UNDERSTANDING STATISTICS & EXPERIMENTAL DESIGN

Content

Lesson #1

Be aware of partial information!

Bayes Equation

The HIV test has a sensitivity of 0.9999 and a specificity of 0.9999. The incidence rate is 0.0001 in the normal population.

Now, a person is tested and the test is positive.

What is the probability that the person is infected by HIV?

It is 50%!

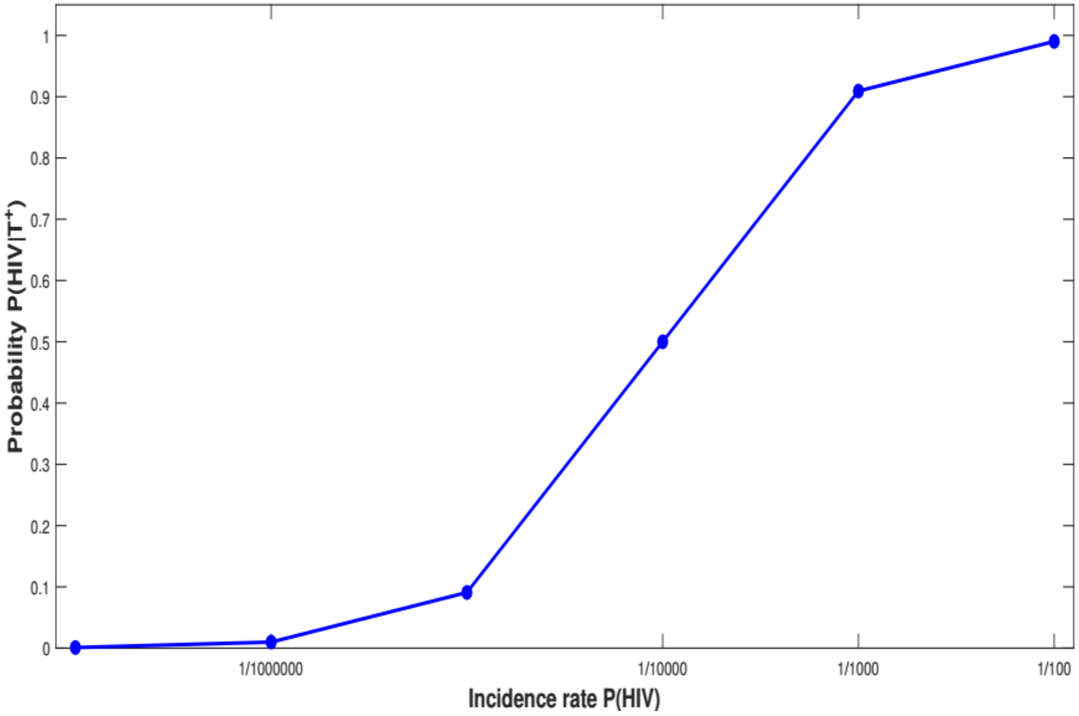
$$p(A|B) * p(B) = p(B|A) * p(A) \rightarrow p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

$$p(HIV \mid T+) = \frac{p(T+ \mid HIV) * p(HIV)}{p(T+)}$$

$$p(HIV \mid T+) = \frac{p(T+ \mid HIV) * p(HIV)}{p(T+ \mid HIV) * p(HIV) + p(T+ \mid \sim HIV) * p(\sim HIV)}$$

$$p(HIV \mid T+) = \frac{0.9999 * 0.0001}{0.9999 * 0.0001 + (1-0.9999) * 0.9999} = 0.5$$

Incidence Rate



Incidence Rate

An increase of x% is a void statement in general

2. Signal Detection Theory

The Basic Scenario

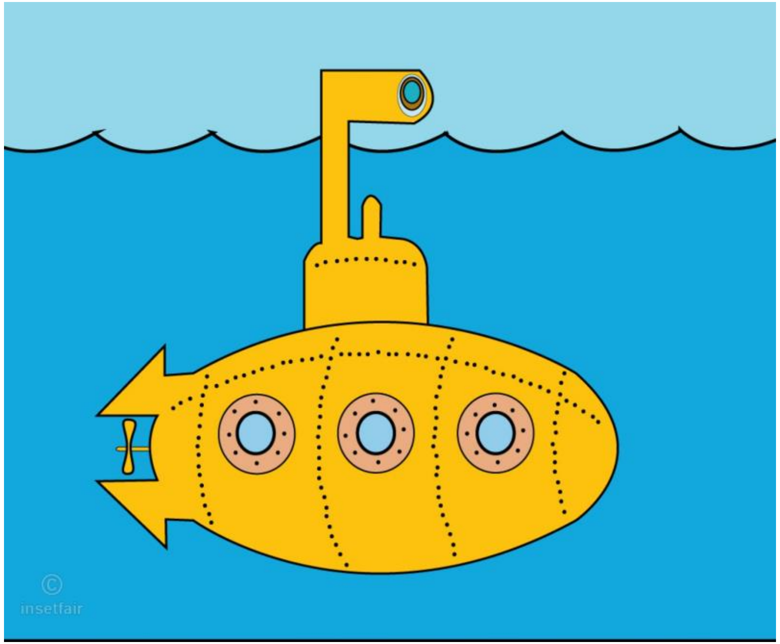
Terminology

EPFL

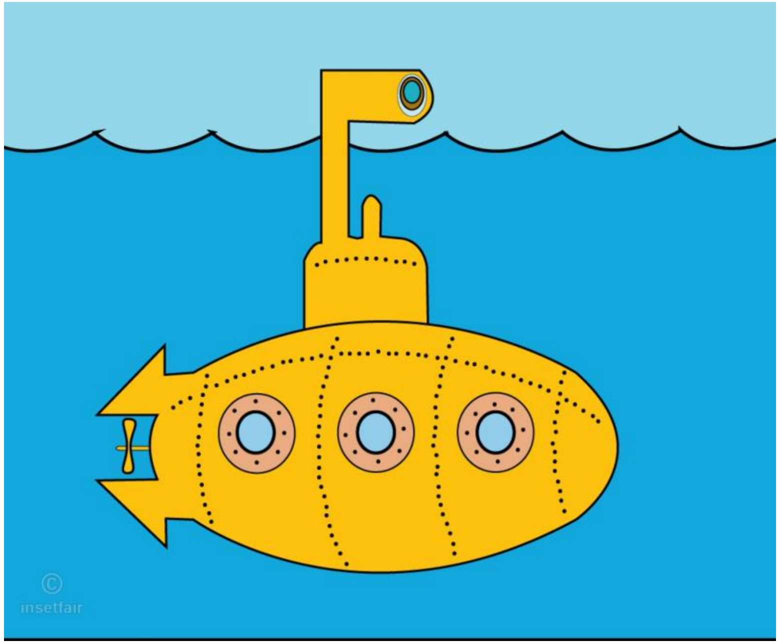
Experiment: is there a light patch (S) on the computer screen?

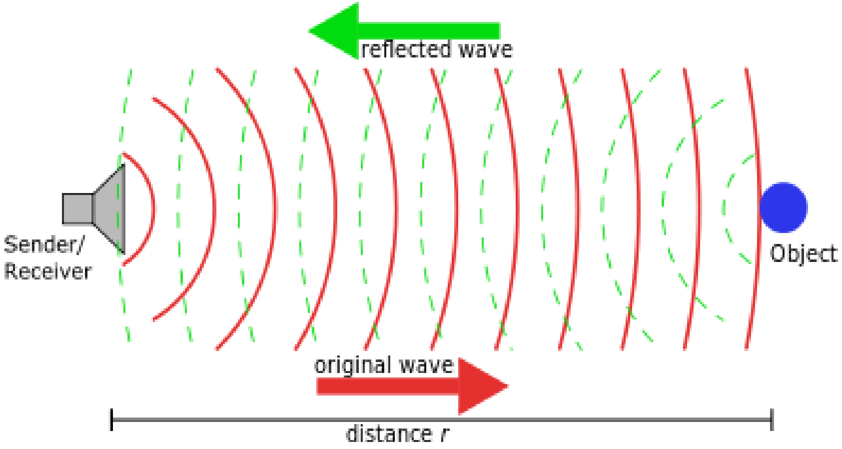
Response		Stimulus present	Stimulus absent
	Present	Hit	False Alarm (FA)
	Absent	Miss	Correct Rejection (CR)

Decisions: The submarine

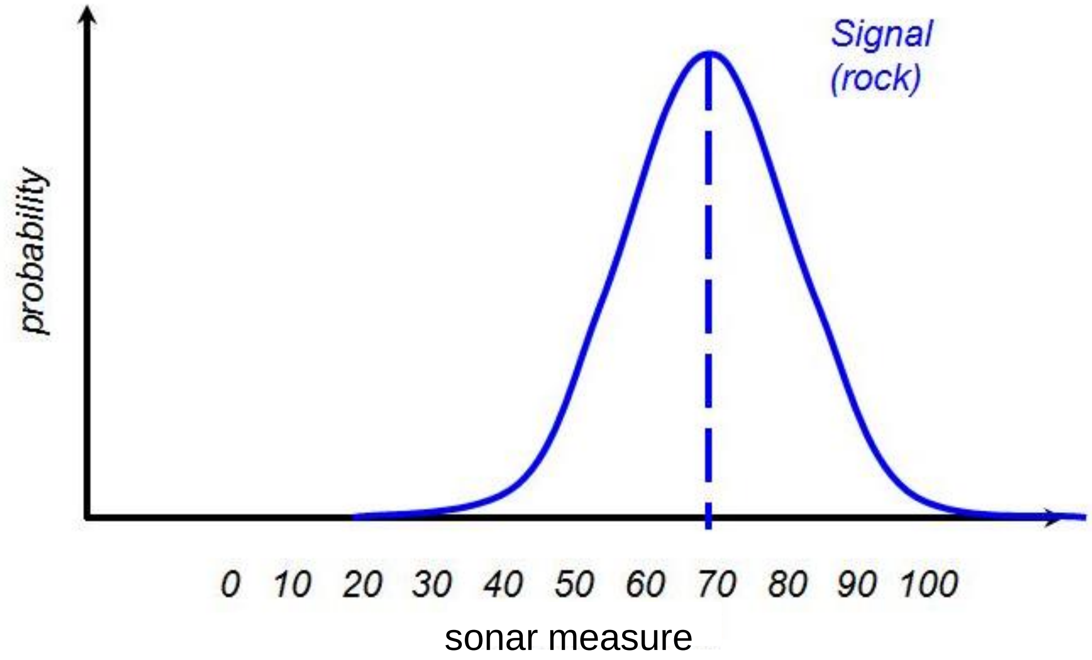


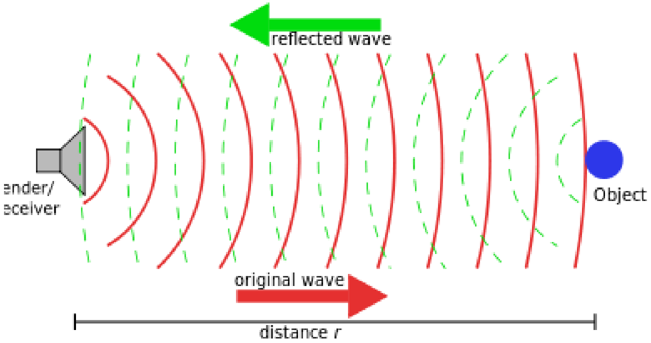
Decisions: The submarine ride



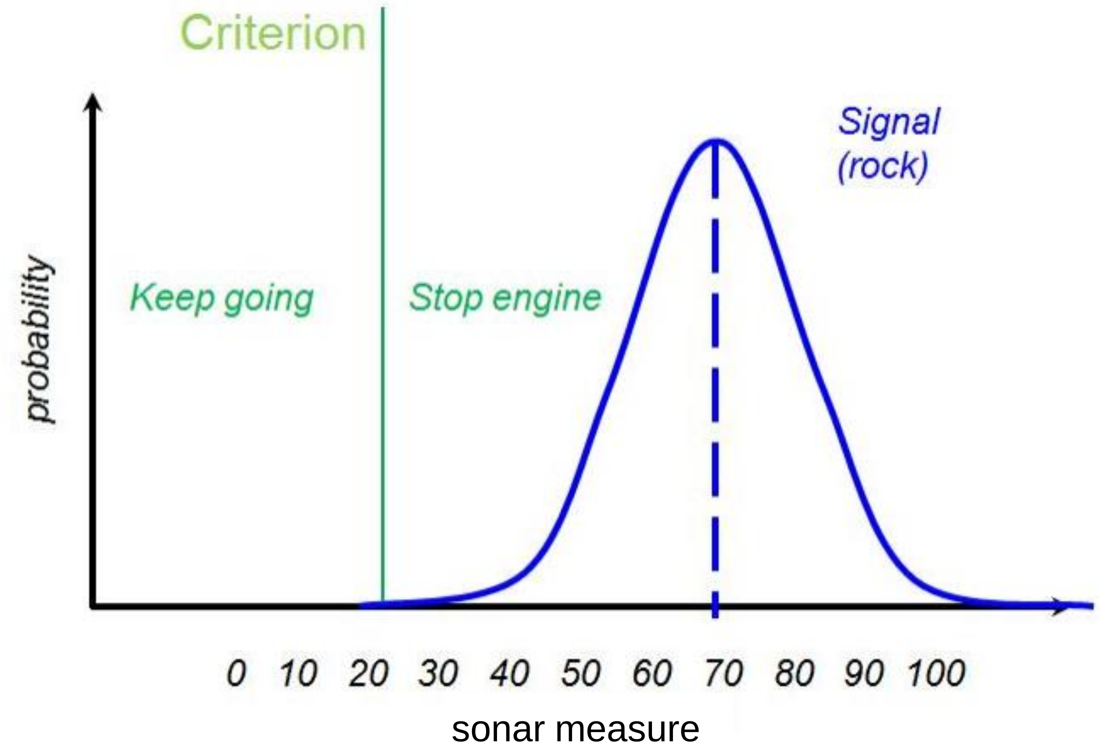


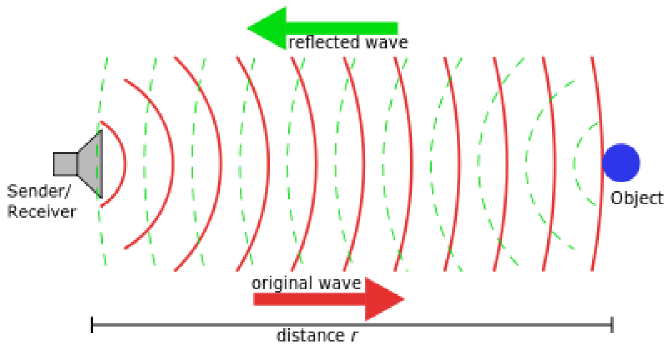
Conditional Probability: Rock



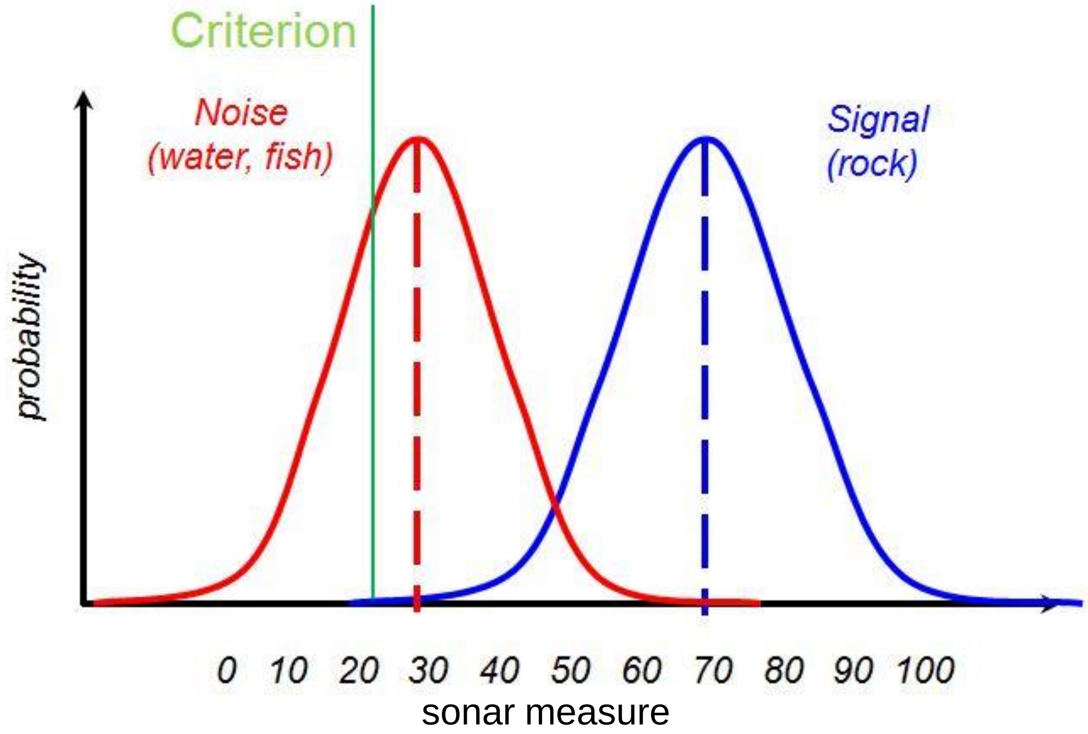


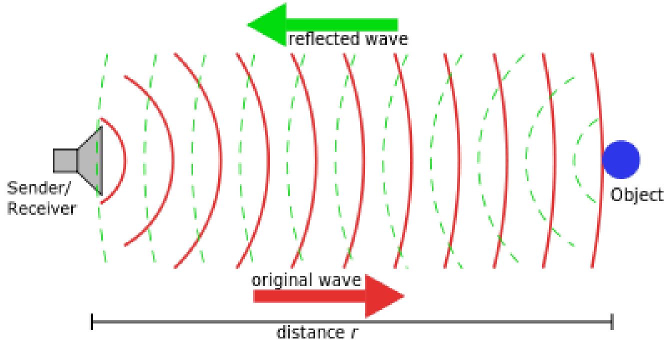
Conditional Probability: Rock



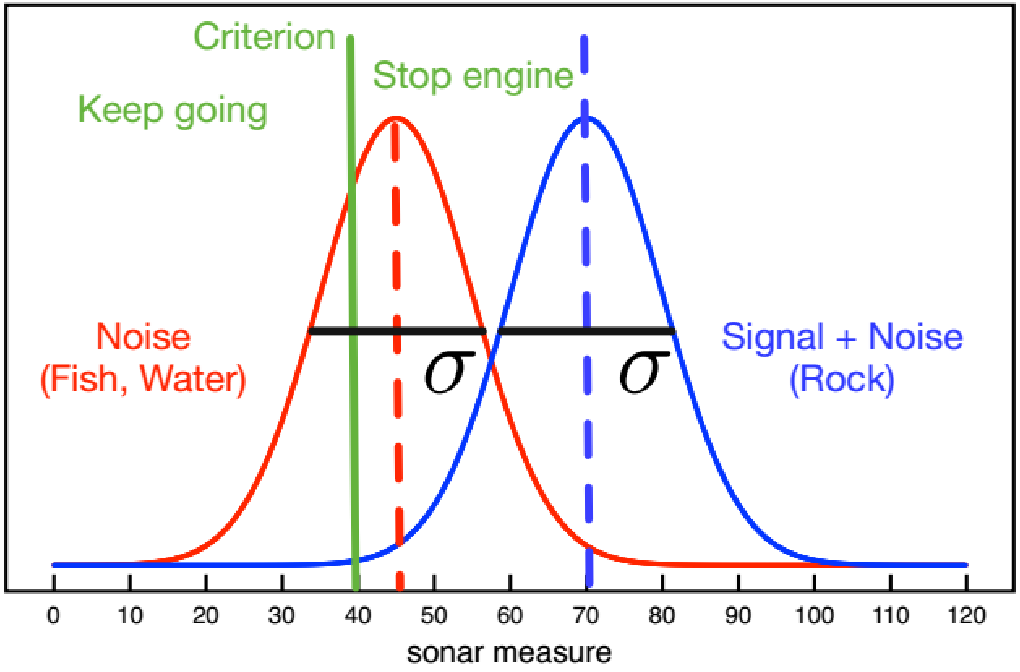


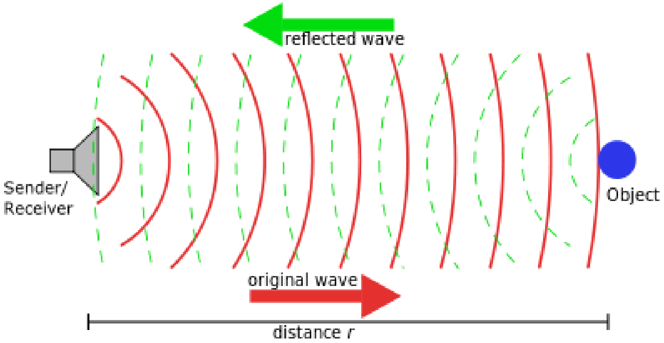
Conditional Probability: Noise (fish, water)





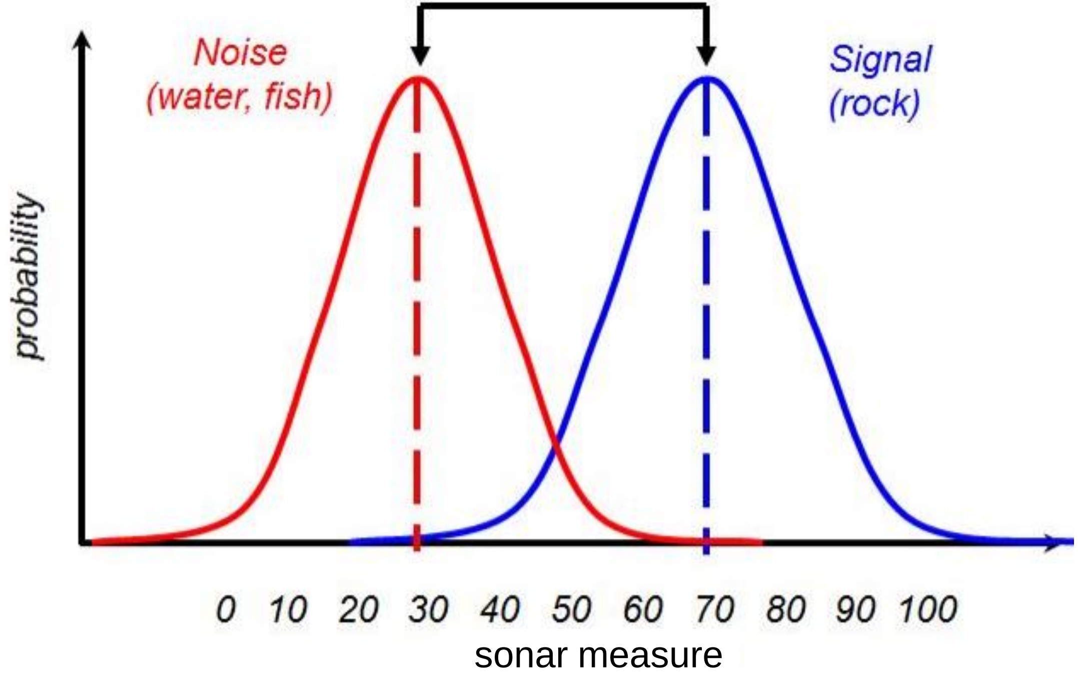
Conditional Probability: Noise (fish, water)

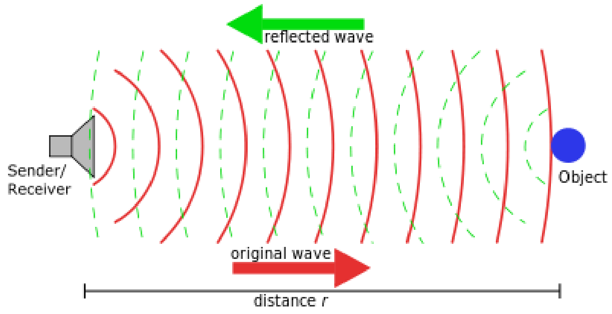




Conditional Probability: Noise (fish, water)

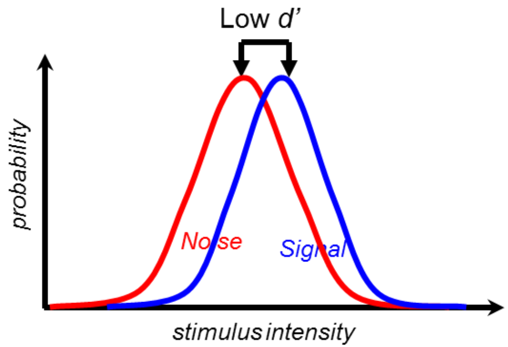
$$\text{Discriminability } (d') = \frac{\mu_S - \mu_N}{\sigma}$$

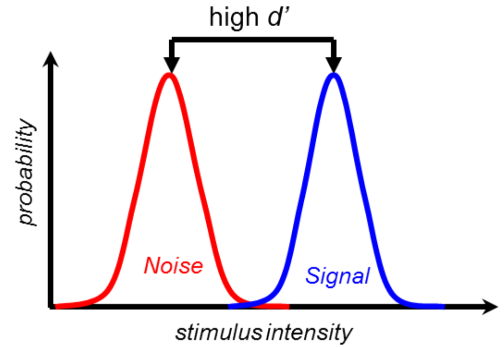




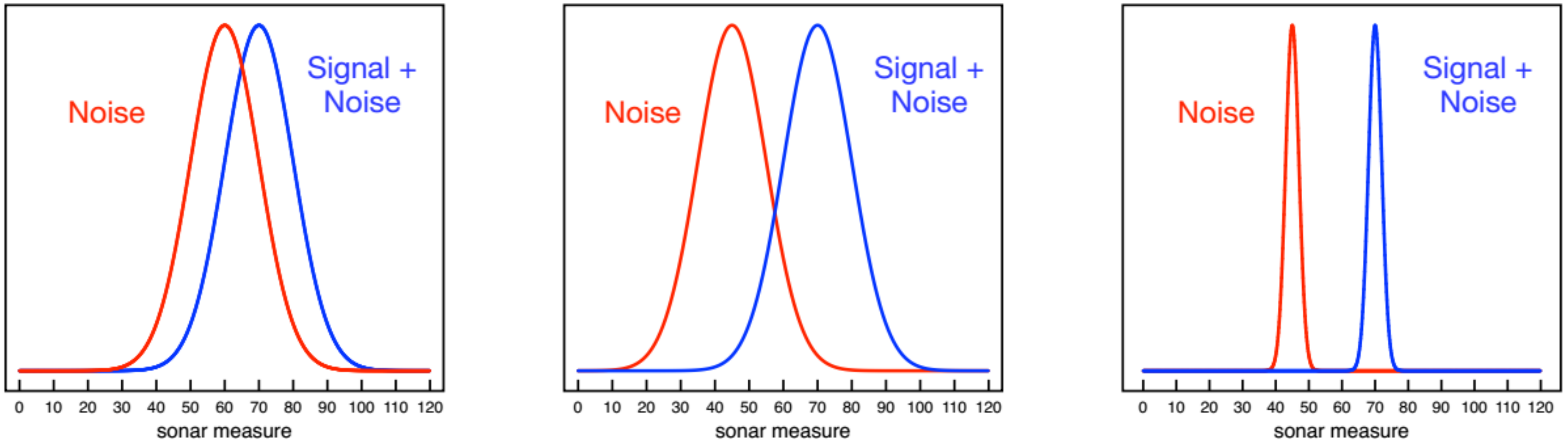
Interpreting d’

- If *d’* is low, discriminability is low
  - The noise and signal + noise distributions are highly overlapping
  - *d’* = 0: chance level
- If *d’* is high, discriminability is high
  - *d’* = 1: moderate performance
  - *d’* = 4.65: “optimal” (corresponds to hit rate=0.99, false alarm rate=0.01)



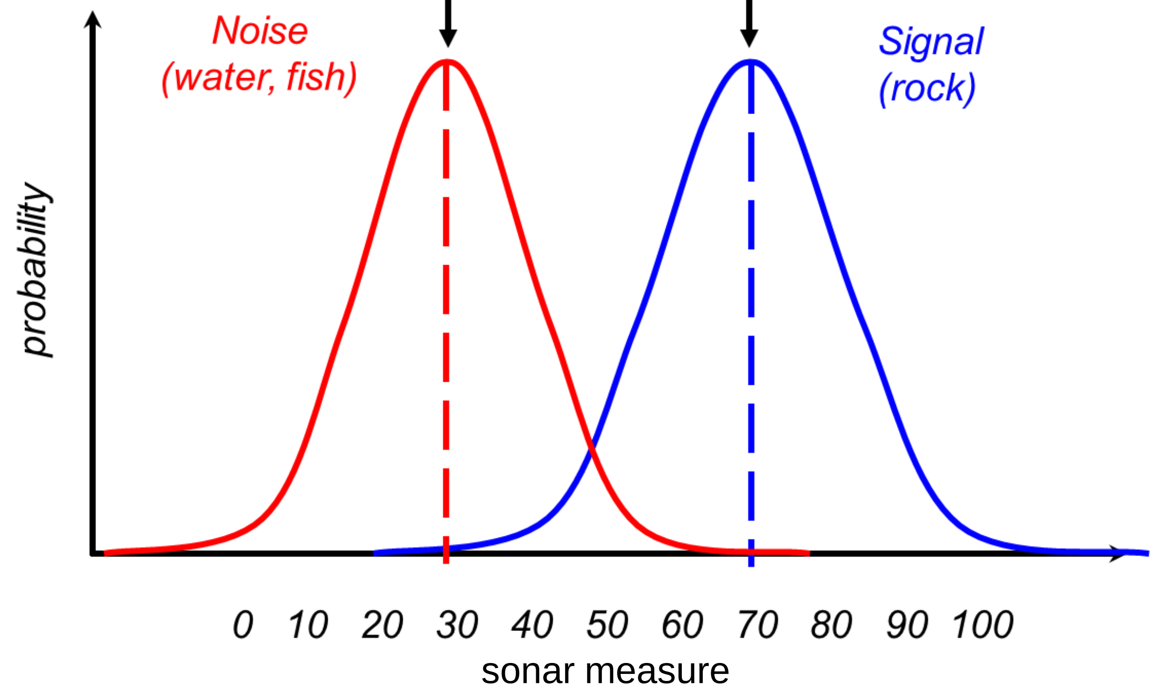


Interpreting d’



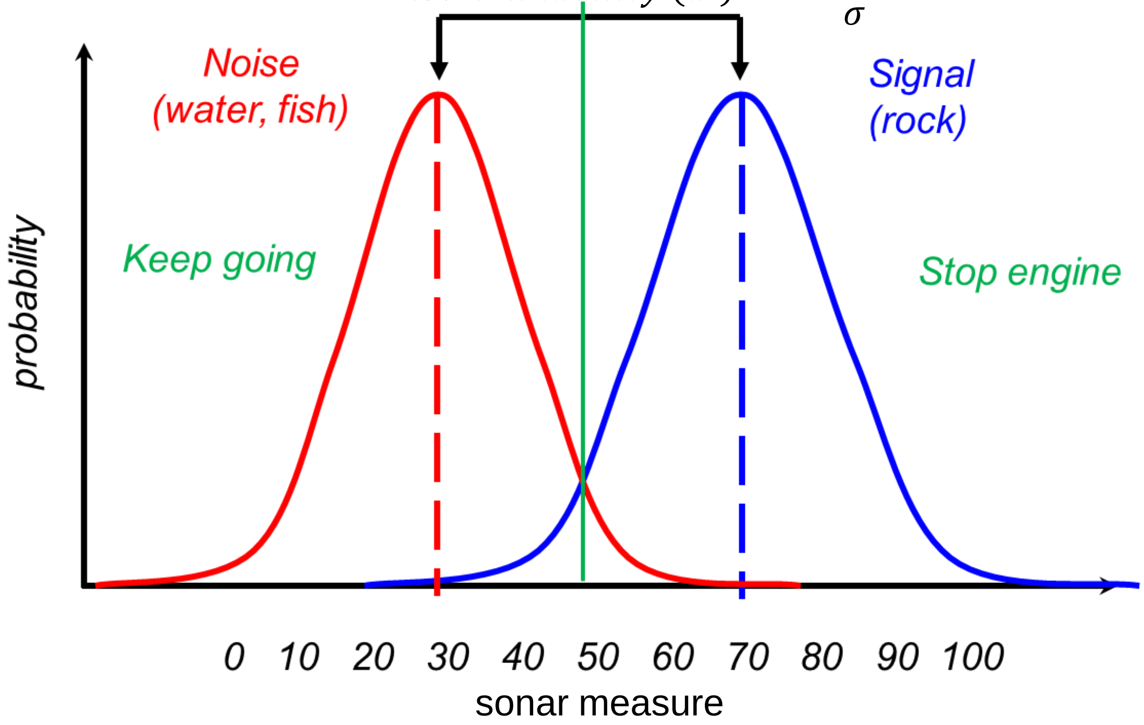
Discriminability is independent of Criterion

$$Discriminability\ (d') = \frac{\mu_S - \mu_N}{\sigma}$$



Discriminability is independent of Criterion

$$Discriminability\ (d') = \frac{\mu_S - \mu_N}{\sigma}$$

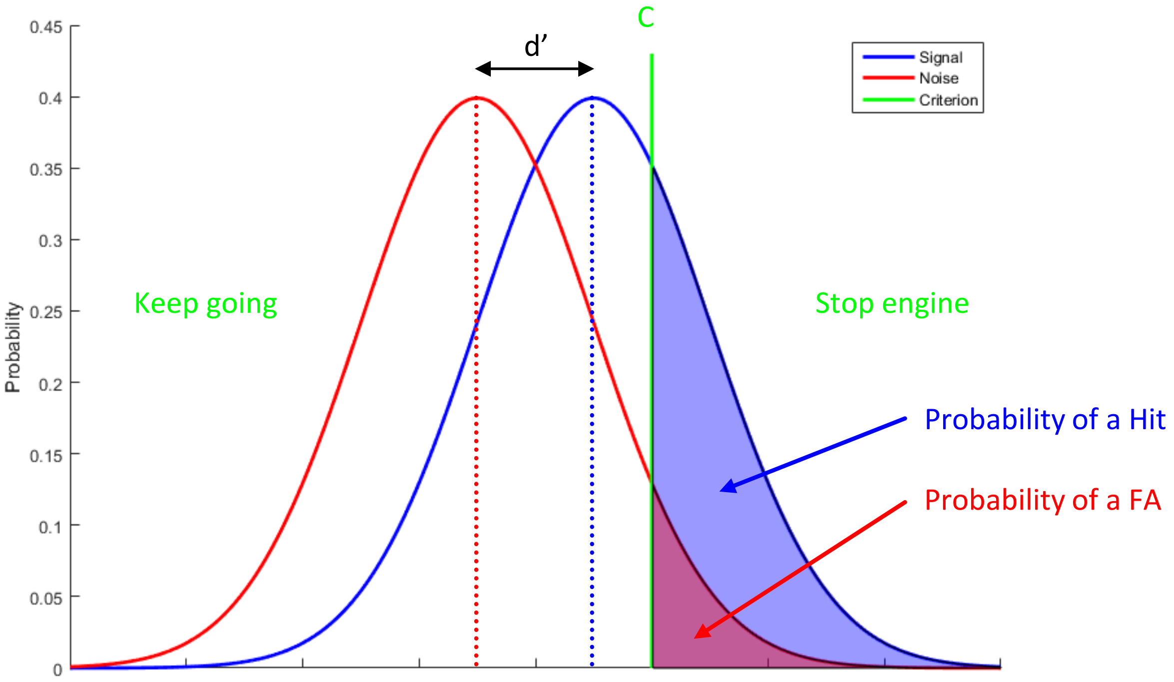


Lesson #1

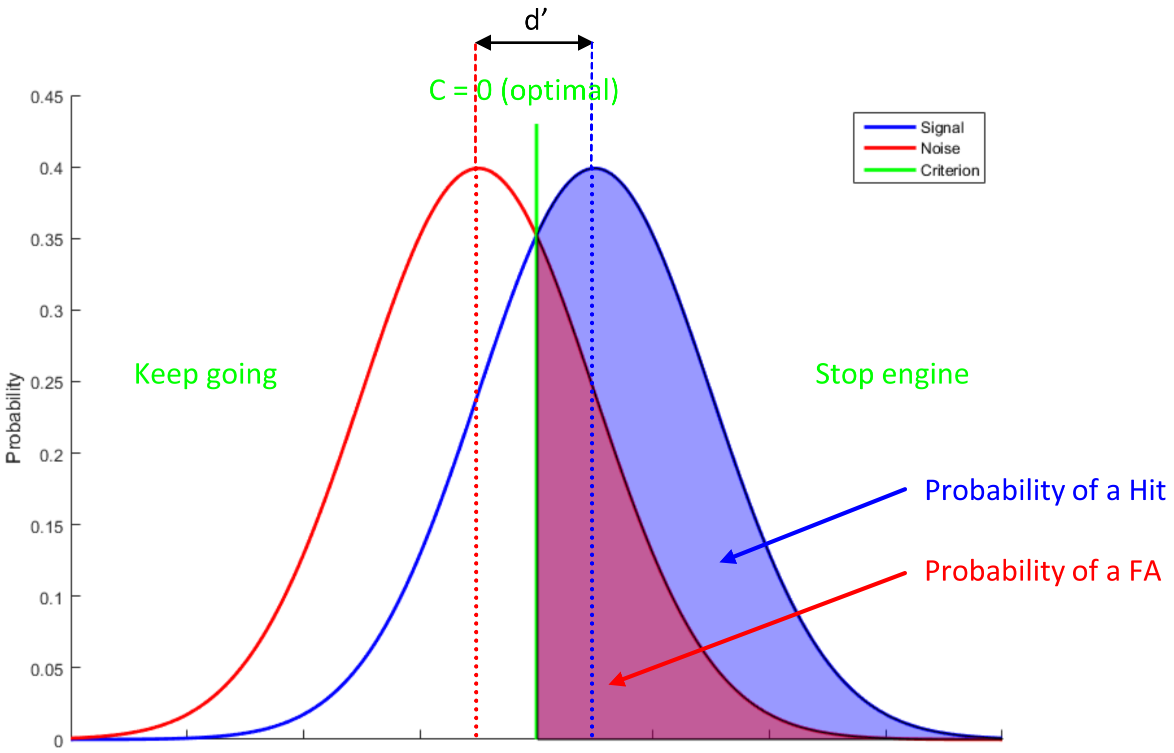
Decisions depend on both discriminability and criterion

SDT & Percentage of correct responses

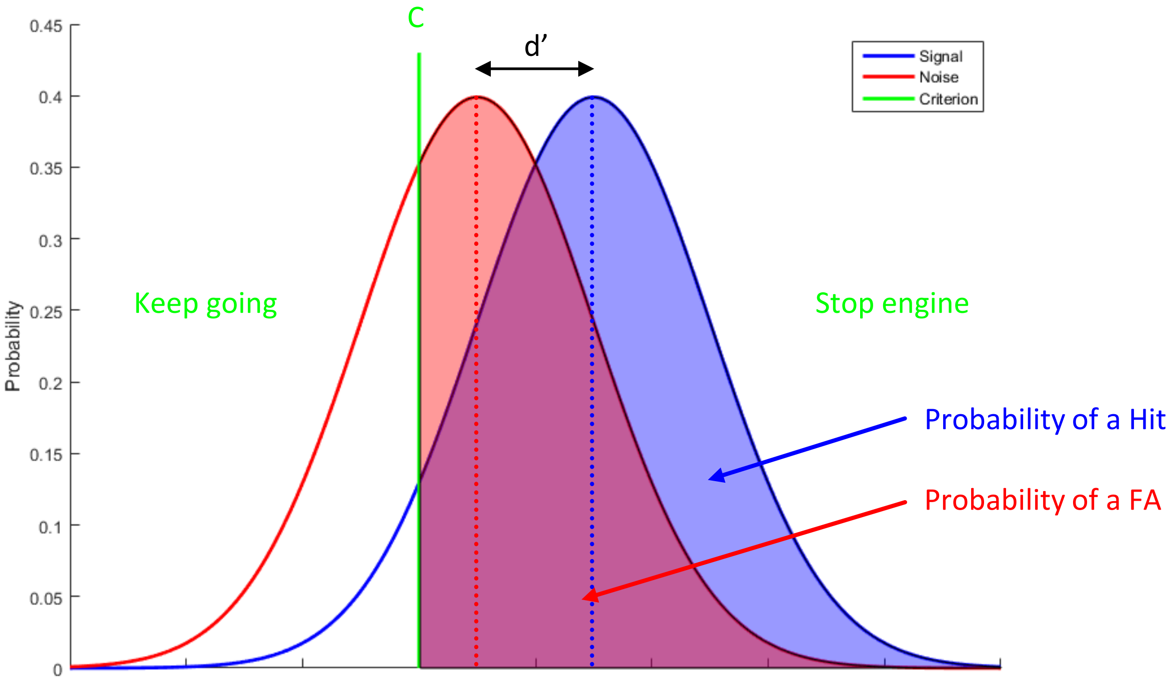
Discriminability & Criterion



Discriminability & Criterion



Discriminability & Criterion



Lesson #2

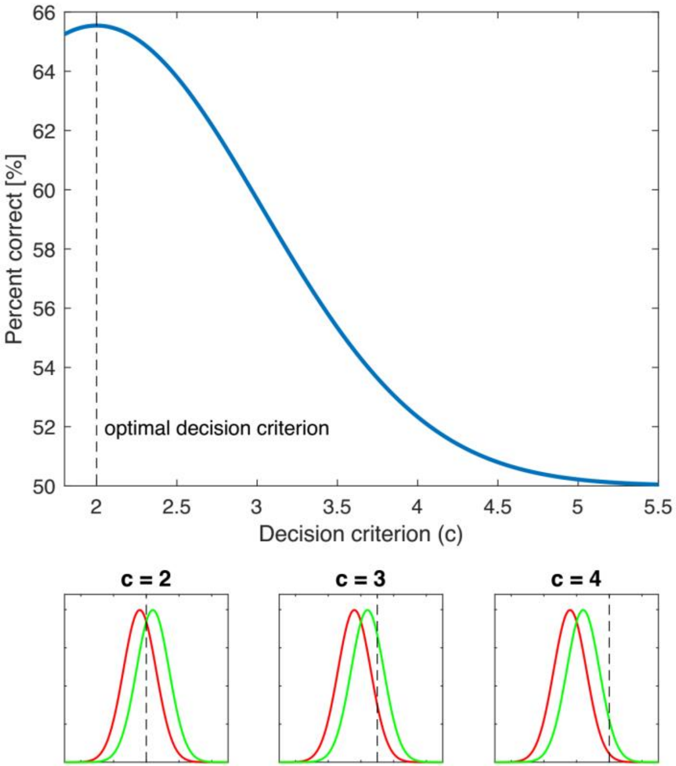
A change in criterion c leads to a response bias-
but not to a change in discriminability

Terminology

Percent correct: $\frac{Hit + CR}{2}$

Response	Stimulus present	Stimulus absent
Present	Hit	False Alarm (FA)
Absent	Miss	Correct Rejection (CR)

Percent correct (pc) depends heavily on c



Percent correct: $\dfrac{Hit + CR}{2}$

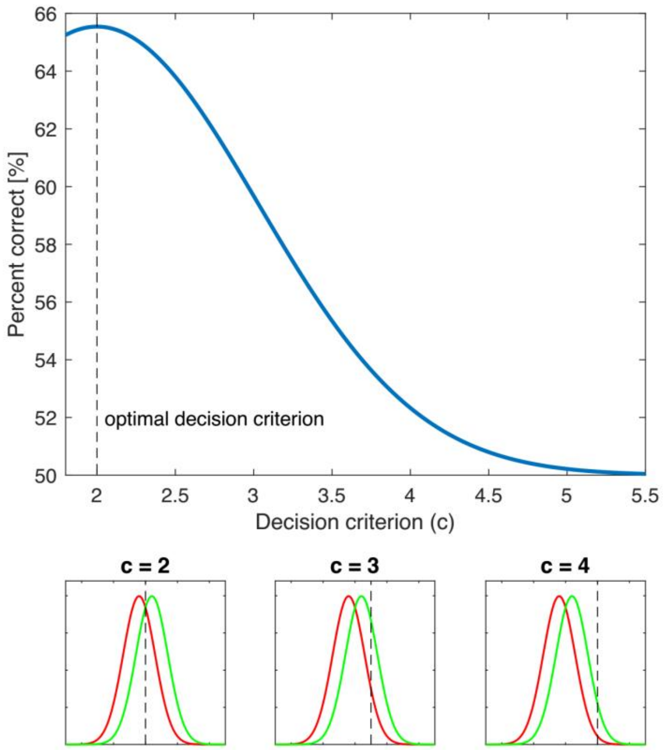
Lesson #3

EPFL

Decisions depend on both discriminability and criterion

Percent correct confounds the two: partial information!

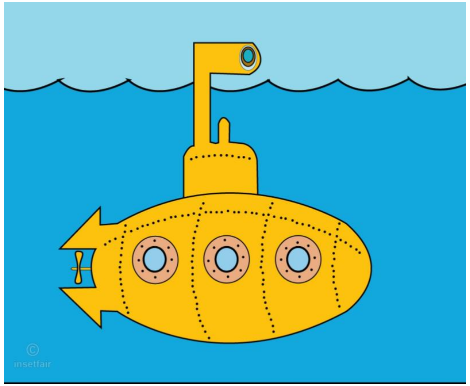
Percent correct depends heavily on c

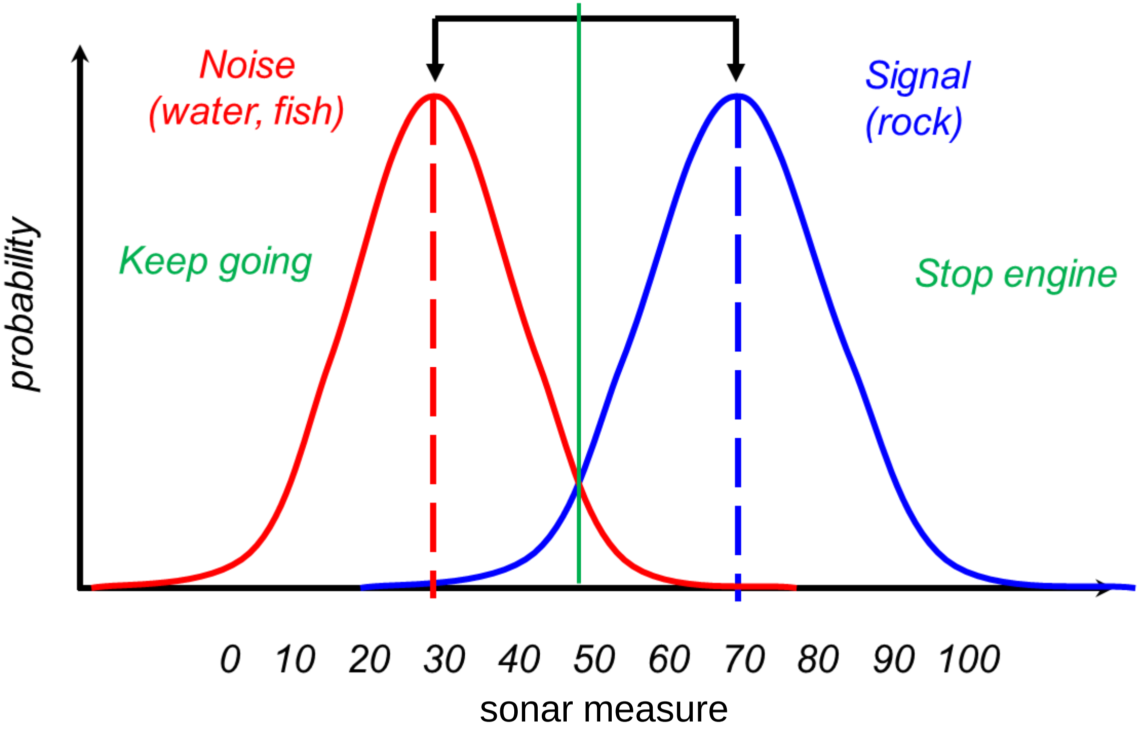


Aren't, then, all behavioral experiments hopeless?

This was about theory: Sensitivity & Criterion

Question: can we infer from a device or human d‘ and c?





The empirical d'

Estimation of d’

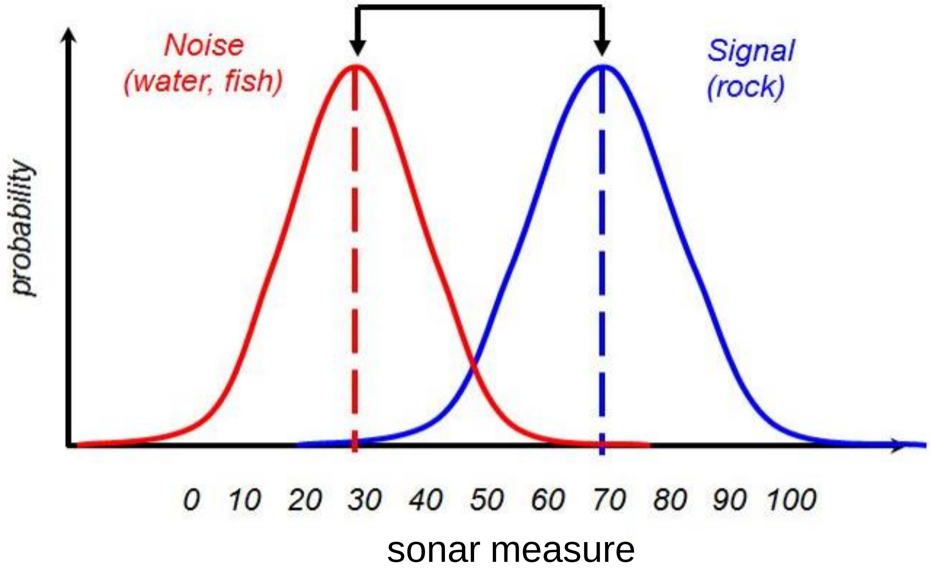
- ***Theoretical*** $(d') = \frac{\mu_S - \mu_N}{\sigma}$
- d'can be estimated from the **experimental** Hit and False Alarm rates: $d' = \mathrm{z(Hit)} - \mathrm{z(FA)}$

Detour

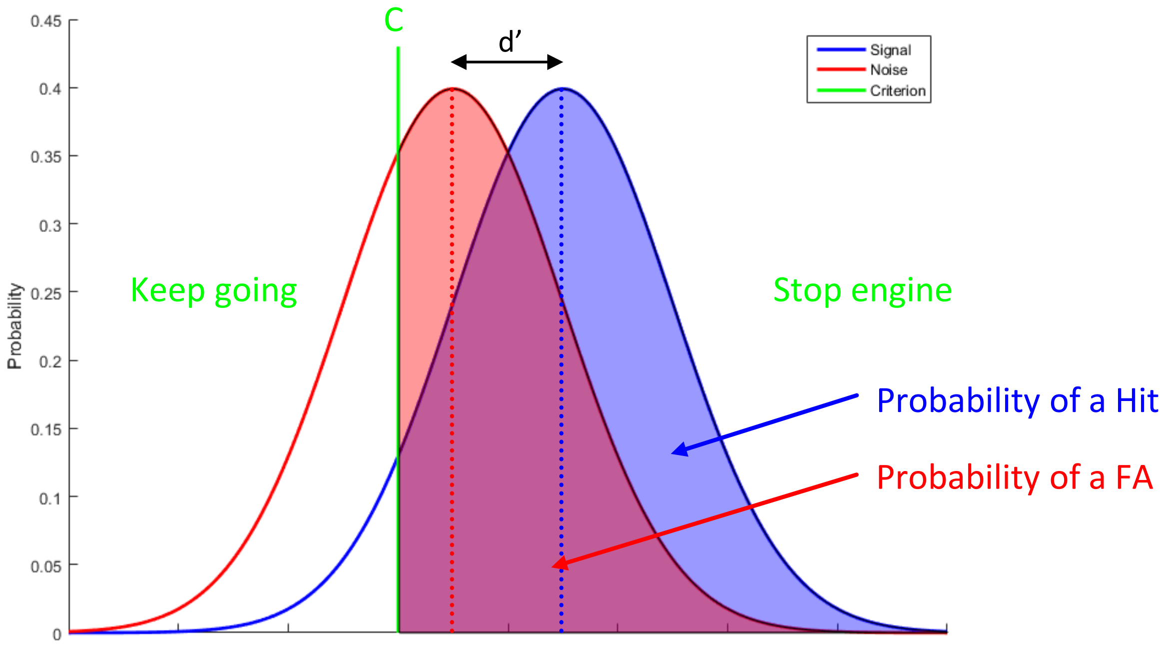
Detour

$$Discriminability\ (d') = \frac{\mu_S - \mu_N}{\sigma}$$

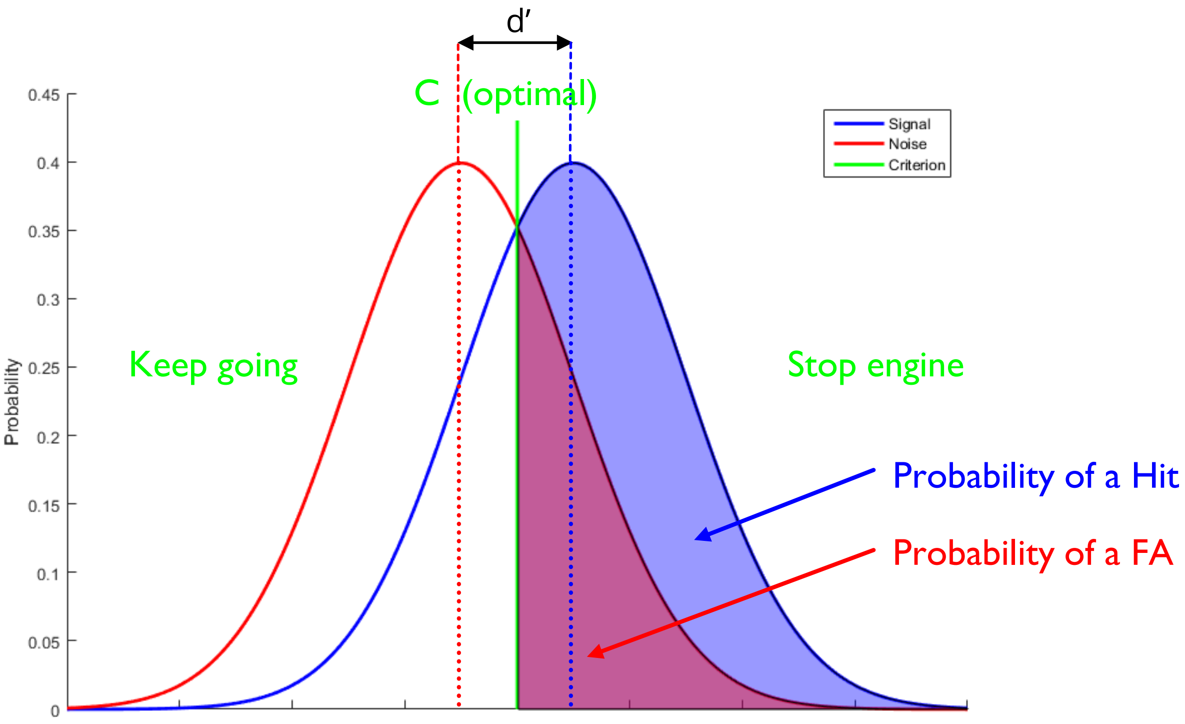
Assumptions:
1) Equal Variance
2) Gaussians



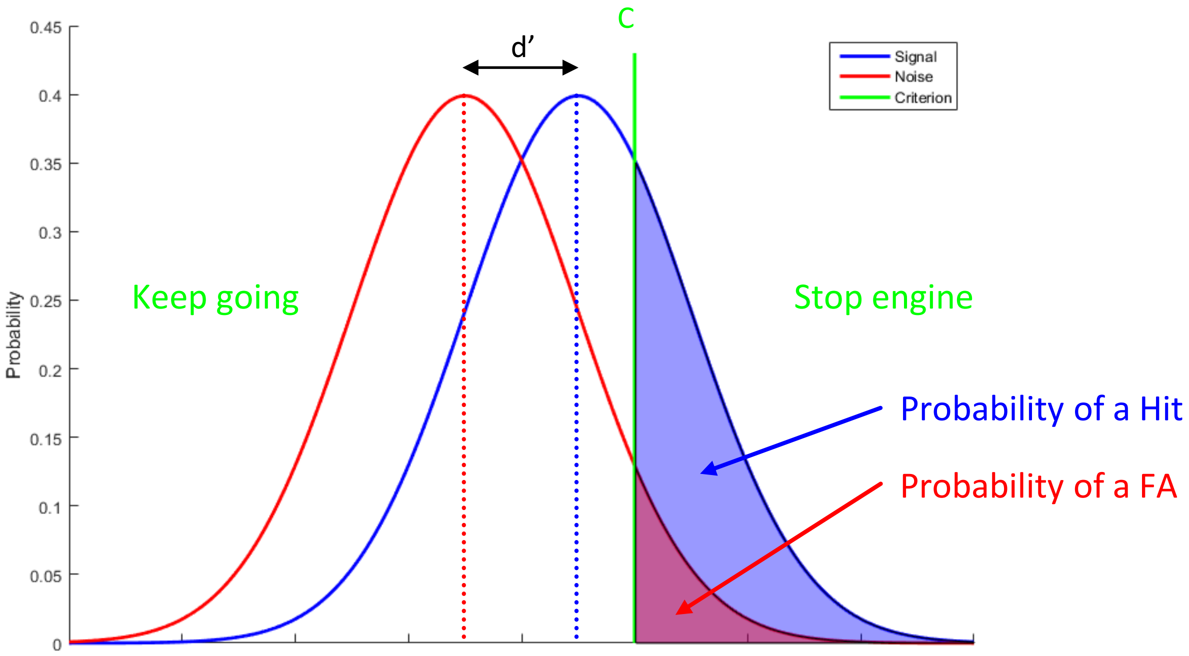
Discriminability & Criterion



Discriminability & Criterion

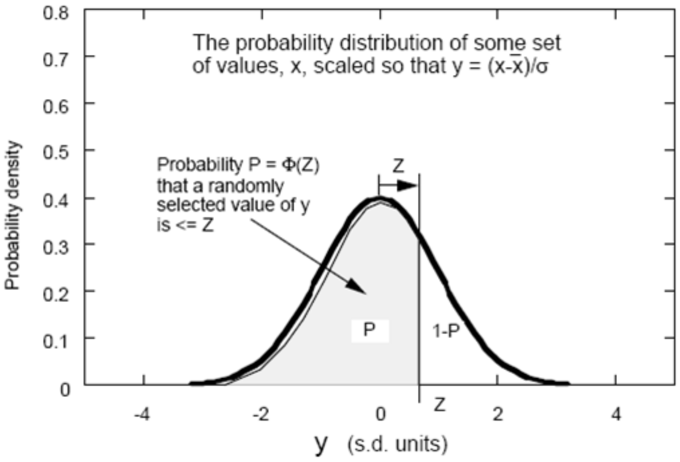


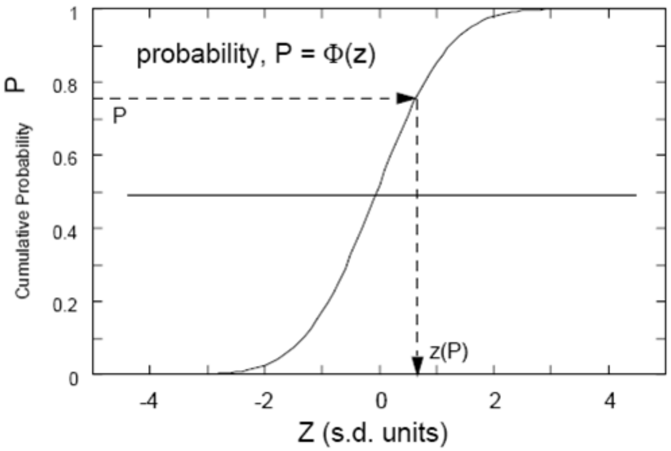
Discriminability & Criterion



Detour

Recap: Standard normal distribution function





$$\mathrm{FA} = \frac{1}{\sigma\sqrt{(2\pi)}} \int_c^\infty e^{-(x-\mu_n)^2/(2\sigma^2)} dx$$

$$\mathrm{HIT} = \frac{1}{\sigma\sqrt{(2\pi)}} \int_c^\infty e^{-(x-\mu_s)^2/(2\sigma^2)} dx$$

$$z(1-P) = -z(P)$$

Detour

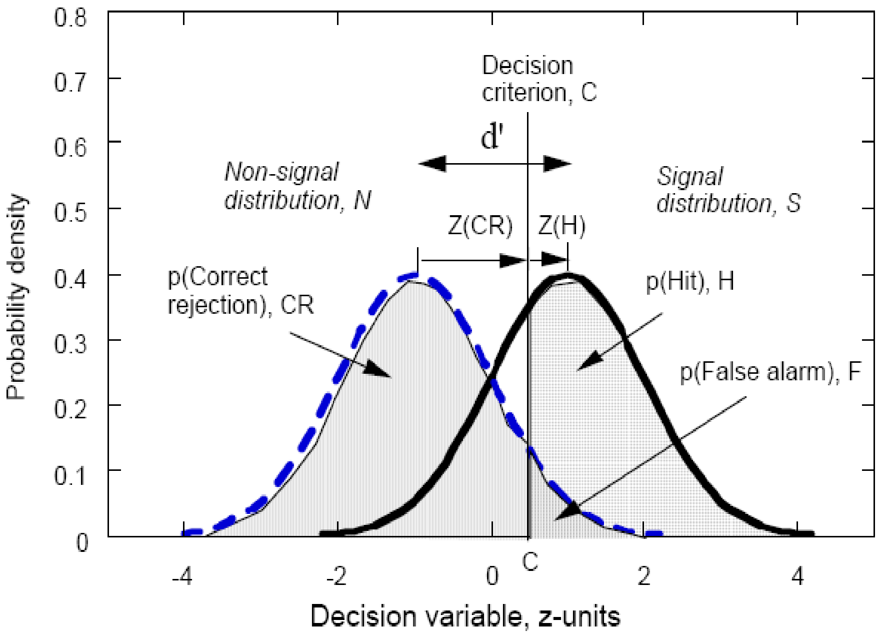
$$d' = z(CR) + z(Hit)$$

Since $CR + FA = 1$ hence $CR = 1 - FA$

and so $z(CR) = z(1 - FA)$

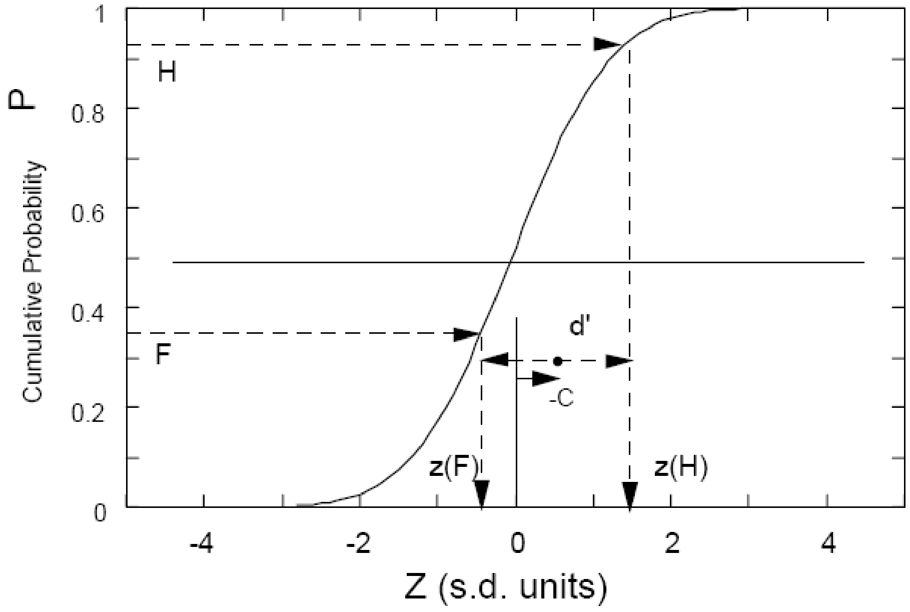
Since $z(1 - FA) = -z(FA)$, therefore
$z(CR) = -z(FA)$

Hence: $d' = z(Hit) - z(FA)$



Detour

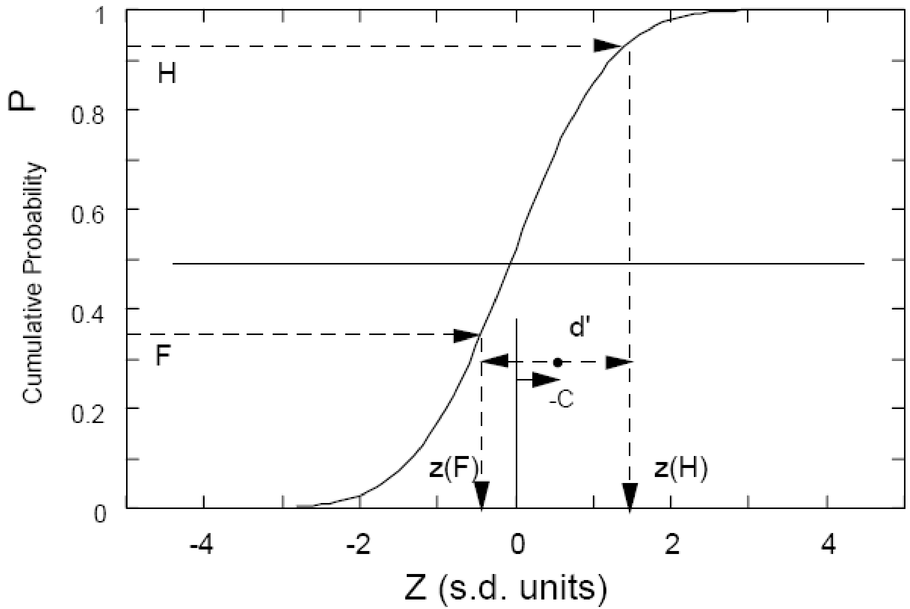
$\boldsymbol{d'}$ is independent of the criterion: **sensitivity**



Courtesy: Mark Georgeson

Detour

Understanding the criterion c in SDT



$$c_{emp} = -\frac{[z(H) + z(F)]}{2}$$

End of Detour

Estimation of d'

- ***Theoretical*** $(d') = \frac{\mu_S - \mu_N}{\sigma}$
- *d'* can be estimated from the **experimentally** Hit and the False Alarm rate:

$$\boldsymbol{d'_{emp} = z(Hit) - z(FA)}$$

- The criterion *bias* can be computed by

$$\boldsymbol{c_{emp} = -\frac{[z(H) + z(FA)]}{2}}$$

Examples

Example 1

Doctors' performance		
Signal	Present	Absent
Yes	80	20
No	20	80

Automated recognition		
Signal	Present	Absent
Yes	98	38
No	2	62

Example 1

Doctors' performance			Automated recognition		
Signal	**Present**	**Absent**	**Signal**	**Present**	**Absent**
Yes	80	20	Yes	98	38
No	20	80	No	2	62
	P	**Z**		**P**	**Z**
Hit	0.8	0.842	Hit	0.98	2.054
FA	0.2	-0.842	FA	0.38	-0.305
Sensitivity, d'	1.683		Sensitivity, d'	2.359	
Bias, b	0.000		Bias, b	-0.874	
P(correct)	0.800		P(correct)	0.800	

Example 2

Pergamon

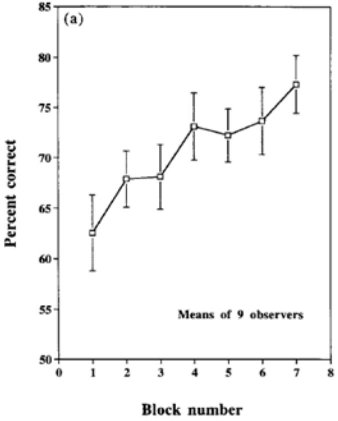
PII: S0042-6989(97)00043-6

Vision Res., Vol. 37, No. 15, pp. 2133–2141, 1997
© 1997 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0042-6989/97 $17.00 + 0.00

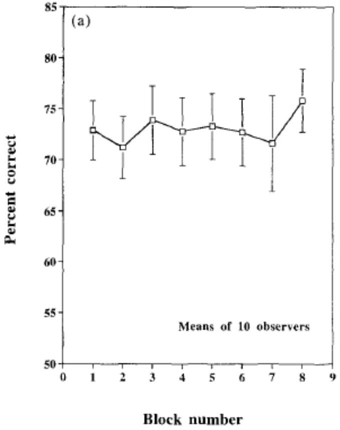
The Role of Feedback in Learning a Vernier Discrimination Task

MICHAEL H. HERZOG,* MANFRED FAHLE†‡

Received 30 May 1996; in revised form 21 January 1997



with Feedback

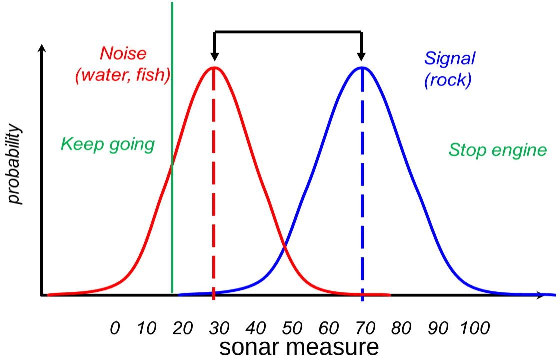


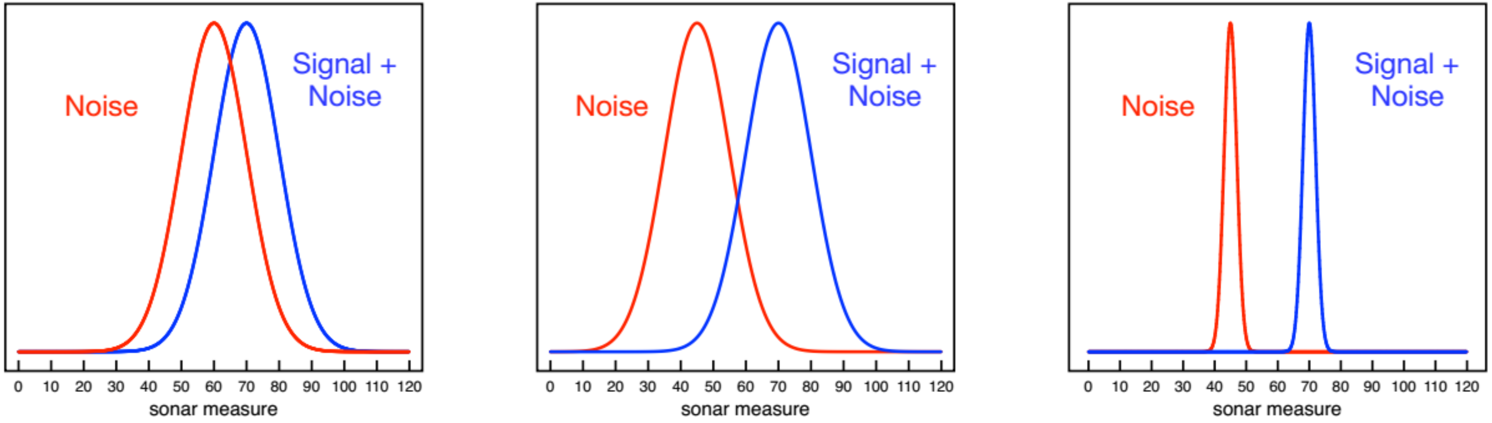
without Feedback

Example 2: what may change during learning

Improvement of performance- in terms of % correct:

- Optimize criterion
- Increase d' by "moving" means or decrease σ



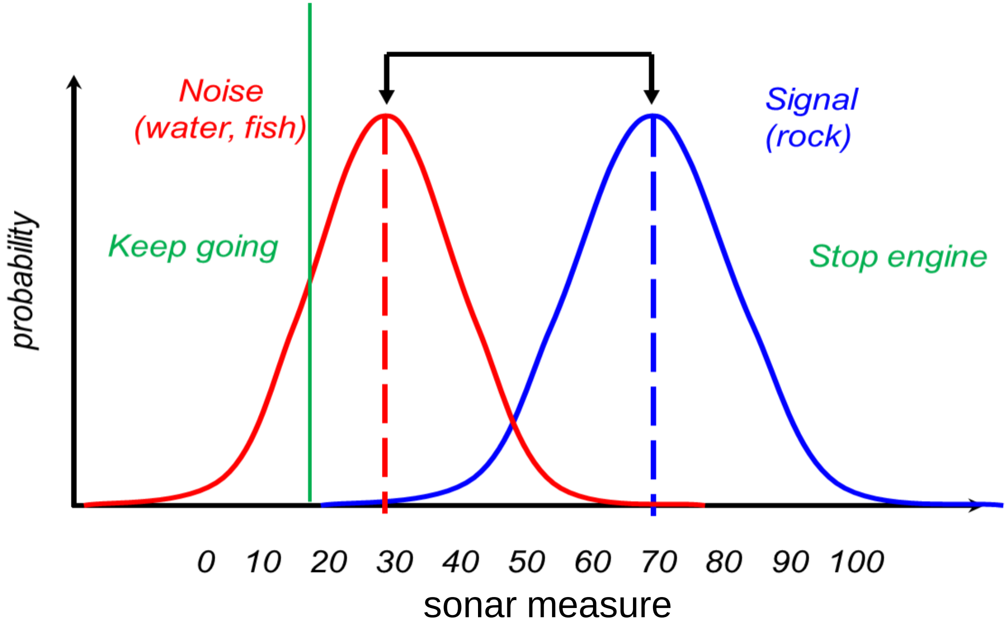


Example 3

The HIV test depends on a criterion and so do sensitivity and specificity in general

Example 3

The HIV test depends on a criterion and so do sensitivity and specificity in general



Assumptions

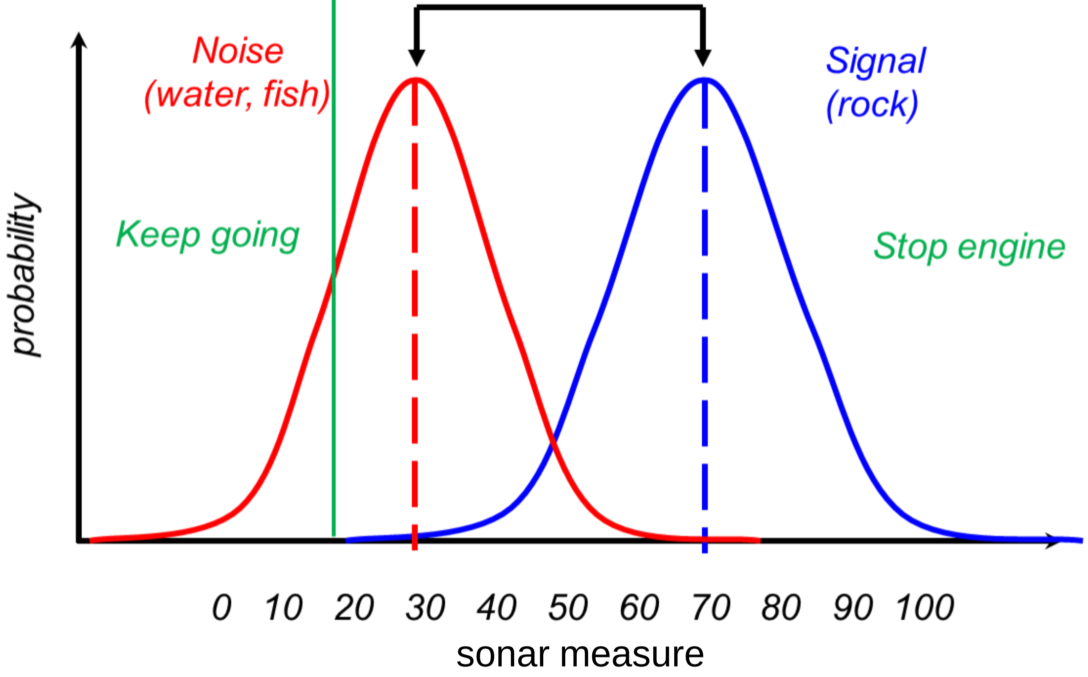
Assumptions: There is no free lunch

How reasonable are these assumptions?

- Gaussian functions
- Equal Variance $\sigma_n = \sigma_s$
- Criterion is constant

Basic Scenario

$$Discriminability\ (d') = \frac{\mu_S - \mu_N}{\sigma} = \mathrm{z(H)} - \mathrm{z(FA)}$$



Assumtions: There is no free lunch

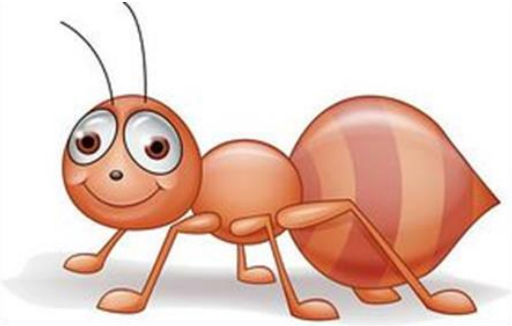
How reasonable are these assumptions?

- Gaussian functions
- Equal Variance: $\sigma_n = \sigma_s$
- Criterion is constant

Assumtions: There is no free lunch

How reasonable are these assumptions?

- Gaussian functions
- Equal Variance: $\sigma_n = \sigma_s$
- Criterion is constant


Terminology

Experiment: is there a light patch (S) on the computer screen?

Response	Stimulus present	Stimulus absent
Present	Hit	False Alarm (FA)
Absent	Miss	Correct Rejection (CR)

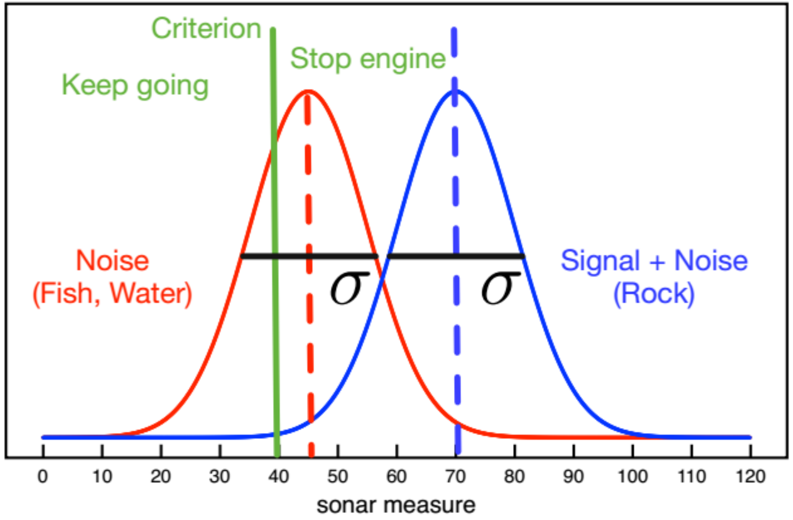
Attention

Sensitivity = Discriminability = Effect Size

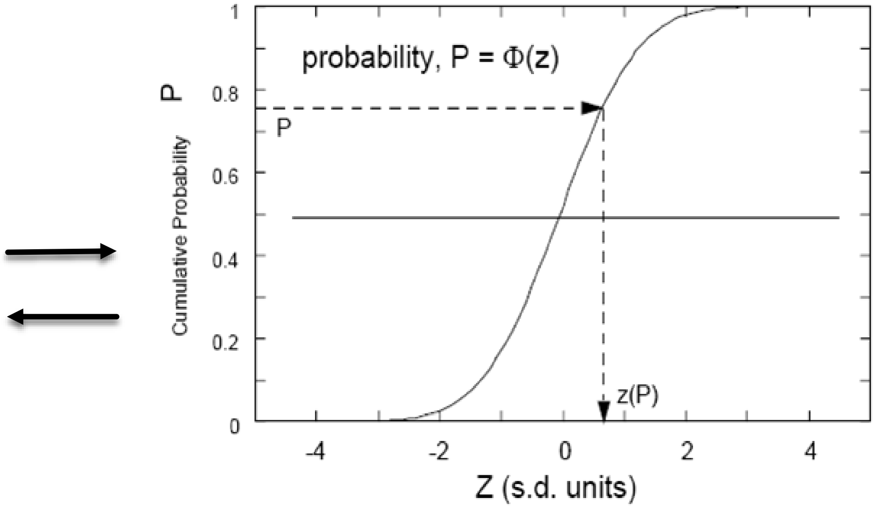
≠

Sensitivity = Hit Rate

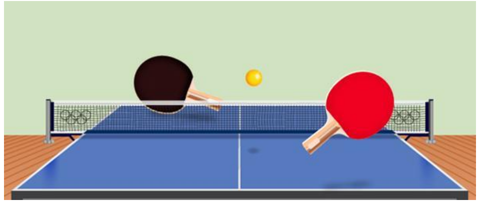
Ping-Pong



Criterion
Stop engine
Keep going
Noise
(Fish, Water)
σ
σ
Signal + Noise
(Rock)
0
10
20
30
40
50
60
70
80
90
100
110
120
sonar measure



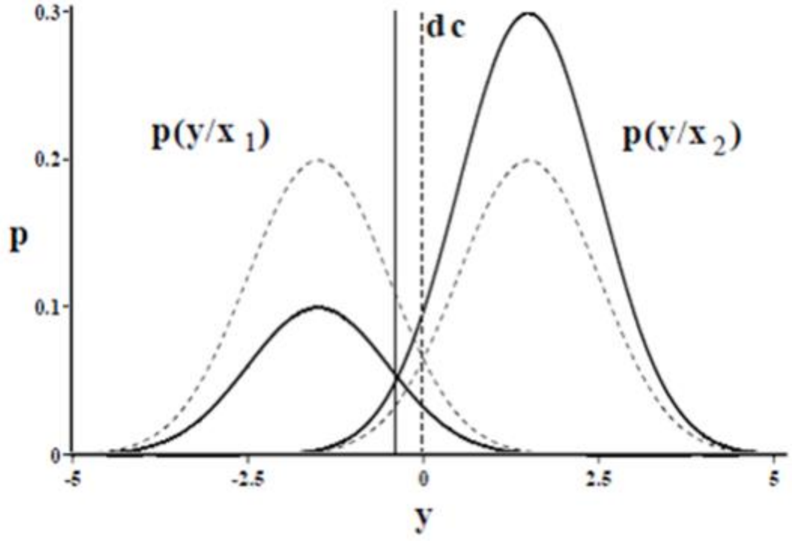
probability, P = Φ(z)
P
Cumulative Probability
1
0.8
0.6
0.4
0.2
0
P
z(P)
-4
-2
0
2
4
Z (s.d. units)



Extensions & Comments

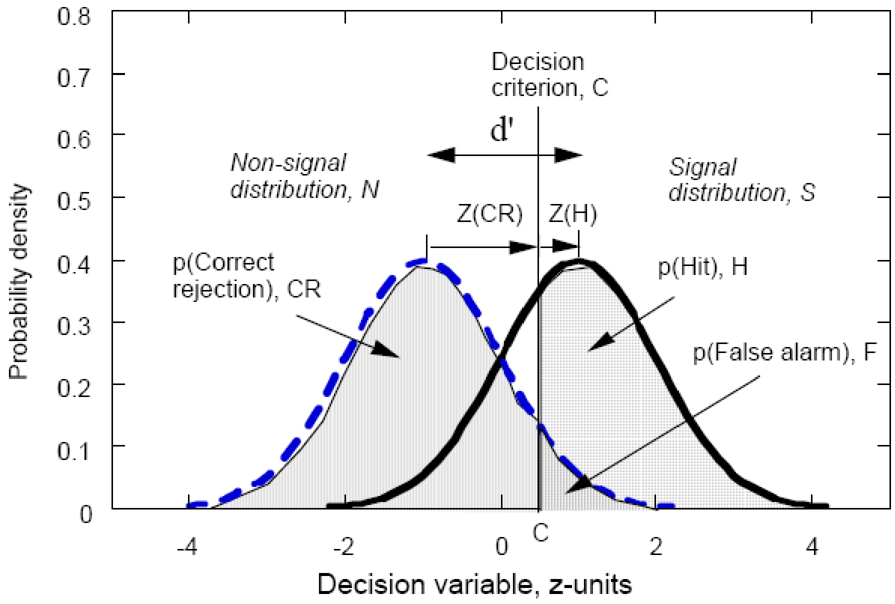
Probabilities

Up to now we assumed probabilites are the same



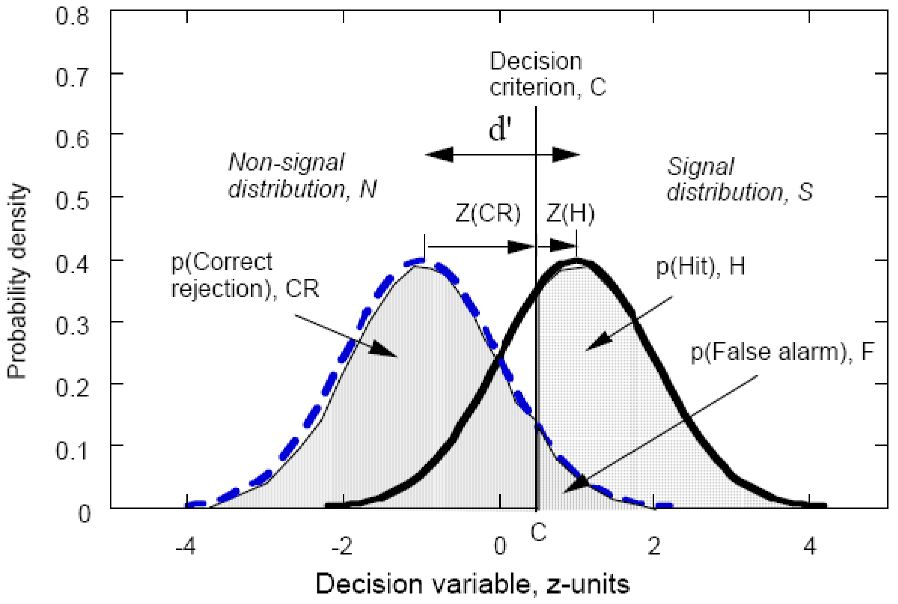
What is the optimal criterion?

Minimize FA and Misses, depending on incidence rate and costs



What is the optimal criterion?

Minimize: $p(S) * c(S) * Miss + p(N) * c(N) * FA$



Example 2

Pergamon

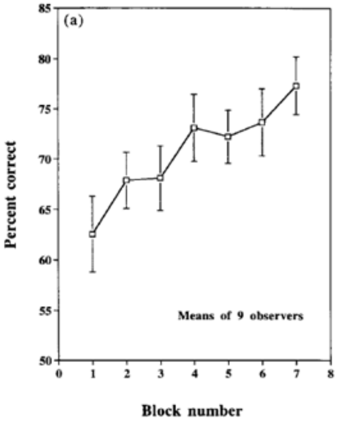
PII: S0042-6989(97)00043-6

Vision Res., Vol. 37, No. 15, pp. 2133–2141, 1997
© 1997 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0042-6989/97 $17.00 + 0.00

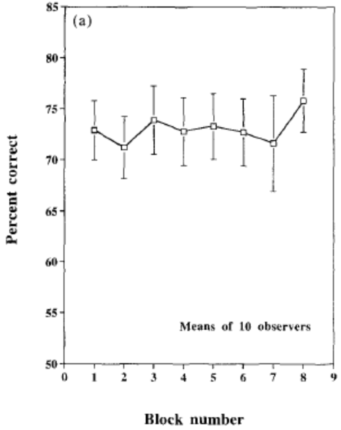
The Role of Feedback in Learning a Vernier Discrimination Task

MICHAEL H. HERZOG,* MANFRED FAHLE†‡

Received 30 May 1996; in revised form 21 January 1997

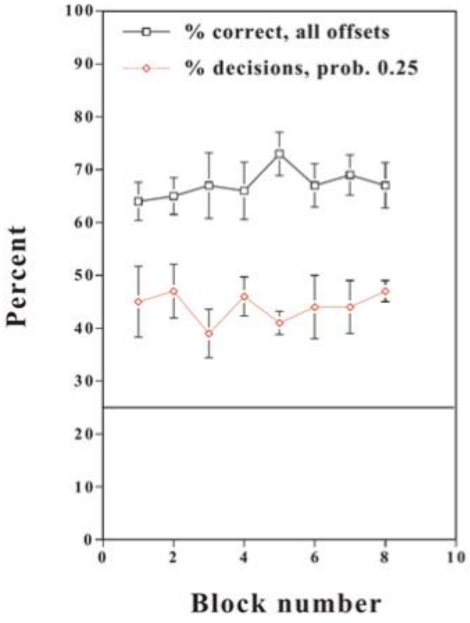


with Feedback

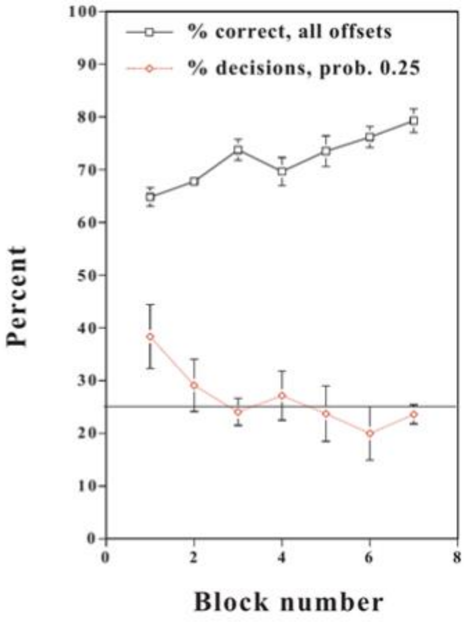


without Feedback

Example 2



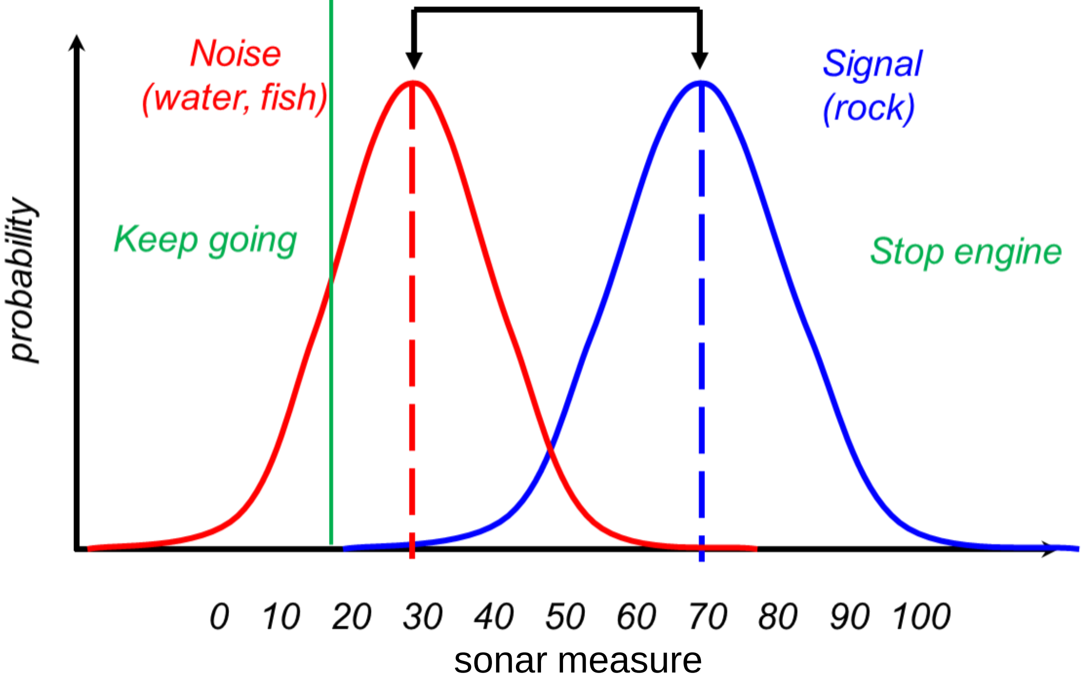
without Feedback



with Feedback

What to do if the Gaussian assumption is not met?

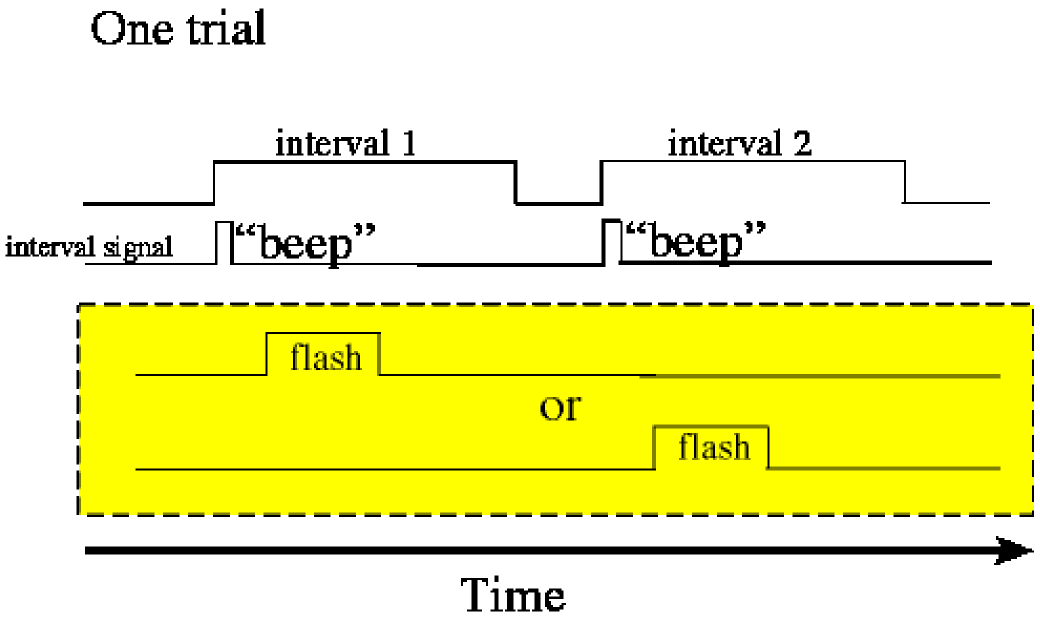
$$Discriminability\ (d') = \frac{\mu_S - \mu_N}{\sigma} = \mathrm{z(H)} - \mathrm{z(FA)}$$



ROC analysis

Otherwise you may compute a ROC analysis

2 Interval forced choice (IFC) Task & bias

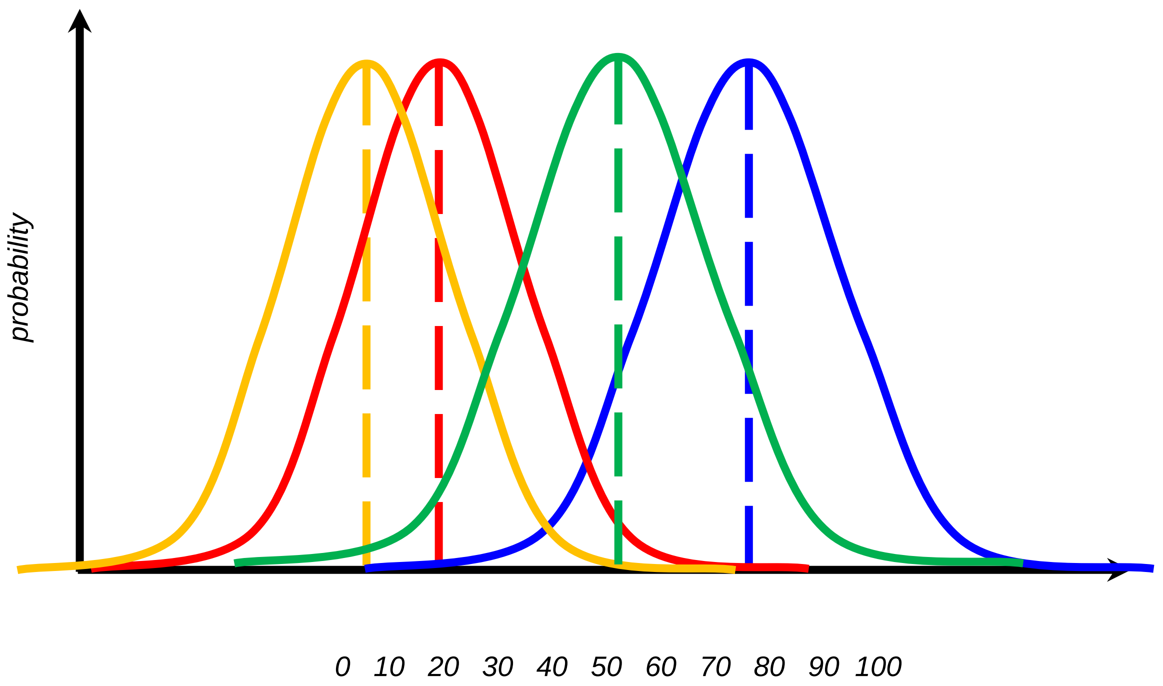


What about more alternatives?

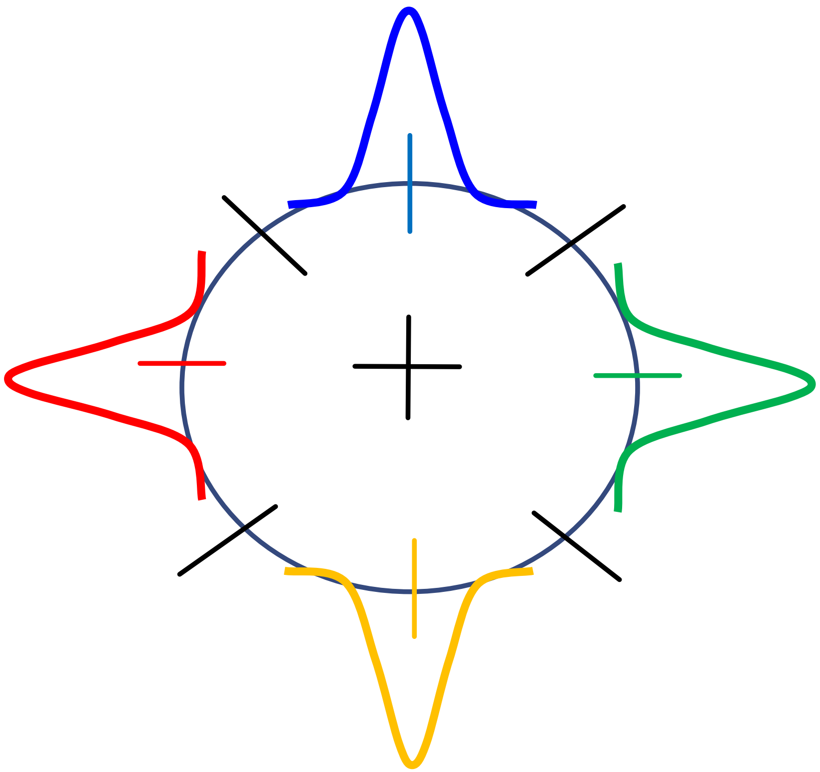
50% chance level is a waste of resources

Why not using more stimulus alternatives?

Discriminability



Experiment with 4 stimulus alternatives



Lesson #4

Keep your design simple

Speed accuracy trade off

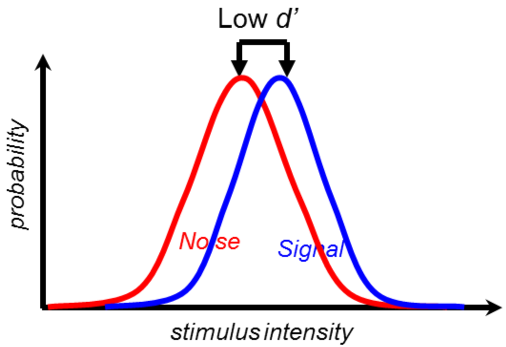
Elderly often perform slower than younger subjects

However, they are more accurate

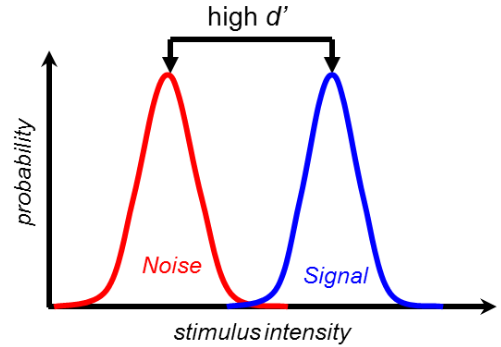
Often, the decision space is high dimensional and the decision criterion a hyperplane, e.g., accuracy and reaction times
Hence, good knowledge about the experimental situation is needed

Floor and Ceiling effects

- If *d’* is low, discriminability is low.
    - The noise and signal + noise distributions are highly overlapping.
    - *d’* = 0: chance level



- If *d’* is high, discriminability is high.
    - *d’* = 1: moderate performance
    - *d’* = 4.65: “optimal” (corresponds to hit rate=0.99, false alarm rate=0.01)



d’ is all over the place

Statistic	Description
Cohen’s d or Hedges’ g	Estimated standardized effect size
t	Test statistic
p	p-value for a two-tailed t test
d_{95}(lower) or g_{95}(lower)	Lower limit of a 95% confidence interval for d or g
d_{95}(upper) or g_{95}(upper)	Upper limit of a 95% confidence interval for d or g
Post hoc power from d or g	Estimated power for experiments with the same sample size
Post hoc v	Proportion of times OLS is more accurate than RLS
Λ	Log likelihood ratio for null and alternative models
ΔAIC, $\Delta AICc$	Difference in AIC for null and alternative models
ΔBIC	Difference in BIC for null and alternative models
JZS BF	Bayes Factor based on the Jeffreys-Zellner-Siow prior

Table 1: For known sample sizes of a independent two-sample t-test, each these terms is an equivalent sufficient statistic for the standardized effect size of the population. For given sample sizes, it is possible to transform any value to all the others. OLS – ordinary least square; RLS – randomized least squares.

Question 5 Which one is the correct equation for the empirical d' as used in SDT?

- $(\mu_1 - \mu_2)/\sigma$
- $(z(\mu_1) - z(\mu_2))/\sigma$
- $z(Hit) - z(FA)$
- $-[z(Hit) + z(FA)]/2$
- None of them
- $z(Hit) - z(CR)$
- $(Hit + CR)/2$

Question 5 Which one is the correct equation for the empirical d' as used in SDT?

- [ ] $(\mu_1 - \mu_2)/\sigma$
- [ ] $(z(\mu_1) - z(\mu_2))/\sigma$
- [x] $z(Hit) - z(FA)$
- [ ] $-[z(Hit) + z(FA)]/2$
- [ ] None of them
- [ ] $z(Hit) - z(CR)$
- [ ] $(Hit + CR)/2$

Exam

Question 22 Which of the following statements about the criterion in SDT is (are) correct?

- It depends on the incidence rate
- The criterion may reflect the response strategy of participants
- None of them
- The criterion changes when lowering the standard deviation
- If sensitivity and specificity are equal, the criterion is optimal

Exam

Question 22 Which of the following statements about the criterion in SDT is (are) correct?

- [ ] It depends on the incidence rate
- [x] The criterion may reflect the response strategy of participants
- [ ] None of them
- [ ] The criterion changes when lowering the standard deviation
- [x] If sensitivity and specificity are equal, the criterion is optimal

Take home messages

EPFL

Take Home Messages

1. Be aware of partial information. The percentage of correct responses confounds discriminability d' and decision criterion c.
2. Be aware of partial information. The same is true for many other measures such as Sensitivity and Specificity in medical tests.
3. You can disentangle discriminability and criterion by using d'_{emp}.
4. d'_{emp} is criterion free but not model free. There is no free lunch.

END Class 2