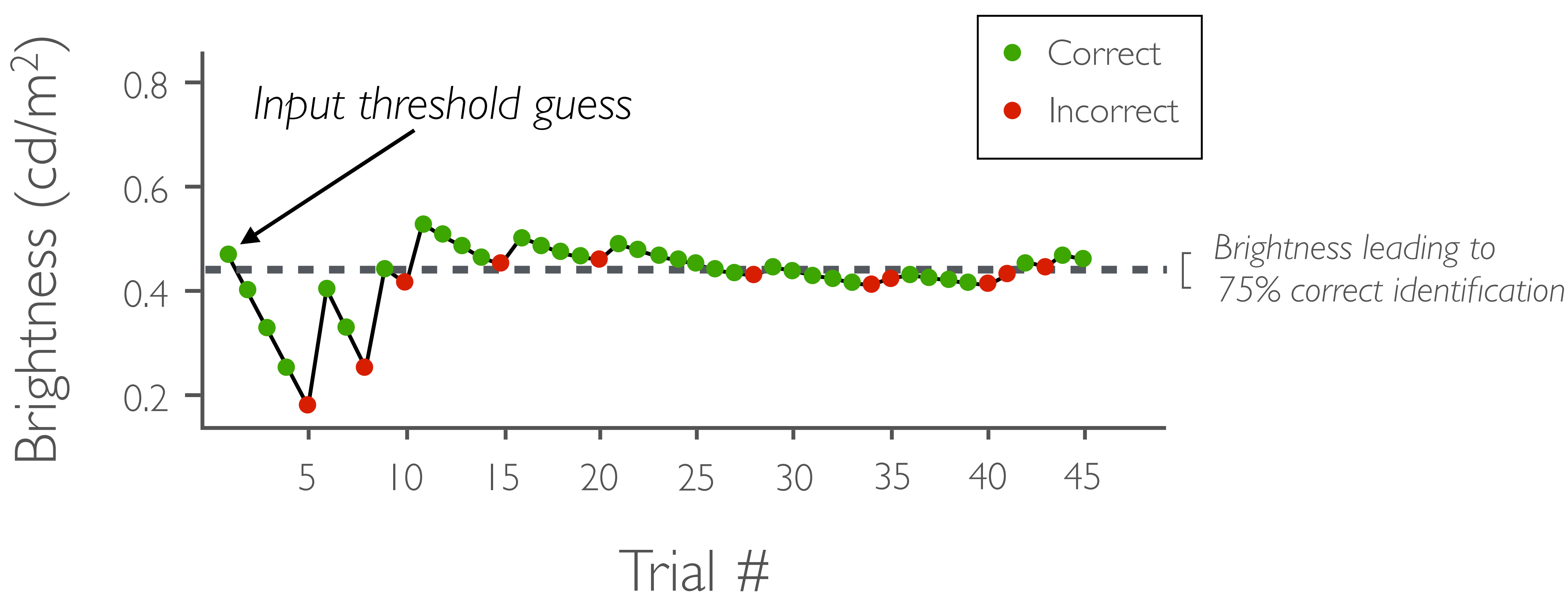
'Running Fit' Methods

- We know the pattern that responses usually take (a psychometric function)
- Can we use this to guide the choice of intensity on each trial?
- QUEST (Quick Estimation) takes this approach: places trials near the threshold of a psychometric function with parameters that depend on an initial guess plus observer responses (Watson & Pelli, 1983)



QUEST

- Input a guess threshold and its standard deviation (your certainty) to build an *a priori* distribution of potential threshold values
- Bayesian estimation used to select the most likely threshold estimate as the intensity for the next trial (based on both initial inputs and responses made throughout the experiment)
- Very quick to converge & give a final output of the most likely threshold



Adaptive procedures: issues

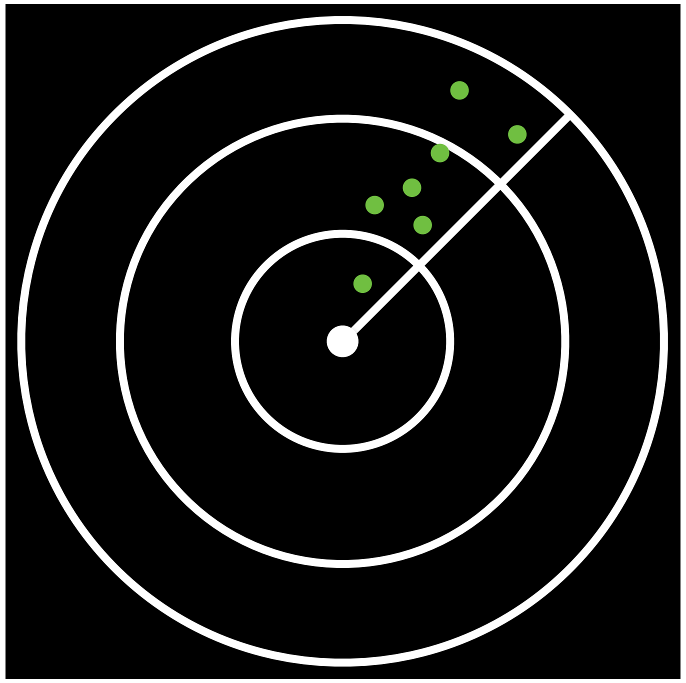
- Advantages
  - Very quick - threshold estimate in 40-50 trials or less
  - Avoids issues of the Method of Limits (esp. with interleaved staircases)
  - Can target different points on psychometric function (so we can measure *either* PSE or threshold)
- Disadvantages
  - Rapid drop to threshold can be difficult for observers, particularly clinical populations or children
  - Only returns a single performance level - not ideal if both threshold and PSE are of interest
  - Method of Constant Stimuli preferable in these circumstances

From behaviour to function

- **Stimulus**
  - Luminance patch
  - Gabor
  - etc.
- **Task**
  - detection
    - yes/no
    - forced choice
  - discrimination
    - yes/no
    - forced choice
- **Method**
  - Limits
  - Adjustment
  - Constant stimuli
  - Adaptive procedures
  - SDT
- **Outcome**
  - performance
    - Reaction times
    - Percent correct
    - Thresholds
    - d' and c
  - appearance
    - PSEs
    - Matching
    - Scaling

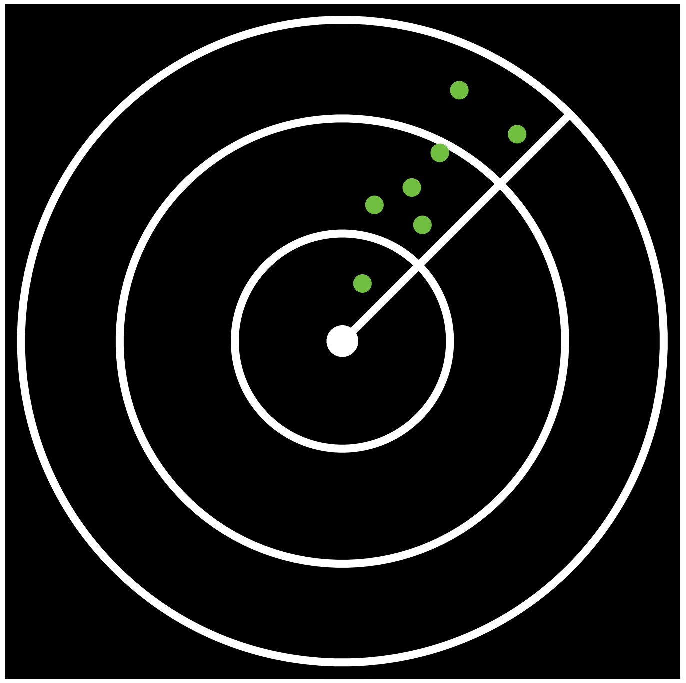
Signal Detection Theory

- Derives from radar operators during World War II
  - Radar antenna direction given by line
  - Dots trailing this visible only briefly and could arise from objects in environment, weather patterns, noise, or enemy aircraft
  - Upon seeing a dot: should you raise the alarm or not?



Signal Detection Theory

		Signal: Is it actually an enemy plane?	
		Yes	**No**
Decision: Is there an enemy plane?	**Yes**	Hit	False Alarm
	No	Miss	Correct Reject



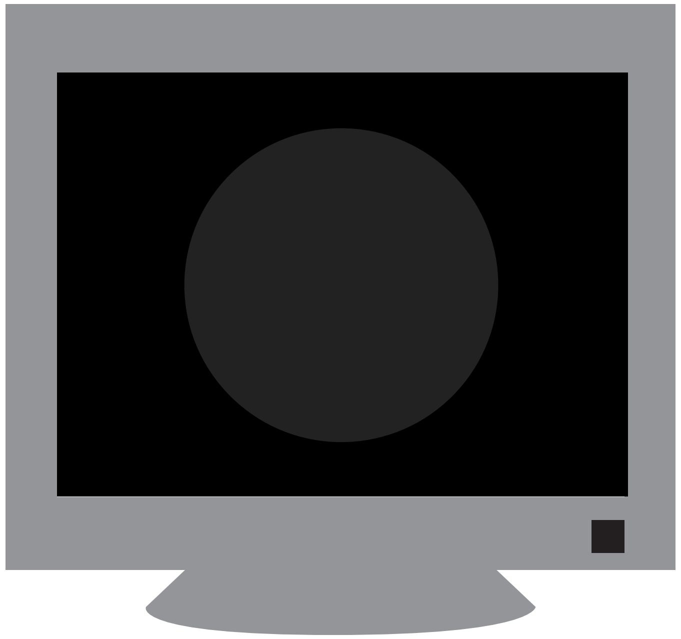
- Consequences:
  - **Hit:** Enemy are engaged and turned away
  - **Miss:** Enemy attack their target unscathed
  - **False alarm:** Aircraft take off for nothing, fuel wasted, pilots fatigued
  - **Correct rejection:** Crew able to rest and fuel is not wasted

SDT for brightness

Signal:
Is there a luminance patch?

Decision:
Is there a luminance patch?

	Yes	No
Yes	Hit	False Alarm
No	Miss	Correct Reject



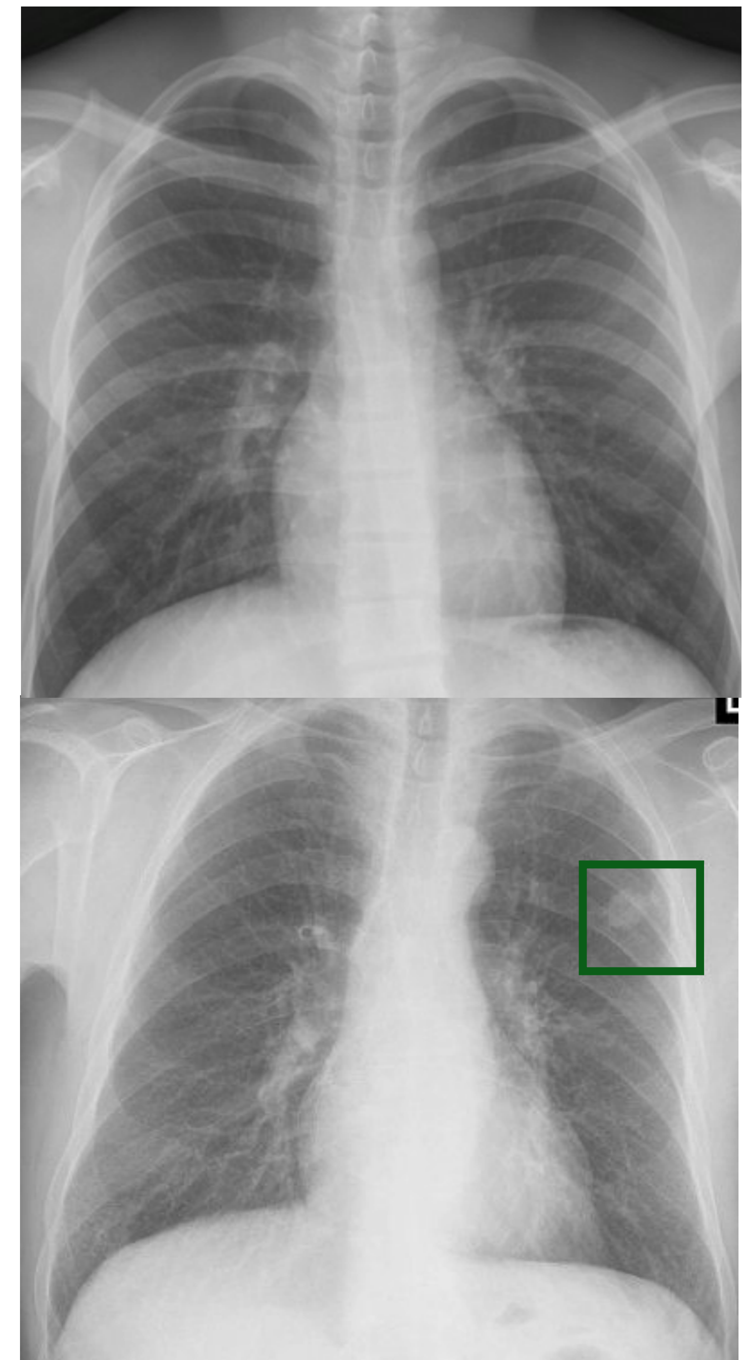
- Formalised for psychophysics by Green & Swets (1966)
- Easy to transpose this situation into a yes/no decision task, e.g. with our luminance patch
- Here we need two types of trials: signal present or absent
  - Decisions in each case: yes/no for each type of trial

SDT and X-ray diagnosis

Signal:
Is there a tumour?

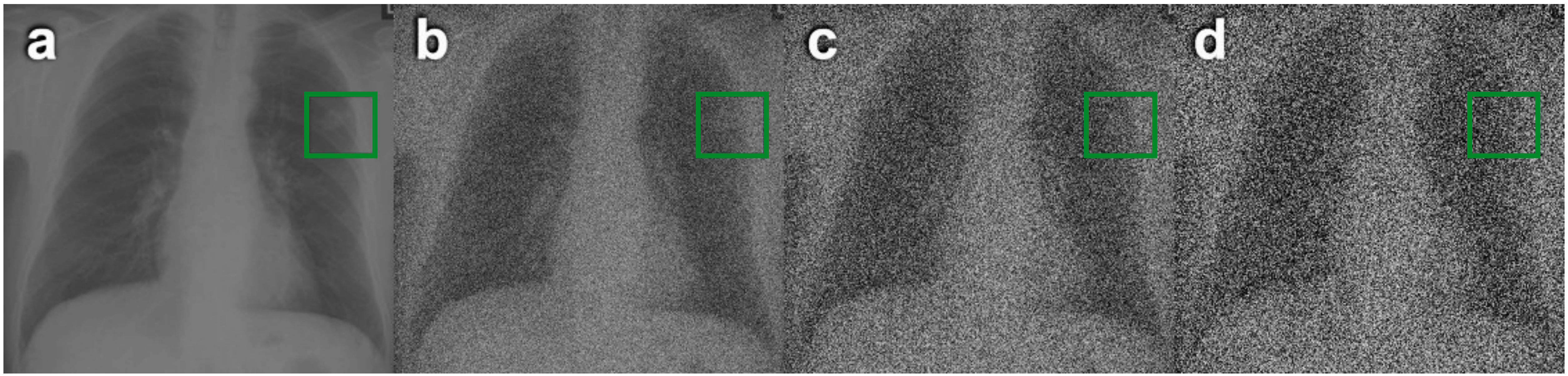
Decision:
Is there a tumour?

	Yes	No
Yes	Hit	False Alarm
No	Miss	Correct Reject
	signal+noise	*noise*



- Radiologists examine chest x-rays and asked "is a tumour present or absent?" (Kundel & Nodine, 1975)
- What limits performance and how can we characterise this?

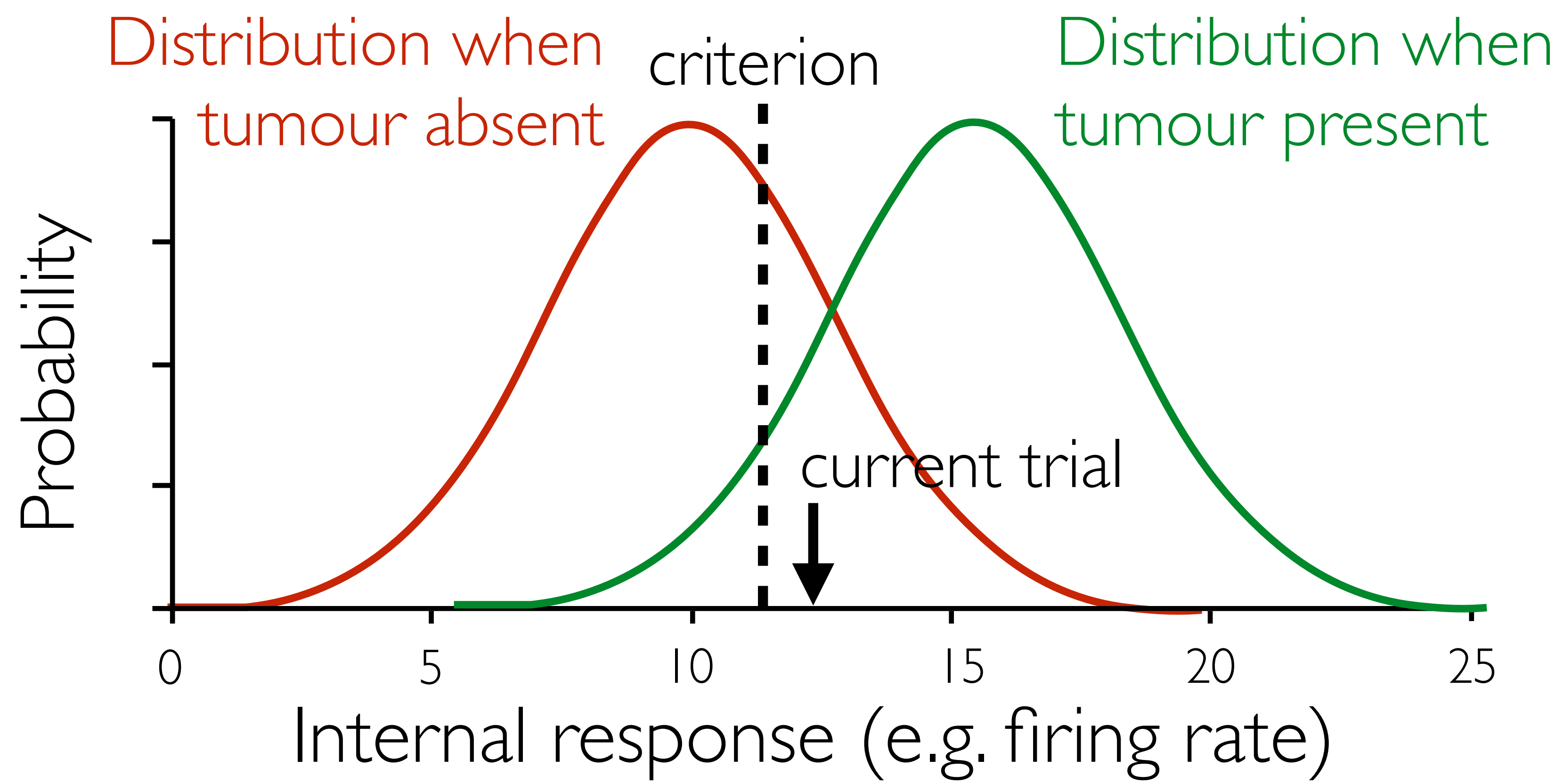
Noise



Increasing external noise →

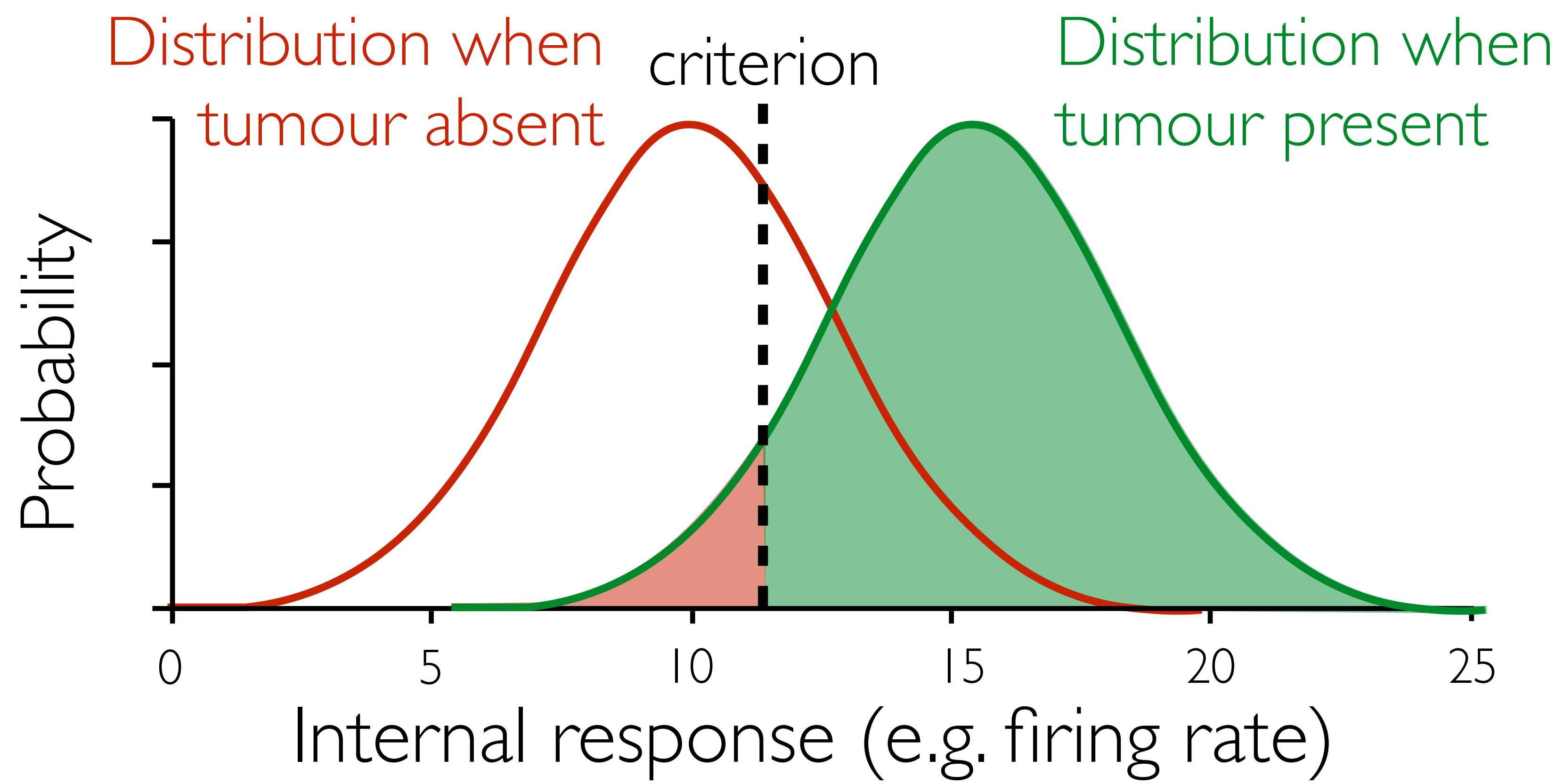
- Uncertainty on these tasks arises from two types of noise
- **External noise:** e.g. imaging errors, variation in lung tissue
- **Internal noise:** radiologist uses some neural response to detect a tumour - these responses are variable

Internal distributions



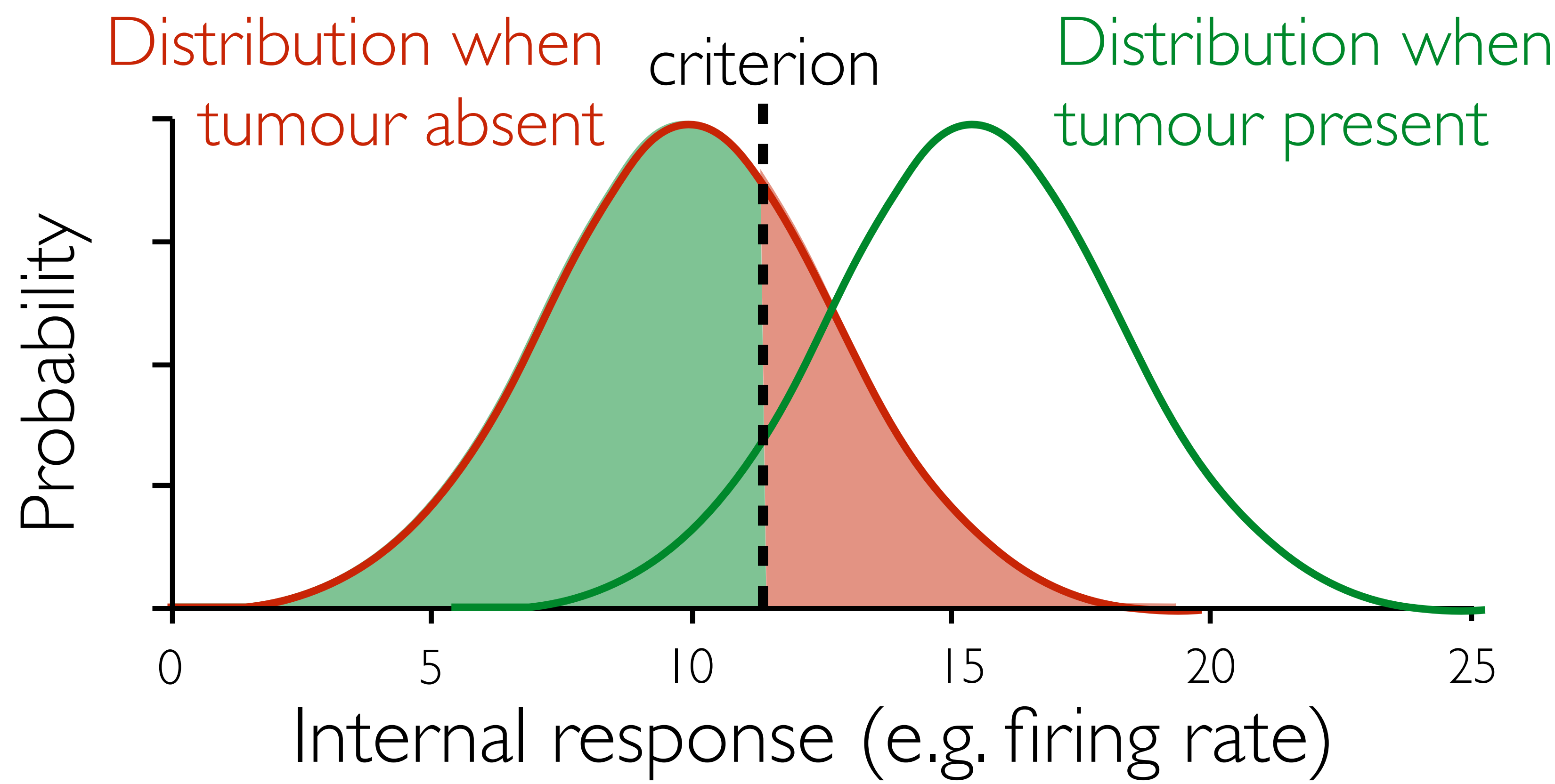
- Compare internal response probability of occurrence curves for noise alone vs. signal+noise trials
- Discriminability of the two possibilities set by separation/breadth of curves
- But decision also requires that we set a criterion value

Distributions to responses



- Signal present trials:
  - Response above the criterion = hit
  - Response below the criterion = miss

Distributions to responses

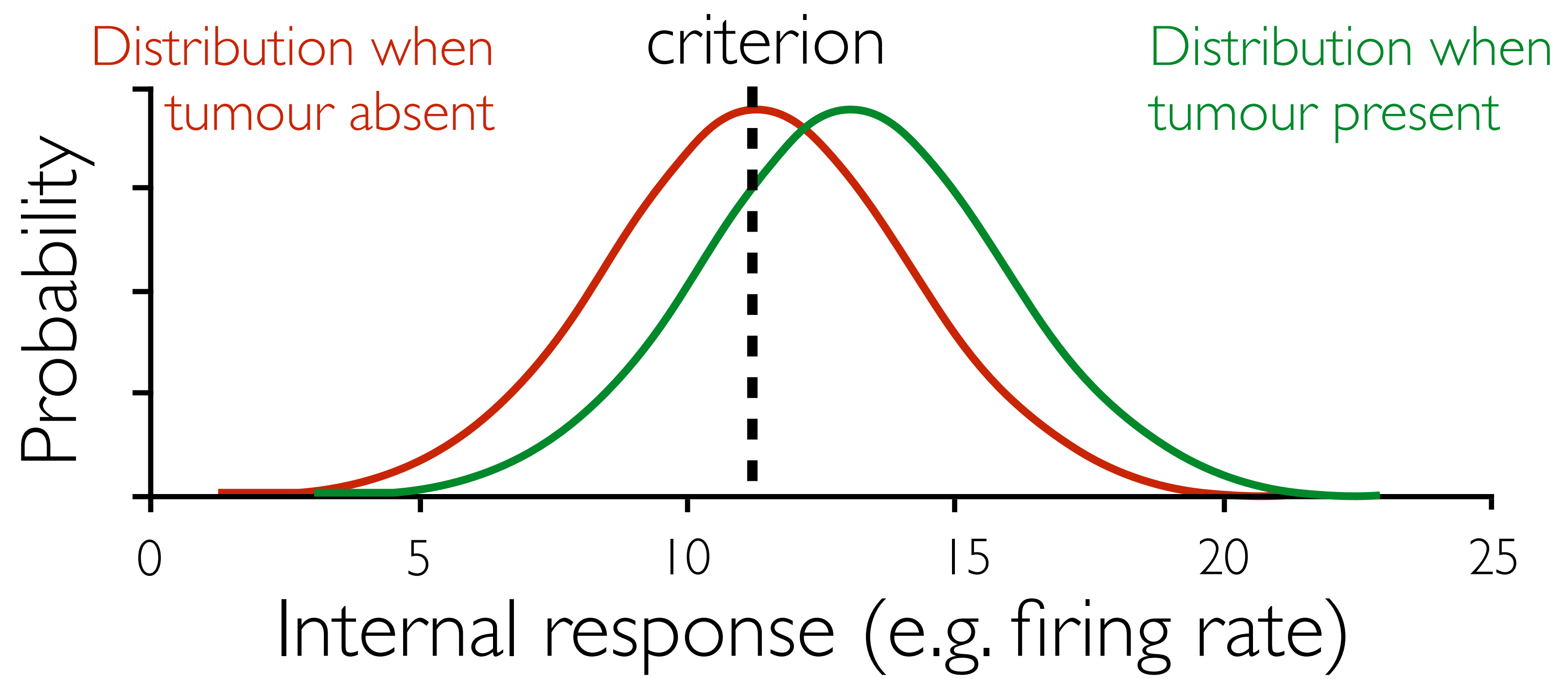


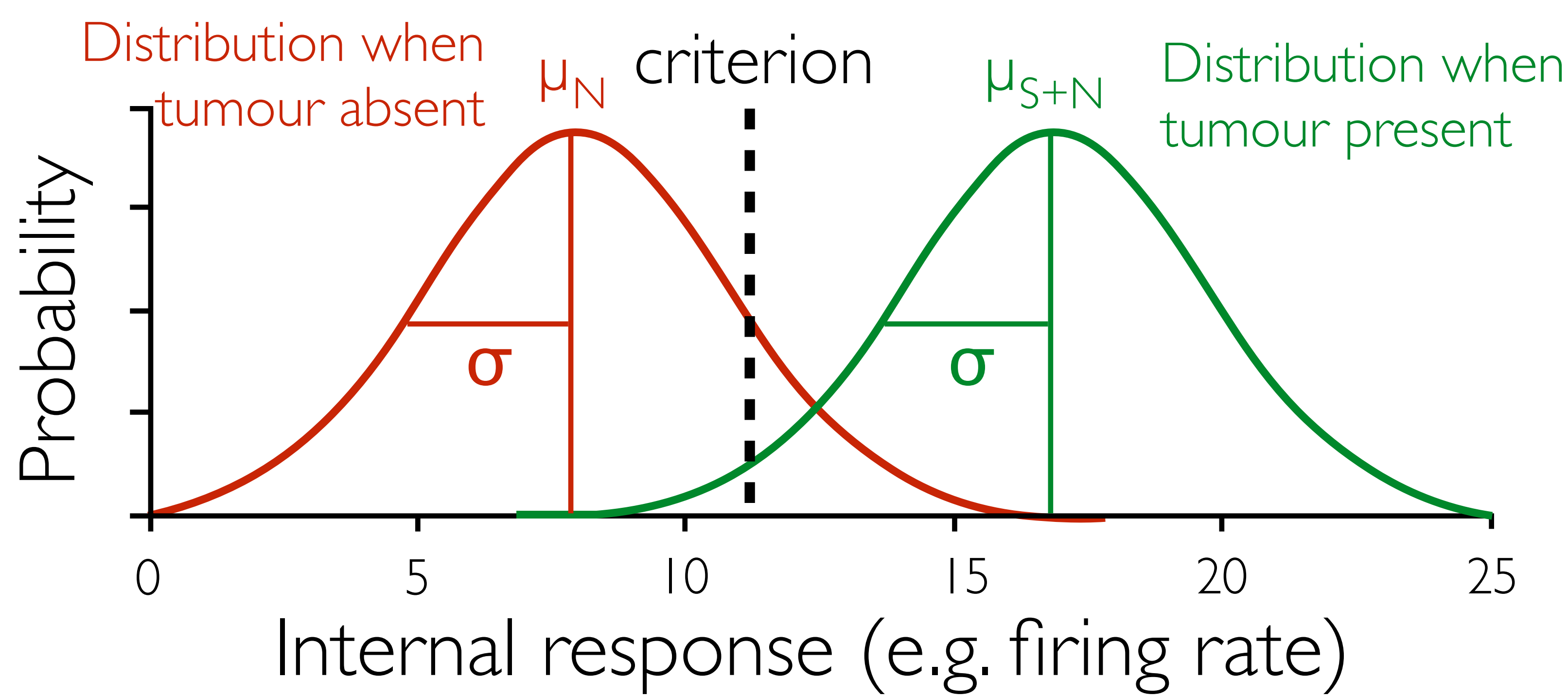
- Signal absent trials:
  - Response below the criterion = correct rejection
  - Response above the criterion = false alarm

Measuring sensitivity

- Sensitivity is characterised by d' (d prime)

$$d' = \frac{\mu_{S+N} - \mu_N}{\sigma}$$





Calculating d'

		Signal: Is there a tumour?	
		Yes	**No**
Decision: Is there a tumour?	**Yes**	Hit	False Alarm
	No	Miss	Correct Reject
		signal+noise	*noise*

- Sensitivity is characterised by d' (d prime)

$$d' = \frac{\mu_{S+N} - \mu_N}{\sigma}$$

- d’ = z(Hit) - z(FA)

d' examples

Early stage tumour

Signal: Is there a tumour?

Decision: Is there a tumour?

	Yes	No
Yes	0.84	0.50
No	0.16	0.50

Late stage tumour

Signal: Is there a tumour?

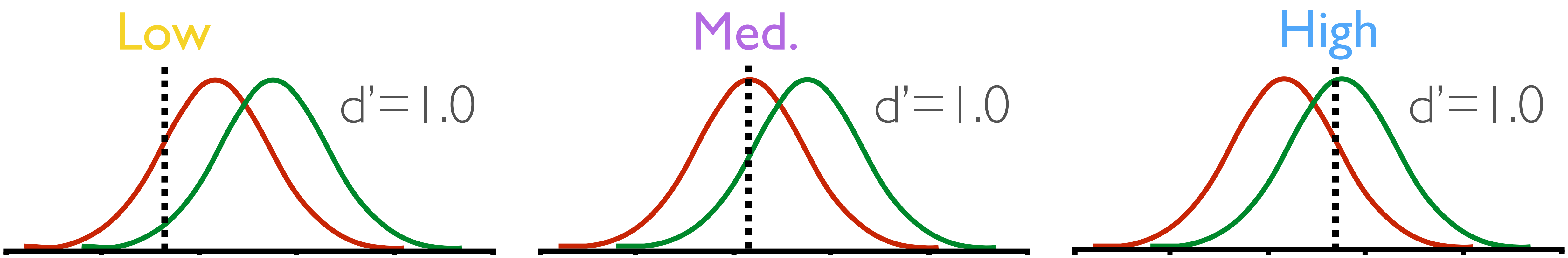
Decision: Is there a tumour?

	Yes	No
Yes	0.98	0.33
No	0.02	0.77

- Early stage tumour: d' = z(0.84) - z(0.5) = 1
- Late stage tumour: d' = z(0.98) - z(0.33) = 2.5

Criterion effects

- The criterion can also alter performance drastically
  - e.g. Radiologists may weigh errors differently - one considers missed diagnoses fatal, another minimises unnecessary procedures
  - Note there is no point that completely removes false alarms without missing many 'signal present' trials



Low	Med.	High
Hits = 98%	Hits = 84%	Hits = 50%
False Alarms = 84%	False Alarms = 50%	False Alarms = 16%

Measuring the criterion

Low

		Signal: Is there a tumour?	
		Yes	**No**
Decision: Is there a tumour?	**Yes**	0.98	0.84
	No	0.02	0.16

High

		Signal: Is there a tumour?	
		Yes	**No**
Decision: Is there a tumour?	**Yes**	0.50	0.16
	No	0.50	0.84

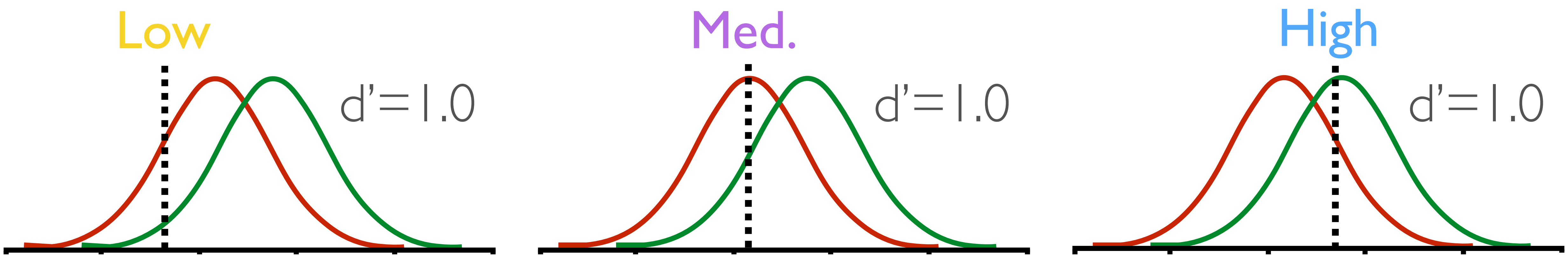
- Is there a way to characterise this criterion?

$$c = \frac{-(z(\text{Hit}) + z(\text{FA}))}{2}$$

- Negative means many 'yes' responses; positive means 'no'

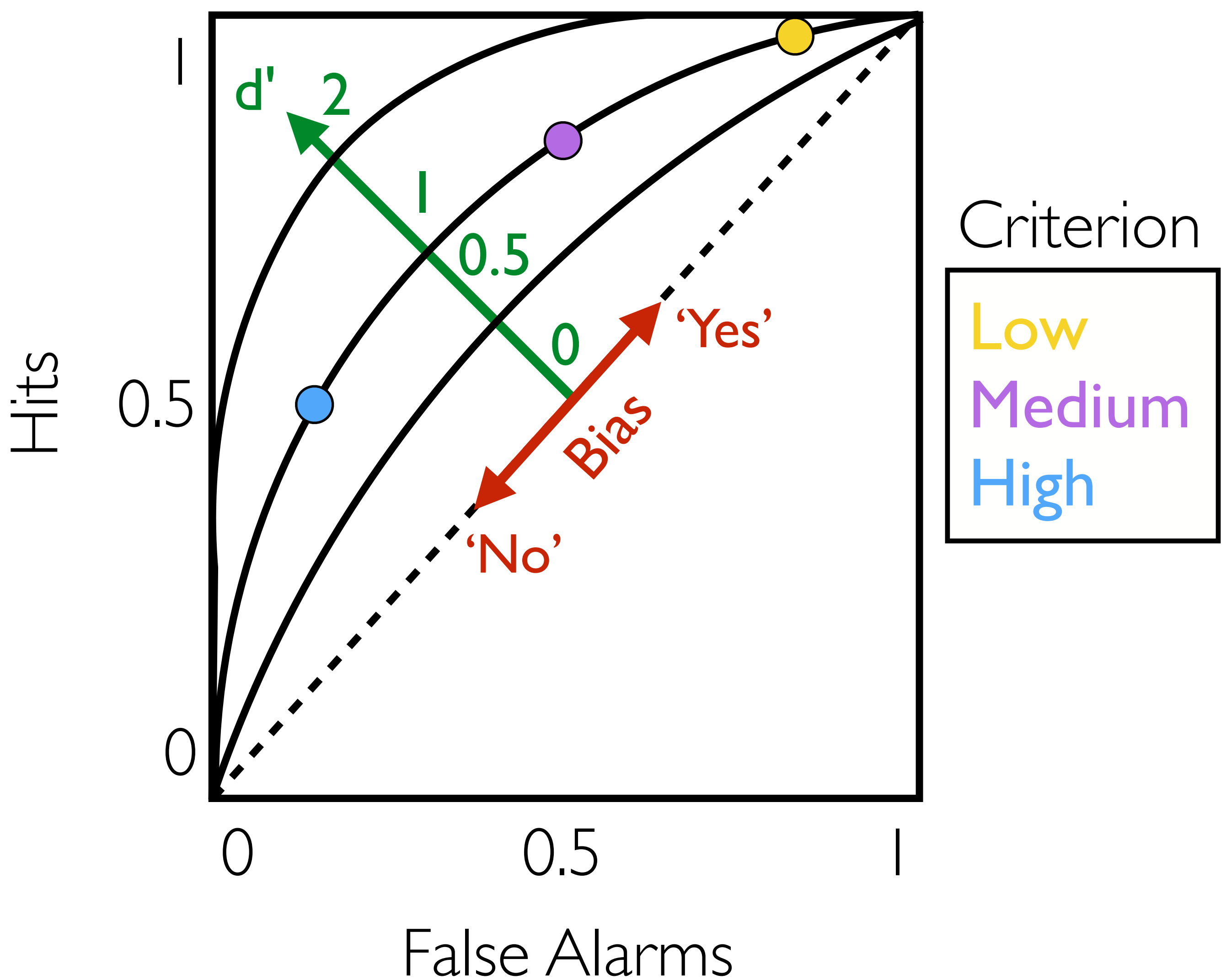
Criterion examples

$$c = \frac{-(z(\mathrm{Hit}) + z(\mathrm{FA}))}{2}$$



Hits = 98%	Hits = 84%	Hits = 50%
False Alarms = 84%	False Alarms = 50%	False Alarms = 16%
$c = -1.5$	$c = -0.5$	$c = 0.5$

The ROC curve

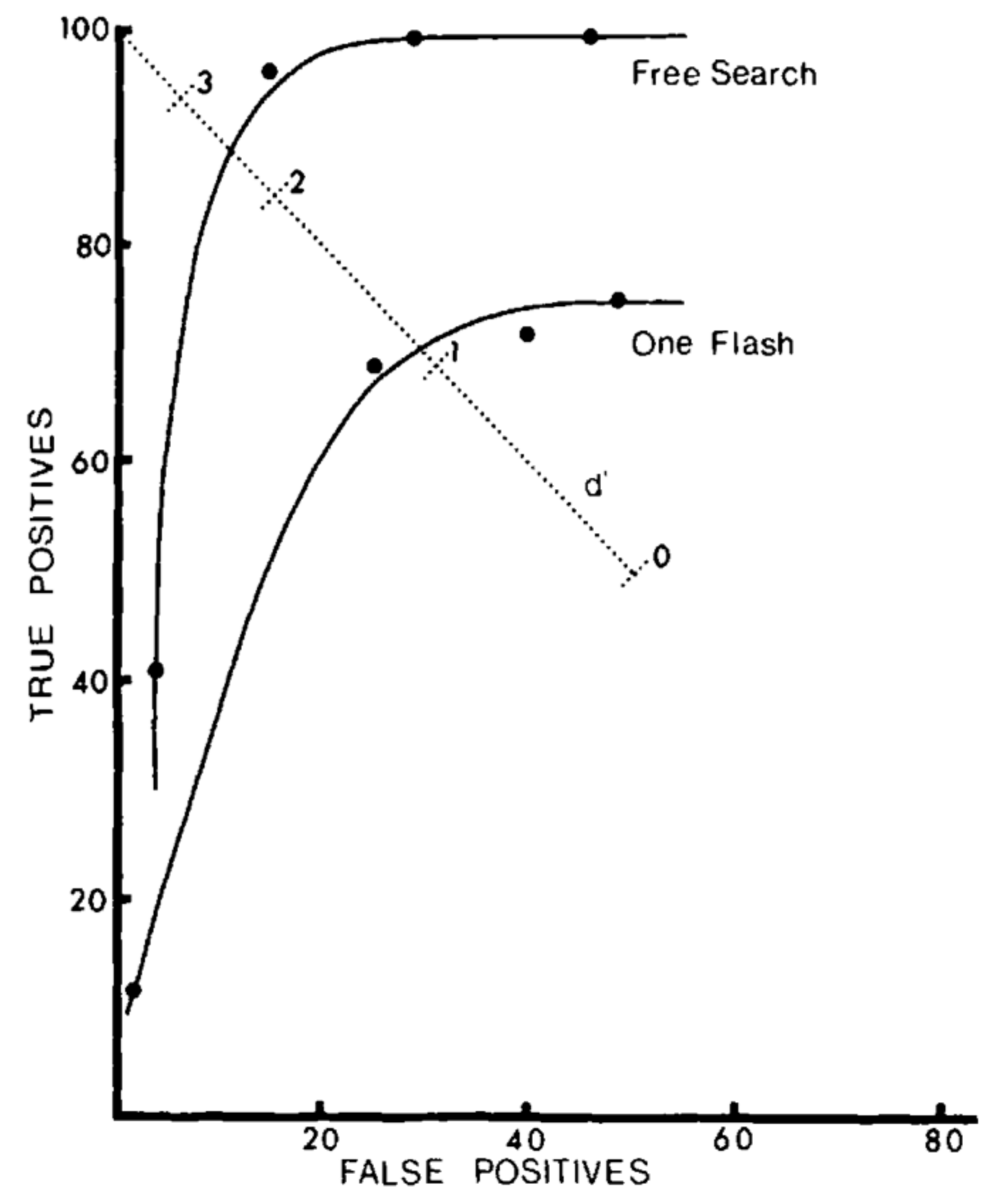


- d' increases along one axis
- Bias shifts points along these constant d' functions
- Allows separation of the two parameters

- Receiver Operating Characteristic (ROC) curves show how Hits and False Alarms relate to d' and criterion

ROCs for radiologists

- Kundel & Nodine (1975) asked 10 radiologists to identify lung abnormalities with flashed images (0.2 sec) or with free viewing
- Single flash d' of around 1.0
- Free viewing d' around 2.5
- Also rated confidence at 4 levels
  - Notice the shifts in criterion with confidence for the two conditions

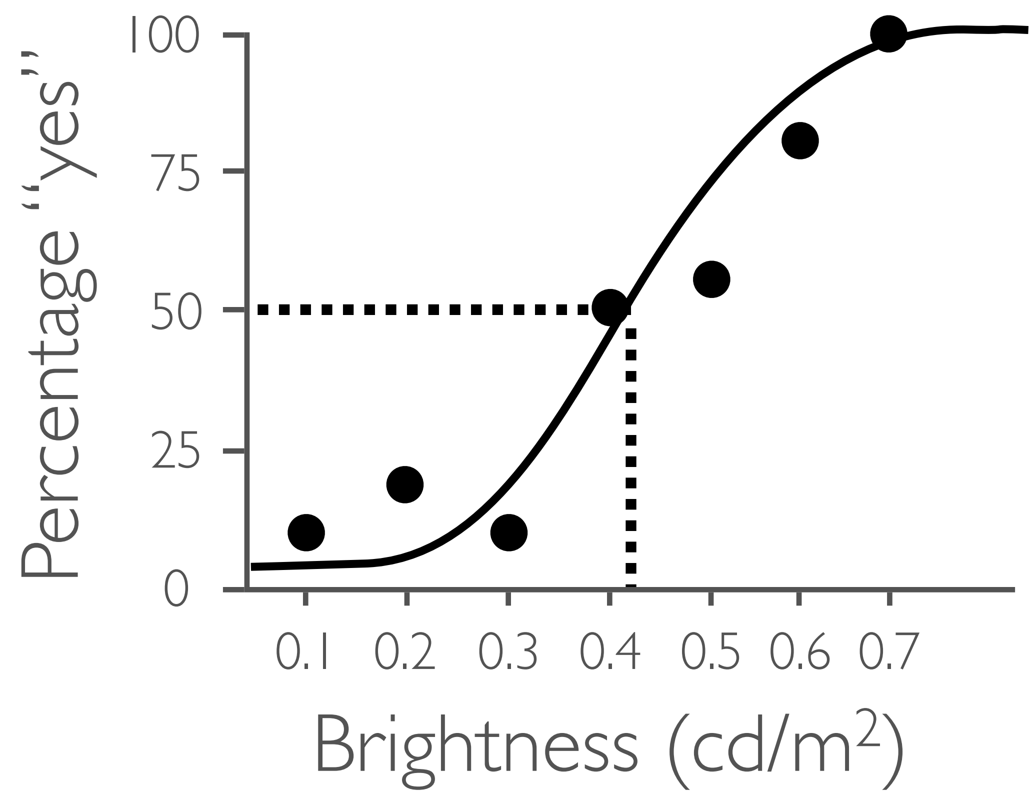


SDT summary

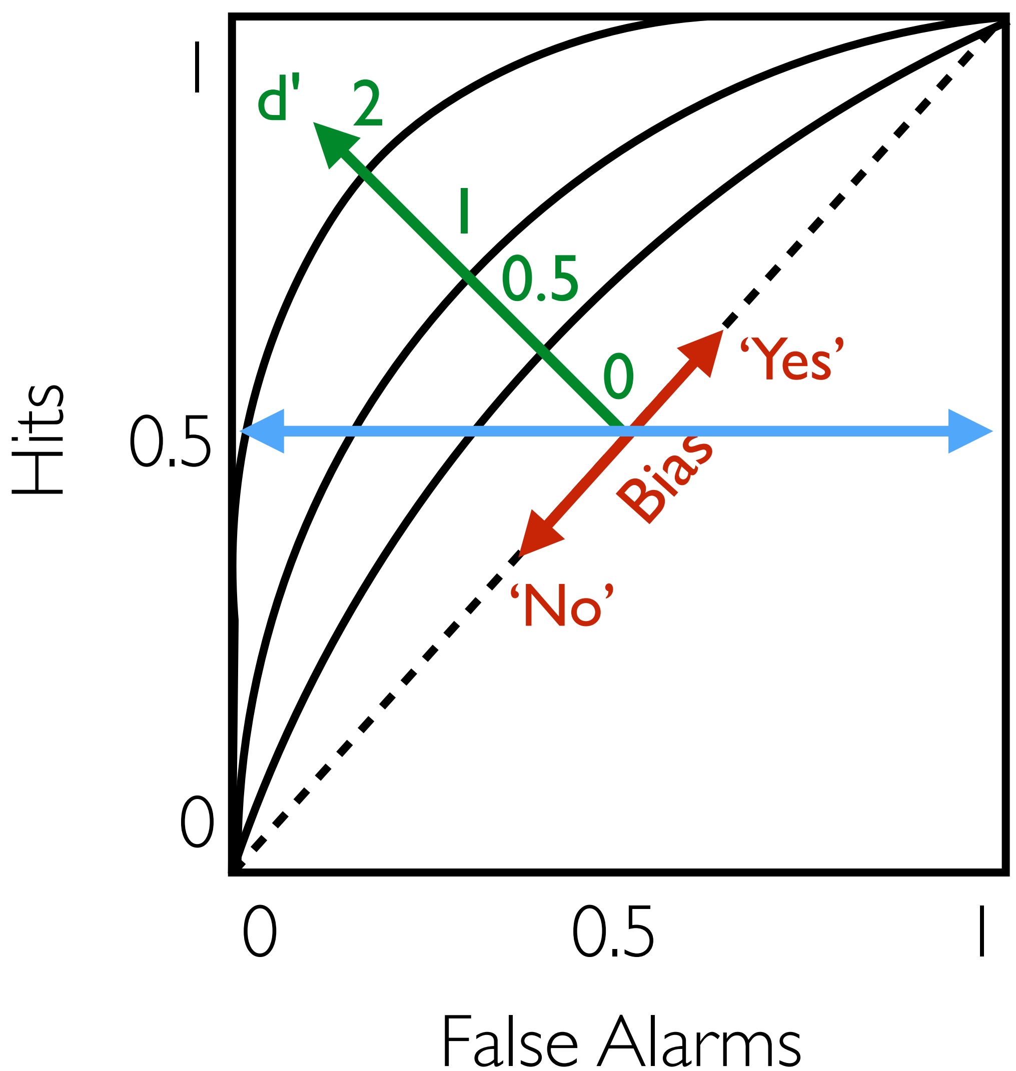
- We can characterise performance using two values
    - d' - sensitivity
    - c - criterion
- Previously we sought to avoid the subjective criterion through the use of forced choice procedures
- SDT allows us to measure it
    - Through the separation of 'signal present' and 'signal absent' trials
- ROC curves allow depiction of the full variation in performance when either sensitivity or criterion changes

SDT vs. the threshold

- How do we reconcile the SDT approach with what we've been describing thus far?
  - Much of what we were measuring earlier concerned finding a threshold point at a given performance level
  - SDT argues that this 'threshold' does not exist



Where is the threshold?



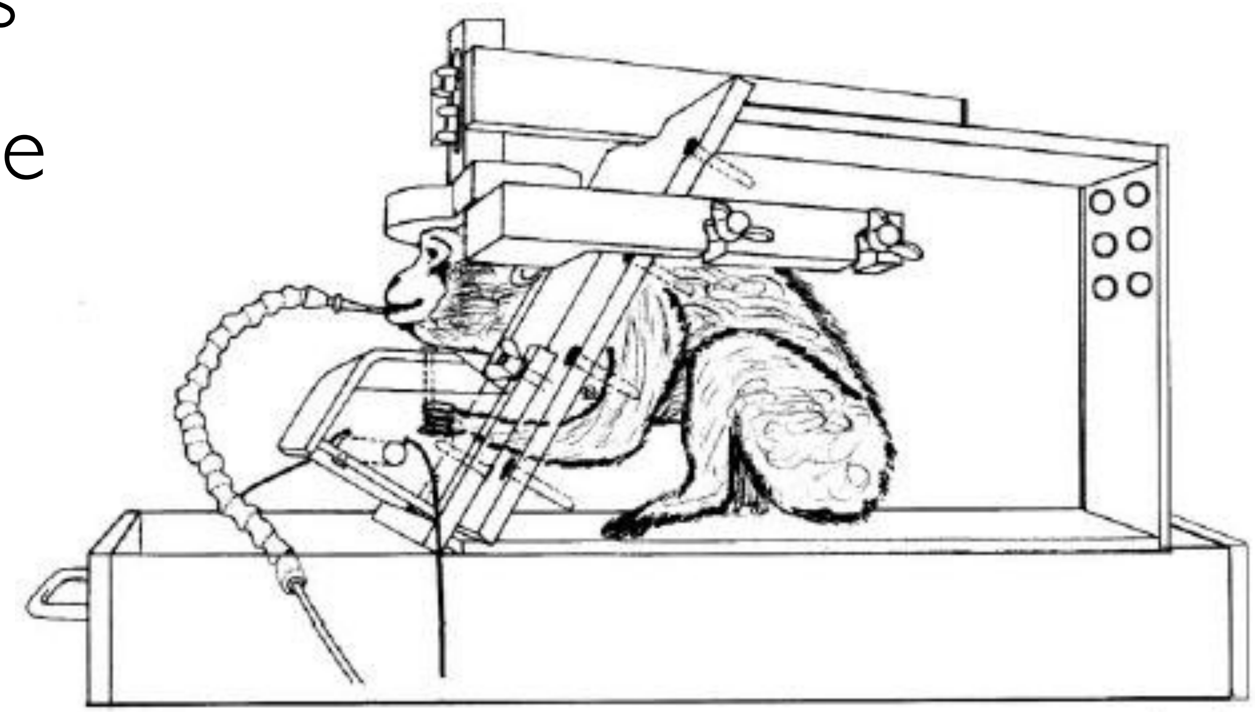
- Percent correct performance can be the result of a range of 'sensitivity' values, depending on the criterion of the observer
- So there is no 'threshold' intensity where a stimulus will always shift from unseen to seen (Swets, 1961)

Do we abandon thresholds?

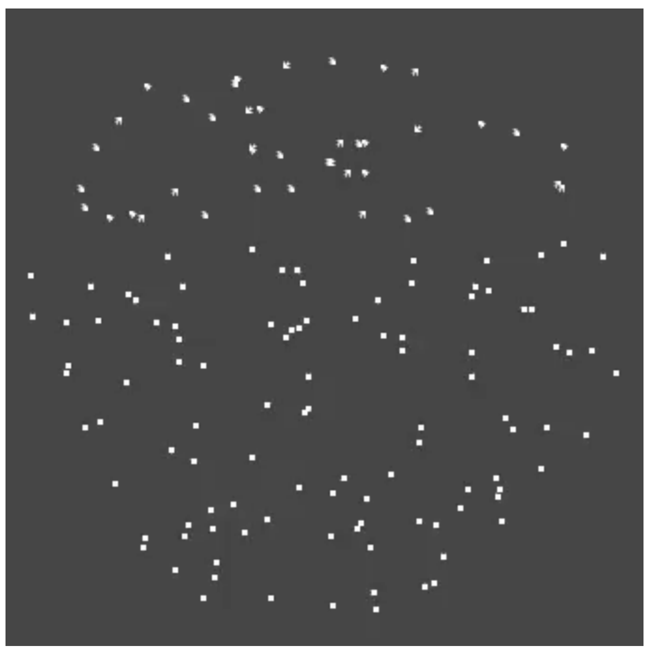
- Perception feels binary but there is unlikely to be a single intensity at which the unseen/seen transition always occurs
- Nonetheless, thresholds do capture a meaningful aspect of our sensory experience:
  - Some stimulus values/differences are harder to see than others
  - Thresholds give us this in meaningful units (brightness, speed, etc.)
- If we minimise criterion effects (e.g. with forced choice procedures) then we can measure sensitivity more closely
- But it is worth keeping in mind that there is likely no ‘neural correlate’ of a threshold - it’s more an explanatory concept

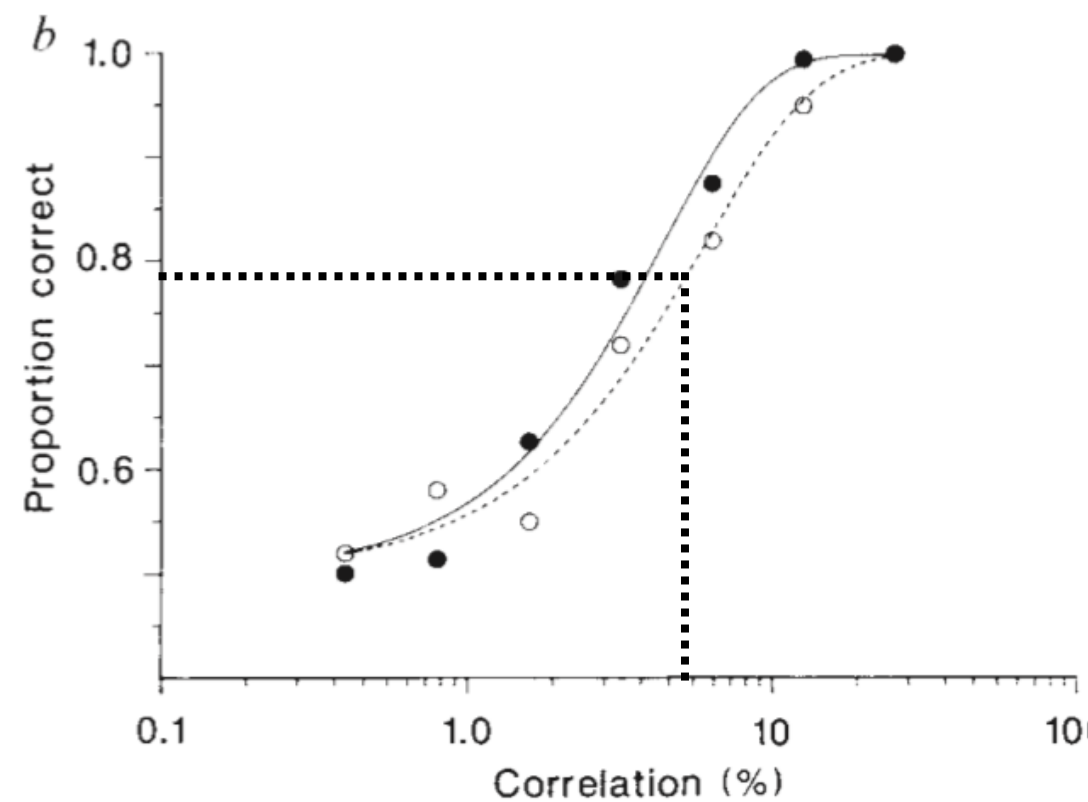
Psychophysics and the brain

- We've seen how performance varies with stimulus intensity, but can we link this with neural activity?
- Newsome, Britten & Movshon (1989) compared psychophysics and single-cell recordings in brain region MT/V5 of macaque monkeys
- Recall that with forced choice procedures we can test perception in animals as well as humans



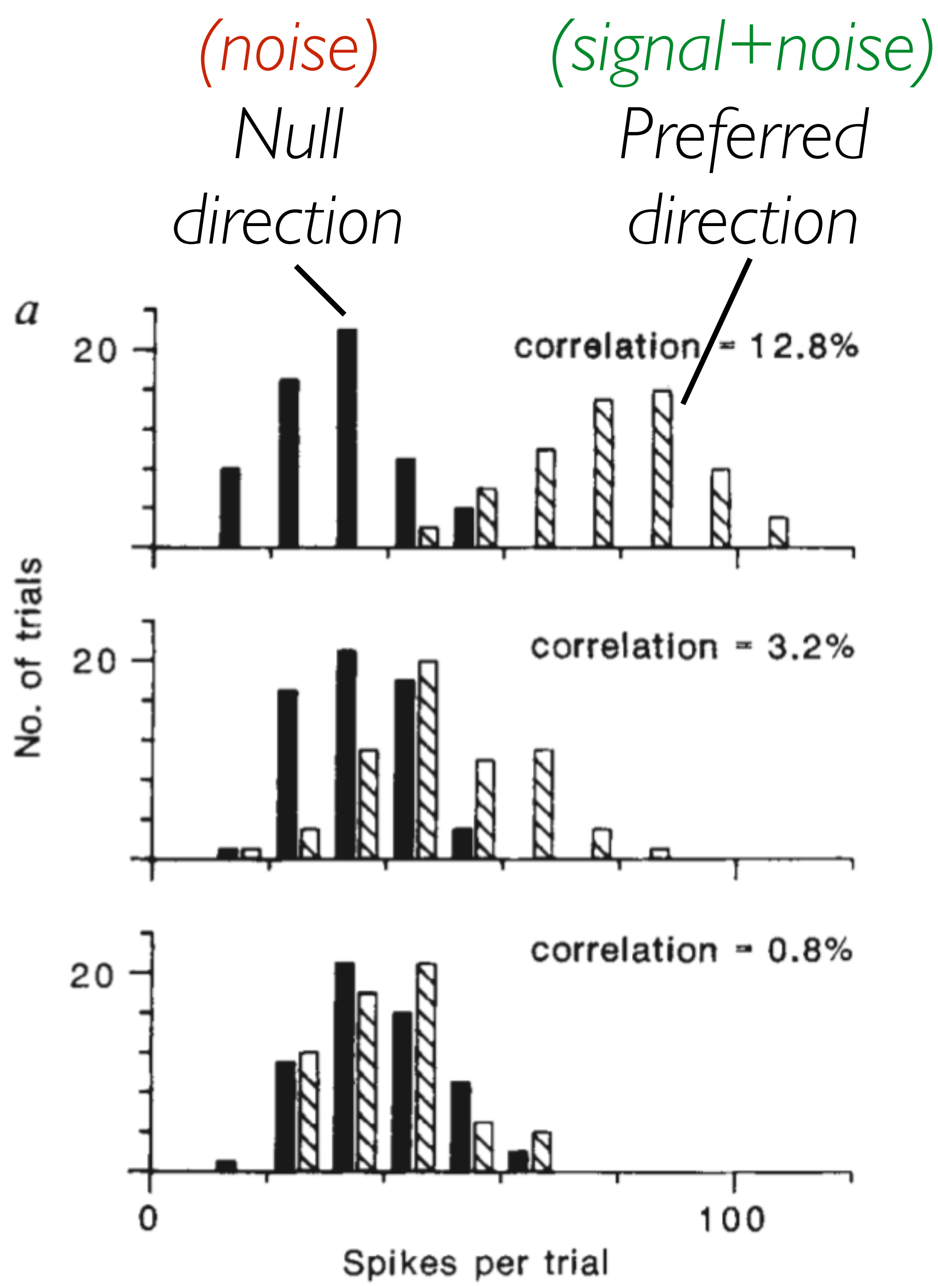
Newsome *et al* (1989)





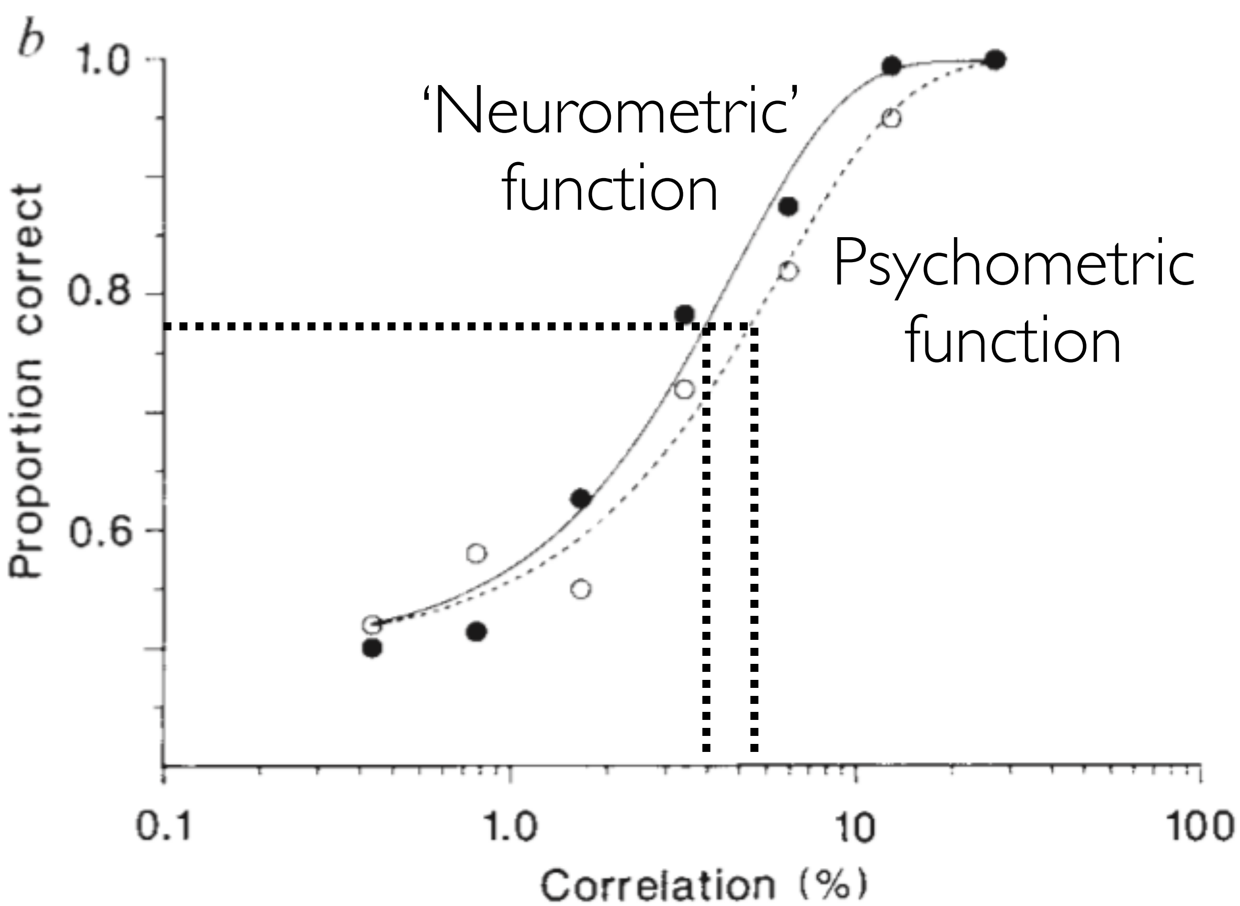
- Measure detection thresholds for 'global motion'
  - Random dot kinematograms with varied 'correlation' in direction
  - Psychometric function using a 2AFC MCS task gives threshold ~6%

From neurons to behaviour



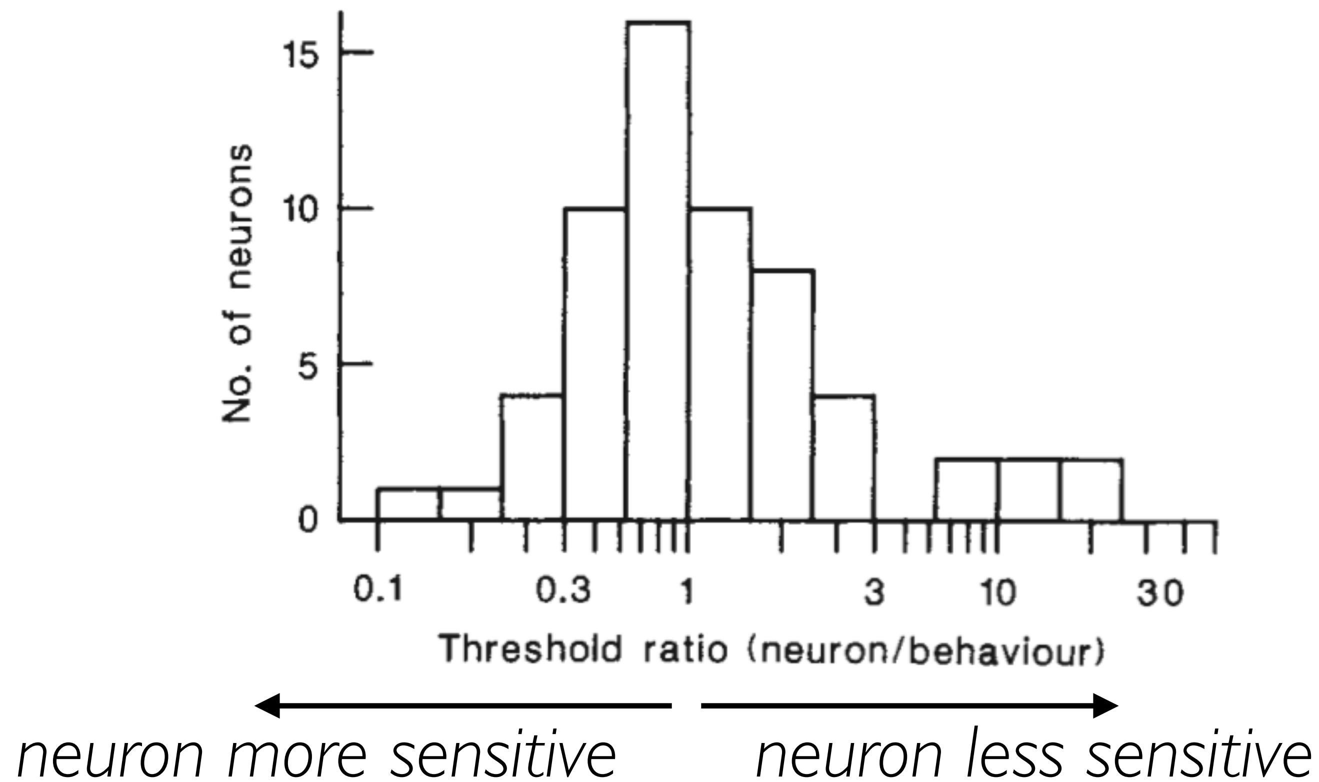
- Cells in MT/V5 have a preferred direction of motion and produce more spikes with increasing correlation
  - Use this to construct response distributions for the preferred vs. opposite ('null') direction
  - Use SDT to simulate a 2AFC response:
    - On each trial, take one spike rate from the preferred direction distribution and one from the null direction
    - Highest spike rate gives the 'decision'

‘Neurometric’ functions



- Black dots show the ‘neurometric function’ of a single MT/V5 neuron
- If we take thresholds:
  - Behaviour: ~6% coherence
  - Single neuron: ~4% coherence
- Here is a single cell that responds in a similar fashion to the monkey’s behavioural responses

Cells vs. behaviour



- Compare neuronal and behavioural thresholds
  - Some cells are more sensitive, others less sensitive
  - Distribution peaks with the same sensitivity as the individual

Newsome *et al* summary

- Behavioural results show a similar pattern to the responses of individual neurons to the same stimulus
- Extreme view: perception is derived from a single neuron
  - But many neurons were more sensitive than the monkey
  - Why didn't the monkey 'listen' to the more sensitive neurons if they were better at indicating the presence of motion?
  - Some pooling of responses must be occurring
- More likely that motion derives from a 'population code', e.g. select the mean response out of many cells
  - More evidence for population coding in future (e.g. the Spatial Vision lecture)

Overall summary

- Using our knowledge of psychophysics we can design more efficient perceptual tests
  - Staircase and QUEST procedures
- Signal Detection Theory allows separation of sensitivity from criterion (which we previously sought to avoid)
  - Measurement of d' and c with plotting in ROC functions
  - Calls into question the notion of a threshold (though it is a useful explanatory concept nonetheless)
- Psychophysical responses can be linked with the sensitivity of individual neurons (though we likely pool from many)

Reading

- Some background:
  - SDT overview: Wolfe et al. *Sensation & Perception*. Ch 1.
- Further reading (if interested / completely confused):
  - Levitt (1971). Transformed up-down methods in psychoacoustics. The Journal of the Acoustical Society of America. Vol 49 (Issue 2): pp467-477.
  - Swets (1961). Is there a sensory threshold? *Science*. Vol 134 (Issue 3473): pp168-177.
  - Newsome, Britten & Movshon (1989). Neuronal correlates of a perceptual decision. *Nature*. Vol 341: pp52-54.