



HUMAN PERFORMANCE

M6X00915

LECTURE 4: SIGNAL DETECTION

DR KATE WILLIAMS

TODAYS SESSION:

- Background to Signal Detection
 - Focus on Signal Detection Theory (SDT)
 - Receiver Operator Characteristic (ROC) Curves
- Practical Applications of SDT
- Link in to Vigilance.

BACKGROUND

- Detection in the world is important.
 - At an individual level – e.g., moving out of the way of an object being thrown at you.
 - At a broader level – e.g., in industrial settings.
- However, humans are inconsistent with detection.
- Even when levels of stimulation remain constant.
- In exp. research, Ps usually required to detect sound of varying volume, light varying in intensity, or object recognition.

BACKGROUND

- How can we account for performance and ability on detection tasks?
- Through the use of **Signal Detection Theory** (Green & Swets, 1966; Macmillan & Creelman, 2005).
- When observer classifies in to 1 of 2 states:
 - Either present or absent
- However, detection decisions often take place in presence of uncertainty (due to internal & external factors).

WHAT IS SIGNAL DETECTION?

- Detection is modelled within framework of SDT
- Provides precise language & graphic notation for analysing decision making in presence of uncertainty
- SDT is applicable in any situation where there are 2 discrete states:
 1. Signal
 2. Noise
- The process of signal detection results in 2 response categories:
 1. Yes (the signal is present)
 2. No (the signal is not present)

WHAT IS SIGNAL DETECTION?

- Combination of these state & response categories produces a 2 x 2 table.

		Perceiver's response	
		"Yes"	"No"
Stimulus	Present	Hit	Miss
	Absent	False alarm	Correct rejection

Generates
4 classes of
joint events.

WHAT IS SIGNAL DETECTION?

- Perfect performance: No misses or FAs.
- But, not always easy to distinguish signal from noise.
- Target detection latencies also taken into consideration.
 - Can be critical in real-life scenarios
- Premise of signal detection:
 - When a stimulus or signal occurs, it creates a neural response.

WHAT IS SIGNAL DETECTION?

- Neural noise fluctuates constantly, so the brain must decide whether neural activity reflects noise alone or also a signal.
- When a stimulus is difficult to detect, cognitive factors become critical.
- Nevertheless, there will always be a trade-off between hits and false alarms.
- We must consider number of times a person claims to detect a target when it is not there (FA rate), just as much as correct detection of a target that is present (Hit rate).
 - WHY?.....

WHAT IS SIGNAL DETECTION?

- Detection ability is calculated as a probability from the hit and false alarm rates:
 - 20 signals presented, 5 hits & 15 misses = $5/20 = 0.25$ **Hit rate**
 - 10 noise trials, $1/2$ were “yes” response = $5/10 = 0.50$ **FA rate**
- By combining hit & FA it produces an index of **response criterion** placement.....
 - Observers willingness to respond affirmatively.

WHAT IS SIGNAL DETECTION?

- It determines observers response characteristic
 - Conservative
 - Liberal
- Low hit & false alarm rates =
 - *More conservative* in response.
- High hit & false alarm rates =
 - *More liberal* in response.
- Reduction in hit rate but steady/increase FA rate =
observer is not good at distinguishing between 2 types of events.

RESPONSE CHARACTERISTICS

- The application of SDT to vigilance situations provides an advantage in the assessment of detection efficiency.
- It determines perceptual sensitivity – which shows the efficiency of the observers perceptual system.
- Sensory evidence about signals is received by a noisy nervous system.
- A separate decision process acts on such evidence to determine whether or not a detection response should be made.

RESPONSE CHARACTERISTICS

- A signal detection response is the joint outcome of the operation of sensory and decision processes.
- However, it is difficult to discriminate signals from noise in many situations
 - There is considerable overlap between the two.
- Difficult to tell whether strength of sensory evidence is indicative of the signal or attributable to noise.

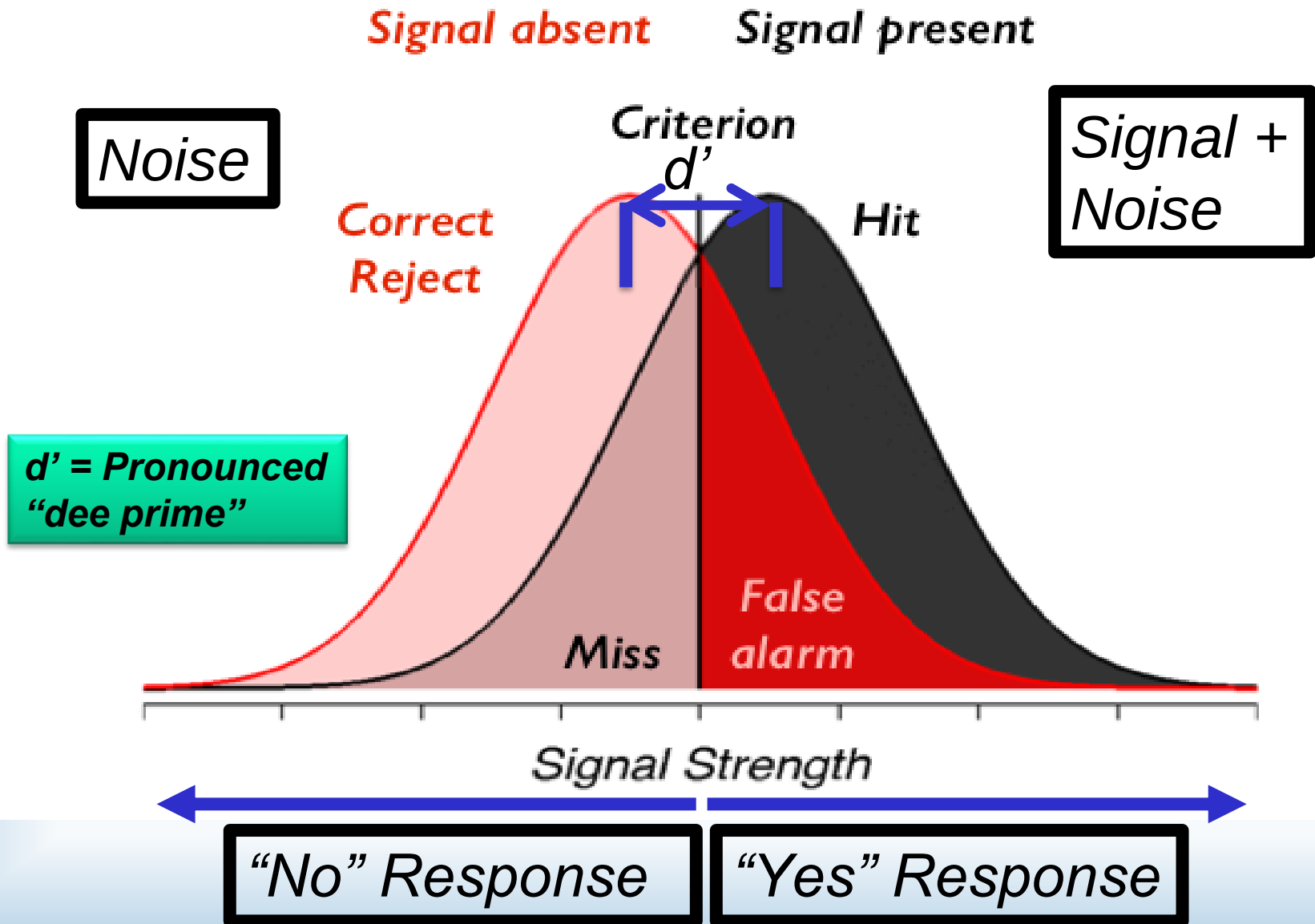
SIGNAL DETECTION THEORY

- Observers are thought to resolve this by setting a criterion to report a target as being there if they judge the level to have been exceeded.
- So, signal and noise are represented as a pair of normal distributions
- This shows the distinction between sensory and decision processes.

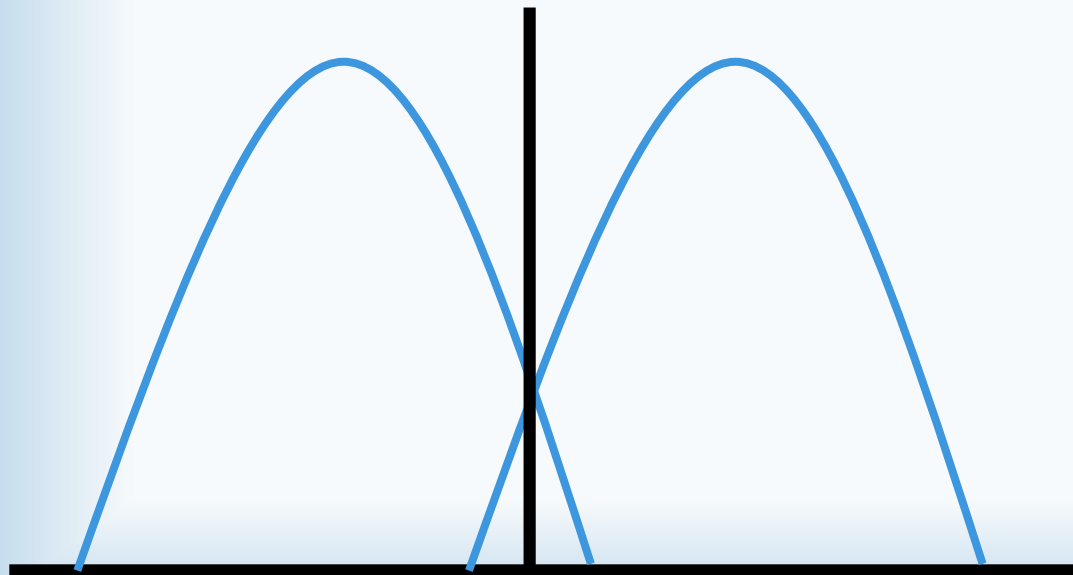
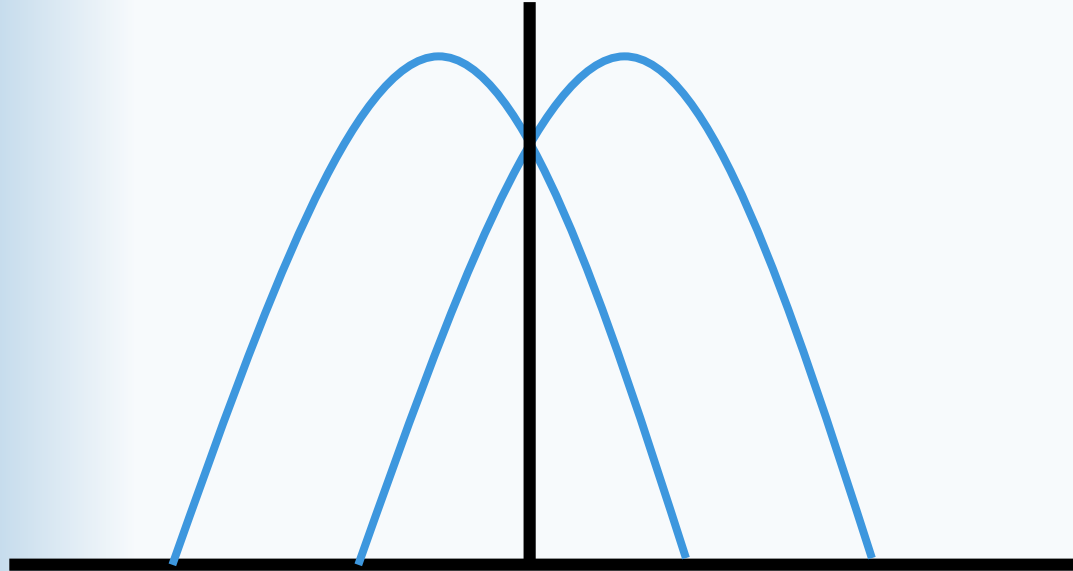
SIGNAL DETECTION THEORY

- The detectability of the signal depends on the difference between the means of the noise plus signal and the noise alone distributions.
- This is known as **d' (d prime) – the index of sensitivity.**
 - The greater the overlap between the distributions, the lower the d value & the more difficult it is to discriminate signal from noise.

SIGNAL DETECTION THEORY



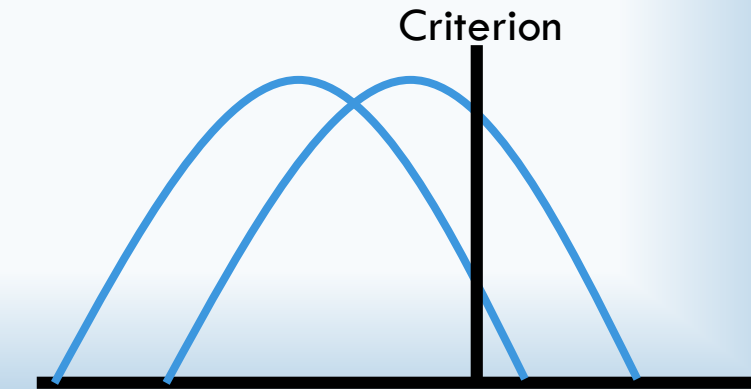
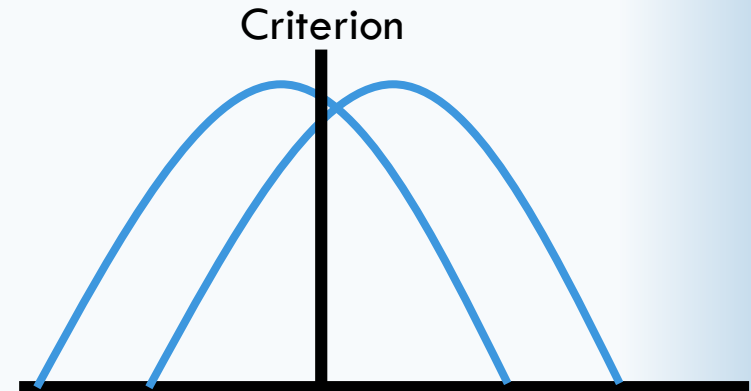
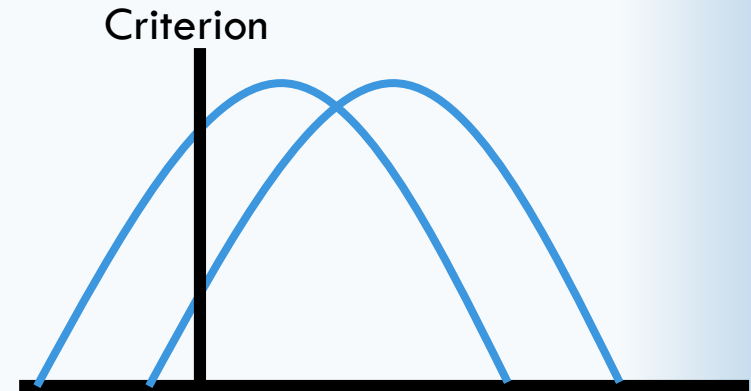
SIGNAL DETECTION THEORY



Which graphical representation indicates a more accurate performance?

SHIFTING THE RESPONSE CRITERION

- People may set themselves specific criterion levels during detection decision scenarios.
- Only when that threshold is reached will a person decide to respond “yes” to the possibility of a signal.
- Notice there is no way of setting the criterion to achieve only hits & no FAs.



BETA AND SIGNAL PROBABILITY

- Another important factor in SDT is the beta (β):
 - The ratio of neural activity produced by signal and noise.
 - SDT can prescribe where the optimum beta should be given:
 1. Likelihood of observing the signal
 2. The cost and benefits of the 4 possible outcomes.

BETA AND SIGNAL PROBABILITY

- When a signal occurs just as often as it does not, optimal performance will occur when the criterion is placed at the intersection of the two curves.
 - $\text{Beta} = 1$.
 - Any other placement would produce more errors.
- If a signal is more likely to occur than not, then the criterion should be lowered.
- If the signal probability is reduced then the criterion should be increased.

PAYOFFS

- Optimal setting of beta is also influenced by payoffs.
- However, the term *optimal* is now defined as the option that maximises the **expected value**, rather than as the value that minimises errors.
- If it is deemed more important for signals to never be missed, higher rewards will be given for hits and higher penalties will be given for misses
 - This will lead to a lower setting of beta.
 - But what knock-on effect would this have?

PAYOFFS

- So, response bias and response criterion can be influenced by the value of a desirable event or outcome to occur and not just how the individual may respond due to their general ability.
- Due to this, researchers are also interested in how well individuals can set their criteria in response to changes in probabilities and payoffs, relative to optimal beta.

SETTING BETA

- Exp. research has shown that humans are not good at adjusting beta.
- Typically, they are less risky than they should be if the optimal beta is high and less conservative than they should be if the optimal beta level is low.
- Green and Swets (1966): suggest the poor setting of beta is more pronounced when it is manipulated by probabilities than payoffs

SETTING BETA

- Possible reasons for this:
 1. Observers engage in probability matching instead of setting objective criterion (Laming, 2010)
 - Try to balance errors even when probability of signal is unlikely.
 2. Misperceiving probabilities
 - Overestimate probability of rare events & underestimate frequent events (e.g., Erlick, 1964; Hollands & Dyre, 2000).

RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES

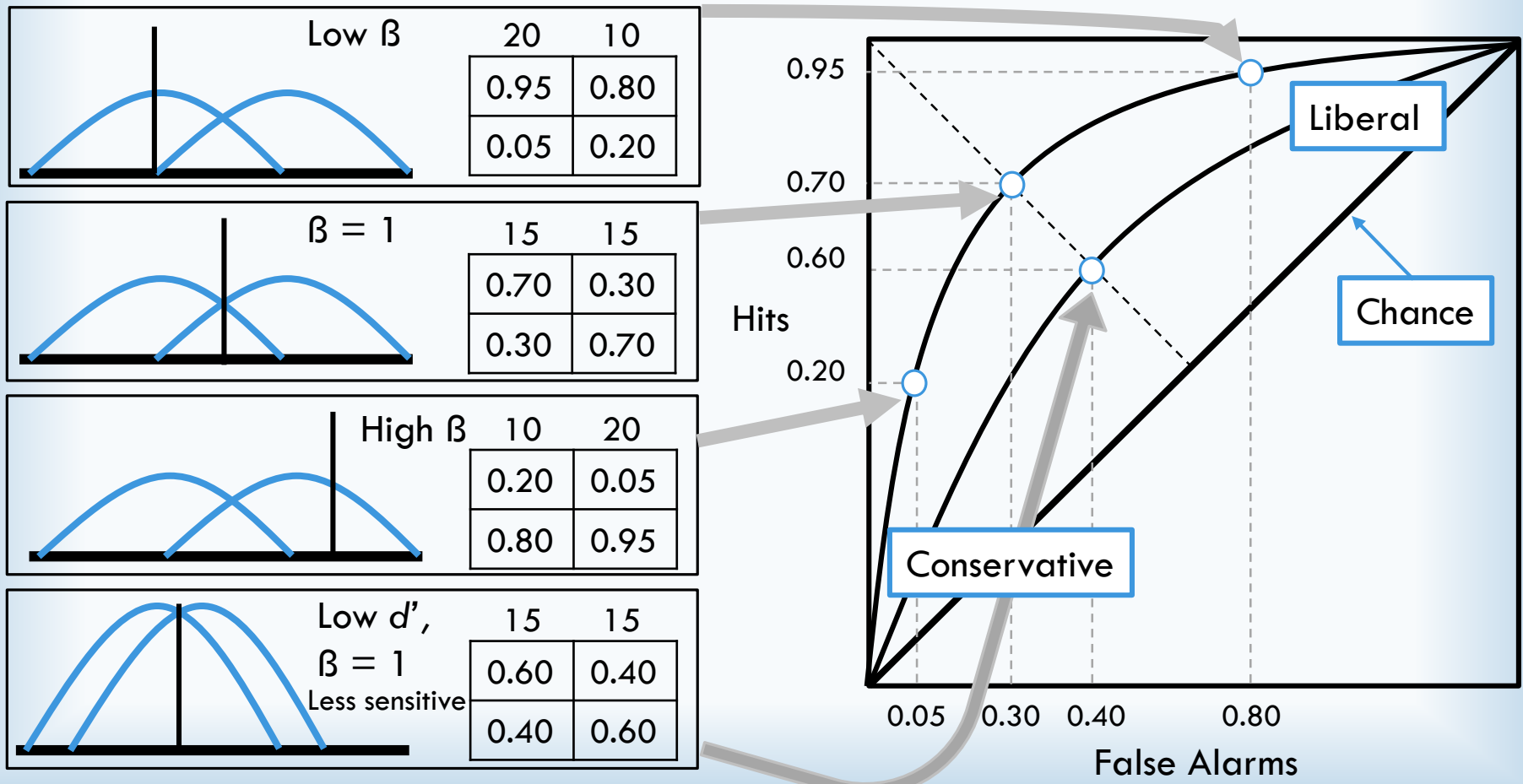
- Used in order to understand the joint effects of sensitivity (d') and response bias (β) on data from signal detection analysis.
- The ROC curve plots hits against FAs for different settings of the response criterion at a constant level of sensitivity.
- So, we can describe a full range of options in a single curve.

RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES

- As criterion is moved to different locations, a different set of values will be generated in the matrix for each of the 4 outcomes (hit, FA, CR, Miss)
- These various alternatives can be displayed as a ROC curve on a single graph.
- Each signal detection condition (matrix) generates 1 point on the ROC
 - ROC curves are plotted with hits on the vertical axis and FAs on the horizontal axis.

RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES

- If the signal and the observer's sensitivity remain the same, changing the beta between conditions/trials (either due to a change in payoffs or signal probability) will produce a curved set of points: ROC curve.



CONFIDENCE RATINGS

- Changing response criterion based on different payoffs and signal probabilities to examine signal detection rates is arduous.
 - Thus, research collects confidence ratings from Ps as a means to examine several criterion settings.
 - Following a trial, Ps asked to indicate for e.g.,
 - 1 = confident signal was NOT present
 - 2 = uncertain
 - 3 = confident signal was present

CONFIDENCE RATINGS

- Data can then be analysed in two ways:
 - Responses 1 & 2 indicating no (conservative bias setting)
 - Responses 2 & 3 indicating yes (liberal bias setting)
 - So, two beta settings are available from one set of trials.
- Any issues with basing outcomes on Ps confidence ratings?????

BENEFITS OF SDT

1. Provides ability to compare sensitivity
 - Compares **quality** of performance between conditions/people that may differ in bias.
 2. Separate sensitivity and bias components
 - Different corrective actions can be implemented depending on whether a change in performance results from sensitivity loss or shift in response bias (Swets & Pickett, 1982).
- Both are important for real-world issues.
 - Can you think why???

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- Theoretical outcomes of SDT and empirical data collected from field research will differ
 - Why do you think this is the case???
- Empirical data will not always fall within the angle of the slope.
- Number of signal and noise conditions cannot be controlled.
- Noise/signal in reality don't follow assumptions of normal distribution & equal variance.

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- Difficult to collect multiple data points to plot on the ROC curve.
 - Perhaps only 1 stimulus-response matrix will be available.
- In this case the non-parametric equivalent called A' (a-prime) should be employed (Macmillan & Creelman, 2005).
 - This is a measure of the **area** under the ROC curve.
 - It is sensitive to response options close to 0 & 1.

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- There are also alternative measures of bias
 - E.g., C bias, which locates the criterion relative to the intersection of the 2 distributions.
 - The intersection point is the zero, & the distance from the criterion is measured in z-units.
 - Positive C values = conservative bias
 - Negative C values = liberal bias
 - Some suggest that the measure of C is better than Beta for measuring bias because it is less sensitive to changes in d' .

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- Why is this important?.....
 - In lab studies, the noise and signal are pre-determined and there is control over the number of each presented during a trial.
 - In the real-world there is vast variation.

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- When definition of a signal or category is not clear, **fuzzy logic** is used to mathematically represent data through SDT.
 - Parasuraman et al. (2000) developed equations for fuzzy SDT by combining SDT & fuzzy logic.
 - Fuzzy logic allows an event to belong to more than one set.
 - Like a grey area – e.g., varying temperatures can belong to varying levels of comfort (not just dichotomous category).
 - **Mapping function**: describes the degree of membership.

CONSIDERATIONS FOR REAL-WORLD APPLICATION

- An event can belong to the set “signal” with some degree between 0 and 1 and the response can belong to the set “response” with membership in the range of 0 and 1.
- Once the set and response are mapped on to the range, event membership to 4 fuzzy outcome categories (hits, miss, FA, CR) can be computed (See Parasuraman et al., 2000 for formulas).

APPLICATION OF FUZZY SDT

- Air traffic control: flight conflict detection.
 - One lab & one field study (Masalonis & Parasuraman, 2003)
 - Fuzzy SDT analysis resulted in lower values of sensitivity & more conservative response bias compared to standard SDT.
 - Event falling outside of criterion in SDT was defined by fuzzy SDT as worthy of attention.
 - Fuzzy SDT provides more complete picture of performance in conflict detection tasks.

APPLICATION OF FUZZY SDT

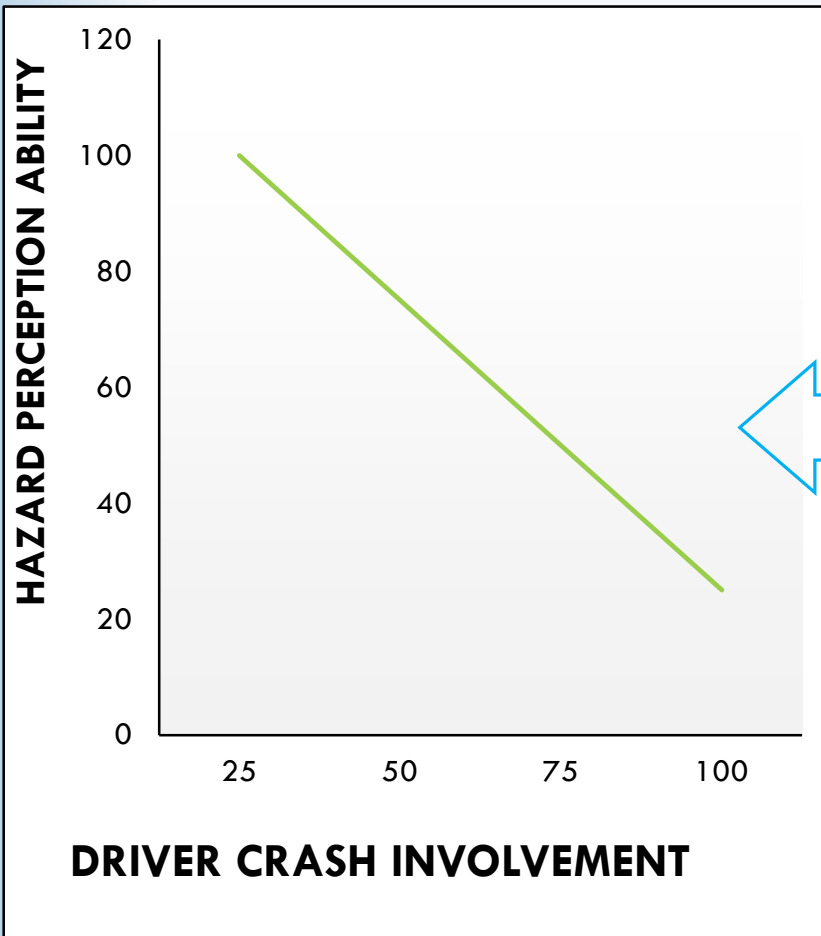
- Sporting decisions (MacMahon & Starkes, 2008).
 - Asked umpires, players, & Ps with no baseball experience to make decisions on strikes.
 - As expected from fuzzy SDT (i.e., signal being context-dependent) strike calls depended on the type of ball being played.

APPLICATION OF FUZZY SDT

- Perception of road hazards and driving safety

BUT

- Evaluation of this type of data needs to be more sensitive to individual differences.
- Experienced vs. learner driver
- Wallis & Horswill (2007)
 - Tested 2 models:
 1. Learners = lower sensitivity discriminating hazards
 2. Learners = higher threshold for perceiving situation as dangerous



APPLICATION OF FUZZY SDT

- Use of fuzzy SDT analysis (which considers degree of membership in safe vs. dangerous driving) supported second model.
- Also, training learners to anticipate potential hazards from environmental cues improved criterion placement
 - Suggests they were then using same model as experienced drivers.

REAL-WORLD APPLICATION

- Detection of weapon by airport security (McCarlet et al., 2004)
- Detection on radar (Mackworth, 1948)
- Detection of malignant tumour on x-ray plate (Swets, 1998)
- Malfunction of system in nuclear power plant (Sorkin & Woods, 1985)
- Identity of a target on a battlefield (Holland & Neyedli, 2011)

REAL-WORLD APPLICATION

- Air traffic control event (Metzeger & Parasuraman, 2001)
- Untruthful statement on a polygraph (Ben-Shakhar & Elaad, 2003)
- Detection of hazard when driving (Wallis & Horswill, 2007)
- Crossing a railway line (Yeh, Multer, & Raslear, 2009)
- Detecting a broken element on an aircraft (Drury, 2001; Swets, 1998)

LET'S LOOK AT AN EXAMPLE IN DETAIL.....

Medical Diagnosis:

- Main application: Detection of a disease (present/absent)
- Considerations:
 - Strength of signal & sensitivity of decision maker
 - Salience of disease
 - Number of other 'related' symptoms
 - Experience of decision maker

LET'S LOOK AT AN EXAMPLE IN DETAIL.....

- Example:
 - Detecting cancerous tumour vs. benign cyst.
- Implications:
 - Not detecting a cancerous tumour that is present or
 - Inaccurately detecting a benign cyst as a cancerous tumour.

MEDICAL DIAGNOSIS SDT MATRIX

Radiologists response

Cancerous
tumour

		Radiologists response	
		"Yes"	"No"
Cancerous tumour	Present	Hit	Miss
	Absent	False alarm	Correct rejection

SDT: MEDICAL DIAGNOSIS (SWETS, 1998)

- Could use the 2nd benefit of SDT to help improve radiologists' sensitivity in discriminating cancerous tumours from benign cysts.
- Checklist developed listing features of the growth to be considered.
- Confidence in presence also assessed.
- Less experienced radiologists showed greater sensitivity when using the checklist compared to not using the checklist.

SDT: MEDICAL DIAGNOSIS (SWETS, 1998)

- For every 100 patients with cancer, checklist increased detection of cancer in ~ 13 additional patients and for patients without cancer it decreased the unnecessary need for biopsies in ~ 12 patients.
- Other factors influencing decision: Point at which patient is examined
 - initial screening (high beta, low probability)
 - or at referral (lower beta, higher probability).(Parasuraman, 1985)

SDT: MEDICAL DIAGNOSIS

- What about the consideration of payoffs (i.e., the values and costs) when making medical decisions?
 - How can the consequences of Hits, FAs, & misses be quantified in terms of decisions that consider costs and values to these when human lives are involved?
 - Although uncomfortable to think about, they have an important influence on detection rate.
 - However, instead of considering all 4 outcomes individually, a criterion limit can be set.
 - Swets (1998) suggests the FA rate is $\sim .10$ in medical contexts.

SDT: MEDICAL DIAGNOSIS

- However, there are also reports of differences in criterion setting between different countries
- E.g., in the US the probability of a signal given a “yes” response is 20-30%, whereas in England it is ~50%.
 - Can you think of any difference between the US & UK healthcare that may influence this?

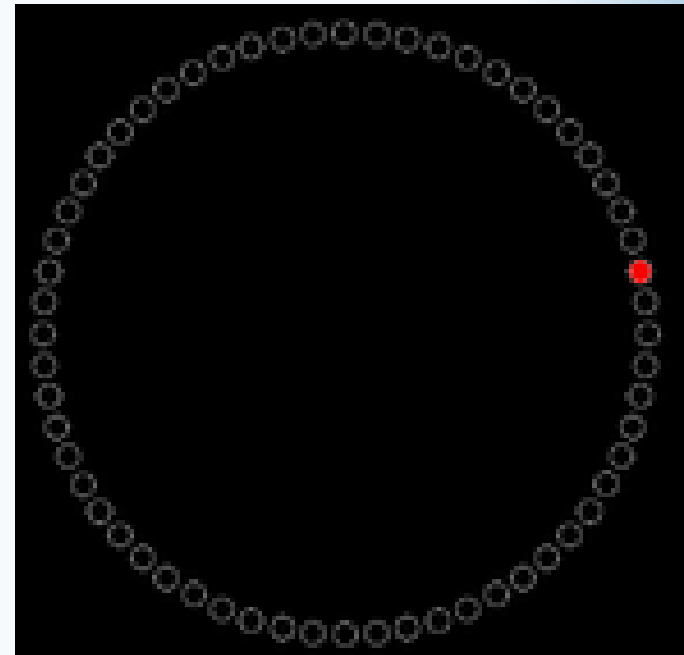
SDT: VIGILANCE

- WW2 highlighted cartridge case inspectors steadily declined noticing faulty cartridges after 45 minutes.
- Other investigations showed high incidences of failures to detect military targets by radar detectors.
- Optimal time radar operators should view radars was determined to be about 30 minutes:
 - After this, efficiency declined even if operators were reporting to be motivated.
 - Also, due to points on radar being so small, only appearing for a few seconds, and having radar display lots of other info, false alarm rates were high.

SDT: VIGILANCE

Mackworth (1948/1950):

- Experienced vs. trainee radar operators.
 - “Clock Test”: Detect “double-jump” of pointer.
 - Decline in efficiency after 30 mins.
 - ***“Breakdown of Vigilance”***
- Vigilance tasks require unbroken direction of attention to one or more sources of information.
 - Sensory vigilance:
 - Difficult to detect due to level of discrimination being so close to threshold of detection.
 - Cognitive vigilance:
 - Stimuli such as letters or numbers are presented above this threshold.



SDT: VIGILANCE

- Performance decline in vigilance task = ***vigilance decrement***.
- Performance decrement less likely to be seen in multiple source monitoring tasks, or where there's complicated signals (Broadbent, 1950; Jerison & Wallis, 1957; Moray et al 1999).
- Signal more likely to be seen if it happens at the start of the task than at the end (Parasuraman & Molloy, 1996).

SDT: VIGILANCE

- Real-life monitoring has very low signal rates and the decrement usually happens after half an hour of work.
- Jerison (1959): performance might begin to decline after presentation of first signal.
 - Even when signals are clearly visible so that all could be detected, there are significant increases in reporting times (Boulter & Adams, 1963).

SDT: VIGILANCE

- Theories of vigilance concerned with explaining vigilance decrement
- But it has been shown that overall levels of detection efficiency are low regardless of whether or not performance deteriorates over time.
 - 90% detection for flaws in rubber seals (Astley & Fox, 1975)
 - 20% detection of faults in electronic equipment (Harris, 1968)

MEASURING VIGILANCE PERFORMANCE

- In SDT, vigilance decrement is either due to decrease in sensitivity or due to shifting to a more conservative criterion.
 - This will depend on the task, context, and situation.
- What influences sensitivity in vigilance performance?
 - Signal strength (change intensity or duration of target)
 - Uncertainty (change time or location of target)
 - Event rate (change number of times target appears)
 - Task type (simultaneous vs. successive)

MEASURING VIGILANCE PERFORMANCE

- What influences bias in vigilance performance?
 - Target probability:
 - High = more hits & FAs (low beta)
 - Low = more CR & misses (high beta)
 - Payoffs (but less consistent & less effective than probability)

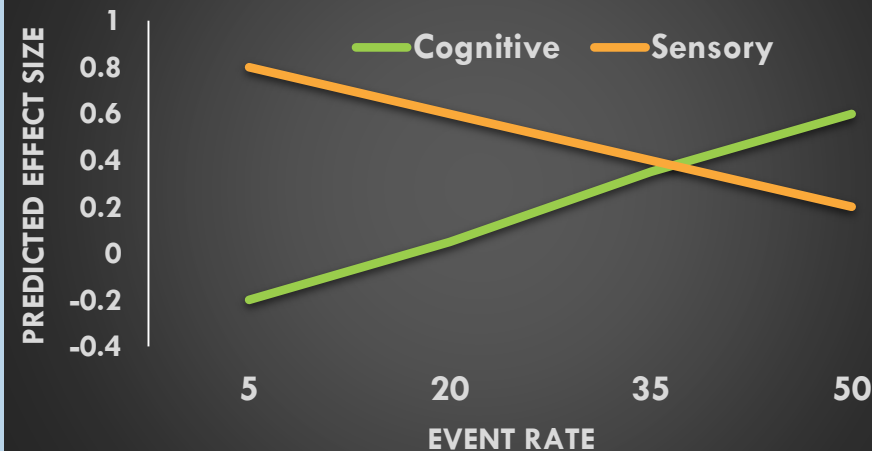
MEASURING VIGILANCE PERFORMANCE

- Meta-analysis of 42 vigilance studies using 138 different conditions.
- Sensitivity decrement found in most with quite a large magnitude.
- Effect sizes were larger for high event rate tasks & tasks requiring a succession of discriminations.
- Also: task difficulty influenced sensitivity decrement independently of other factors –
 - The more difficult the task the greater the sensitivity decrement.

MEASURING VIGILANCE PERFORMANCE

- However, the effects of **event rate** & **information processing** requirements interacted with each other differently depending on whether the vigilance **task was sensory or cognitive**. (See et al., 1995).

Simultaneous task



Successive task



MEASURING VIGILANCE PERFORMANCE

- If we know what influences vigilance, we can manipulate certain task factors to examine the efficiency of vigilance performance.
- Dember and Warm (1979) discussed what they called ***'First order'*** and ***'Second order'*** psychophysiological factors:
- 1st order: physical parameter of signal is altered
 - Increasing intensity, duration, or predictability of signal will increase detection efficiency as perceptual sensitivity is enhanced.
 - But if presentation rate increases (made faster), it reduces perceptual sensitivity and so efficiency decreases.

MEASURING VIGILANCE PERFORMANCE

- 2nd order: probability & frequency signal will occur.
 - Increasing the probability improves the detection efficacy through its effect on the response criterion.
- Signal probability is a major task factor affecting criterion shifts over a vigilance session
- It suggests an observers readiness to say yes to the signal being there is related to their expectancies regarding the frequency of the signal.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Inhibition theory:**
- Vigilance decrement seen in same way as extinction of a conditioned response when that response is no longer reinforced (Mackworth, 1950).
 - Decline in detection attributed to accumulation of inhibition/fatigue.
 - This eventually results in a failure to produce the detection response.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Support for Inhibition theory:**
 - Mackworth's study - found both knowledge of results and introduction of rest pauses abolished the vigilance decrement.
- **Criticisms of Inhibition theory:**
 - Only concerned with the vigilance decrement
 - Did not explain low overall levels of performance in vigilance studies.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Arousal theory:**
 - Vigilance decrement occurs due to decline in CNS arousal
 - In a prolonged low-event environment, the evidence variable shrinks while the criterion stays constant (Welford, 1968).
 - The shrinking results from a decrease in neural activity with decreased arousal – this results in lower hit & FA rates but keeps the separations of the distributions at a constant.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- Mackworth (1950): decrement in vigilance can be reduced (better vigilance) by giving Ps amphetamine
- It could also be increased (so worse vigilance) when they were sleep deprived.



VS.



Any issues here???

- But empirical work suggests arousal may only have an effect on perceptual sensitivity and not response criterion.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- Many studies examining vigilance performance and cortical arousal using measures like EEG and ERP provide support for arousal theory.
 - When frequency of EEG measures falls over course of a vigilance task, this slowing is related to the reduction in detection rate.
 - When theta activity is high, a signal tends to be missed.
 - With ERP, missed signals are preceded by reductions in early ERP components and there is reduction in the amplitude of both early and late ERP components during the course of the vigilance task.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- However, Davies and Parasuraman (1977):
 - Amplitude of late ERP components reflect a decision process in vigilance rather than changes in general arousal.
- Finally, although level of cortical arousal falls during a vigilance task, a reduction in arousal level is not necessary to see a decrement in performance.
- Studies have shown that arousal levels can decline regardless of whether a decrement in performance is seen or not.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Habituation theory**: suggests a reduction in neural responses following repeated presentation of the same stimulus (Mackworth, 1968/69).
 - However, behavioural (Krulowitz, et al., 1975) & neurophysiological (Parasuraman, 1985) evidence fails to provide support.
- **Filter theory**: attending to the same information for long periods of time leads to interruptions that increase with time due to the mechanism that selects information from the environment (the filter), is biased towards selecting novel information (Broadbent, 1958).

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Resource theory**: workload is main factor in controlling vigilance decrements
 - Prolonged performance on task may deplete pool of resources as the person becomes fatigued.
 - More demanding a task, more resources are required.
 - Known as workload-dependent decrements.
 - Theory supported by a generalised decrement being seen across qualitatively different tasks.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Sustained-demand theory**: suggest vigilance tasks place major demands on information-processing resources, leading to increased arousal and stress (Warm, et al., 2008).
 - Maintaining supply of resources over sustained time is difficult, leading to vigilance decrement.
 - Resource-demanding nature of vigilance tasks means they are susceptible to interference from other tasks
 - E.g., when vigilance task & secondary task both draw on spatial working memory resources, a greater decrement is observed (Caggiano & Parasuraman, 2004).

MEASURING VIGILANCE PERFORMANCE: THEORIES

- **Expectancy theory**: observers tend to keep track of past signals in order to predict future ones.
 - If signals are presented frequently and regularly or if there is knowledge given about the accuracy of the signal, then expectancies will be more accurate and detection efficiency will improve.
 - Baker (1959, 1961, 1963b) argued that people perform a continuous averaging process on the time intervals between signal presentations in order to predict the occurrence of future ones.

MEASURING VIGILANCE PERFORMANCE: THEORIES

- Presenting signals at regular time intervals helps the averaging process and results in higher detection efficiency than presenting at irregular time intervals.
- Any problem here? Let's try a quick experiment:
 - Close your eyes and count to 20 in your head
 - When you get to 20 put your hand up and open your eyes.
 - How good are we at estimating time intervals?
- Criticism of expectancy theory is that we are imprecise and the averaging of timings from experimental work doesn't seem to correspond with detection efficiency.

AFTER READING WEEK

- Lecture: Vigilance/Attention and performance.
- Seminar: The Architecture of Attention
- ESSENTIAL READING:
 - See module Handbook