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

9th ICCRTS, Copenhagen
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?
Situational Awareness

**Situation**
- Entities, events, states, actions, environment

**Information**
- Perception
- Comprehension
- Projection + Assessment

**Current Awareness**
- Constituent elements of the situation

**Situational Understanding**
- Global characteristics of the situation

**Operational Appreciation**
- Implications of the situation for one's goals/tasks

**Course of Action**
- Decisions/actions informed by SA

Decision-making

9th ICCRTS, Copenhagen
Situational Awareness

Metacognition
Cognitive gaps, conflicts, uncertainties, confidence

Current Awareness

Situational Understanding

Operational Appreciation

Course of Action
Decisions/actions informed by SA

Decision-making

9th ICCRTS, Copenhagen
<table>
<thead>
<tr>
<th>COGNITION</th>
<th>METACOGNITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Fighting in the city has mostly ceased</td>
<td>• This is certain. Current info, very reliable.</td>
</tr>
<tr>
<td>• Column of red tanks is leaving south of the city</td>
<td>• Not sure about this. Reports may not be from reliable source. Need to check.</td>
</tr>
<tr>
<td>• Enemy is beginning retreat</td>
<td>• Confidence in this -- 50-60% Need to look for evidence.</td>
</tr>
</tbody>
</table>

“Actual SA”  “Perceived SA”
Situational Awareness

Strategic

international situation

strategy

Tactical

local situation

next action
Assessment of Situational Awareness

- **Objective Indicators**
  - Performance Indices
  - Behavioural Markers: SABARS
  - Physiological Correlates

- **Subjective Self-Ratings**
  - Unidimensional: SARS, PSAQ
  - Multidimensional: SART, CARS

- **Direct Probes / Queries**
  - Situation Reports
  - Multi-choice Queries: SAGAT
  - True / false Probes
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 a probe response

**A column of enemy tanks is now leaving the city.**

- True
- False
- Very High
- High
- Moderate
- Low
- Very Low
QUASA procedures

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
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
**QUASA in use**

**LOE 2 SA data collection**

---

**Probe 1**
Explosive materials have been found in a storage container at Xxxxxx

<table>
<thead>
<tr>
<th>(a) True or false?</th>
<th>(b) Level of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>◯ TRUE</td>
<td>◯ Very Low</td>
</tr>
<tr>
<td>◯ FALSE</td>
<td>◯ Low</td>
</tr>
<tr>
<td></td>
<td>◯ Moderate</td>
</tr>
<tr>
<td></td>
<td>◯ High</td>
</tr>
<tr>
<td></td>
<td>◯ Very High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Which teams will mostly answer this probe correctly?</th>
</tr>
</thead>
<tbody>
<tr>
<td>◯ A</td>
</tr>
</tbody>
</table>

- True / false probe
- Subjective confidence level
- Perception of other teams’ SA
Analysis of probes data

Contingency table

<table>
<thead>
<tr>
<th>Enemy forces have captured bridge Charlie.</th>
<th>[ T ]</th>
<th>[ F ]</th>
</tr>
</thead>
</table>

Subject’s response

<table>
<thead>
<tr>
<th>Probe type</th>
<th>[ T ]</th>
<th>[ F ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>HIT</td>
<td>MISS</td>
</tr>
<tr>
<td>False</td>
<td>FALSE ALARM</td>
<td>CORRECT REJECTION</td>
</tr>
</tbody>
</table>
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
Signal Detection Theory

Non-signals (Noise)

Internal response strength

REJECT

ACCEPT
Signal Detection Theory

- **Noise**
- **Signals + noise**

**Internal response strength**

- low
- high
Signal Detection Theory

- Noise
- Signal + noise

- REJECT
- (uncertain)
- ACCEPT

9th ICCRTS, Copenhagen
### Signal Detection Theory

- Contingency table: 4 possible outcomes

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>RESPONSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Accept: HIT</td>
</tr>
<tr>
<td>Non-signal</td>
<td>Accept: FALSE ALARM</td>
</tr>
</tbody>
</table>
Signal Detection Theory

- Contingency table — 4 possible outcomes

```
<table>
<thead>
<tr>
<th></th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Non-signal</td>
<td>12</td>
<td>88</td>
</tr>
</tbody>
</table>
```

- Hit rate = 0.80
- Miss rate = 0.20
- False Alarm rate = 0.12
- Correct Rejection rate = 0.88
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

- **CORRECT REJECTIONS**
  \[ P(CR) = 1.00 \]

- **MISSES**
  \[ P(M) = 0.40 \]

- **HITS**
  \[ P(H) = 0.60 \]
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

Noise

Signal + noise

low

high

9th ICCRTS, Copenhagen
Signal Detection Theory

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

\[ d' = Z(H) - Z(FA) \]
\[ d' = 4.00 \]
Signal Detection Theory

Noise

Signal + noise

low

high
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

\[ C = k - \frac{d'}{2} \]

\[ C = 2.16 - 2.00 = 0.16 \]

9th ICCRTS, Copenhagen
**Signal Detection Theory**

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

- $\beta = f_s(k)/f_N(k)$
- $\beta = \exp^{d'C}$
- $\beta = 1.38$

**Likelihood ratio of probability densities of the two distributions at the criterion**

- $\log \beta = \frac{1}{2}(Z^2(FA) - Z^2(H))$
- $\log \beta = d'C$
- $\log \beta = 0.32$

9th ICCRTS, Copenhagen
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
Example

- Compare two subjects (LOE 2)

![Diagram showing responses for two subjects: Subject A and Subject B. The diagram includes categories for True and False responses, with counts for Hits, Misses, False Alarms, and Correct Rejections.](image-url)
Receiver Operating Characteristic (ROC)

Hit rate = 0.80
FA rate = 0.10

Hit rate = 0.90
FA rate = 0.75

9th ICCRTS, Copenhagen
Subject A criterion = close to neutral
Subject B criterion = strong liberal bias
**ROC — Sensitivity**

The graph illustrates the Receiver Operating Characteristic (ROC) curve, which is a graphical plot used inbinary classification problems. It shows the trade-off between the true positive rate (hit rate) and the false positive rate (false alarm rate) for different thresholds.

- **Point A**: Hit rate = 0.80, FA rate = 0.10, $d' = 2.00$
- **Point B**: Hit rate = 0.90, FA rate = 0.75, $d' = 0.60$

The isosensitivity contour map helps in visualizing different sensitivity levels, with $d'$ representing the distance between the normal distribution curves.
Team A has highest hit rate ...
QUASA data – LOE 2

SA probe sensitivity

But team A is no more accurate overall at discriminating true from false probes.
Team A is very liberal when uncertain (inclined to accept probes as true) -- hence the high hit rate.
QUASA data - LOE 2

ROC curve: grouped by teams

Hit rate \( P(H) \)

False Alarm rate \( P(FA) \)

LOE 2 teams
- Team A
- Team B
- Team C
- Team D
- Team E

9th ICCRTS, Copenhagen
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
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 correlate with actual accuracy.

- **Intelligence analysis**
  - Don’t want overconfident intelligence reports based on dubious data

- **Situational awareness**
  - Accidents attributed to overconfidence in poor/inaccurate SA
Calibration Curve

- Actual accuracy (%)
- Perceived accuracy (%)

- Under-Confident
- Well-calibrated
- Over-Confident

9th ICCRTS, Copenhagen
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).

Source
QUASA data – LOE 2

SA response confidence ratings

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

ROC curve: hypothetical confidence levels

Hit rate
\[ P(H) \]

False Alarm rate
\[ P(H) \]

Confidence ratings:
- very high
- high
- moderate
- low
- very low

9th ICCRTS, Copenhagen
ROC / confidence calibration

- **ROC curve**: Idealised confidence levels
- **Hit rate (P(H))** vs. **False Alarm rate (P(H))**
- **Confidence ratings**:
  - very high
  - high
  - moderate
  - low
  - very low

- **Certain**
- **Guessing**
ROC/confidence calibration

ROC curve: observed confidence levels

Suggestive of overconfidence when guessing

Confidence ratings
- very high
- high
- moderate
- low
- very low
**LOE 2 calibration analysis**

**Calibration scores**
- using hit + correct rejection rates as actual accuracy

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived accuracy</strong></td>
<td>0.716</td>
<td>0.795</td>
<td>0.803</td>
<td>0.832</td>
<td>0.774</td>
</tr>
<tr>
<td><strong>SA accuracy (correct responses)</strong></td>
<td>0.647</td>
<td>0.691</td>
<td>0.656</td>
<td>0.706</td>
<td>0.692</td>
</tr>
<tr>
<td><strong>Calibration bias</strong></td>
<td>+0.07</td>
<td>+0.11</td>
<td>+0.15</td>
<td>+0.13</td>
<td>+0.08</td>
</tr>
</tbody>
</table>

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.
LOE 2 calibration analysis

Calibration scores

Mean SA probe hit rates per team in LOE 2.
LOE 2 calibration analysis

Calibration curve

Actual SA (P[correct])

Perceived SA (confidence ratings)

Under-Confident

Over-Confident

LOE 2 teams
- Team A
- Team B
- Team C
- Team D
- Team E

9th ICCRTS, Copenhagen
IOE 2 calibration analysis

Individual Calibration: team E

- Actual SA vs Perceived SA
- Data points for individuals C1 to C13

ICCRTS, Copenhagen
QUASA - 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
Conclusions

- QUASA yields potentially insightful quantitative results
- T/F probes analysed with SDT provide a measure of actual SA
- Probe confidence ratings provide a measure of perceived SA
- Calibration analysis compares actual SA with perceived SA
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
BACKUP SLIDES
Lessons learned

- **T/F probes need objective referent (‘ground 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
## LOE 2
### Information Sharing Agreements

<table>
<thead>
<tr>
<th>Country</th>
<th>ML</th>
<th>TL</th>
<th>BL₁</th>
<th>BL₂</th>
<th>Coalition</th>
<th>Private</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
</tbody>
</table>