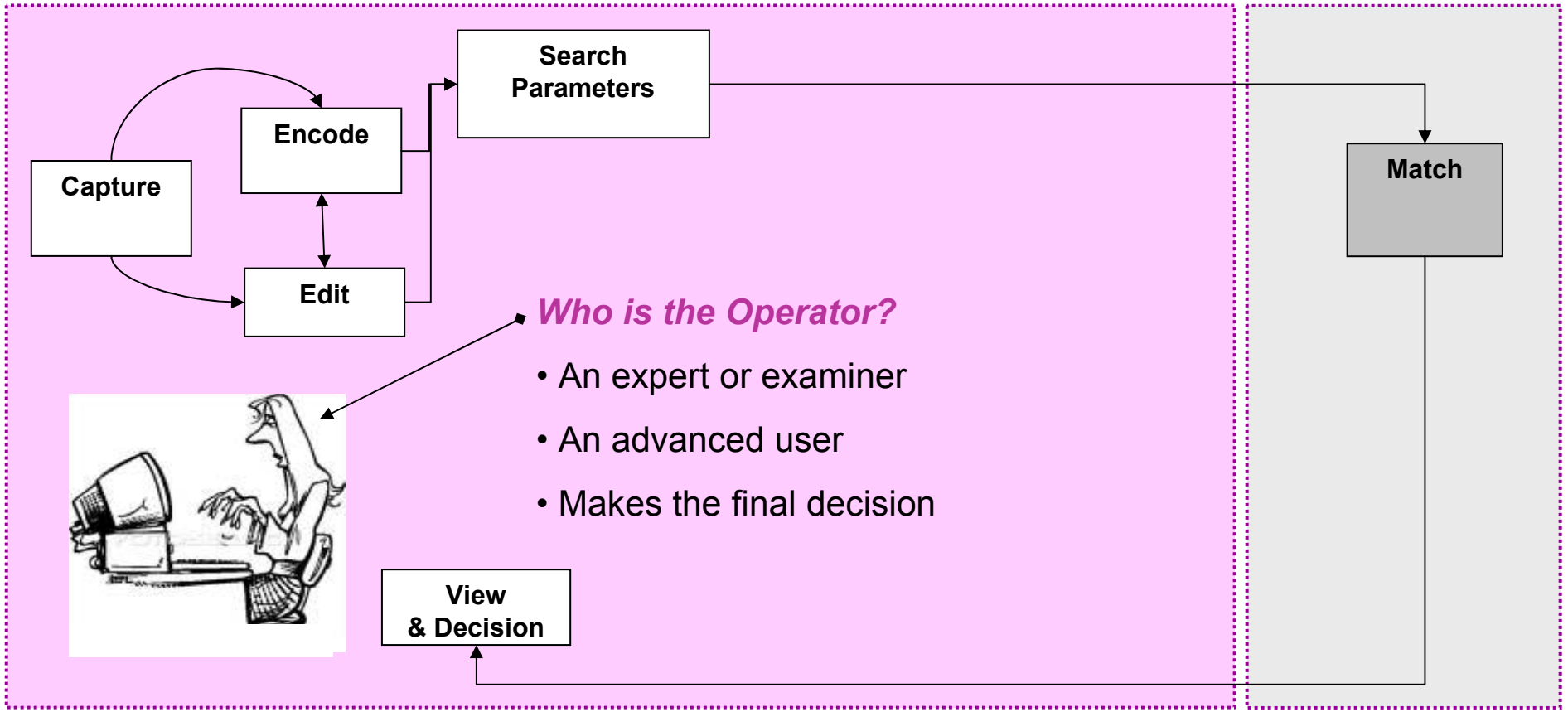
Evaluating Examiner/Operator-Led Biometric Applications

Ambika Suman
Biometric Technologies Manager
National Policing Improvement Agency
March 2010

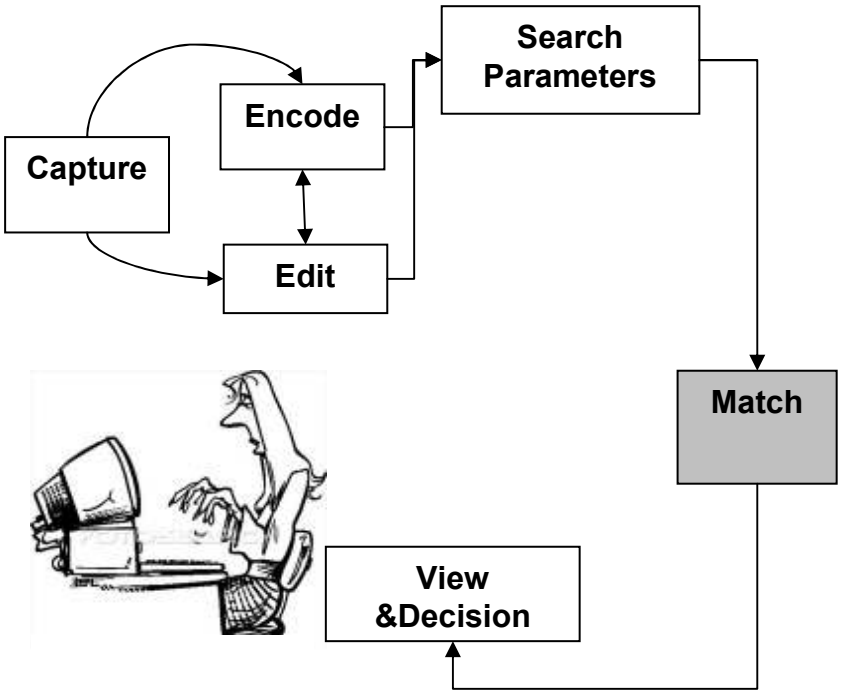
What is an Examiner/Operator-Led Application?

— *A biometric system that is reliant on the interaction and skill of a human operator ror one or more stages of the overall biometric search process – that is, from data capture, through to enrolment, template generation and final decision*

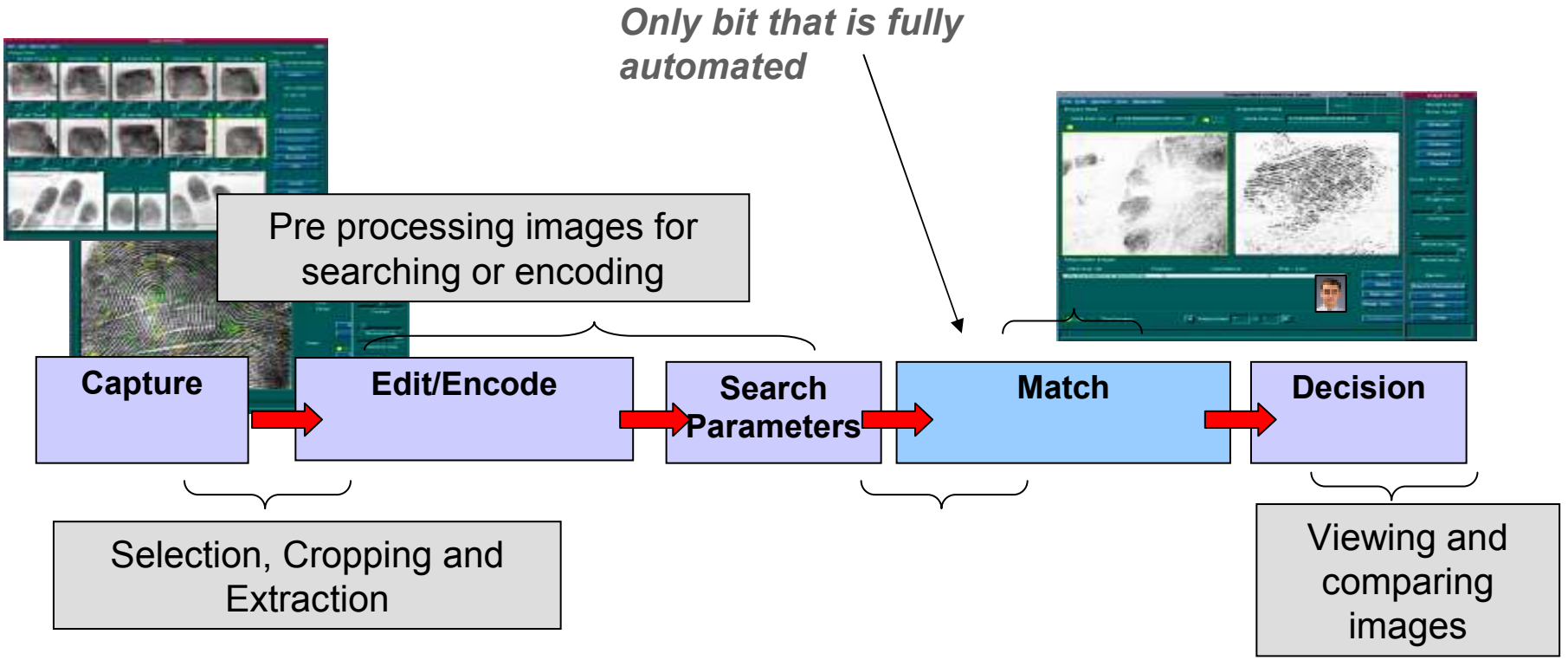


Where would these be used?

- Non real time /offline
- Back Office Checking Systems
- Forensic Applications
  - E.g. Fingerprint, Face, Voice/Speaker, Signature analysis
- 1:many or 1:1 searches



Example: Forensic Fingerprint Systems



All fingerprint evidence that goes to court must be validated by at least 2 human experts

A Real World Example:
IDENT1 – National Palm and Fingerprint searching across UK

- 8.1 million subjects
- 17.8 million fingerprint sets (rolls, flats and multiples)
- 7.9 million palm prints
- 1.8 million un-id crime scene marks, 160,000 from palms
- ***48 000 crime scene mark id's within last 6 months***
- **Over 1200 fingerprint expert operators across the UK**
- All fingerprint evidence used for court checked by at *least* 2 experts

What is the value of such an approach?

NPIA
National Policing Improvement Agency

Minimises the labour of manual processes

Help manage and present the data

Compensates for poor quality data

Improves and compliments the performance of both

Expertise – leverages human perception/ brain power!

All fingerprint evidence used for court checked by at *least* 2 experts

Political/ legally motivated requirement?

Why is it important to address such systems specifically?

- Understand the role of the operator together with the system to **influence** its **design** and **performance**.
  - *What value does the system provide to the operator?*
  - *What value does the operator's skill provide?*
  - *Not measuring their skill but how they employ it!!*

- "***Educated Users***" - Expert's perception of system performance may affect their own **decision** making

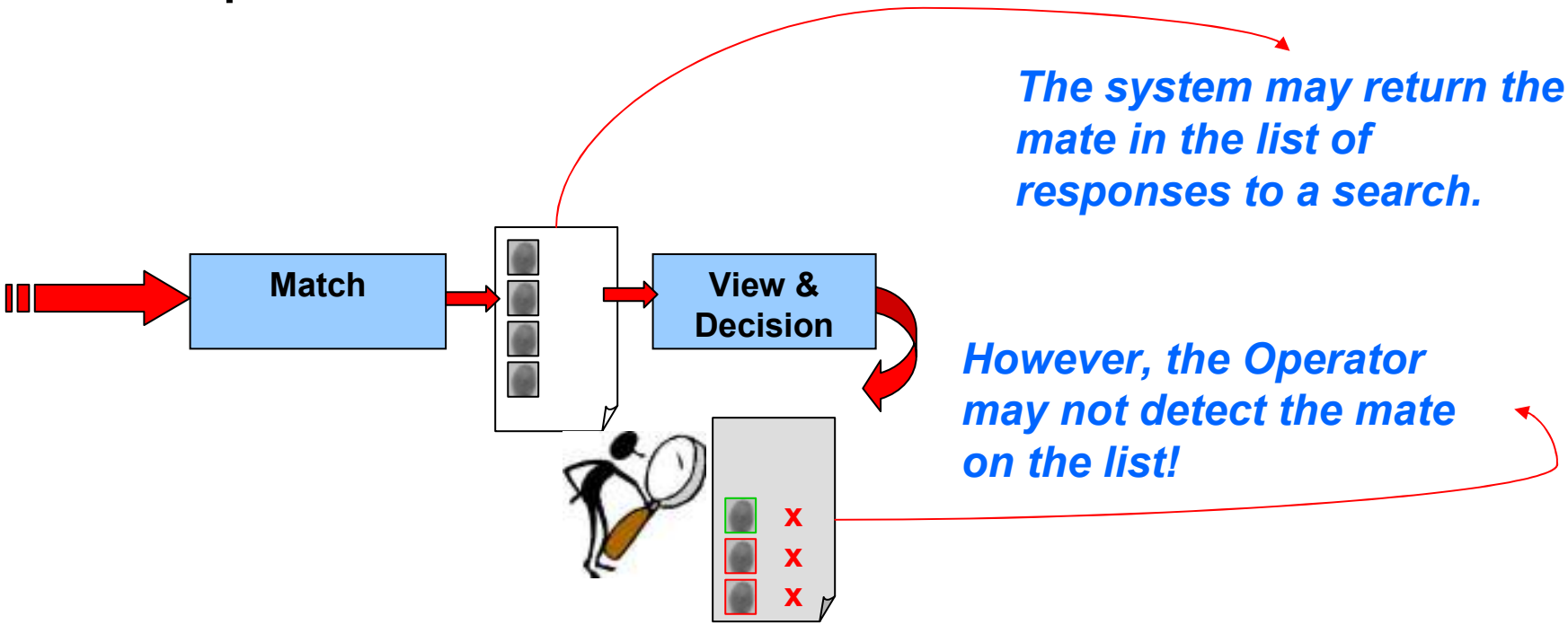
- Key to obtaining **user acceptance** of the technology
  - No confidence – no use for the application!

Examples of unique factors to consider

- ***Measuring Accuracy Performance***
- ***Reporting Accuracy Performance (for 1:Many applications)***
- ***Controlling Test Variables***
- ***Example of Designing a test***

1. How Accuracy Performance is Measured

Measuring the true accuracy must include the operator's decision!

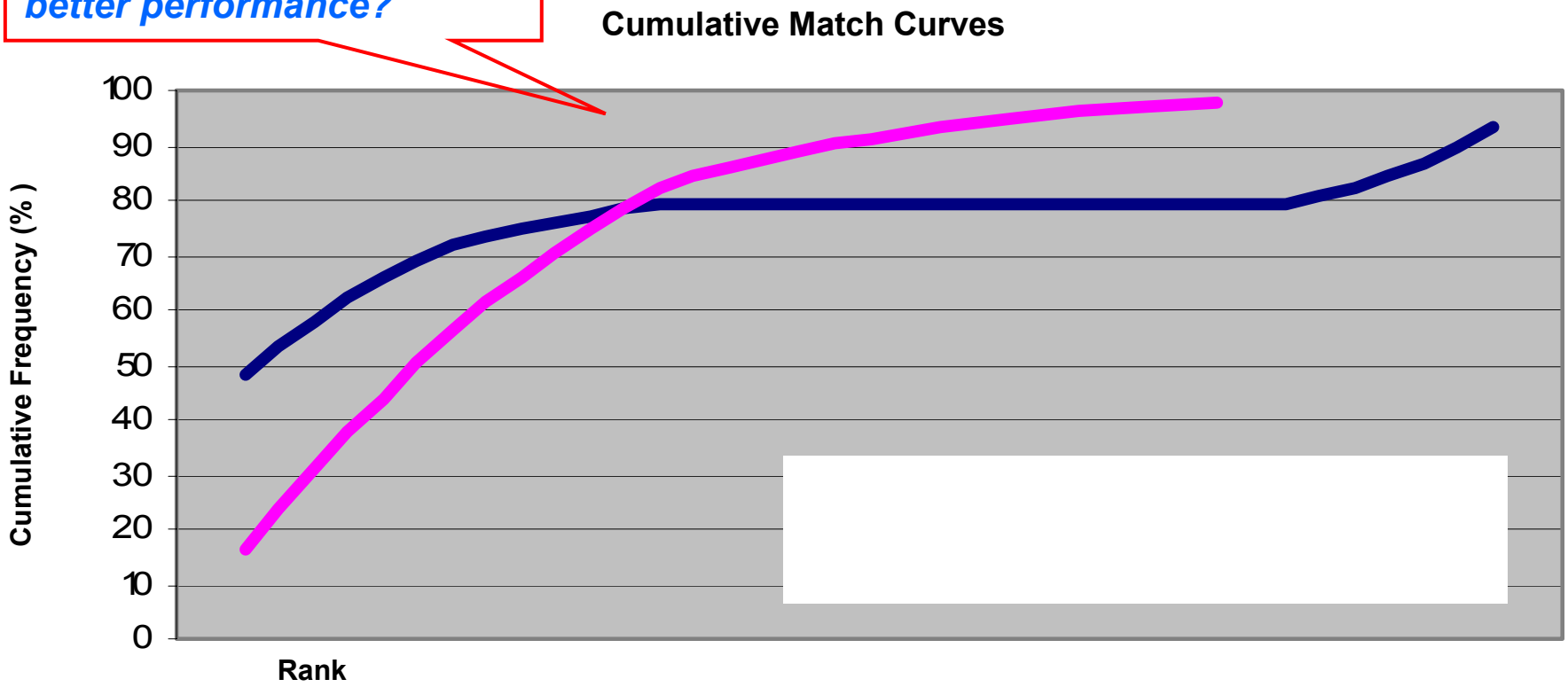


The system may return the mate in the list of responses to a search.

However, the Operator may not detect the mate on the list!

2. How Accuracy Performance is Reported
(1:Many Closed set)

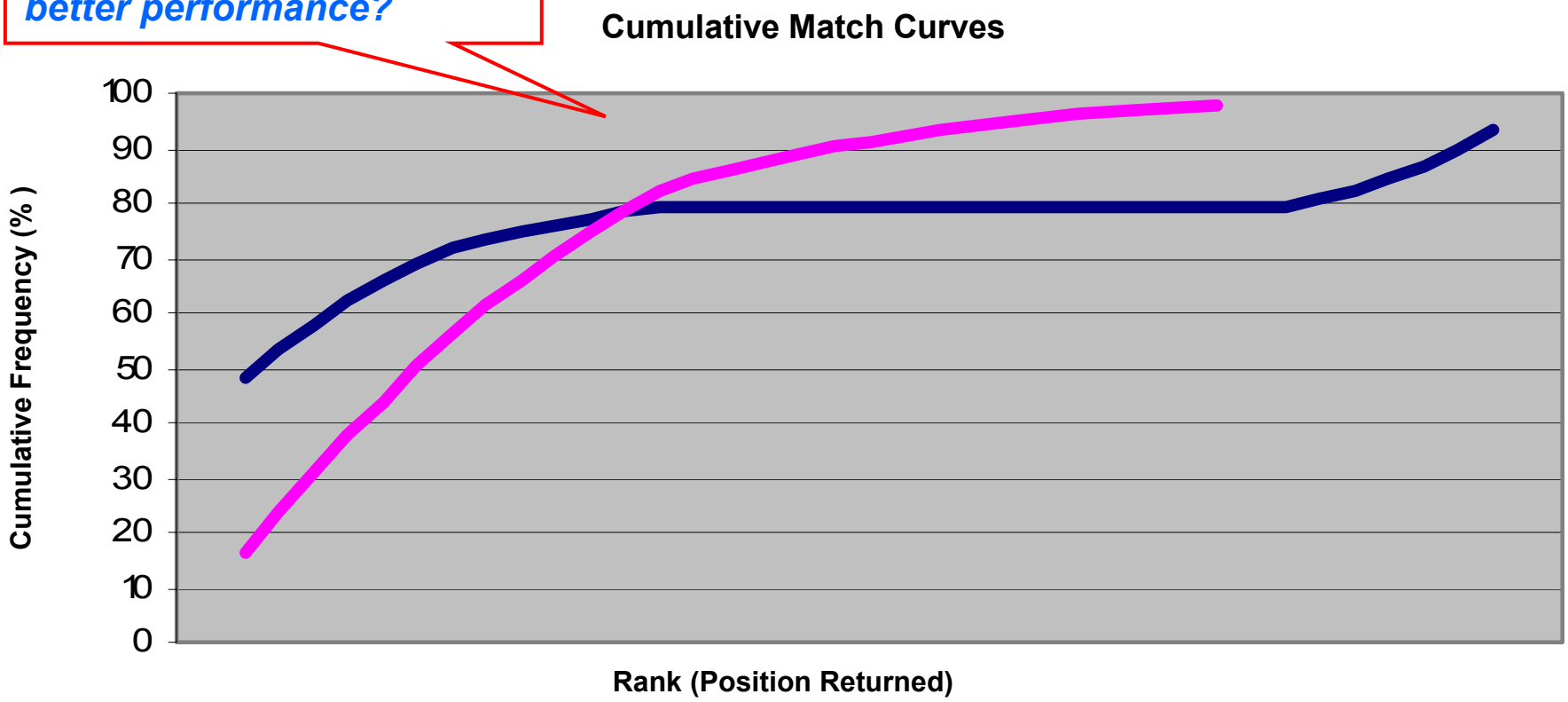
Q: Which CM Curve shows better performance?



2. How Accuracy Performance is Reported

(1:Many Closed set)

Q: Which CM Curve shows better performance?

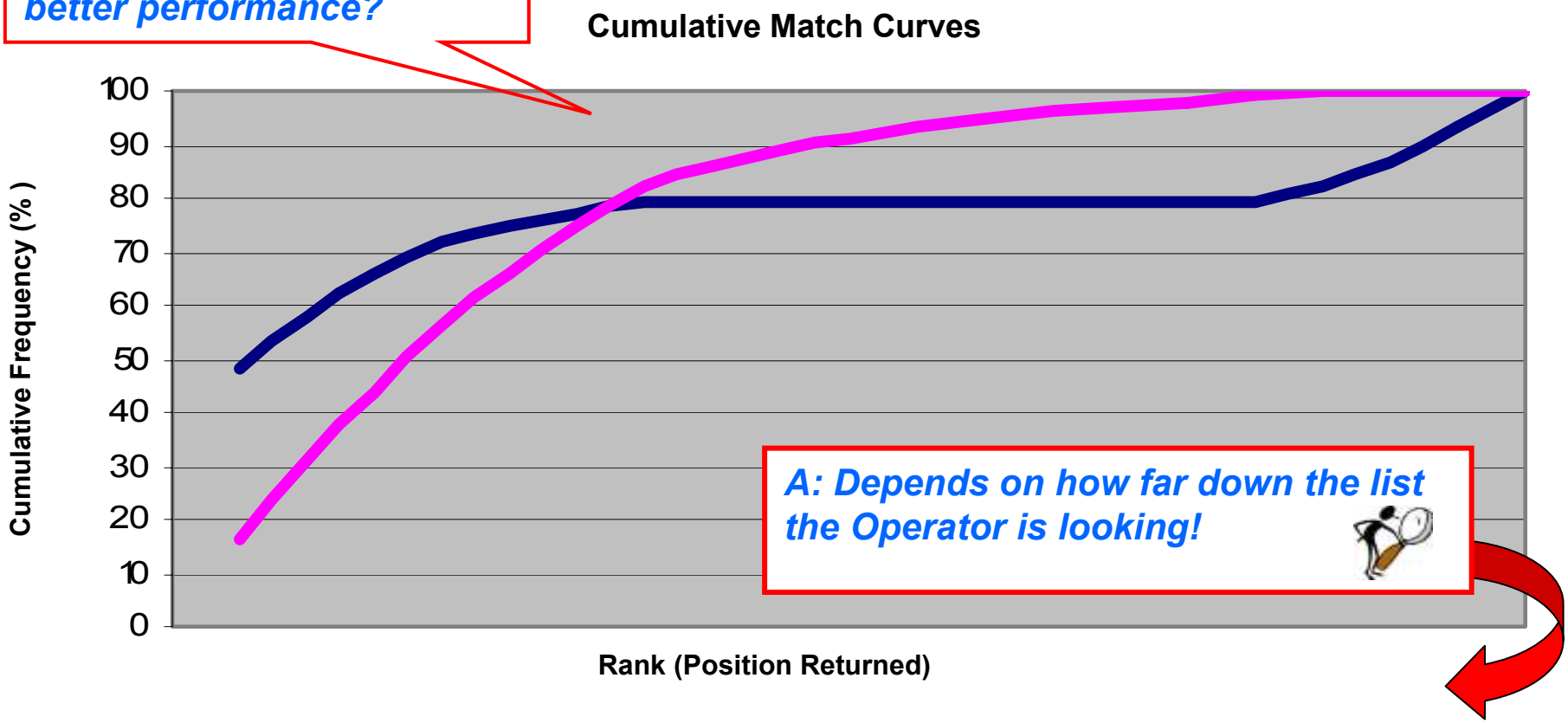


2. How Accuracy Performance is Reported

(1:Many Closed set)

NPIA National Policing Improvement Agency

Q: Which CM Curve shows better performance?



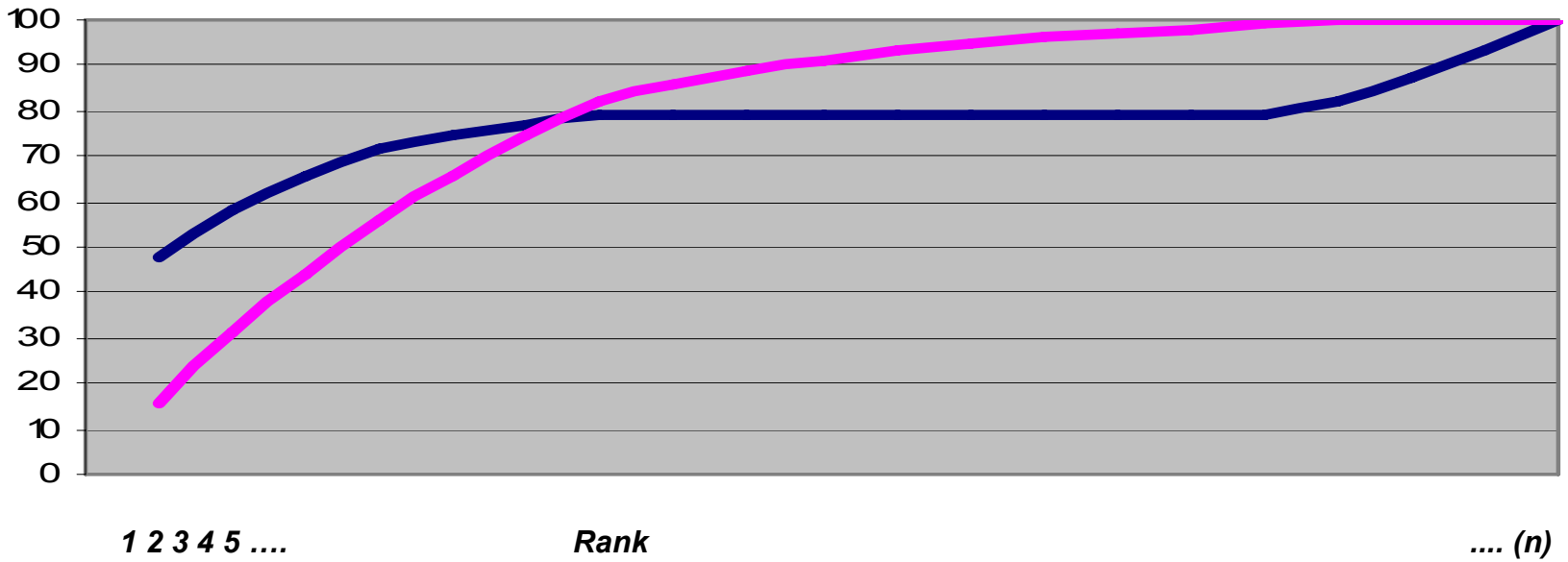
2. How Accuracy Performance is Reported

(1:Many)

Paired functions = {Rel(n) , Sel(n)}

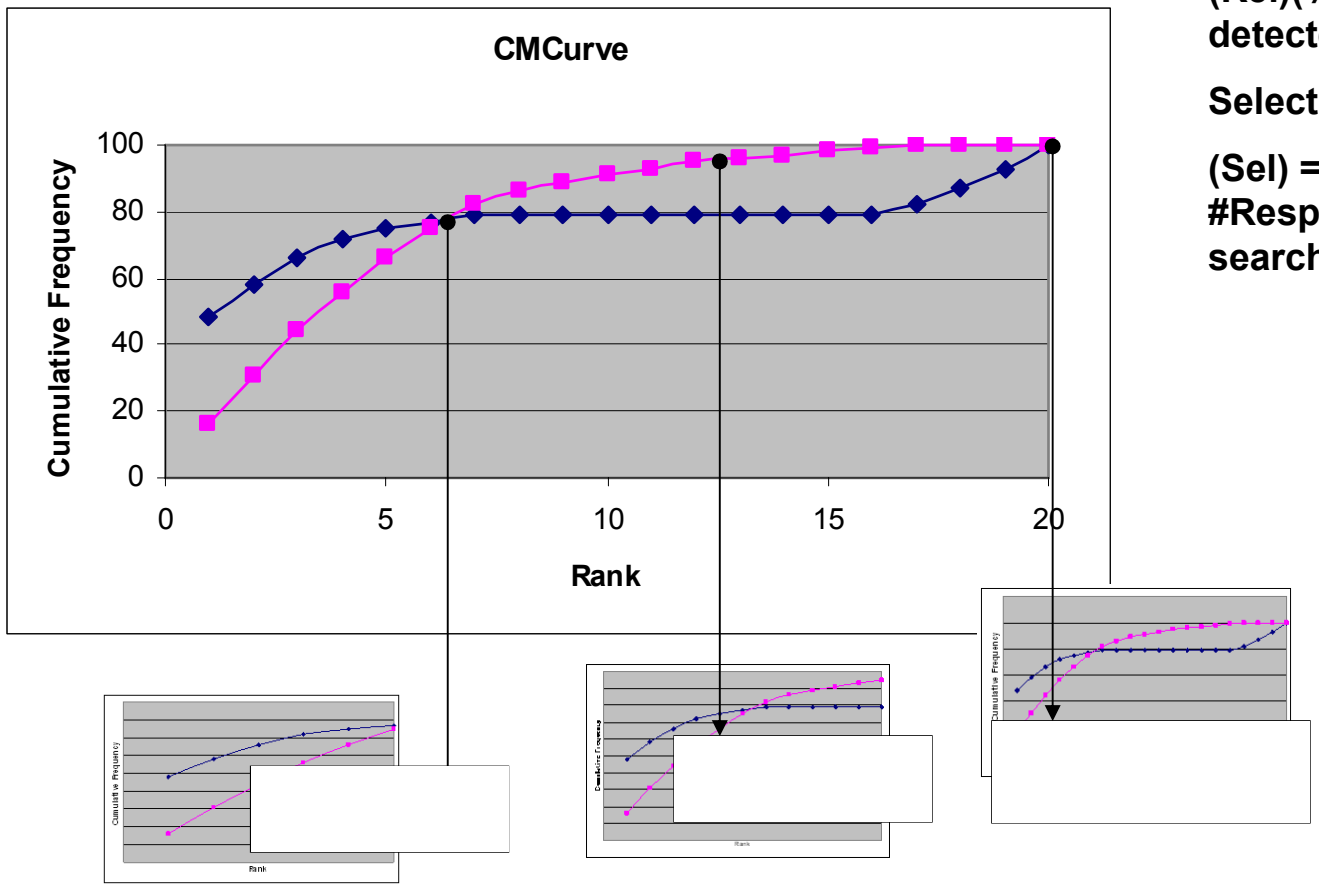
Reliability (Rel)(%) = *True Mates detected within the list of length (n)*

Selectivity (Sel) = *Mean number of responses compared by an Operator within a list of length (n)*



2. How Accuracy Performance is Reported

Example



Reliability:

(Rel)(%) = True Mates detected

Selectivity:

(Sel) = Average #Responses viewed per search

Examples of unique factors to consider

- *Measuring Accuracy Performance*
- *Reporting Accuracy Performance (for 1:Many applications)*
- ***Controlling Test Variables***
- ***Example of Designing a test***

Common Considerations when Controlling Variables

Data:

- –Biometric Search and Match Pairs
- **–Biometric data pre-processed with operator input**
- –Background data

Operators:

- **–Required Expertise**
- **–Available as Test Operators? (Day job?)**
- **–Training - to use the technology**
- **–Perception/Behaviour/ Subjective judgement**

Test Environment

- –Live System? Test Bed?
- –Operational Scale
- **–Operator Behaviour Altered?**

Scope & Design

- –End to end solution
- –Sub-process/component **e.g. HCI**
- **–Testing human or machine or both?**
- –Controlling variables
- – Repeatable/Operationally reflective?
- –White/Black box

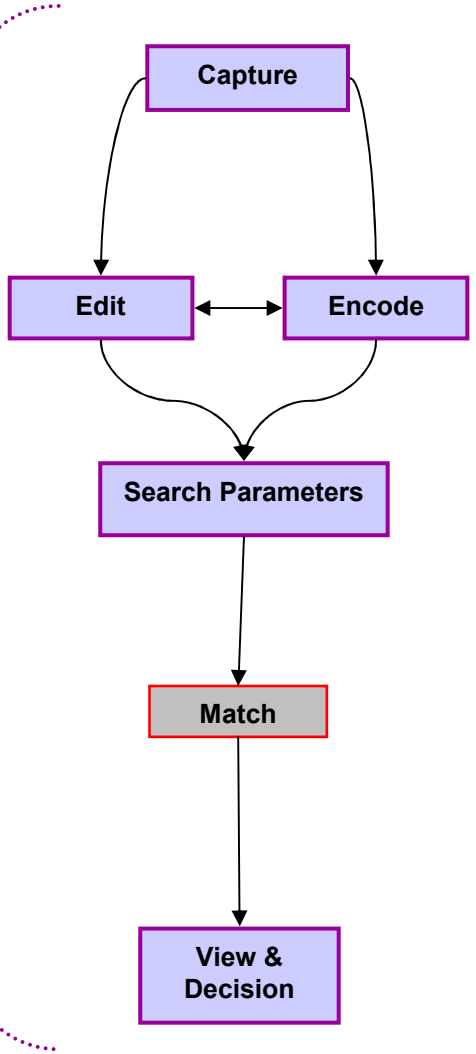
Results Reporting

- **–Interpretation of Data - Operator Decision**
- **–Quantify Value added benefit to manual process**
- –Comparable

Most significant and difficult set of variables to control are those related to the operator

Operator Induced Variables

- Input/interactions will differ between operators as well as for the same operator
- Varying Expertise
  - Specify Search parameters/filters – e.g. Palm ROI, finger mask, pattern
  - Decisions/Judgements – subjective (not always accurate)
  - Make use of other data (e.g. application form, case notes)
- Varying Approaches/Behavior
  - May alter when tested
  - Not testing their expertise but how they employ it!
- Training
  - Proficiency
  - HCI toolsets
- Business procedures
  - Prioritise searches differently
  - Variations in Operator effort
- Test Environments
  - Confidence in system
  - Perception

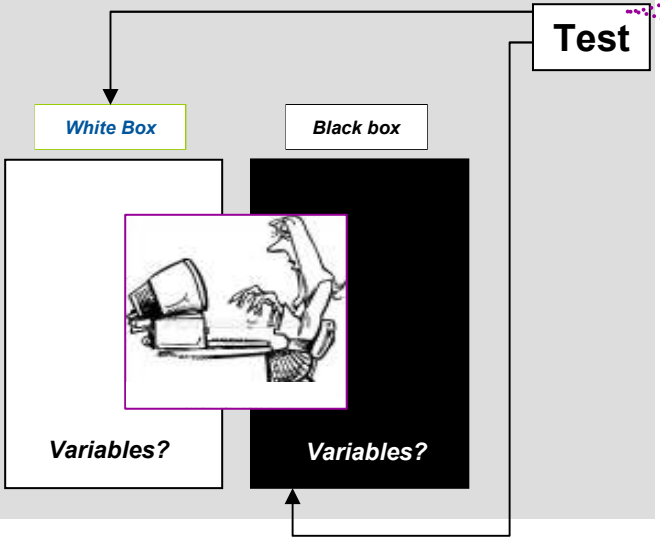


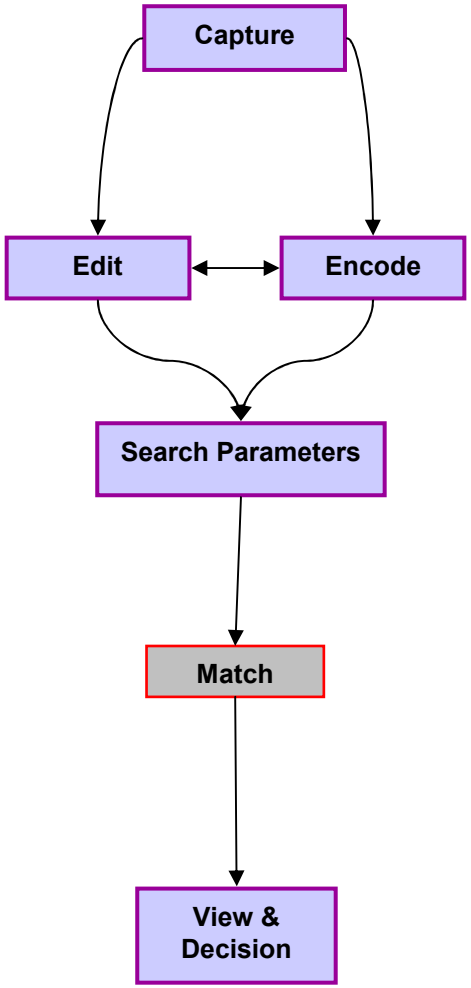
Controlling Variables

Types Testing

- ***Application***
  e.g. Design and develop components or processes within the system?
  - ***Repeatable Tests***
  - ***Biometric data pre-processed with operator input***
- ***Scenarios to define business/operator workflows***
  e.g. Monitoring of operations throughout life of service
  - ***Need operators to reflect reality but the solution and how they behave may change in reality***
- ***Operational performance***
  e.g. benchmarking different solutions
  - ***Realistic***
  - ***Real data***
  - ***Live System***
  - ***Real Operators***

For each testing approach one needs to know what variables to control and how it affects the measured performance.





Examples of unique factors to consider

- *Measuring Accuracy Performance*
- *Reporting Accuracy Performance (for 1:Many applications)*
- *Controlling Test Variables*
- ***Example of Designing a test***

Design a Test!

Background:

An existing Operator led forensic finger and palm searching system is to be replaced by a new solution provided by a different supplier.

Task:

There is a need to demonstrate the accuracy of the replacement solution against the existing AFIS

- *to show no loss in performance to users*
- *And baseline the performance of the new service.*

Approach:

To design and execute a fair benchmark test to compare the accuracy of a legacy system against a new solution that will replace it after contract award.

Design A Test!

Testing New Vs Legacy System

New solution is partially developed only -
Legacy system is complete.
Limited to Algorithm tests?

Design of operator business process to be defined based on new solution
e.g. HCI functions to be developed with user
Operators already familiar with existing toolsets.

Need to maintain a level playing field
Need to discriminate between solutions
Gain user acceptance

Limitations of tests

Algorithm/Application Testing:
Doesn't account for the operators' role.
Cannot guarantee performance.
Unlikely to be based on operationally reflective data
Cannot compare solutions

Scenario Testing:
Must reflect approaches of different between systems
Not testing like for like if tests are not identical
New Solution tested may change from what is delivered
Limited scope

Operational Testing:
Realistic solution - after contract award
Real Data, Real Operators, Live System/Test bed
Everything is Uncontrolled - May not discriminate between systems! What is the difference in performance attributed to? The Operator or the new system?

Volume of Searches depends on…

- *…Constraints*
  - *How many searches are required to significantly discriminate between the performance of the two solutions?*
  - *How many test operators can be spared from their daily work to do these tests?*
  - *How long does it take for the operators to process and view searches?*
  - *How much time is there to execute the tests?*
  - *Cost - proportionate to the purpose/benefit?*

  These constraints limit the scale of the test and what you can infer from the results.

Summary

- The performance of Operator led systems is equally reliant on the algorithm and interaction of a human operator
- Where poor data quality and other algorithmic limitations remain to be a challenge Operator led approaches are necessary.
- Forensic AFIS and Back Office comparisons are key examples
- Addressing the unique challenges described for these applications is important to the design, development and user acceptance of Operator led approaches across modalities.
- Operator induced variables are the most significant and difficult to manage
  - Impact on how performance is measured and reported
    - E.g. Reliability and Selectivity
  - Operators' Perception
  - Managing varying operator behaviours, expertise, training, interactions
- Practical limitations with test approaches – application, scenario and operational tests
- Example of designing test – highlight challenges and give food for thought!
  - though no test is identical. No rule fits all!

Contact Details:

Ambika Suman
Biometric Technologies Manager

National Policing Improvement Agency

New King's Beam House
22 Upper Ground
London SE1 9QY
Tel: +44(0) 20 8358 5446
Mobile :+44(0) 7881 673 890
Email: ambika.suman@npia.pnn.police.uk