Quantitative Analysis of Situational Awareness

(QUASA)

Applying Signal Detection Theory to True/False Probes and Self-Ratings

Barry McGuinness
Principal Scientist
Human Factors Dept
Advanced Technology Centre
BAE Systems
Bristol, UK

Overview

1. Situational Awareness (SA)
2. Assessing SA
3. QUASA Approach
4. Signal Detection Theory
5. Calibration of SA
6. Example: LOE 2 data
7. Further Developments

Situational Awareness

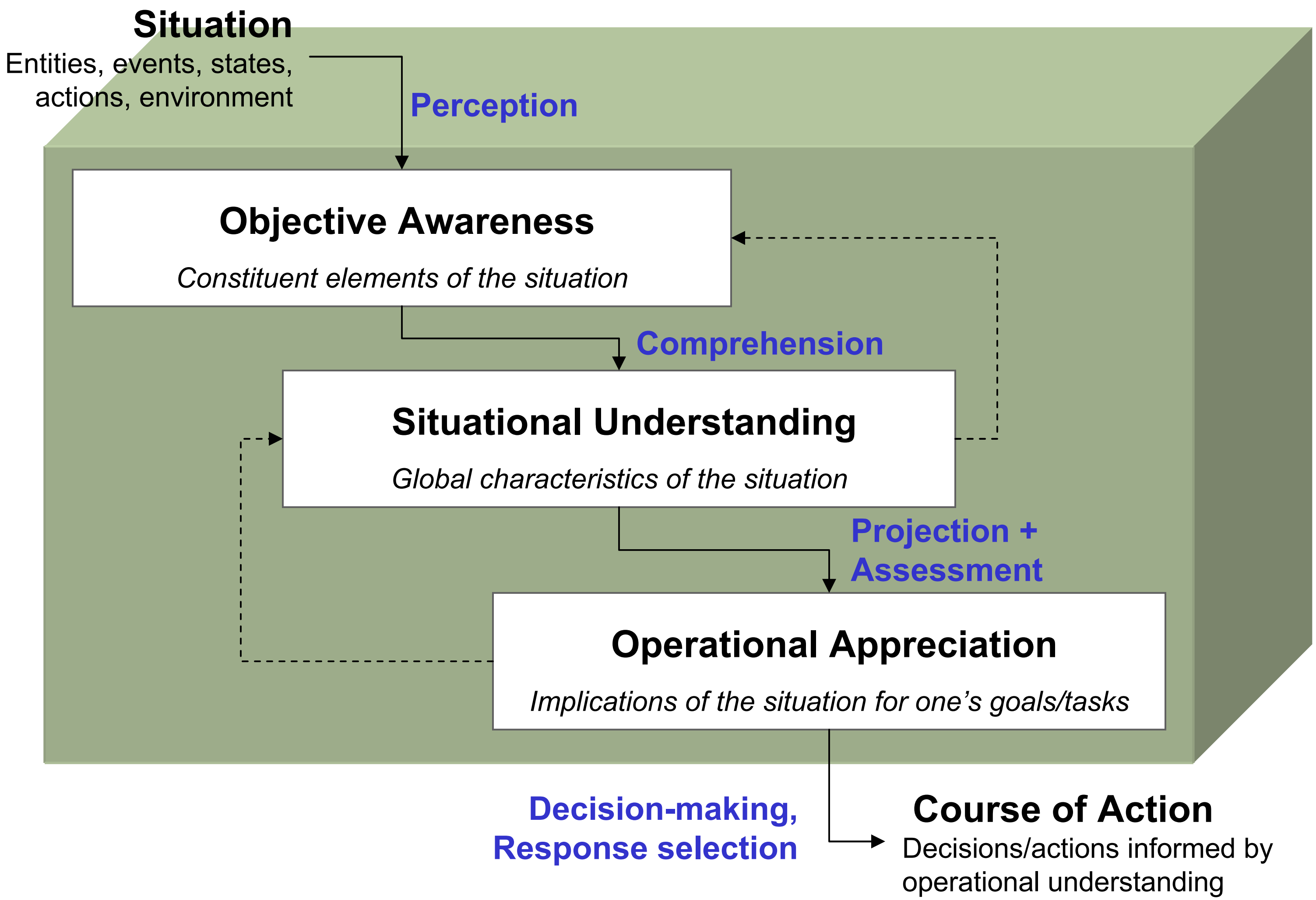
"Knowing what's going on so you can figure out what to do."

"What you need to know not to be surprised."

Who is where? What are they doing?
What's going on? Why?
What will happen next?
What does it mean for my task?

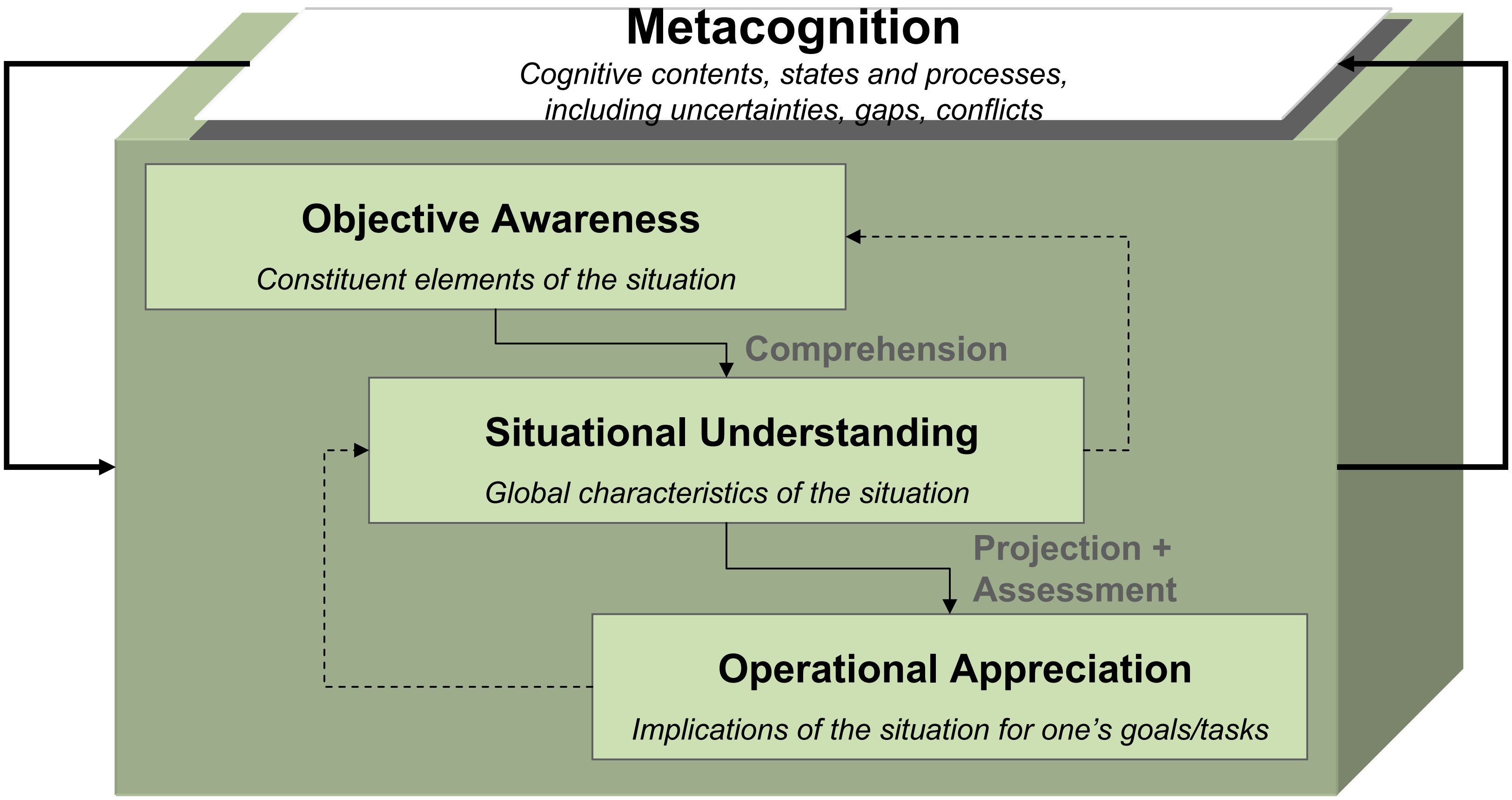
BAE SYSTEMS

Situational Awareness



BAE SYSTEMS

Situational Awareness



Situational Awareness

COGNITION	METACOGNITION
• Fighting in the city has mostly ceased	• This is certain. Current info, very reliable.
• Column of red tanks is leaving south of the city	• Not sure about this. Reports may not be from reliable source. Need to check.
• Enemy is beginning retreat	• Confidence in this -- 50-60% Need to look for evidence.
"Actual SA"	"Perceived SA"

BAE SYSTEMS

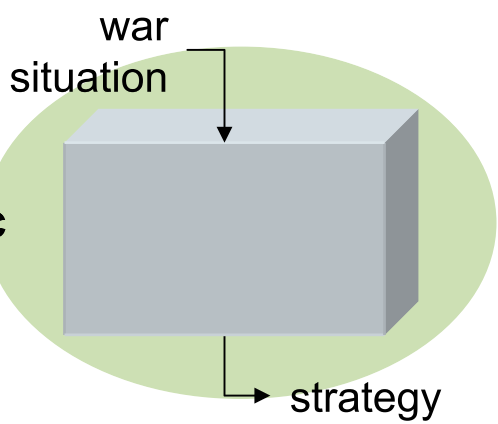
Situational Awareness

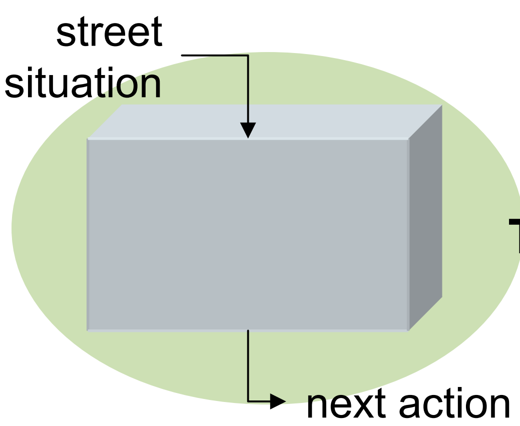
- Hence, SA is not just about having positive knowledge of actual events
- It's also about
  - **Being aware of what is not the case**
  - **Being aware of what we don't know and may need to find out**
  - **Being aware of what others are aware of and unaware of**
- So, SA is a complex, multi-faceted phenomenon

BAE SYSTEMS

Situational Awareness

Strategic





Tactical

BAE SYSTEMS

Assessment of Situational Awareness

- OBJECTIVE INDICES / CORRELATES
  - Performance
  - Behaviours
  - Physiology

SALIENT SABARS

- SELF-RATINGS
  - Unidimensional
  - Multidimensional

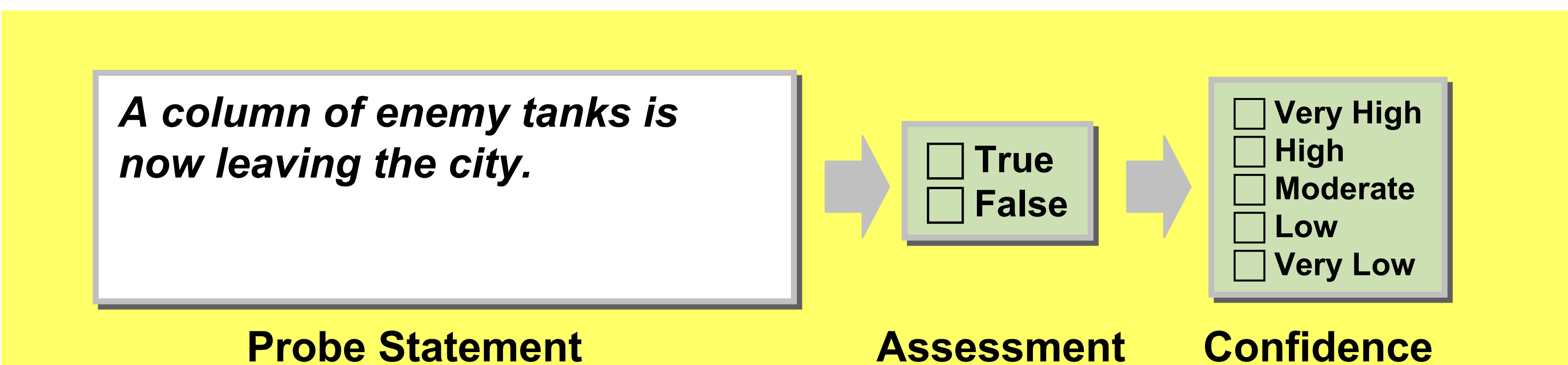
SARS PSAQ SART CARS

- DIRECT PROBES / QUERIES
  - Situation reports
  - Multi-choice questions
  - True / false probes

SAGAT QUASA

QUASA

- Quantitative Analysis of SA
  - Combination of direct probes and simultaneous self-ratings
  - True/false probes
  - Responses analysed using Signal detection Theory
  - Extension of CALIBRATION theory to SA
- Probes and ratings
  - True/false probe = a statement about the situation [a ‘report’] which may or may not be true.
  - Self-rating = indication of confidence in one’s probe response



QUASA

SA Requirements Analysis

- A form of Cognitive Task Analysis with SMEs to capture SA contents
  - Generic for the role/task
  - Specific to the scenario

Probe construction

- Formulate equal numbers of true & false probes
- Ensure that probes are
  - relevant to the subject's task
  - plausible as potentially 'true' descriptions when in fact false

- Process of checks & iterations:
  - independent 'blind' assessment of true/false likelihood
  - assessment of intelligibility
  - assessment of plausability w.r.t. the scenario
  - assessment of relevance to the subject's task

BAE SYSTEMS

QUASA in use

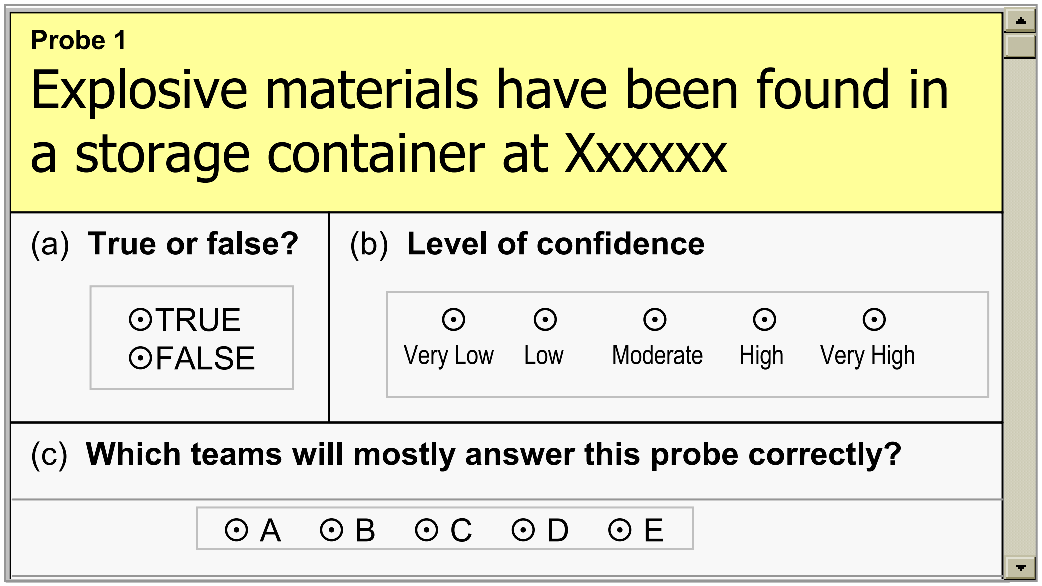
MN LOE 2 experiment

- 5 nations + NATO
  - US lead (JF COM)
- Collaborative planning
  - distributed teams
  - network
  - information sharing agreements
  - ONA process
- 46 subjects in 2 roles
  - Analysts vs Planners
- 2 conditions (methods of online collaboration), each lasting 1 week
- 50 T/F probes per subject per condition
  - 5 at a time every few hours

BAE SYSTEMS

QUASA in use

LOE 2 SA data collection



- True / false probe
- Subjective confidence level
- Perception of other teams' SA

Analysis of probes data

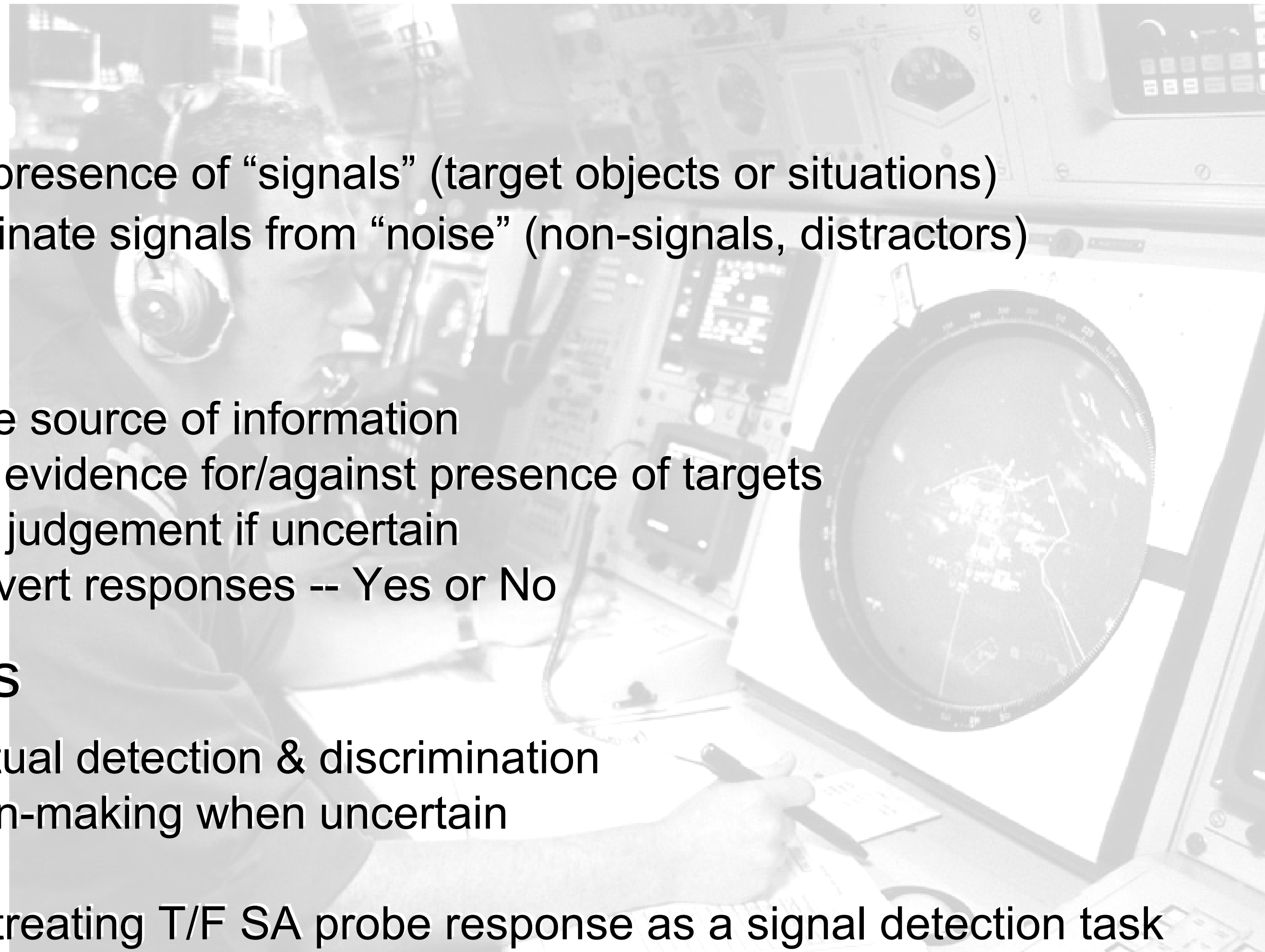
Contingency table

Probe type	Subject's response [T]	Subject's response [F]
True	HIT	MISS
False	FALSE ALARM	CORRECT REJECTION

Enemy forces have captured bridge Charlie. **[T] [F]**

Signal Detection Theory

Signal Detection Theory



Goal

- Detect presence of "signals" (target objects or situations)
- Discriminate signals from "noise" (non-signals, distractors)

Task

- Observe source of information
- Assess evidence for/against presence of targets
- Make a judgement if uncertain
- Make overt responses -- Yes or No

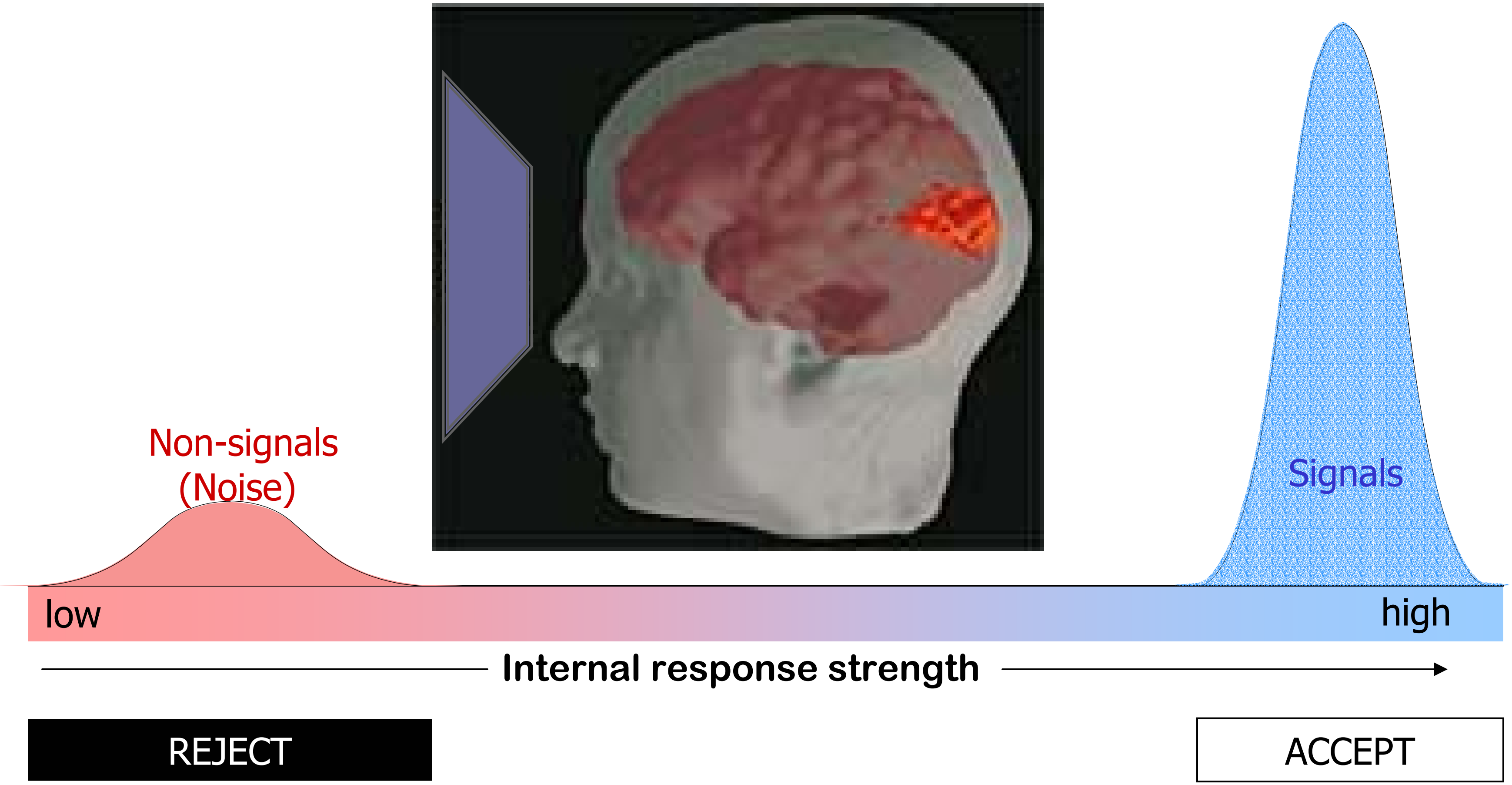
Processes

- Perceptual detection & discrimination
- Decision-making when uncertain

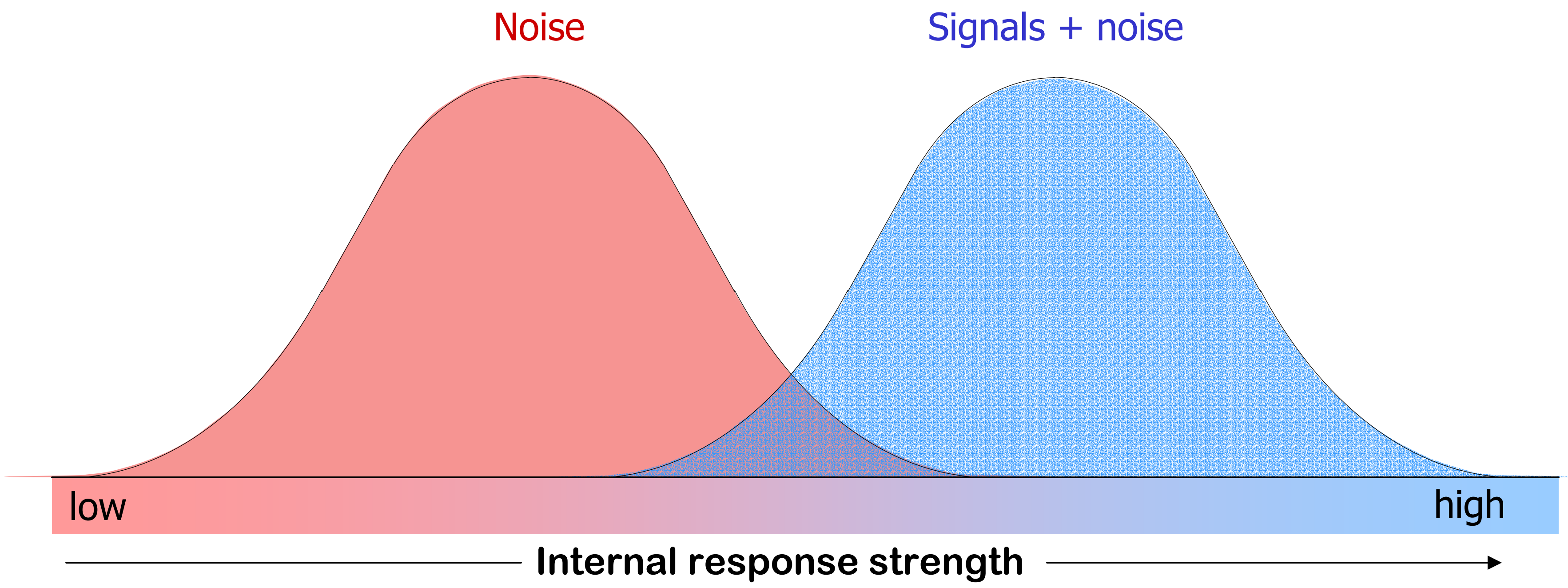
… We're treating T/F SA probe response as a signal detection task

BAE SYSTEMS

Signal Detection Theory

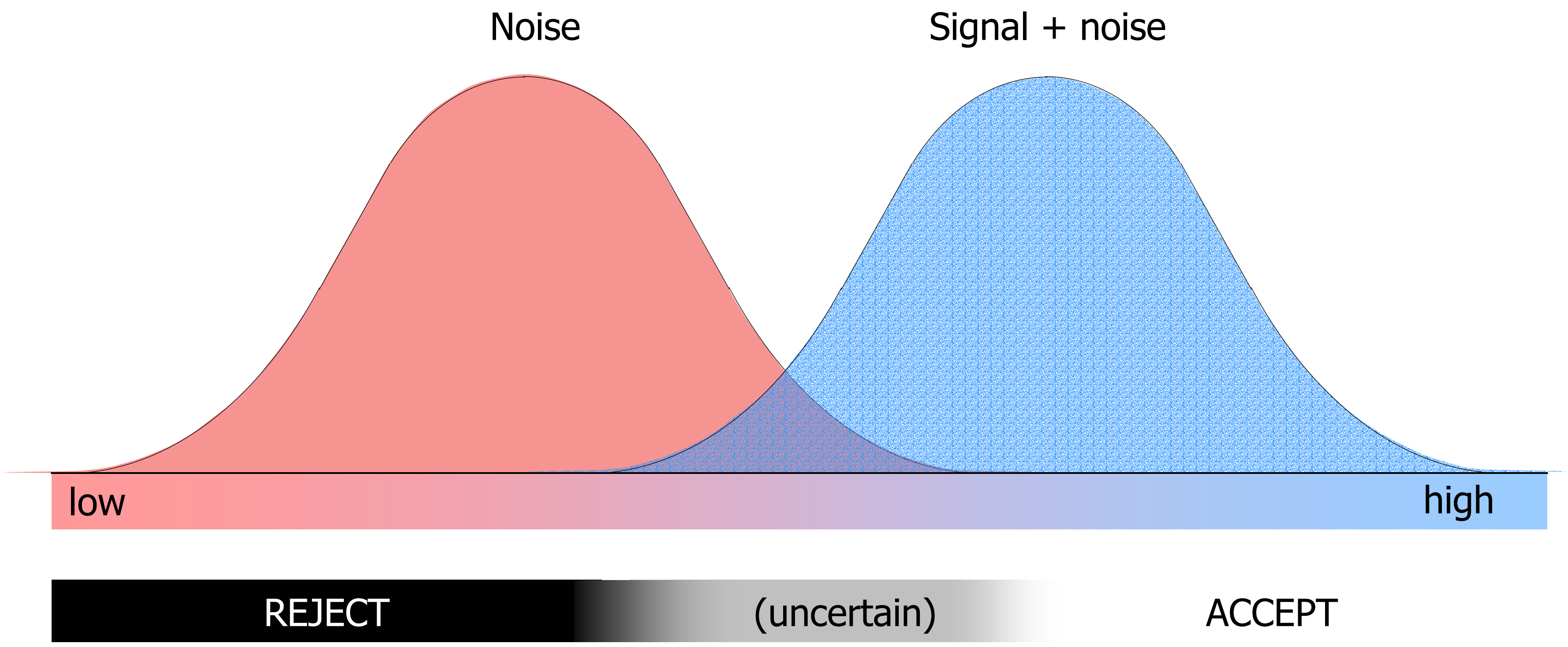


Signal Detection Theory



BAE SYSTEMS

Signal Detection Theory



BAE SYSTEMS

Signal Detection Theory

- Contingency table -- 4 possible outcomes

		RESPONSE	
		Accept	Reject
STIMULUS	Signal	HIT	MISS
	Non-signal	FALSE ALARM	CORRECT REJECTION

Signal Detection Theory

- Contingency table -- 4 possible outcomes

RESPONSE

STIMULUS	Accept	Reject	
Signal	80	20	100
Non-signal	12	88	100

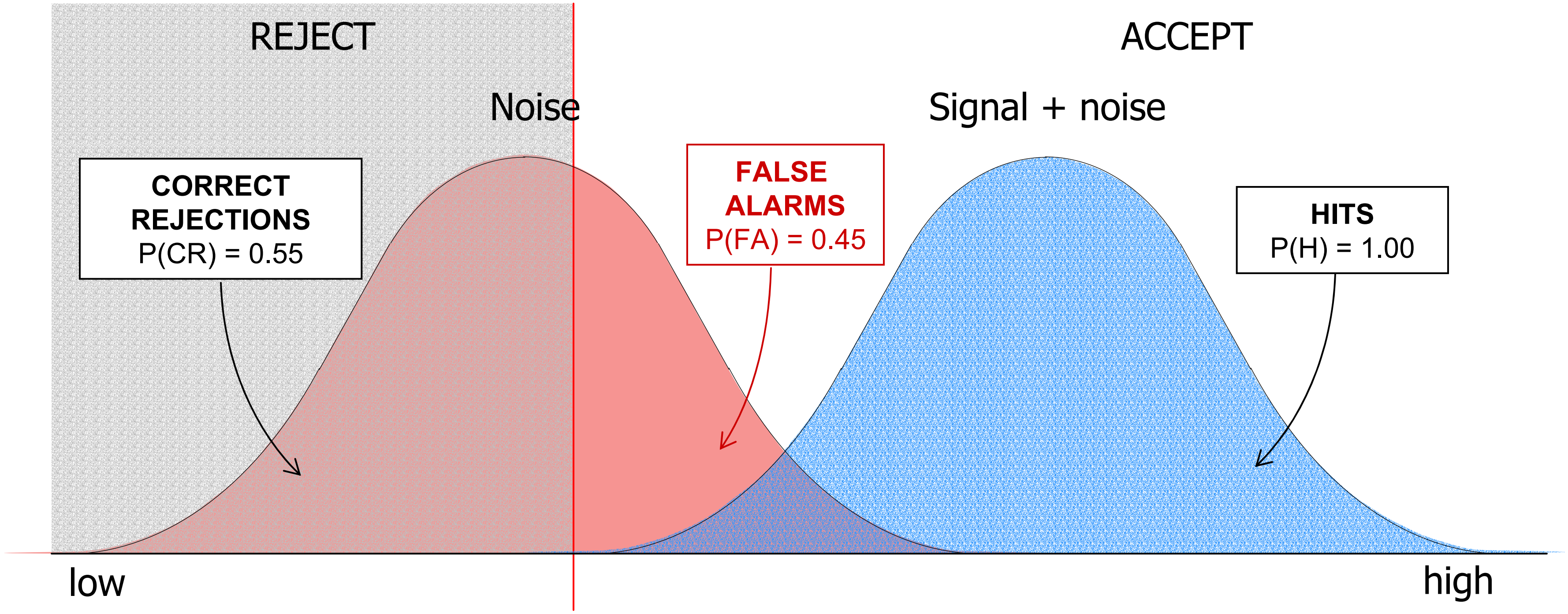
Hit rate = 0.80

Miss rate = 0.20

False Alarm rate = 0.12

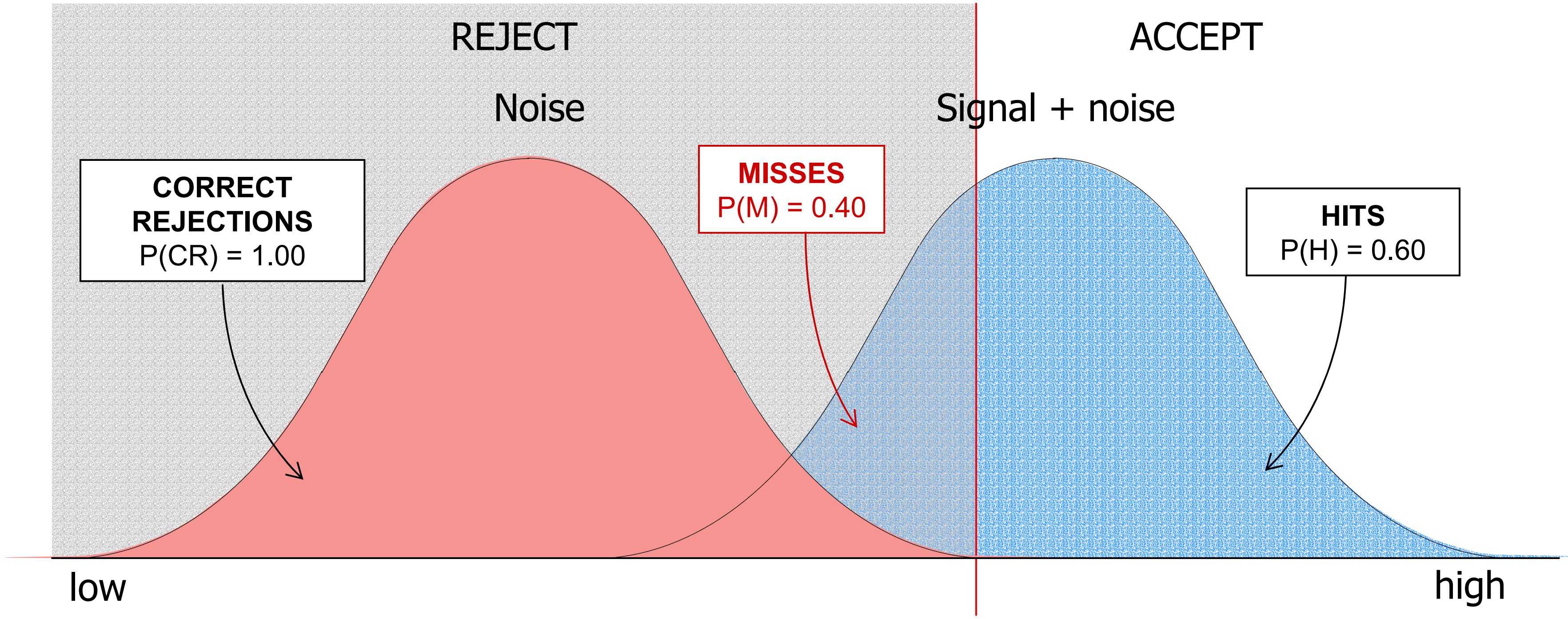
Correct Rejection rate = 0.88

Signal Detection Theory



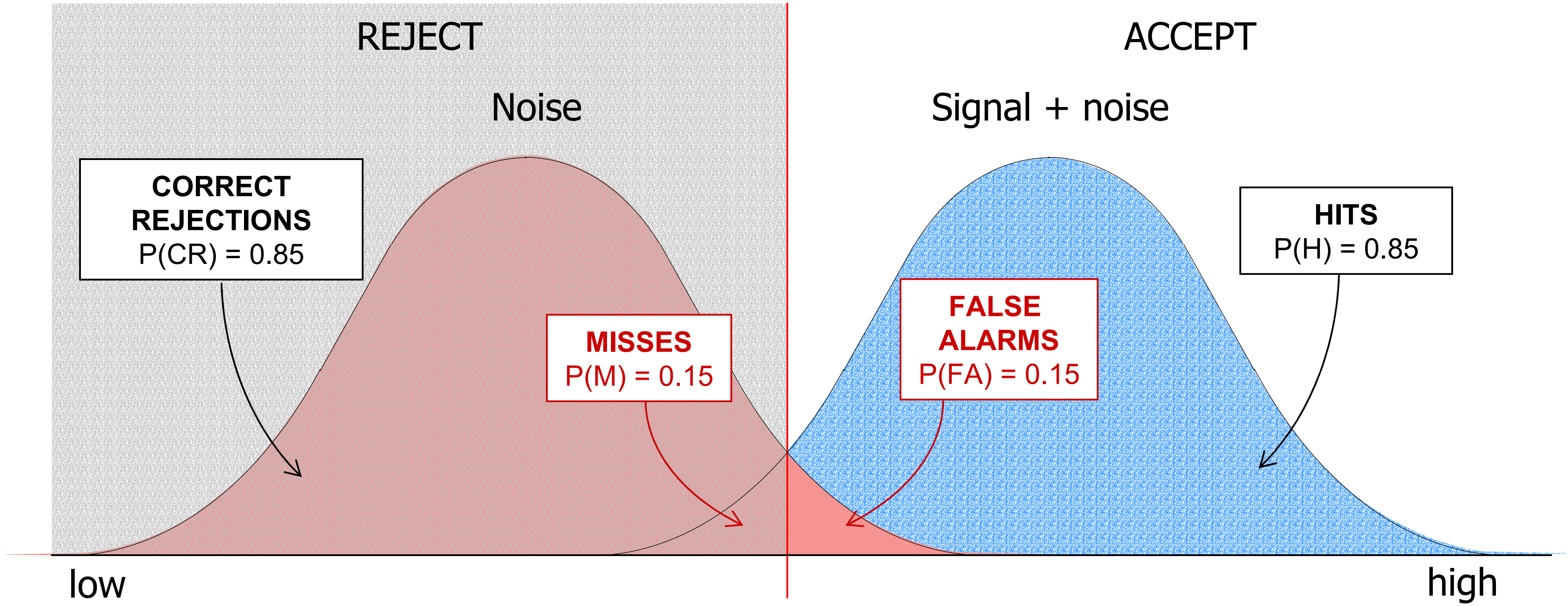
Low criterion (liberal, inclusive)
Letting no true signal slip through the net
Maximum hits, no misses
Prone to false alarms

Signal Detection Theory



High criterion (conservative, exclusive)
Accepting nothing but definite true signals
Maximum correct rejections, no false alarms
Prone to misses

Signal Detection Theory



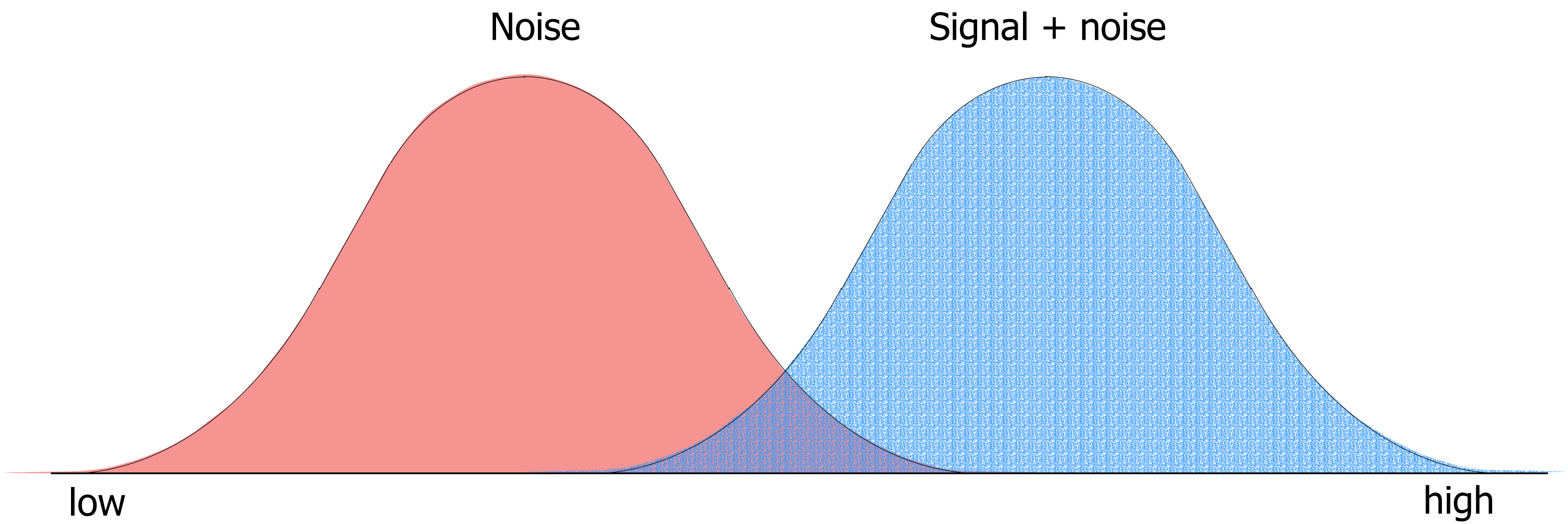
Central criterion (neutral, balanced)

Threshold set at the mid-point of uncertainty

Equal numbers of misses and false alarms

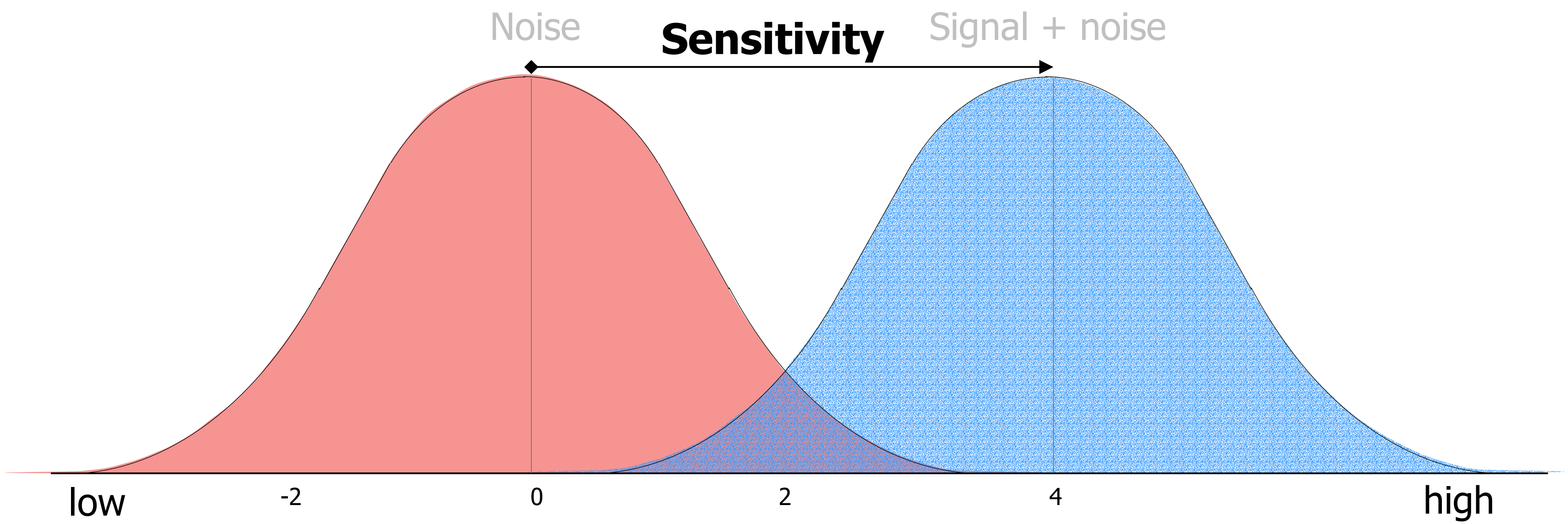
Prone to equal numbers of misses and false alarms

Signal Detection Theory



BAE SYSTEMS

Signal Detection Theory



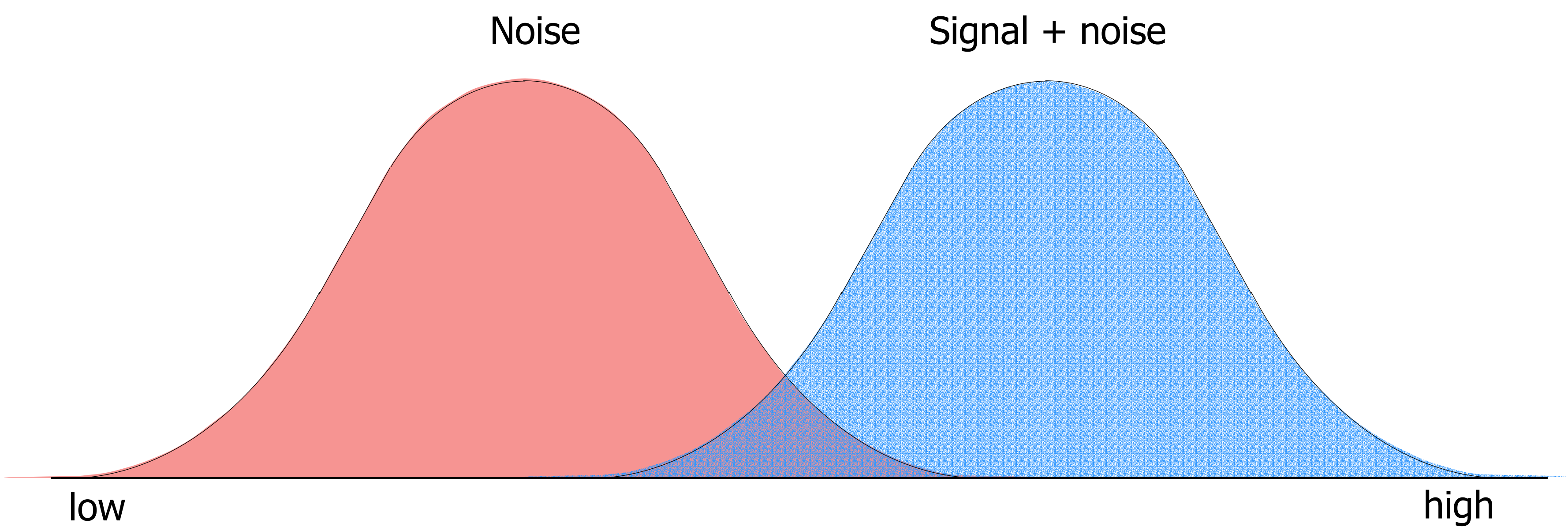
Sensitivity
Difference between noise and signal distributions, relative to their spread (variance)

d' = Z(H)-Z(FA)

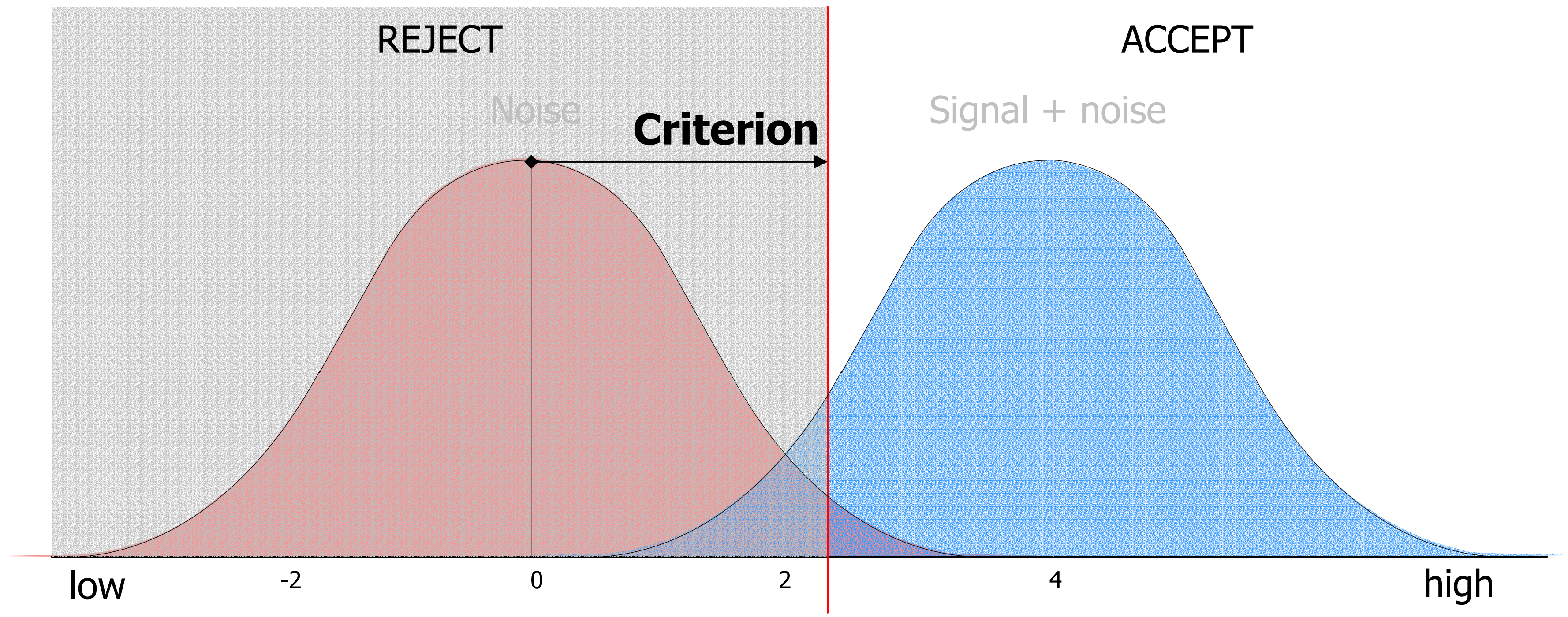
d' = 4.00

BAE SYSTEMS

Signal Detection Theory



Signal Detection Theory

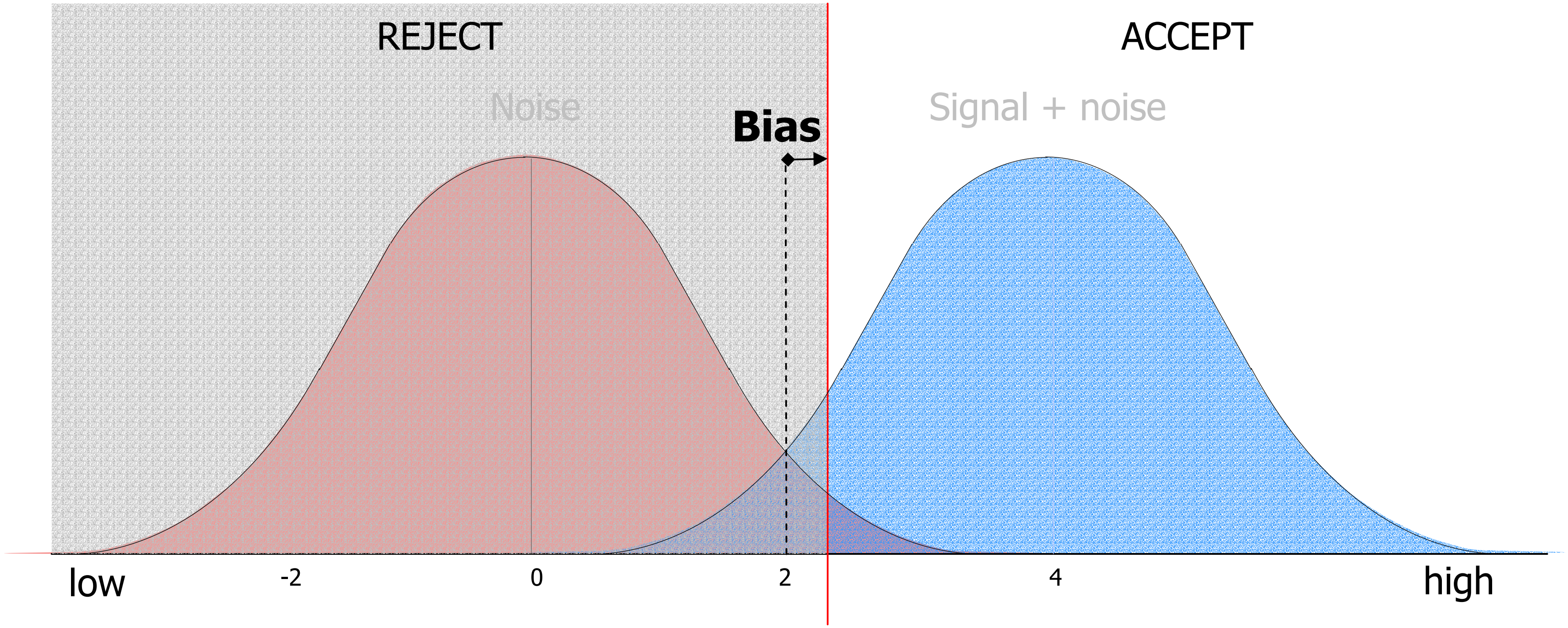


Criterion
Threshold for “accept” response, measured by distance from middle of noise distribution

k = –Z(FA)

k = 2.16

Signal Detection Theory



Bias (1)
Distance of actual criterion from neutral or central criterion

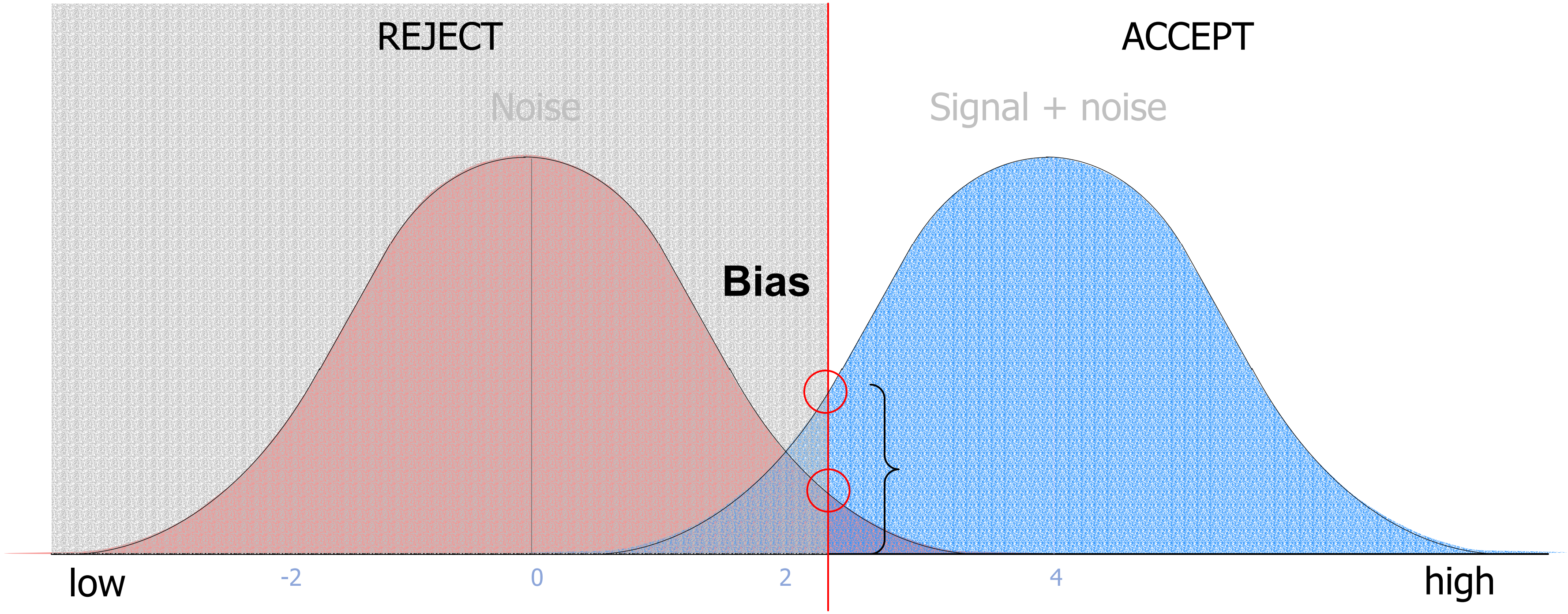
C = k – d′/2

C = 2.16 – 2.00

= 0.16

BAE SYSTEMS

Signal Detection Theory



Bias (2) and (3)
Likelihood ratio of probability densities of the two distributions at the criterion

$ß = f_S(k)/f_N(k)$
$ß = \exp^{d'C}$
$ß = 1.38$

$\log ß = ½(Z^2(FA)–Z^2(H))$
$\log ß = d'C$
$\log ß = 0.32$

Signal Detection Theory

Basic findings

– Perceptual performance depends upon

STIMULUS DISCRIMINABILITY

- **Stimulus quality**
- **Actual signal-noise ratio**

OBSERVER SENSITIVITY

- **Ability to detect signals**
- **Ability to discriminate signals from noise (distractors)**

OBSERVER RESPONSE STRATEGY IN UNCERTAINTY (CRITERION / BIAS)

- **Perceived signal probability**
- **Motivation to maximise hits or minimise false alarms**

– SDT has established that individuals are not just mechanical information processors but also make conscious judgements in conditions of uncertainty

Signal Detection Theory

- SDT in the real world
  - Early studies of radar observer performance
  - More recently:
    - **Recognition memory**
      - eyewitness memory
      - remember / know paradigm
    - **Diagnostic tasks**
      - medical tests
      - weather forecasting
      - psychometric tests
      - polygraph lie detectors
      - forensic tests
  - In principle, any situation that calls for judgement in uncertainty

SDT and Situational Awareness

- Assessing SA with T/F probes
  - Why use them?
  - Output of T/F probes = contingency table
    - **HITS / MISSES**
    - **FALSE ALARMS / CORRECT REJECTIONS**
  - Traditionally, we have assessed SA using % correct responses to questions about the situation
  - This tells us little or nothing about
    - **What the subject knows is not the case**
    - **What the subject wrongly believes is the case**
  - SDT provides separate measures of SENSITIVITY and CRITERION / BIAS

Results

– Compare two subjects (LOE 2)

SUBJECT A

Probe type	Responses: "True"	Responses: "False"
True	HITS 0.80	MISSES
False	FALSE ALARMS 0.10	CORRECT REJECTIONS

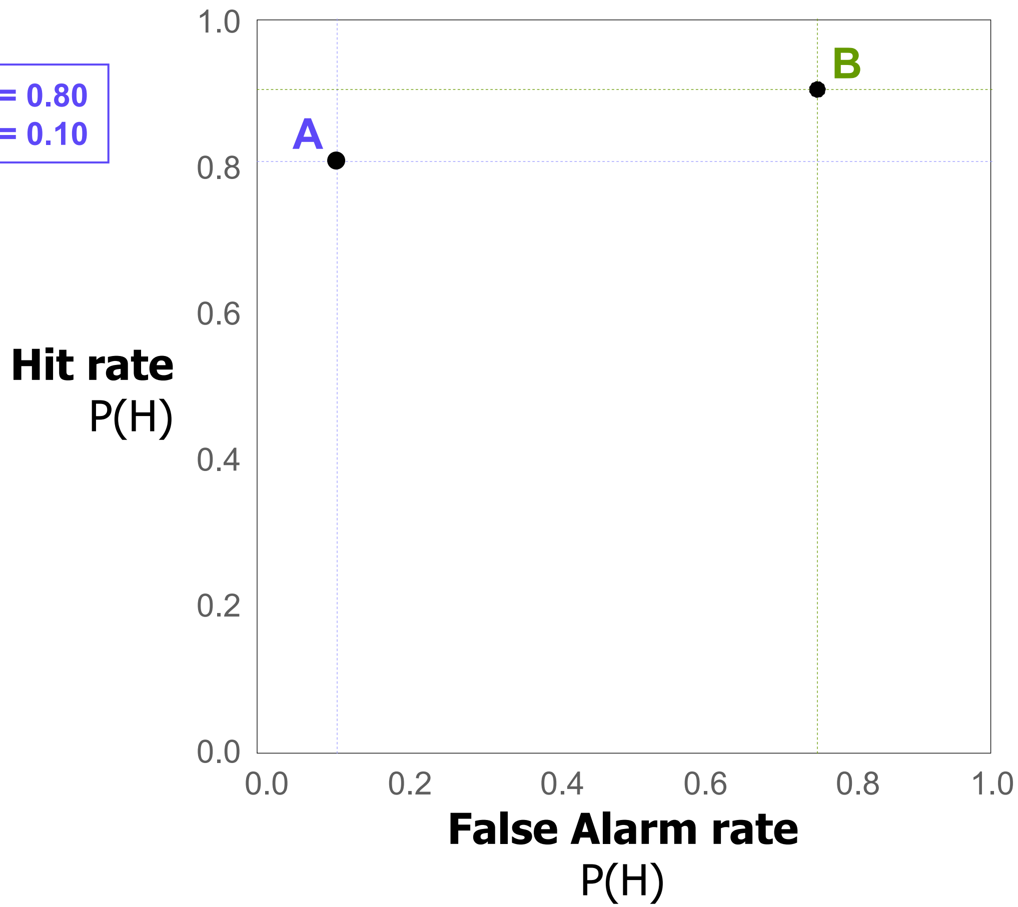
SUBJECT B

Probe type	Responses: "True"	Responses: "False"
True	HITS 0.90	MISSES
False	FALSE ALARMS 0.75	CORRECT REJECTIONS

Reciever Operating Characteristic

A

Hit rate = 0.80
FA rate = 0.10



B

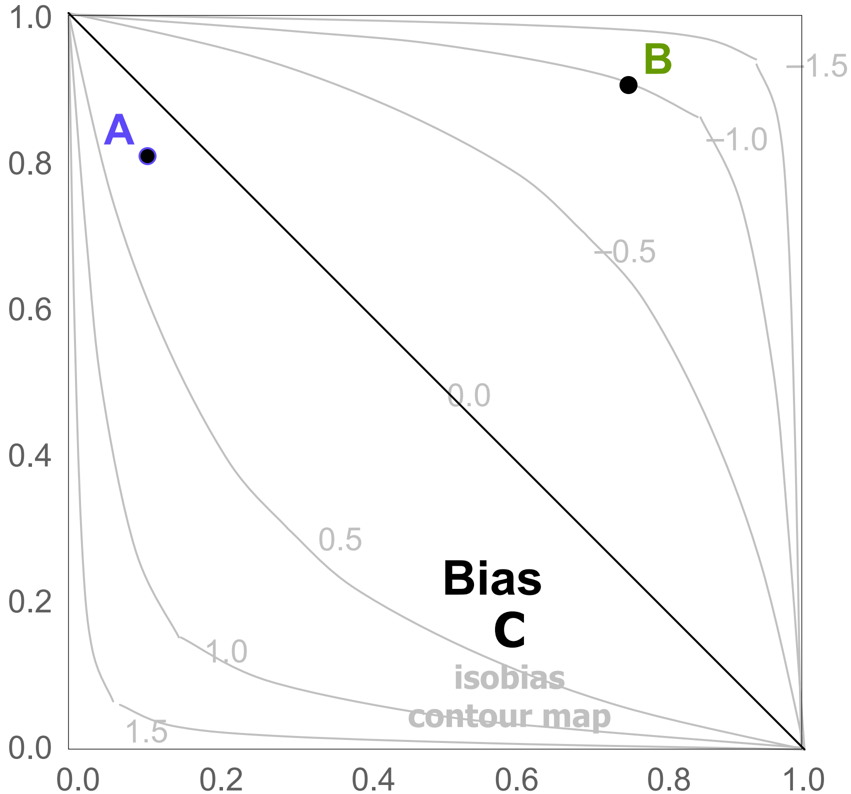
Hit rate = 0.90
FA rate = 0.75

ROC – Criterion / Bias

A

Hit rate = 0.80
FA rate = 0.10

C = 0.16



Hit rate
P(H)

False Alarm rate
P(H)

B

Hit rate = 0.90
FA rate = 0.75

C = -0.98

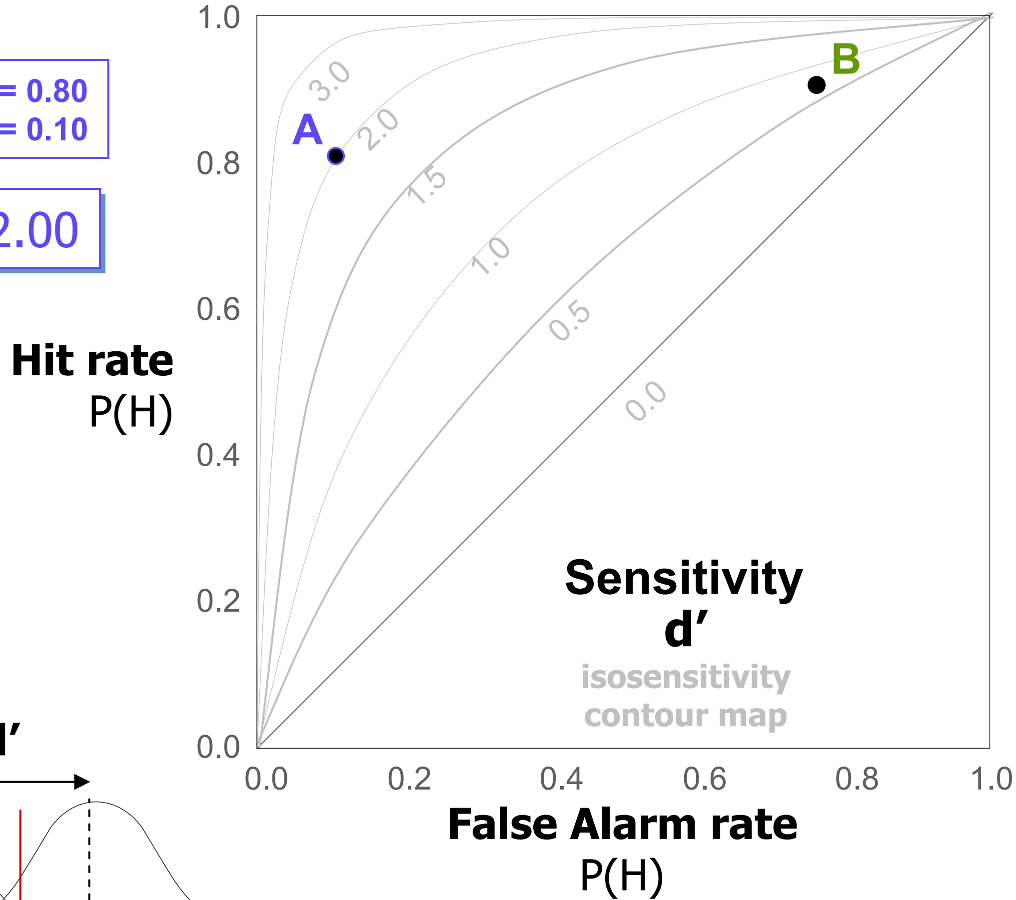
Subject A criterion = close to neutral
Subject B criterion = strong liberal bias

ROC – Sensitivity

A

Hit rate = 0.80
FA rate = 0.10

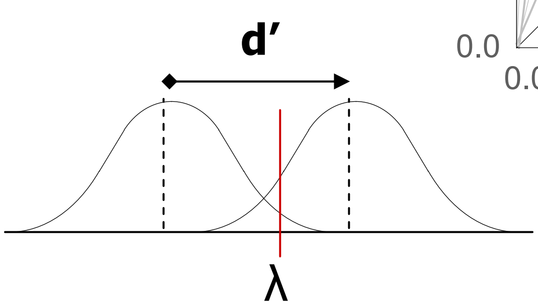
d′ = 2.00

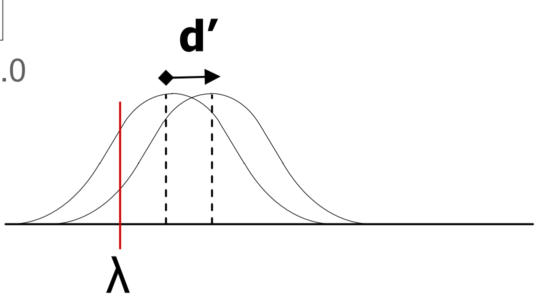


B

Hit rate = 0.90
FA rate = 0.75

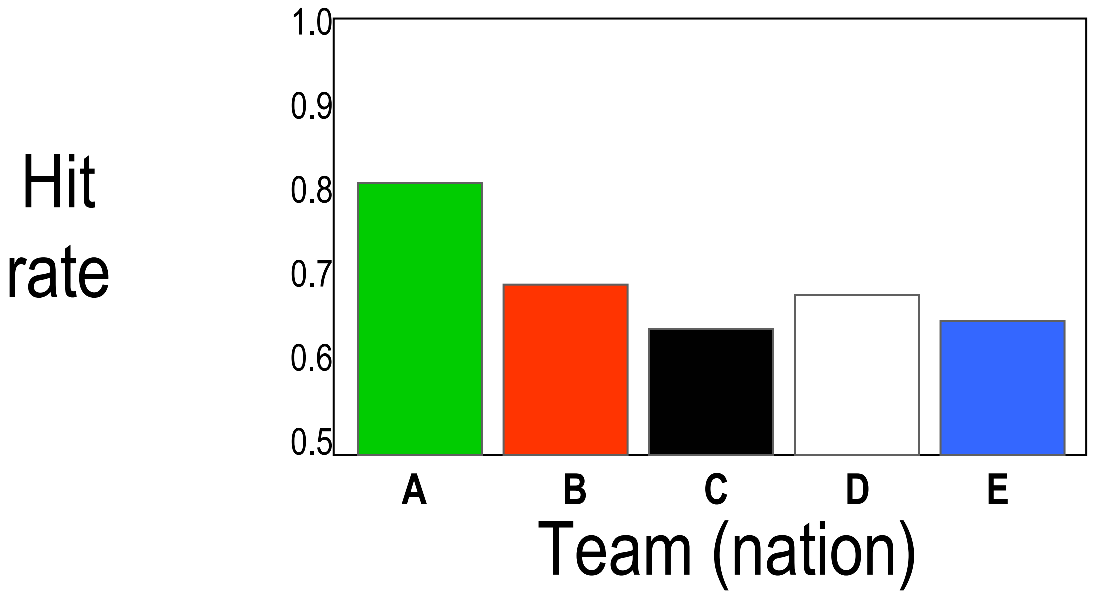
d′ = 0.60





QUASA data - LOE 2

SA probe hit rates

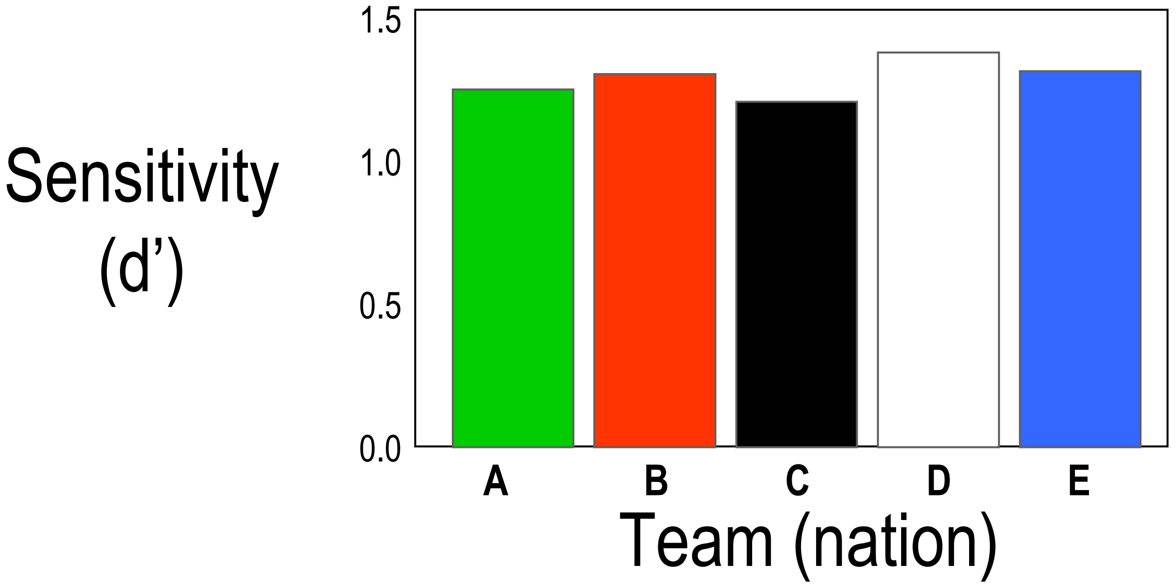


Team A has highest hit rate ...

BAE SYSTEMS

QUASA data - LOE 2

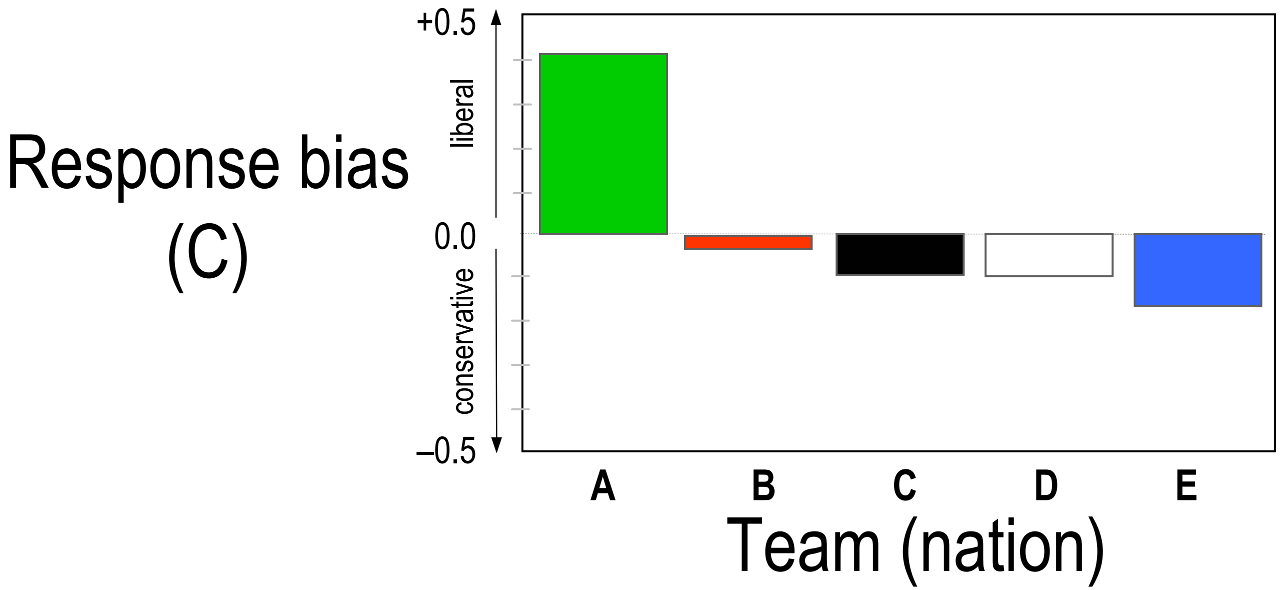
SA probe sensitivity



But team A is no more accurate overall at discriminating true from false probes

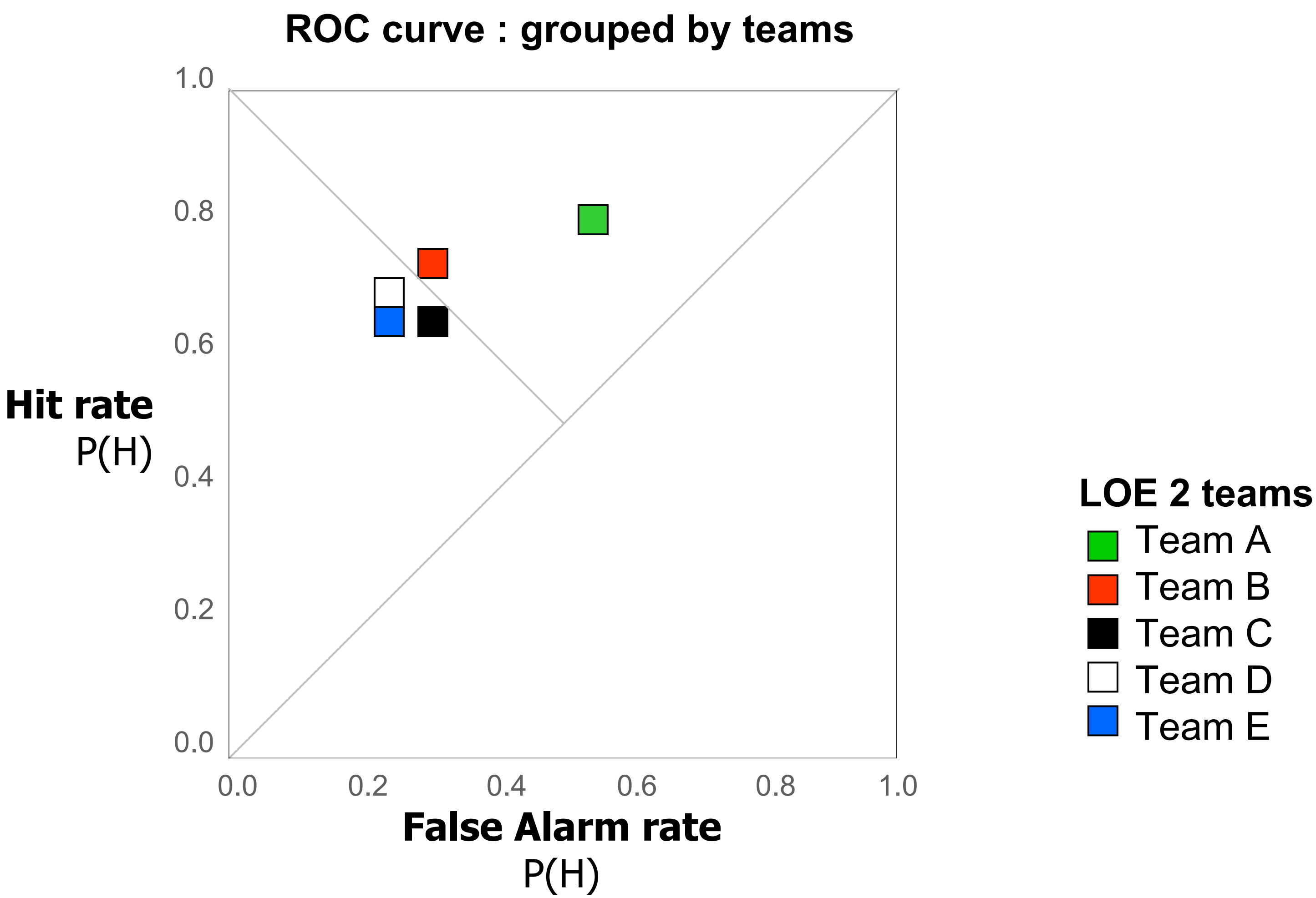
QUASA data - LOE 2

SA probe response bias



Team A is very liberal when uncertain (inclined to accept probes as true) -- hence the high hit rate

QUASA data - LOE 2



QUASA data - LOE 2

Summary so far

- Team A has highest hit rate on SA probes
- But SDT analysis shows all teams are only moderately accurate
- Team A's hit rate due to very liberal response bias when uncertain
- Other teams are neutral or slightly conservative

BAE SYSTEMS

Calibration

Concept

- Overconfidence / underconfidence
- The extent to which people are able to judge the correctness of their own observations or decisions

Method

- Obtain a judgement, then obtain self-rating of confidence in that judgement
    - binary ratings | continuous scales | ordinal ratings
- A well-calibrated person gives low ratings on incorrect / chance-level judgements (i.e. when uncertain) and high ratings on correct judgements (when certain)
- Calibration analysis quantifies this relationship in some way

Calibration

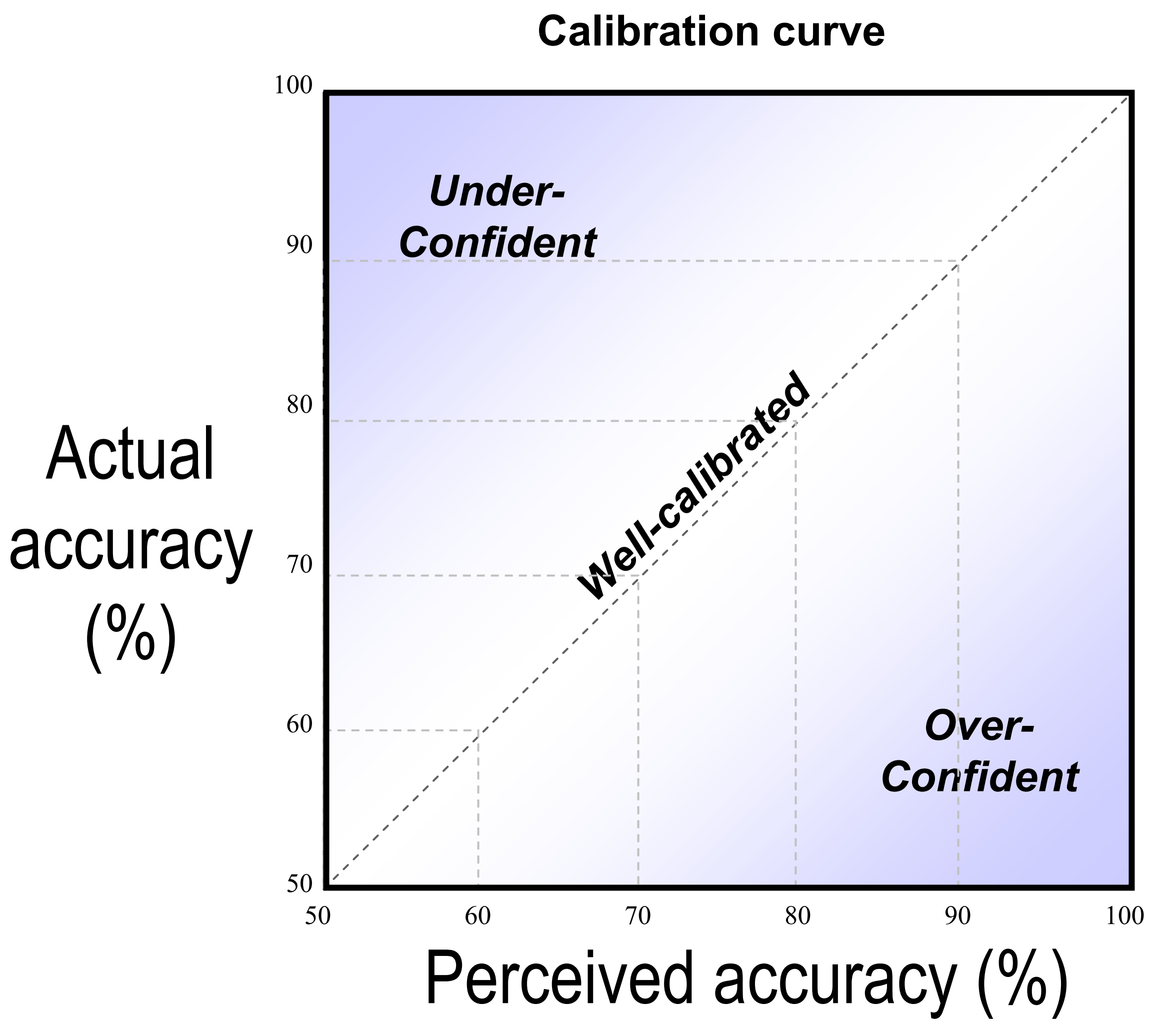
Findings

- Overconfidence common for cognitive tasks
- Underconfidence common for sensory tasks
- (May be an artefact of experimental methods)

Applications

- Eyewitness reports
  - Juries and police tend to be persuaded by highly confident witness reports, but these don't always correlkate with actual accuracy.
- Intelligence analysis
  - Don't want overconfident intelligence reports based on dubious data
- Situational awareness
  - Accidents attributed to over onfidence in poor/inaccurate SA

Calibration



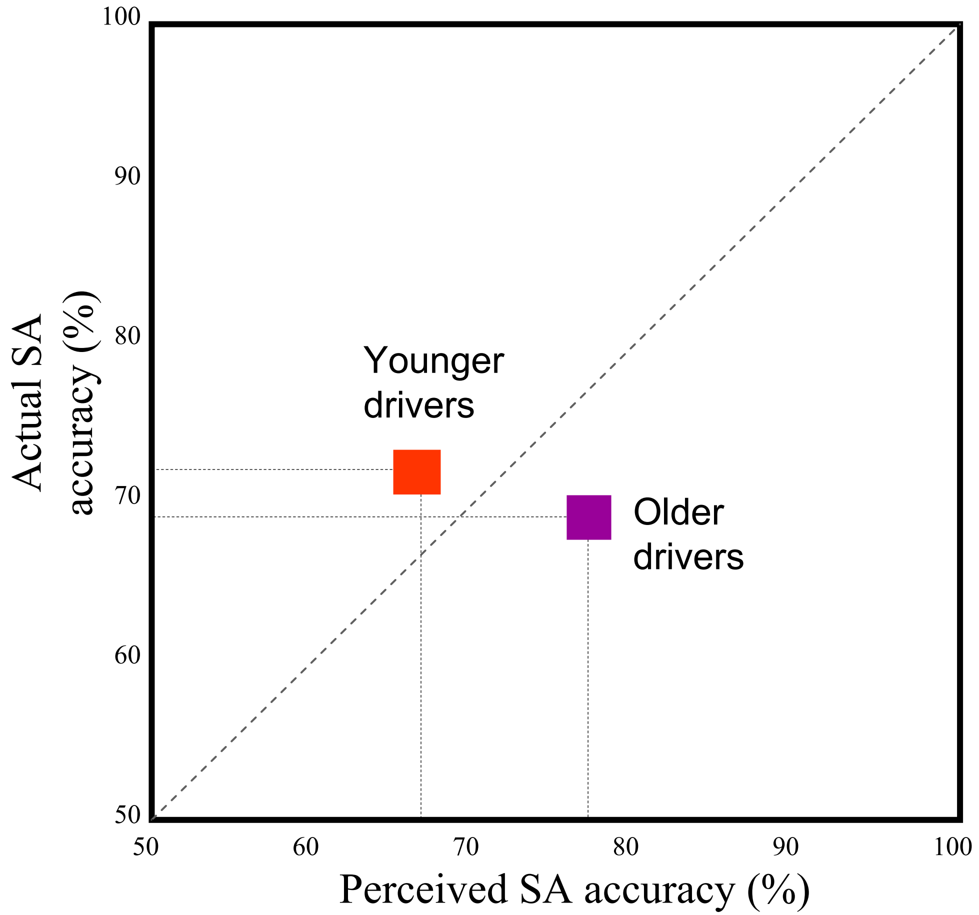
Calibration

SA of car drivers presented with safety-related electronic messages by an Advanced Traveller Information System (ATIS).

SA measured using a 2AFC version of SAGAT.

Confidence in each probe response rated on a continuous scale (50%-100

Calibration curve



Source

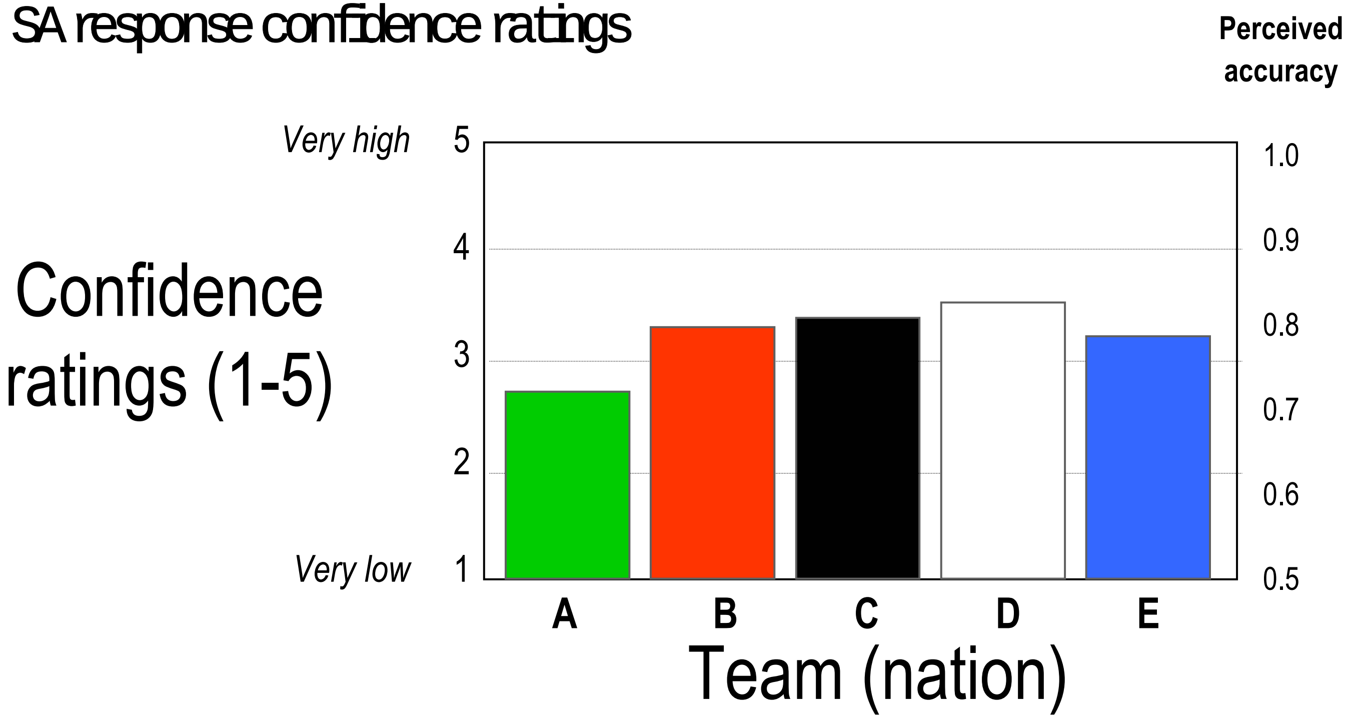
Lee, J.D., Stone, S., Gore, B.F., Colton, C., Macauley, J., Kinghorn, R., Campbell, J.L., Finch, M. & Jamieson, G. (1997).

Advanced Traveller Information Systems and Commercial Vehicle Operations Components of the Intelligent Transportation Systems: Design Alternatives for In-Vehicle Information Displays.

U.S. Federal Highway Administration technical report FHWA-RD-96-147. McLean, Virginia.

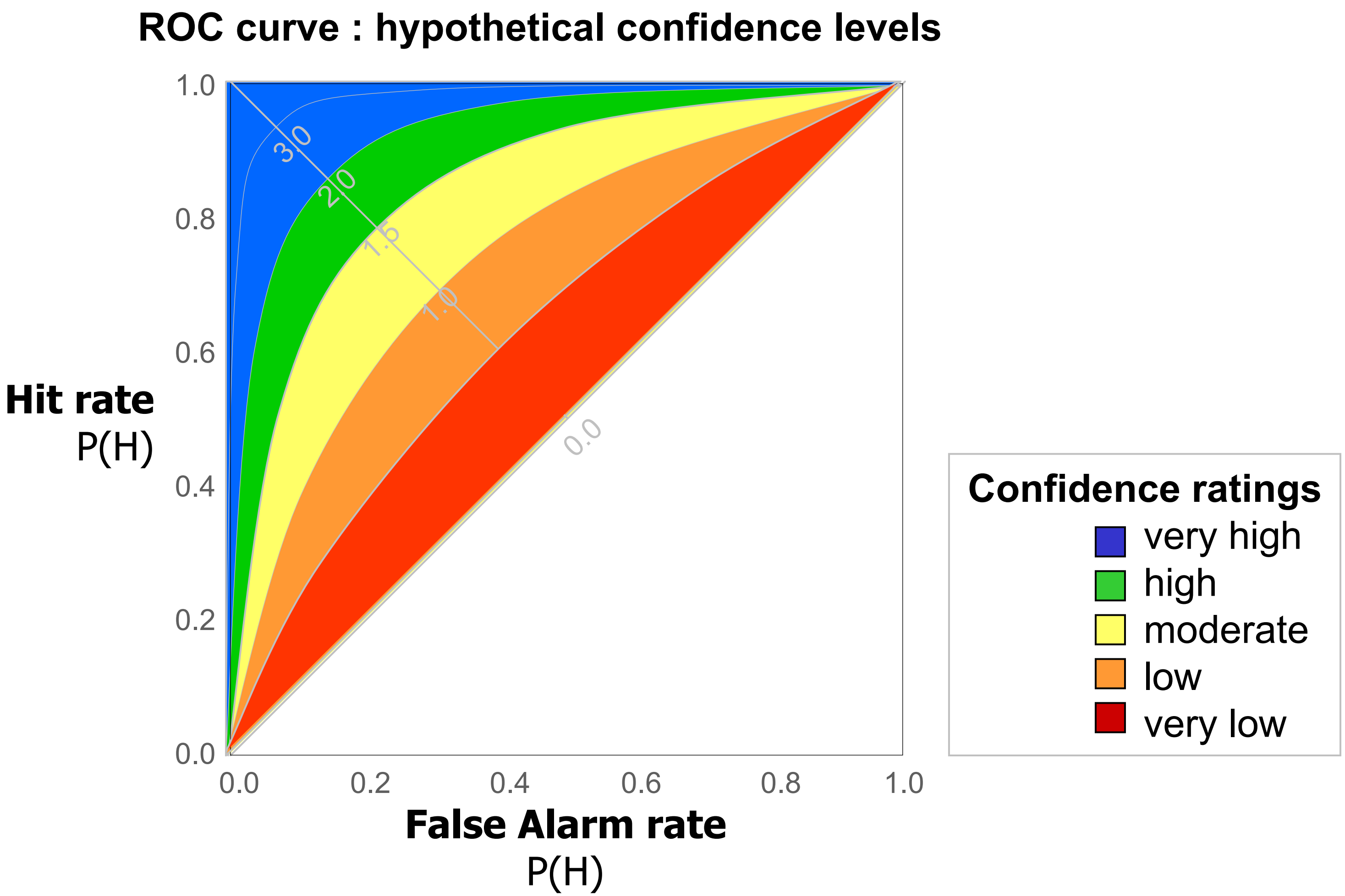
QUASA data - LOE 2

SA response confidence ratings



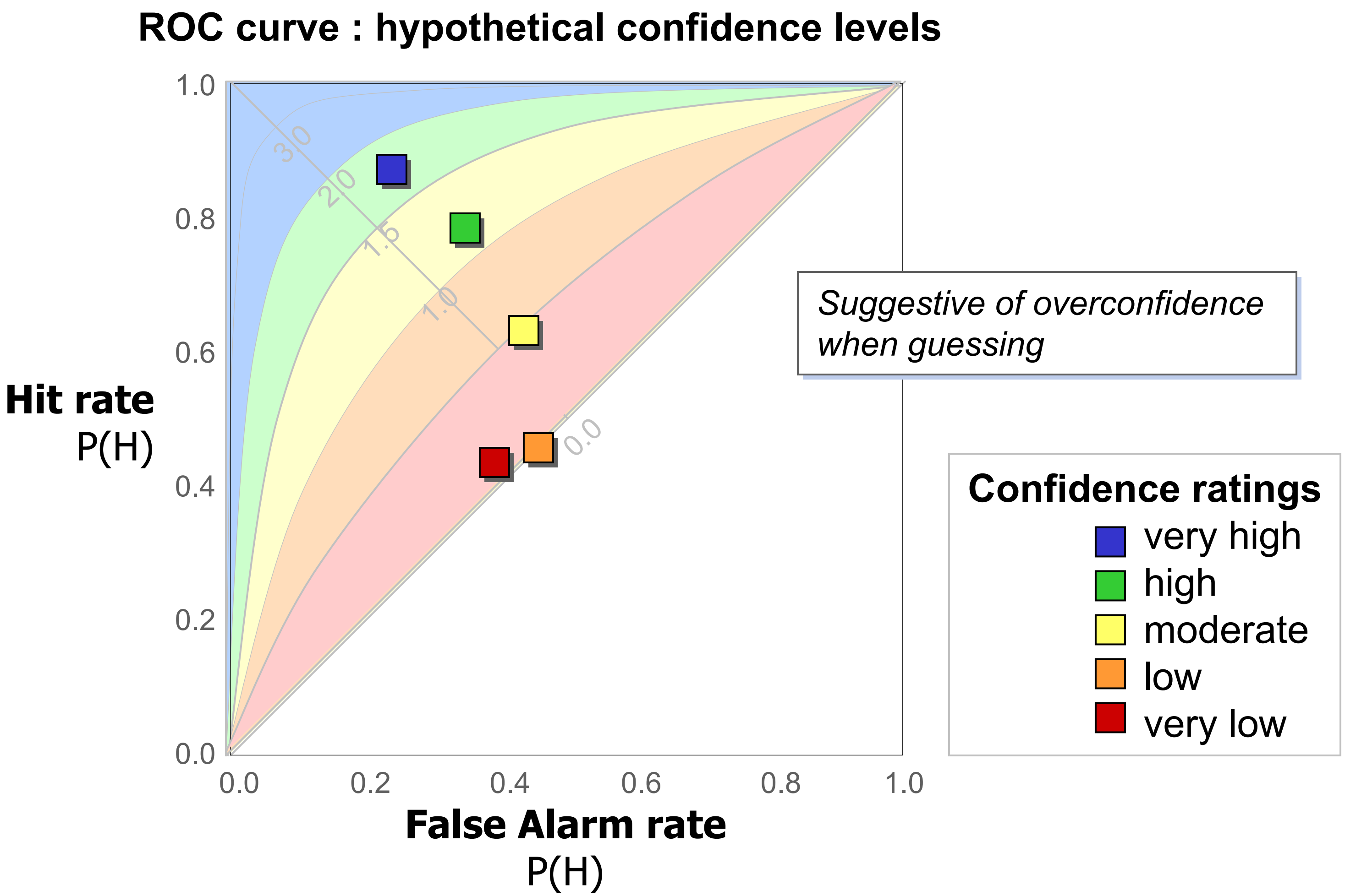
Mean SA probe response confidence ratings per team in LOE 2.

QUASA data - LOE 2



BAE SYSTEMS

QUASA data - LOE 2



QUASA data - LOE 2

Calibration scores

– using hit + correct rejection rates as actual accuracy

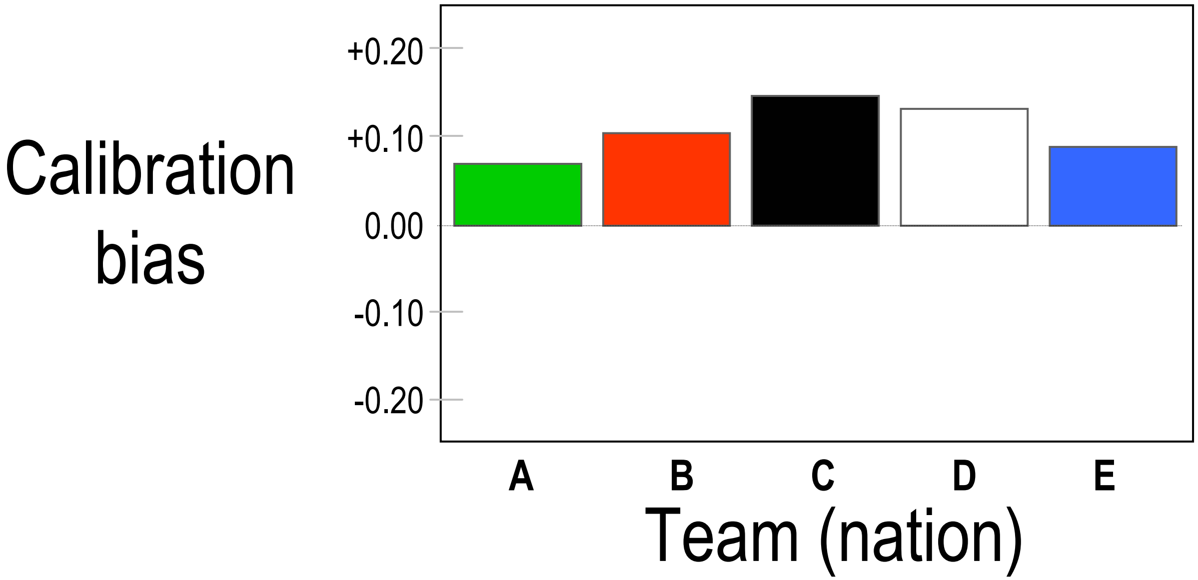
	Team (nation)				
	A	B	C	D	E
Perceived accuracy	0.716	0.795	0.803	0.832	0.774
SA accuracy (correct responses)	0.647	0.691	0.656	0.706	0.692
Calibration bias	**+0.07**	**+0.11**	**+0.15**	**+0.13**	**+0.08**

To assess SA calibration, average confidence ratings were transformed (0.5-1.0) and probe accuracy scores (proportion of hits plus correct rejections) were subtracted from the result to provide a calibration bias statistic.

BAE SYSTEMS

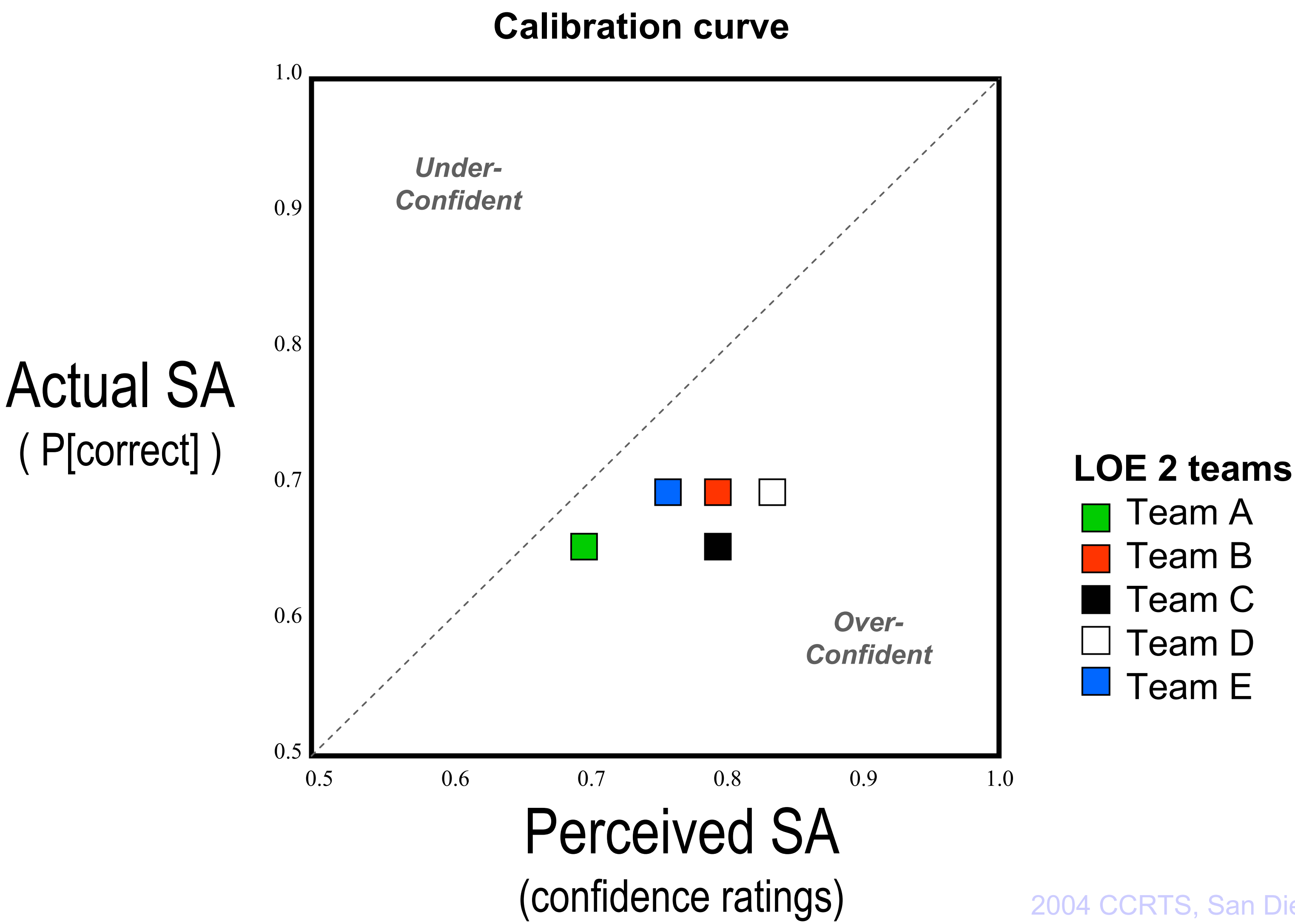
QUASA data - LOE 2

Calibration scores



Mean SA probe hit rates per team in LOE 2.

QUASA data - LOE 2



QUASA data - LOE 2

Summary

- Team A had lowest overall confidence ratings in their SA responses
- Confidence ratings were transformed into “perceived SA” scores and calibrated with actual SA scores
- Calibration analysis revealed general overconfidence
- Team A was actually best calibrated

Summary & conclusions

QUASA

– Technique for SA assessment
– Combines true/false SA probes with simultaneous self-ratings of confidence for each probe response.
– SDT analysis is applied to probe responses
  – Differentiates between actual SA accuracy (sensitivity) and response bias when uncertain
– Calibration analysis examines the relationship between actual SA and perceived SA.

Conclusions

– **QUASA yields potentially insightful quantitative results**
– **SDT statistic can be used as measure of actual SA accuracy.**
– **Subjects appear to be generally well-calibrated for SA**

Lessons learned

- **T/F probes need objective referent (‘groud truth’)**
  - Can be used to assess awareness of empirical information (objective environment & features, type of situation, actions)
  - Cannot be used to assess awareness of non-empirical information (future possibilities, intentions)
- **T/F probes need very careful construction & pre-testing**
  - Avoid ambiguity in language
  - Avoid bias in likelihood
- **In a dynamic situation, T/F probes may need to be constructed on the fly**

Outstanding issues

– Does response criterion/bias obtained with probes reflect a similar criterion/bias of the subject in assessing the real situation?

– How many probes / responses needed?

– How does this compare with other metrics?

– What about time to respond to probe? (= distance from criterion?)

Research directions

- Perform calibration analysis with Fuzzy SDT and/or Type 2 SDT
- Address team / shared SA

BAE SYSTEMS

Quantitative Analysis of Situational Awareness (QUASA)

Applying Signal Detection Theory to True/False Probes and Self-Ratings

Barry McGuinness

Principal Scientist
Human Factors Dept
Advanced Technology Centre
BAE Systems
Bristol, UK

barry.mcguinness @ baesystems.com

BAE SYSTEMS

BACKUP SLIDES

Characteristics of SA

- Mode of <u>cognition</u> that facilitates effective <u>action</u>
  - Critical in situations that are potentially complex, demanding, high-tempo, uncertain and/or unpredictable.
- Consists of <u>mental representations</u> of a situation and its implications:

- **OBJECTIVE AWARENESS :**

  The operational environment and the constellation of elements within it
  - terrain, weather, buildings, platforms, people; locations, movements, actions, states
  - derived from observations or data in context

- **SITUATIONAL UNDERSTANDING :**

  The global characteristics of the situation -- type and status
  - *Hijack situation? Hostage situation? Safe? Problematic? Critical?*
  - inferred from current awareness in context

- **OPERATIONAL APPRECIATION :**

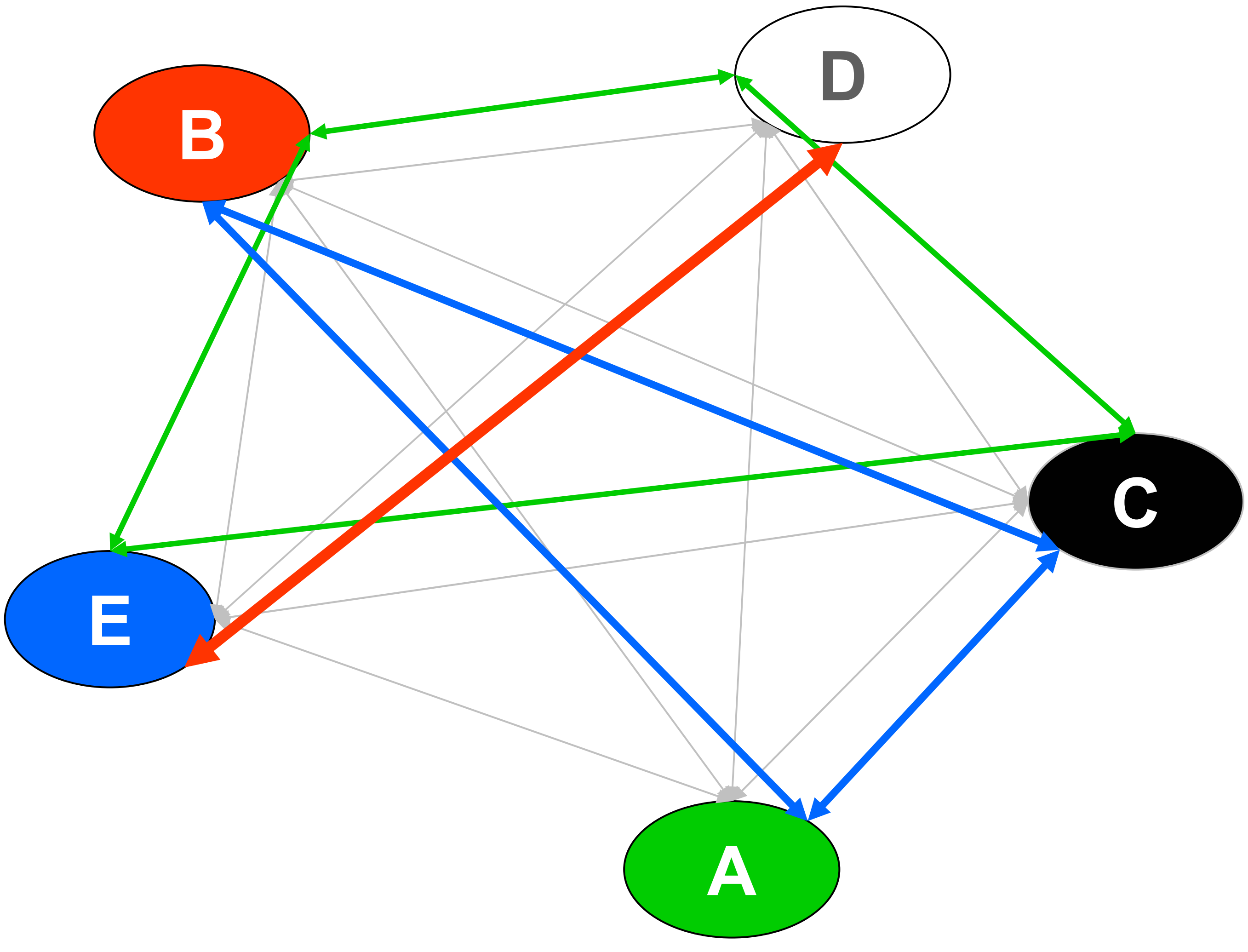
  The implications of the situation w.r.t. one's operational goals / plans / tasks
  - Getting better or worse? Critical points ahead? Need a new course of action?
  - inferred from situational understanding in context

LOE 2 information sharing agreements

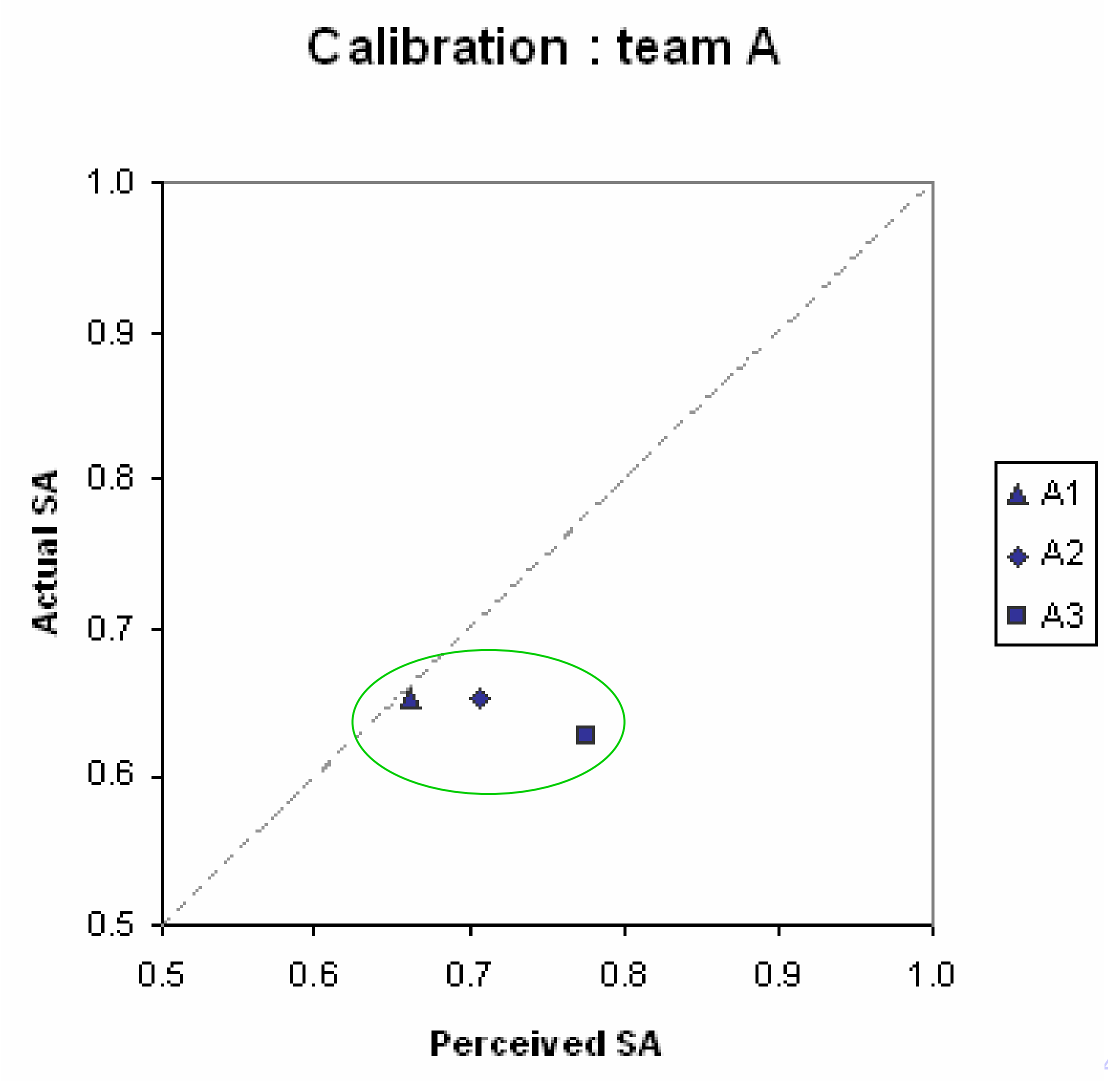
Country	ML	TL	BL_1	BL_2	Coalition	Private	Total
A		X		X	X	X	4
C	X	X			X	X	4
B	X	X			X	X	4
D	X		X		X	X	4
E	X		X		X	X	4

BAE SYSTEMS

LOE 2 information sharing agreements

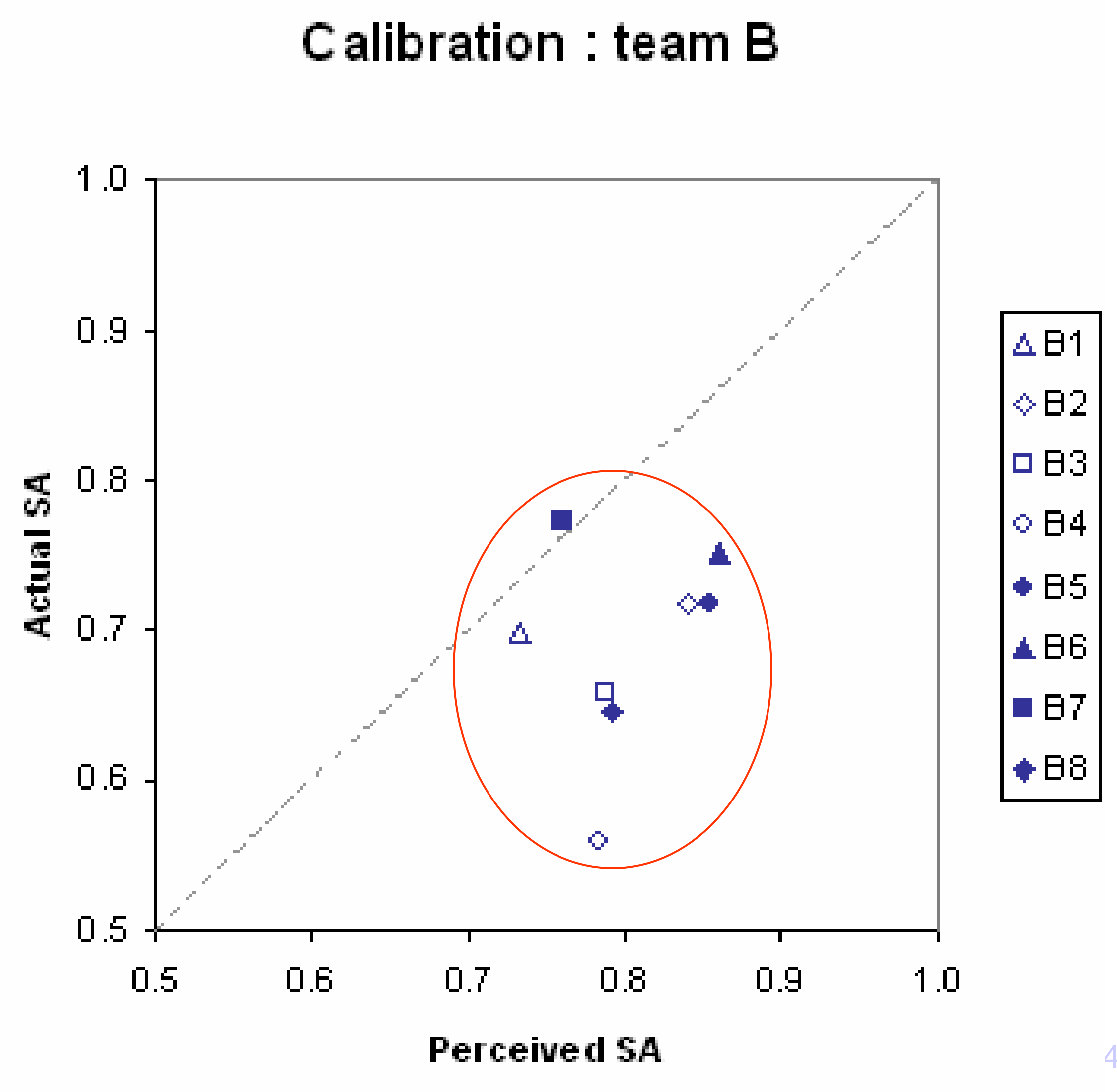


QUASA data - LOE 2



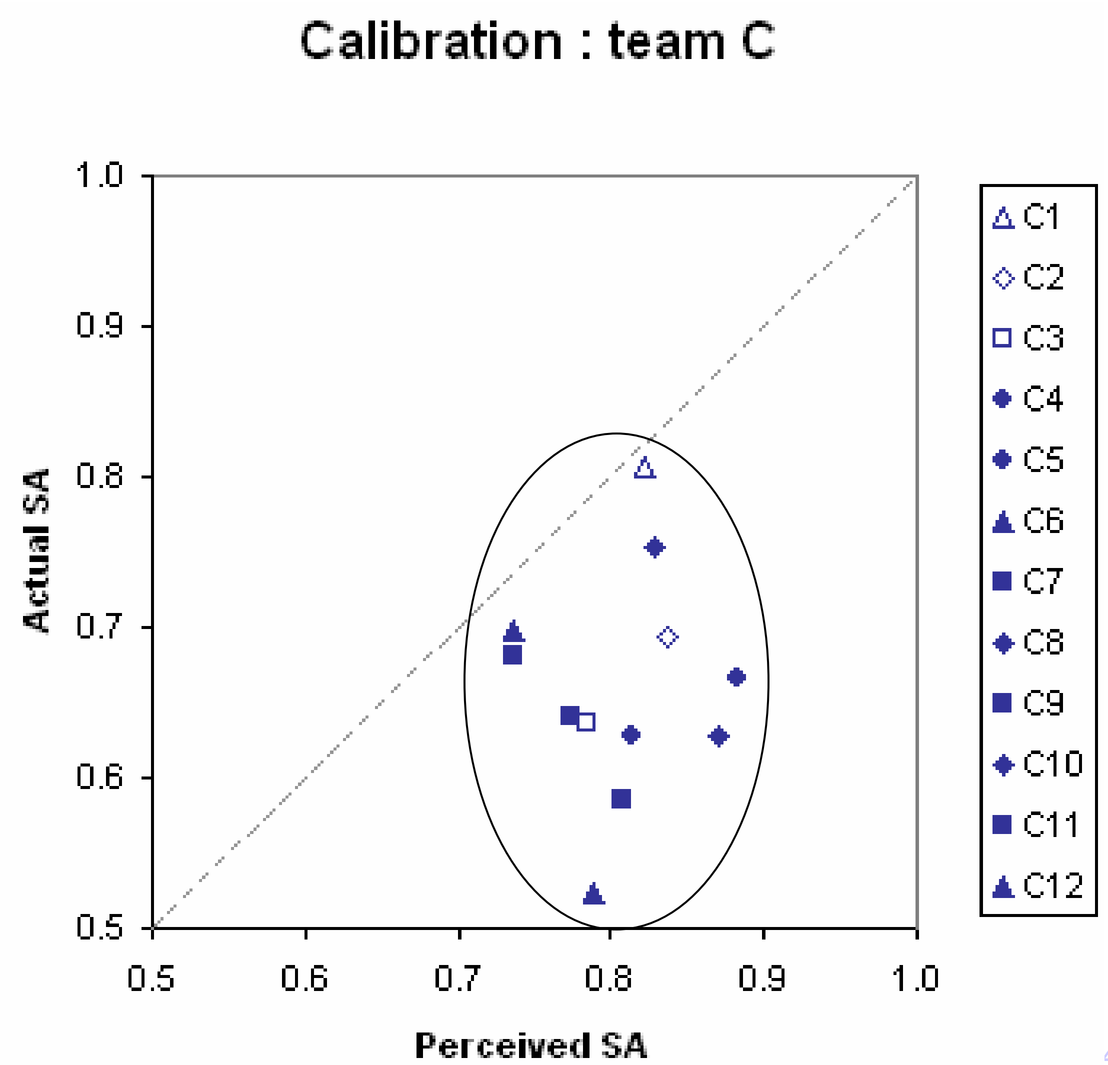
BAE SYSTEMS

QUASA data - LOE 2



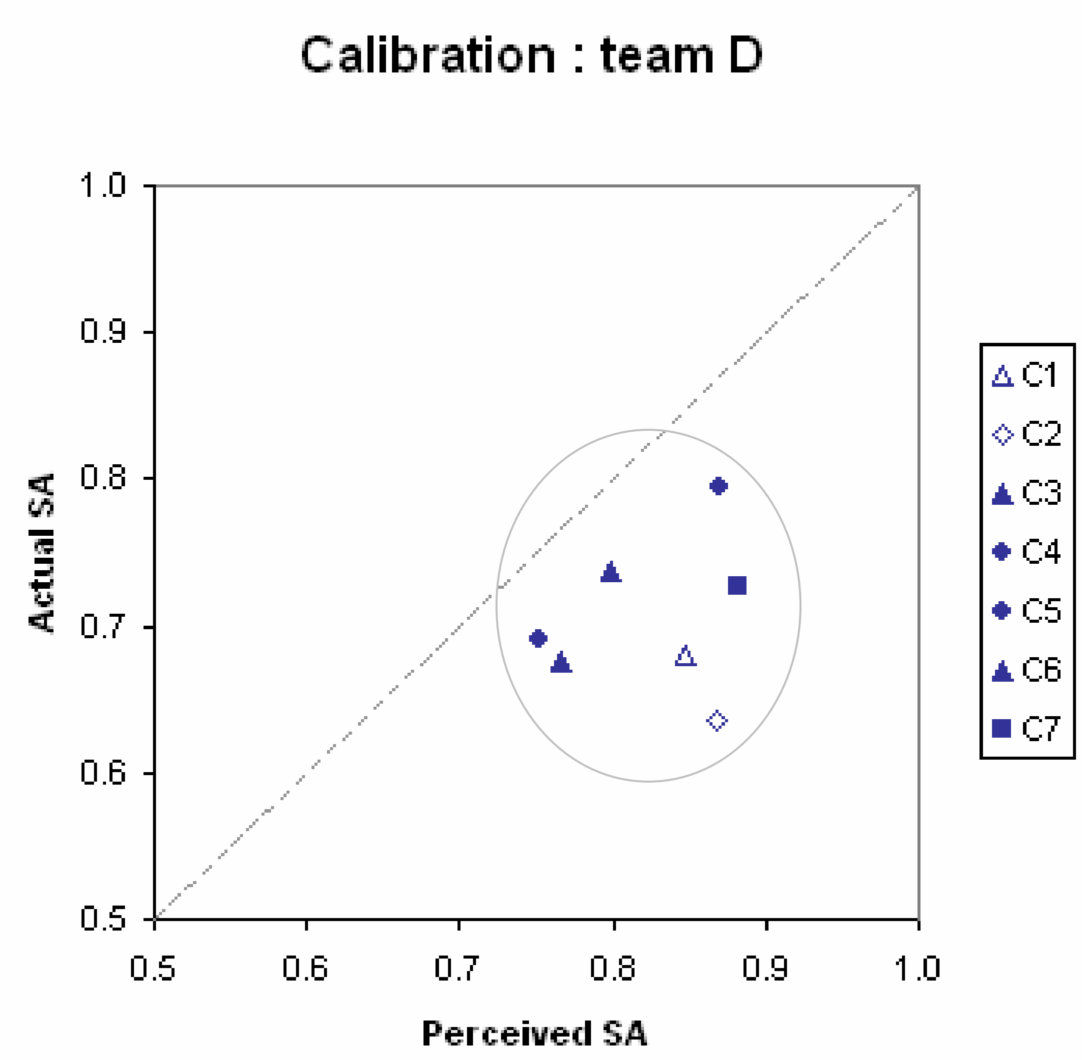
BAE SYSTEMS

QUASA data - LOE 2



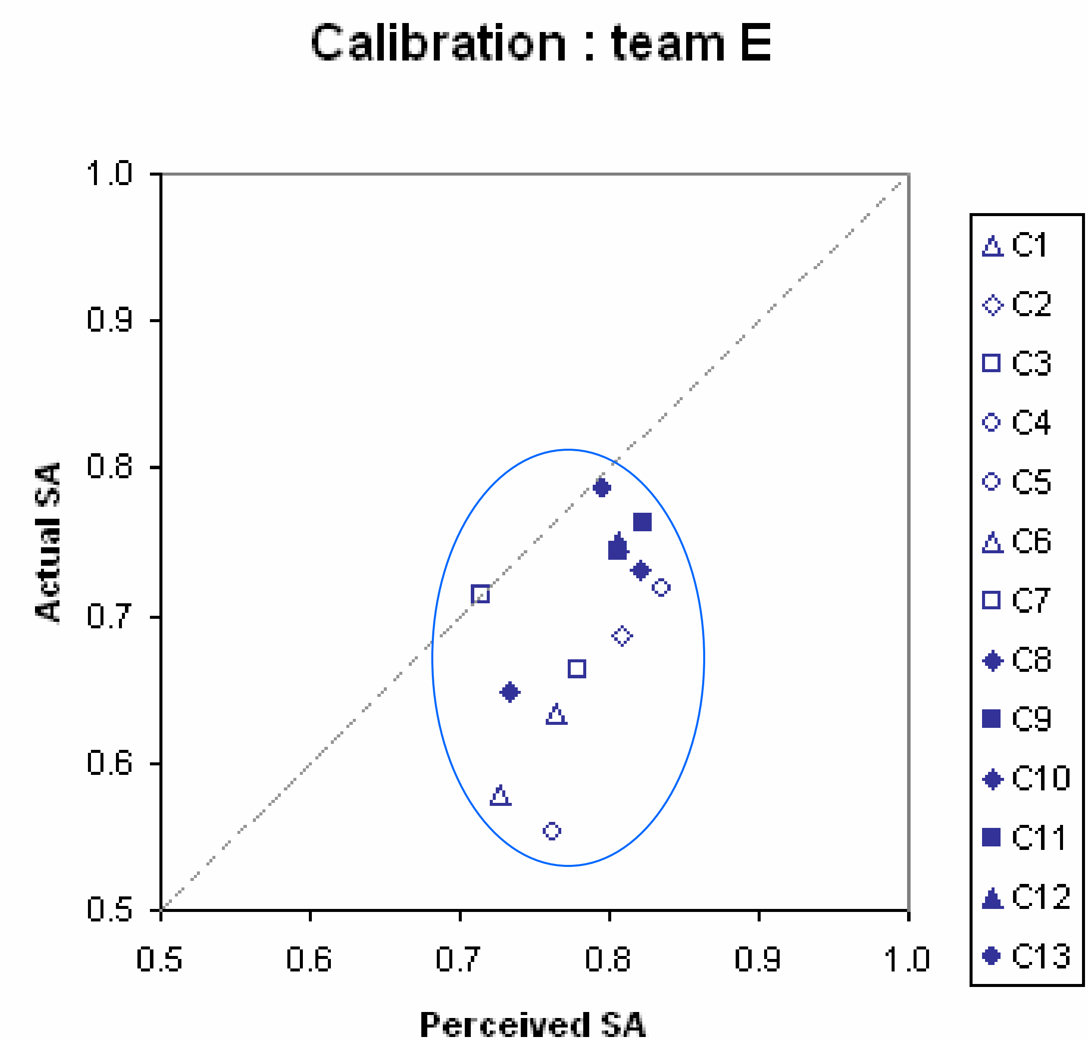
BAE SYSTEMS

QUASA data - LOE 2



BAE SYSTEMS

QUASA data - LOE 2



BAE SYSTEMS

QUASA data - LOE 2

Calibration scores

– using A’ as actual accuracy

	Team (nation)				
	A	B	C	D	E
Perceived accuracy	0.716	0.795	0.803	0.832	0.774
SA accuracy (correct responses)	0.647	0.691	0.656	0.706	0.692
SA accuracy (A’ score)	0.744	0.776	0.737	0.792	0.778
Calibration bias	**- 0.03**	**+0.03**	**+0.07**	**+0.03**	**+0.01**

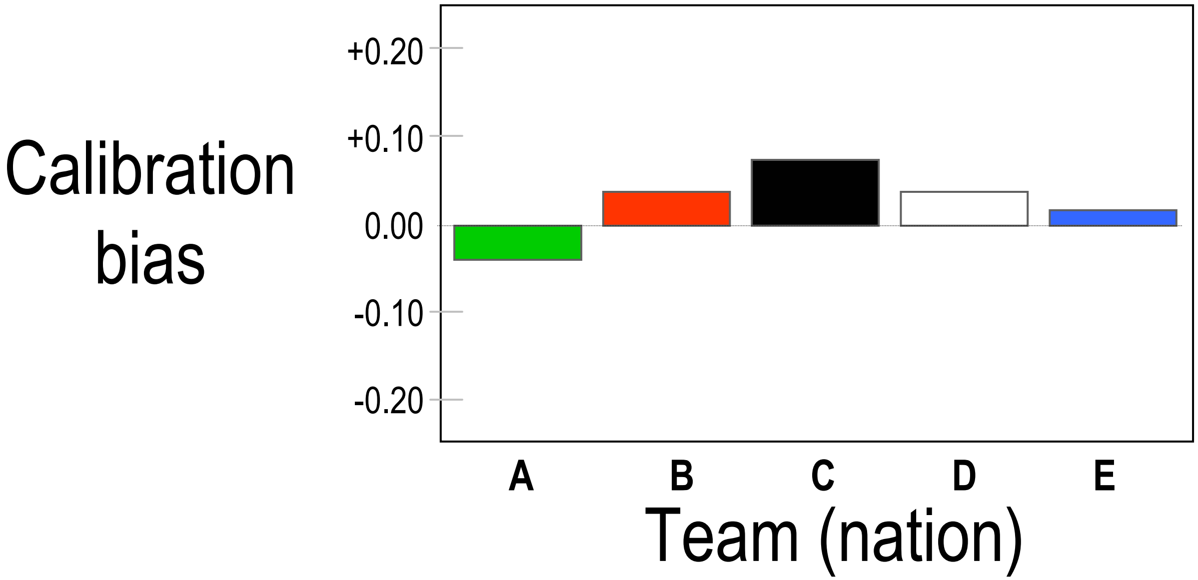
To assess SA calibration, average confidence ratings were transformed (0.5-1.0) and probe accuracy scores (A’, a measure of sensitivity) were subtracted from the result to provide a calibration bias statistic.

BAE SYSTEMS

QUASA data - LOE 2

Calibration scores

– using A’ as actual accuracy



Mean SA probe hit rates per team in LOE 2.

BAE SYSTEMS

QUASA data - LOE 2