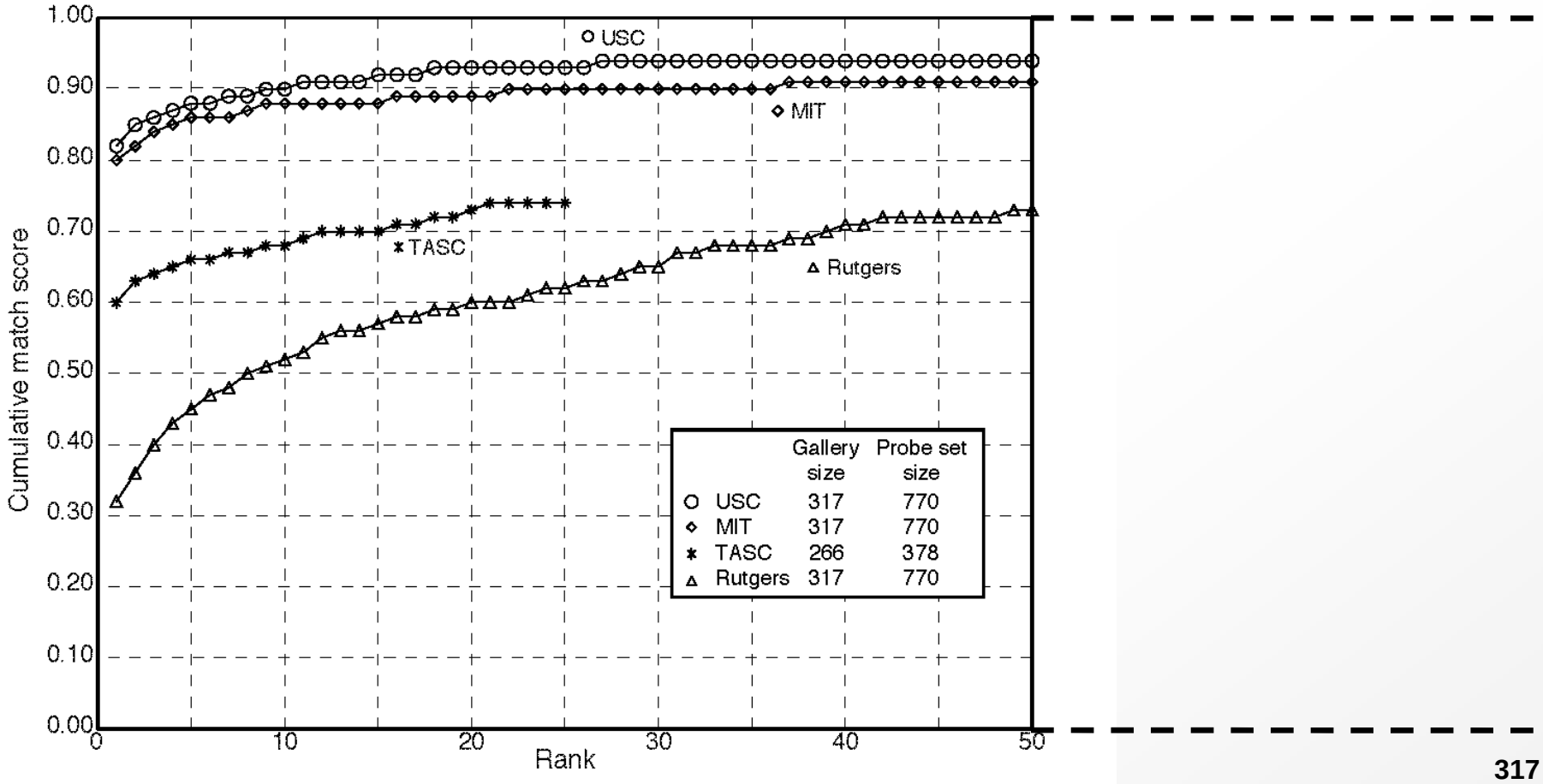
Using Challenge Problems to advance the Technology and Science of Biometrics

Dr. P. Jonathon Phillips

National Institute of Standards and Technology

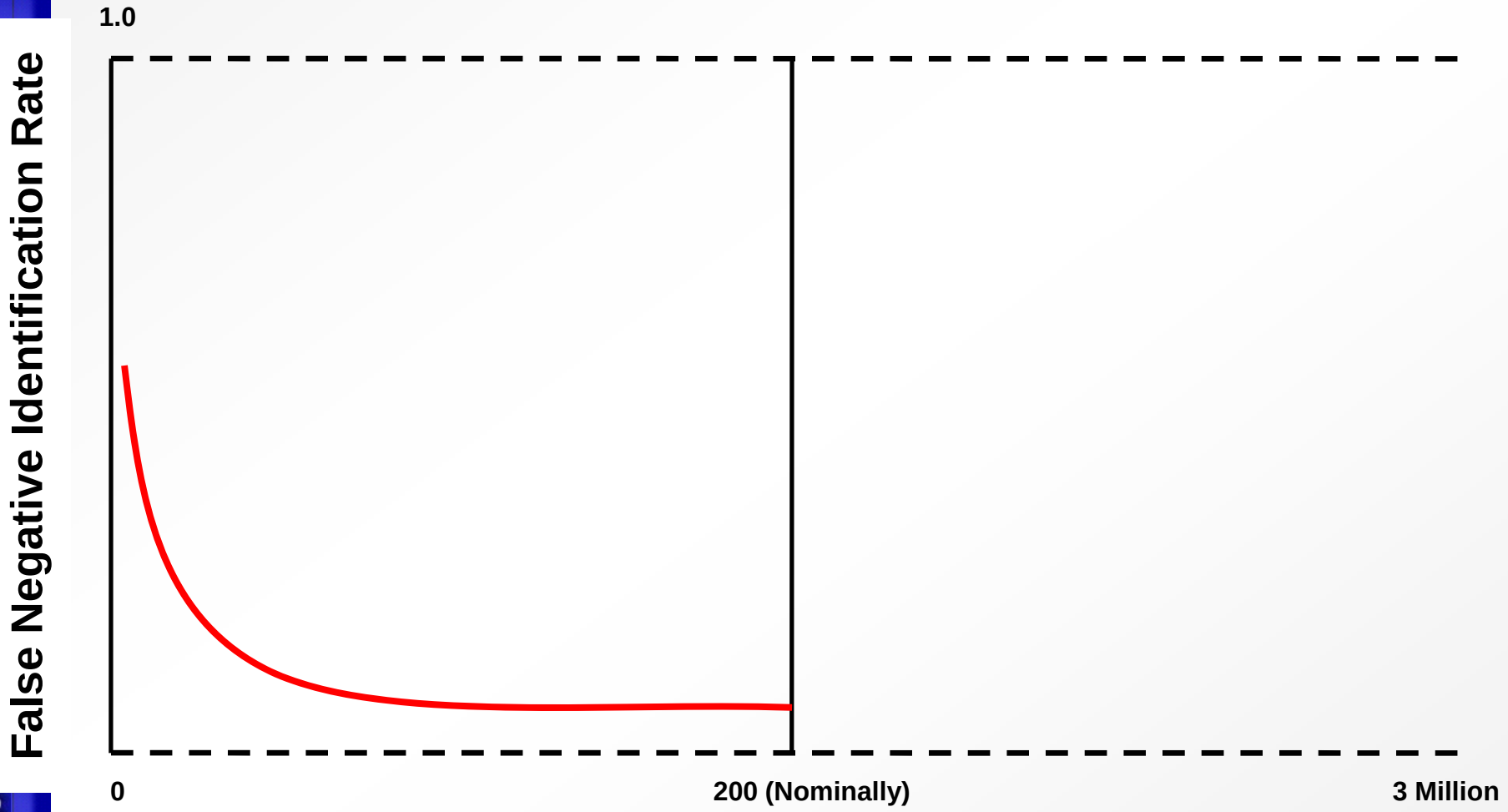
August 1994
FERET Large Gallery Results



317

Performance for top 50 matches out of 317

Sixteen Years Later
MBE 2010 Still Face Track



Performance for top 200 matches out of 3 million

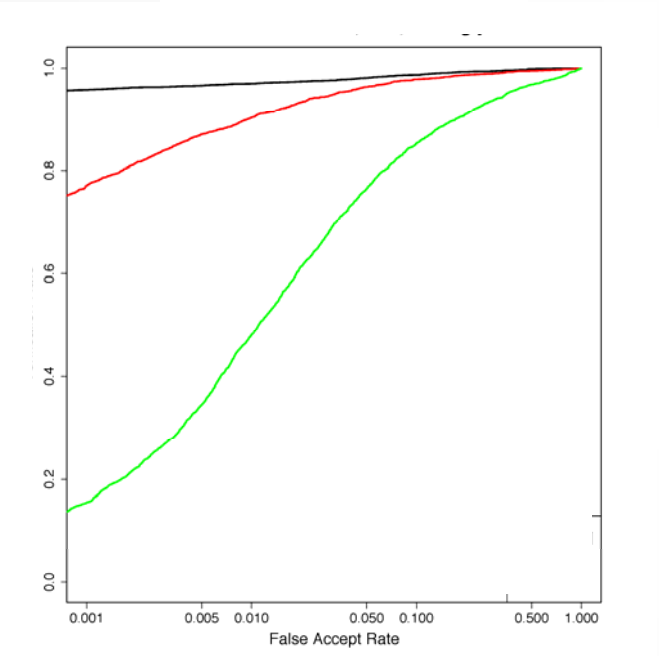
Orders of Magnitude Increase in Knowledge

- Number of people in the field
- Sophisticated test protocols
- Regular Challenge Problems
- Regular Evaluations
- 200,000+ Biometric samples available
- Conferences, Workshops, Journals
- Statistical methods
- Standards
- Human performance comparisons and baselines
- National (US) and International participation

Predicting Performance

Target Set

Query Set



New Target Set

New Query Set

Levels of Predictions

- General Assessment

Levels of Predictions

- General Assessment
- Measuring Improvement

Levels of Predictions

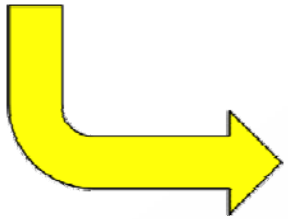
- General Assessment
- Measuring Improvement
- Ranking of Algorithms
  - Relative performance
  - Ranking stable across data sets
  - Limited success

Levels of Predictions

- General Assessment
- Measuring Improvement
- Ranking of Algorithms
  - Relative performance
  - Ranking stable across data sets
  - Limited success
- Predict Performance

Levels of Predictions

- General Assessment
- Measuring Improvement
- Ranking of Algorithms
  - Relative performance
  - Ranking stable across data sets
  - Limited success
- Predict Performance



Now that's a Challenge (Problem)

Formulated as a Challenge Problem

Twenty Questions

Examples of Questions

- Distribution of
  - Gender
  - Race
  - Age
- Distribution of quality measures
  - Focus
  - Illumination
  - Iris area
  - Number of minutia

A Challenge Problem in the Form of Twenty Questions

- Set up:

- Evaluator
  - Algorithm A
  - Set of sequestered images
    - Divided into target and query
  - ROC; FRR and FAR at a fixed threshold

- Participant
  - Algorithm A

A Challenge Problem in the Form of Twenty Questions

- *k* – Questions:
- Participant
  - Asks *k* questions
- Evaluator
  - Provides answers

A Challenge Problem in the Form of Twenty Questions

- End:

- Participant
  - Submits estimated $\widehat{\text{ROC}}$; $\widehat{\text{FRR}}$ and $\widehat{\text{FAR}}$ at fixed threshold

- Evaluator
  - Measure accuracy of estimated $\widehat{\text{ROC}}$; $\widehat{\text{FRR}}$ and $\widehat{\text{FAR}}$

Types of Questions I

- Questions allowed I:

$g($  $) \rightarrow R$

- Answer provided:
  - Empirical Cumulative Distribution Function (ECDF) over target and query

$\hat{F}_g(x)$

Types of Questions II

- Questions allowed II:

- Answer provided:
  - Joint ECDF between target and query

$$\hat{F}_g(x_T, x_Q)$$

Query Set

Target Set

Types of Questions III

- Questions allowed III:

g() → R

h() → R

- Answer provided:
  - Joint ECDF between *h* and *g*

$$\hat{F}_{g,h}(x,y)$$

Example from FRVT 2006 Uncontrolled vs. Uncontrolled

Target

9803 Images

Query

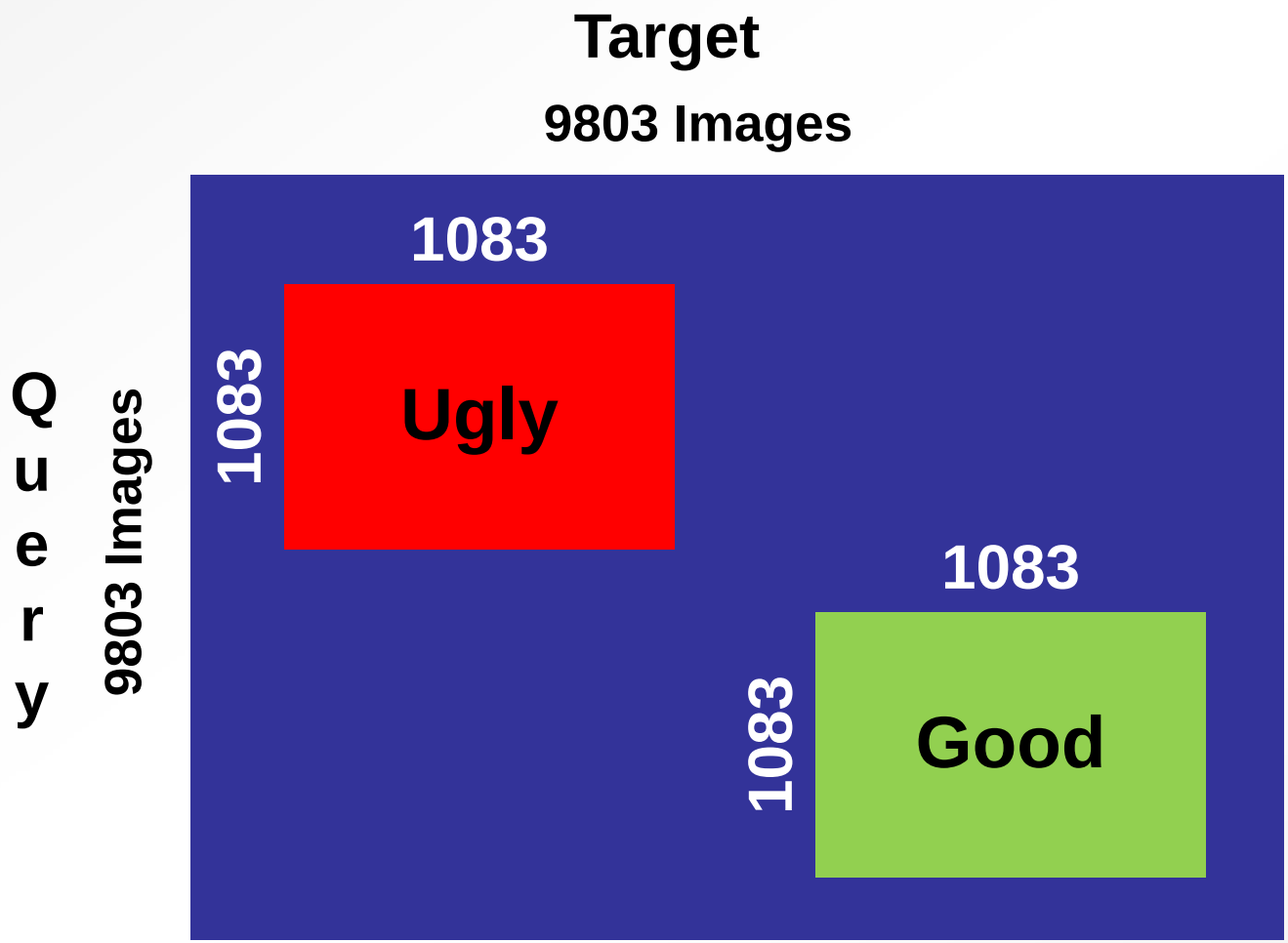
9803 Images

Overall Performance

VR = 0.80

@ FAR = 0.001

What Lurks Within



Face Pairs

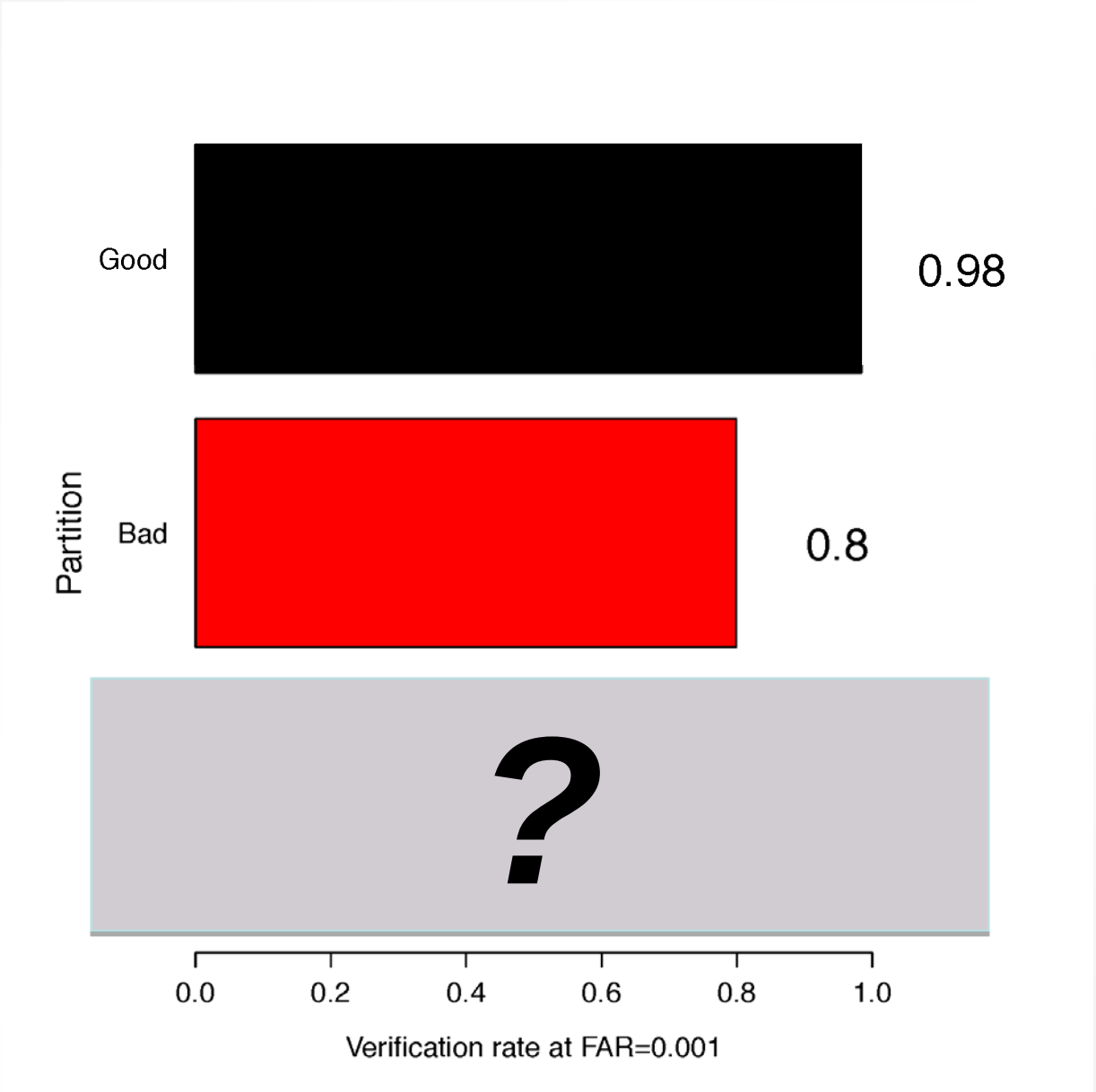
Good Face Pairs




Challenging Face Pairs

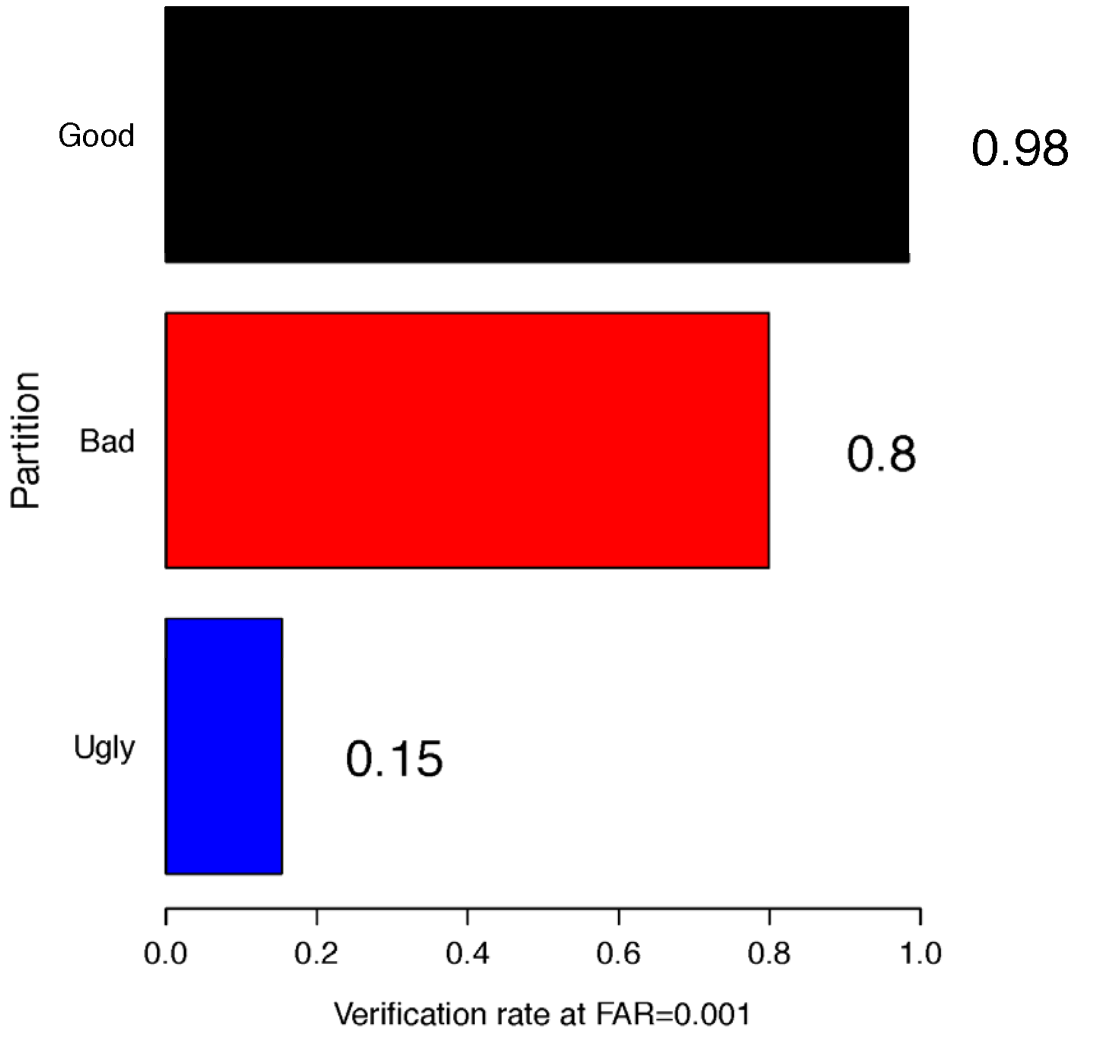



Good, Bad, Ugly Performance



Good, Bad, Ugly Performance

Performance ranges over an order of magnitude



Challenge

- What *k* – questions would characterize performance that ranges over an order of magnitude (VR 0.98 to 0.15)

Biometric-completeness

- NP-completeness
  - Formal definition
- AI-completeness
  - Informal definition
- Biometric-completeness
  - Informal definition

A problem is biometric-complete if solving the problem is "equivalent" to solving the general biometric recognition problem

Approximate solutions

- Push too hard on “prediction” and you will find yourself right back around facing the full complexity of the biometric identification problem.

- Approximate solutions count, and prediction is worth pursuing, just as it is still worthwhile to continue work on AI-complete or NP-complete problems.

Conclusion

- Significant progress in principles of evaluation
- Predicting performance is not solved
- Outlined a challenge problem for prediction
  - Based on twenty questions

Thank You!

Questions?

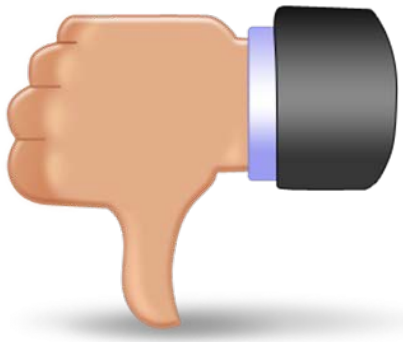
Types of Questions IV

Types of Questions IV

- ??
- ??
- ??

“You can’t always get what you want. But if you try sometime ... You just might find you get what you need!”

- Rolling Stones



An Introduction to Biometric-completeness The Equivalence of Matching and Quality

P. Jonathon Phillips

National Institute of Standards and Technology

J. Ross Beveridge

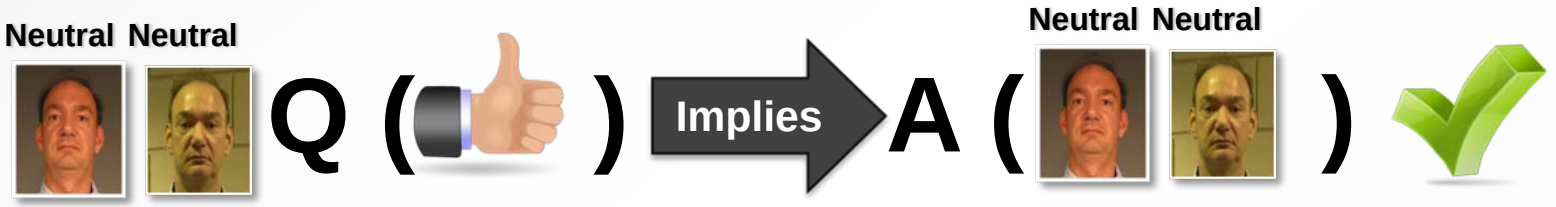
Colorado State University

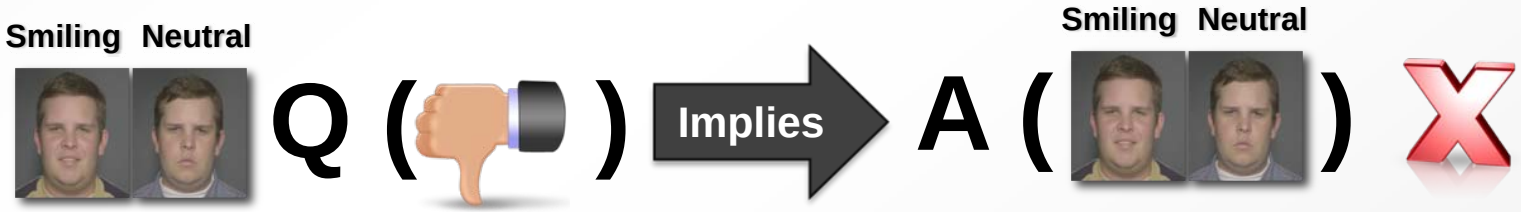
“The Good, the Bad & the Ugly” Still Face Challenge

- Encourage development of face recognition algorithms that work on “hard” to recognize face pairs.

Equivalence of Matching and Quality

Formal Model





“You can’t always get what you want. But if you try sometime ... You just might find you get what you need!”[1]

FRVT 2002 Fine Print

Before summarizing the findings of FRVT 2002, two potentially important issues need to be addressed:

1) Does face recognition work?
2) Which system is best for my application?

The answers to both of these questions are closely related to one another. Face recognition performance, like other biometric types, is application-dependent. Just as there is no best biometric type for all operational applications, there is no best face recognition system for all operational applications. FRVT 2002 was not designed to be a "buyer's guide for face recognition" –where one looks at graphs or scores and selects the best system for installation. Rather, it is a *technology evaluation* that should assist decision-makers in determining (1) if face recognition technology could potentially meet the performance requirements for an operational application, and (2) which systems should be selected for application-specific scenario evaluations.

In order to determine if face recognition works and which system(s) should be deployed, one first needs to properly define the operational application of interest and operational performance requirements. These requirements need to be as specific as possible because even a small change in operational requirements can sometimes significantly alter anticipated performance. Questions to ask when defining an application include:

- Identification, verification or watch list mode of operation?
- The size of the database for identification or watch list?
- Demographics of the anticipated users (age, sex, etc.)?
- Lighting conditions – indoor/outdoor? Supplemental lighting?
- Is the system to be installed overtly or covertly?
- What is the anticipated user behavior?
- How long has it been since the images in the database were taken?
- What is the required throughput rate?
- How many "exception handling" cases can you handle for a given period of time?
- For each mode of operation, which parameter (identification: rank or identification rate; verification: false alarm or probability of verification; watch list: false alarm or correct alarm) is most vital?
- What are the minimum accuracy requirements?

FRVT 2002 can only provide input to several, but not all of these questions. Questions associated with anticipated user behavior, exception handling, human computer interaction, and how a system is integrated into the business model are not addressed in a technology evaluation such as FRVT 2002. Providing answers to these types of questions are the province of scenario