

# Beyond Similarity Scores

*Austin Hicklin*

*Brad Ulery*

*IBPC – March 2010*

# Differing perspectives

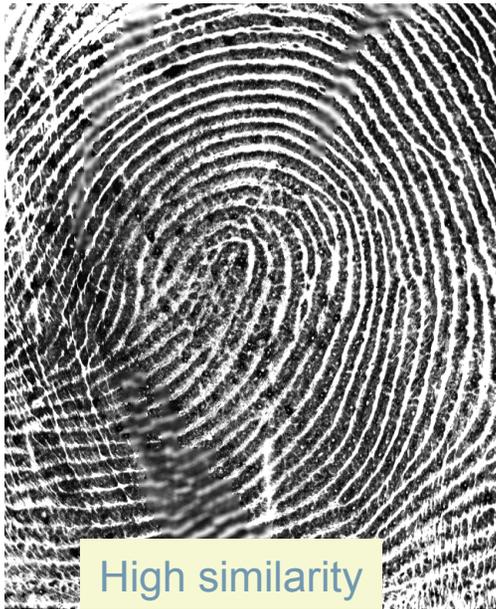
- We have been analyzing automated biometric matching for years
- We have a current focus on assessing human latent examiners
- Perspective makes a difference
  - terminology
  - semantics
  - subtle or substantially different purposes and functions
- In our attempts to model human expert examiners as matchers, we have found the difference in perspective to be enlightening, and we wanted to share some of our observations

# Latent Examiner Studies

- Black box study
- Inter-examiner markup variation
- Extended friction ridge feature set specification (CDEFFS)

# Example

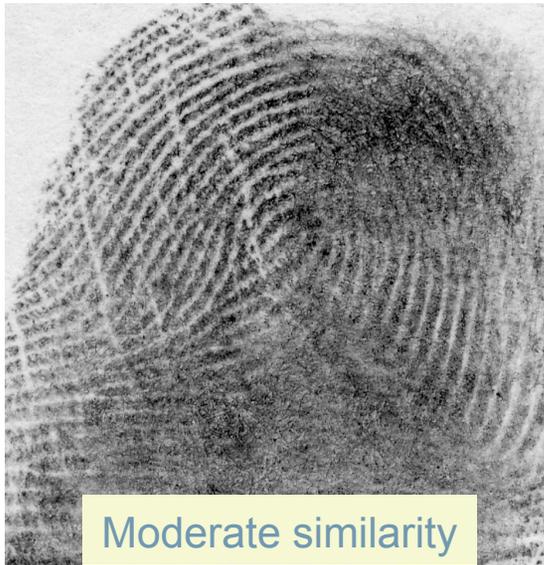


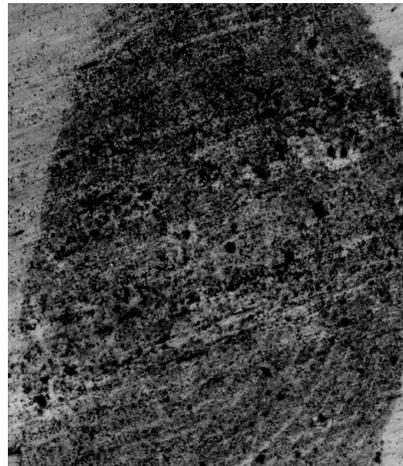


High similarity

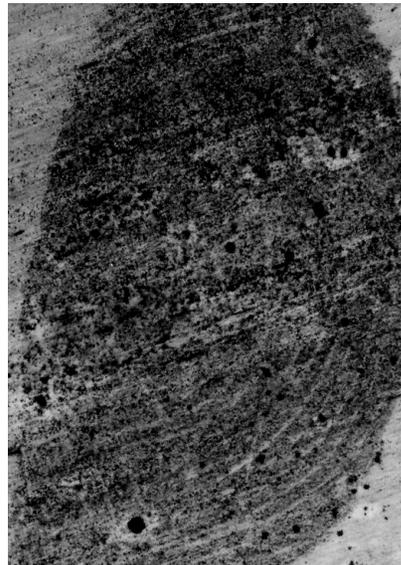


(Exemplar)



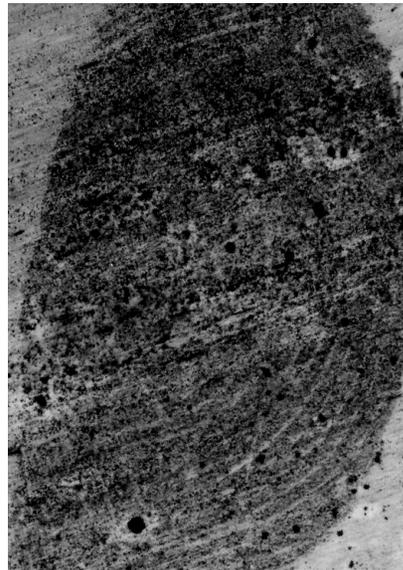


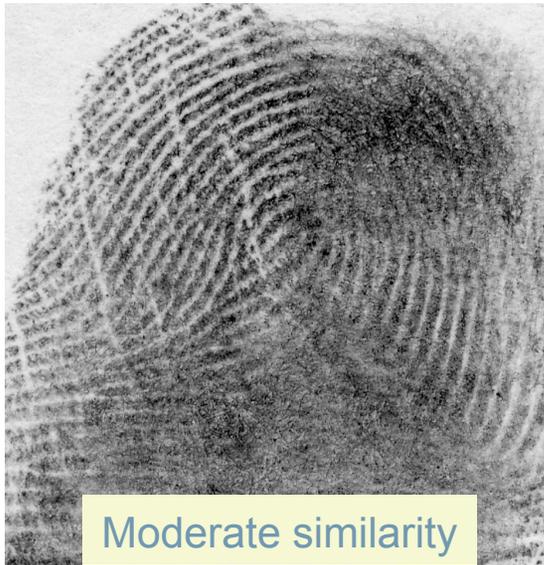
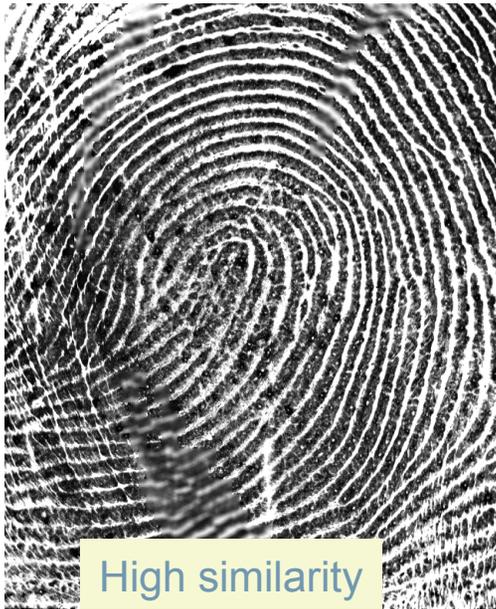
Low/No similarity:  
Mate with poor quality





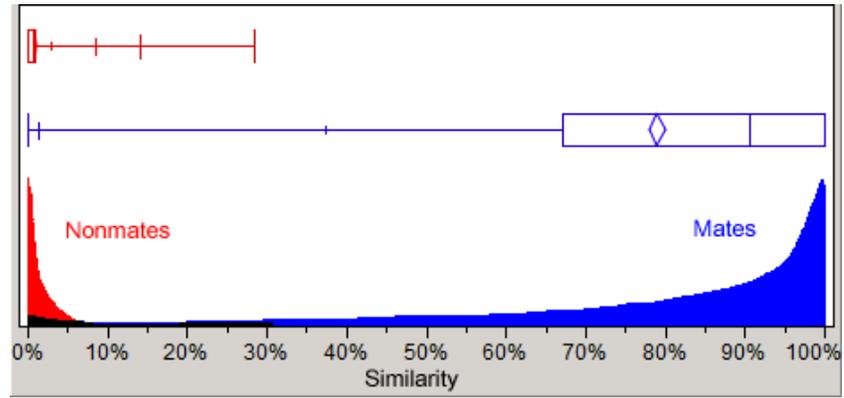
Low/No similarity:  
Mate with no overlap

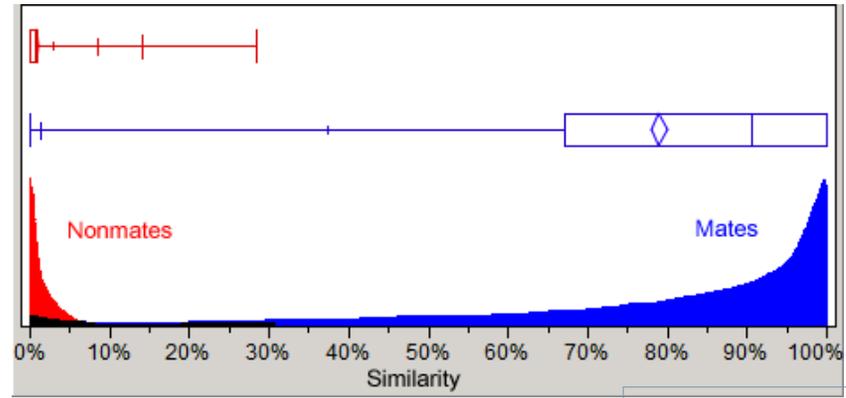




## Automated matching:

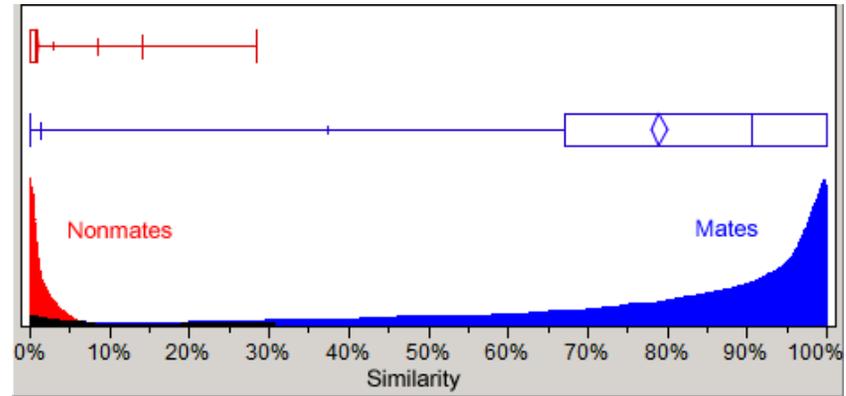
*Similarity scores,  
Probability density functions,  
Receiver operator curves*





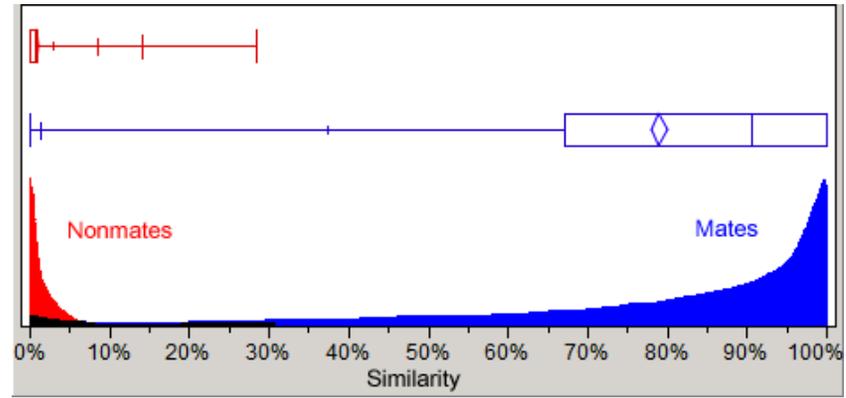
High similarity

□



Moderate similarity

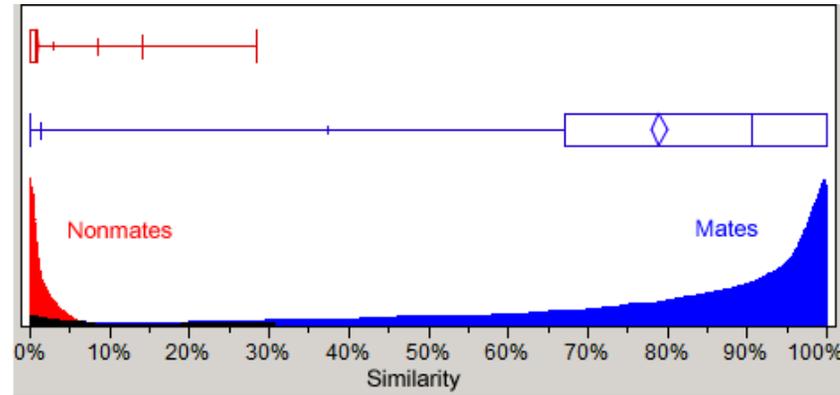
High similarity



Moderate similarity

High similarity

No similarity:  
Nonmate

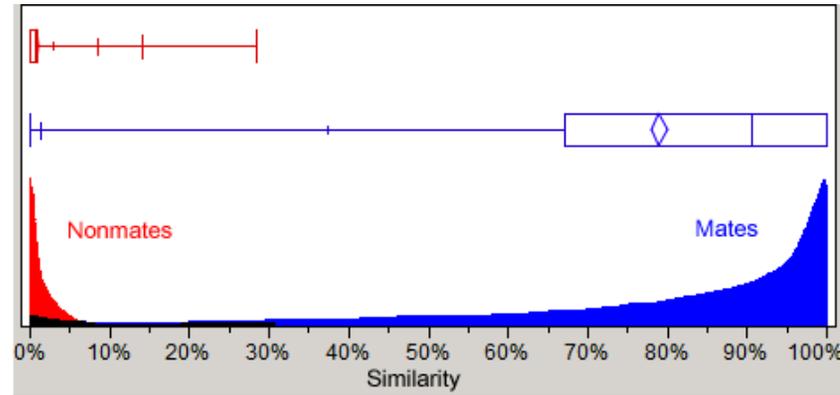


No similarity:  
Mate with poor quality

Moderate similarity

High similarity

No similarity:  
Nonmate



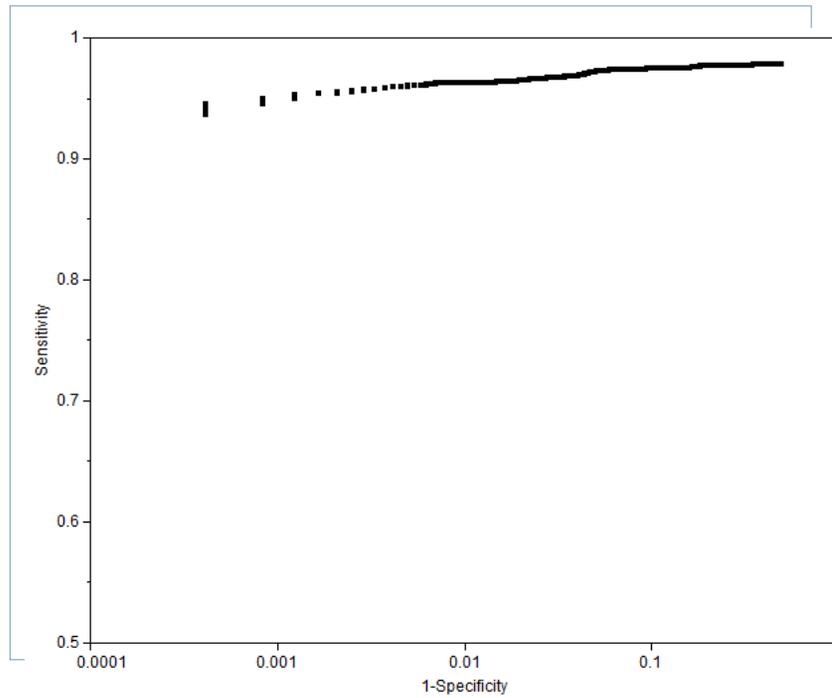
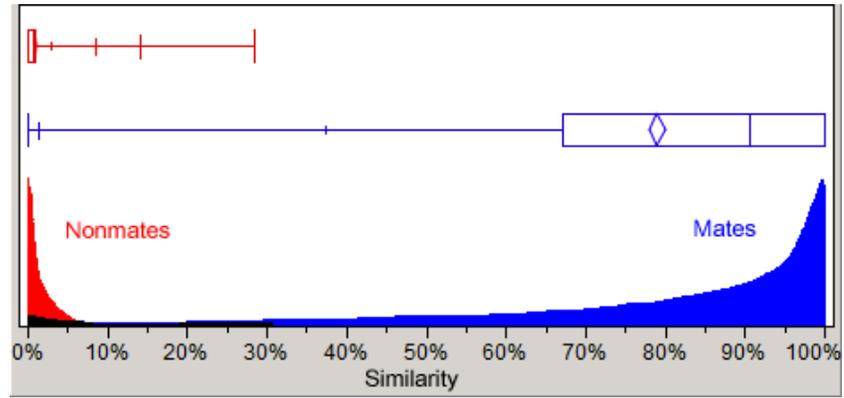
No similarity:  
Mate with poor quality

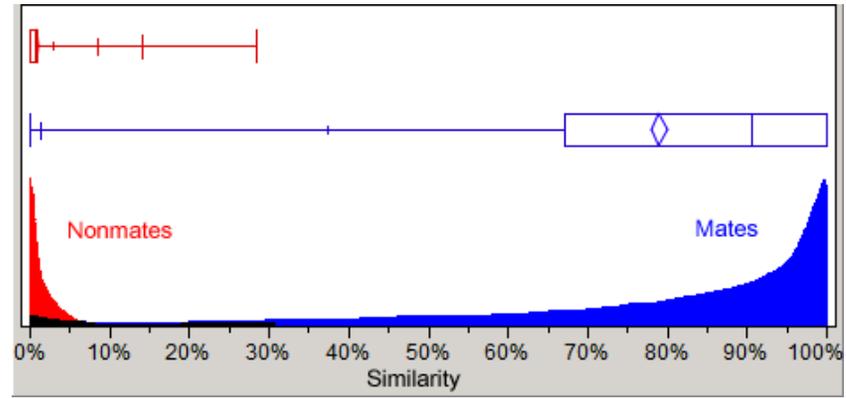
No similarity:  
Mate with no overlap

No similarity:  
Nonmate

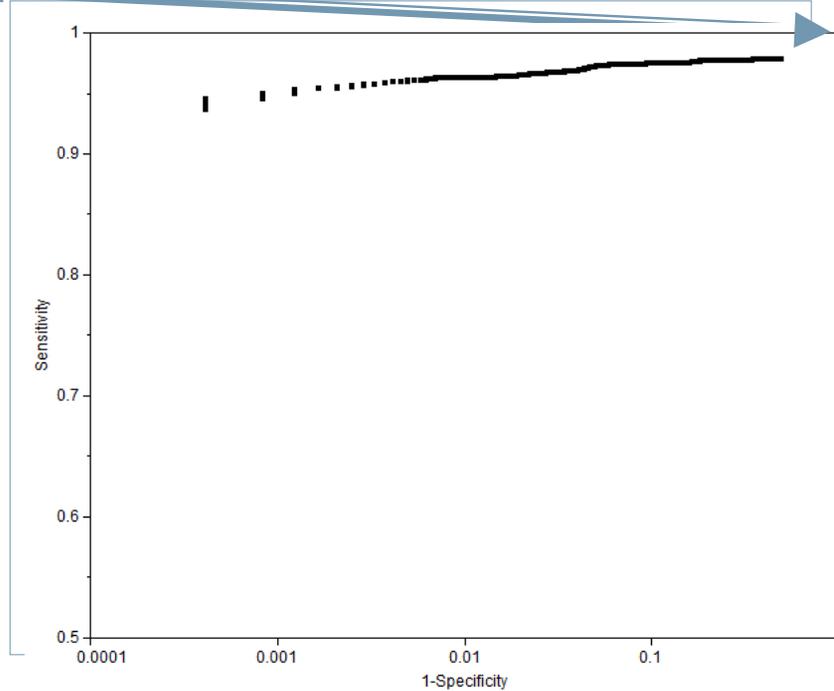
Moderate similarity

High similarity

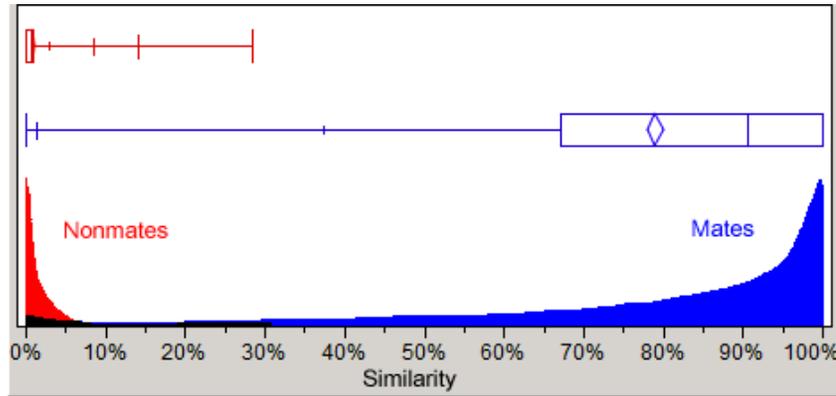




No similarity:  
Nonmate

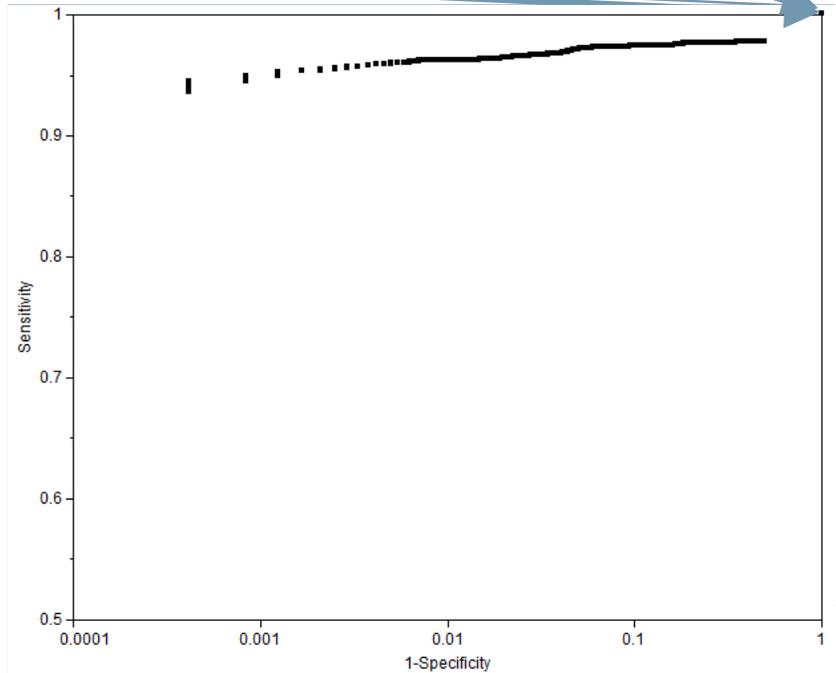


Hypothetical data

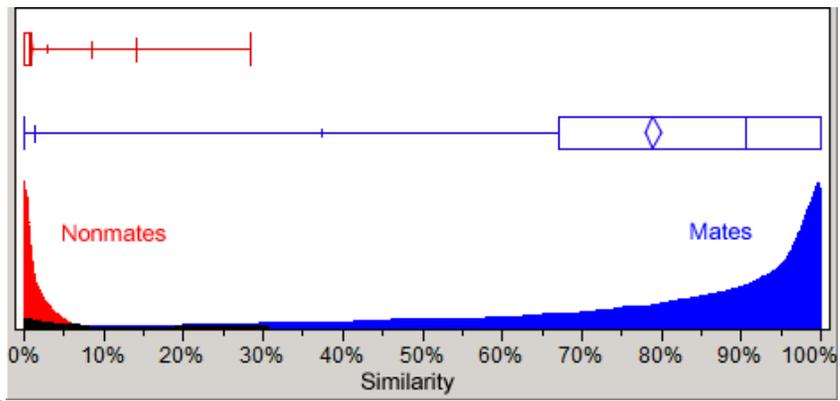


No similarity:  
Mate with poor quality

No similarity:  
Nonmate



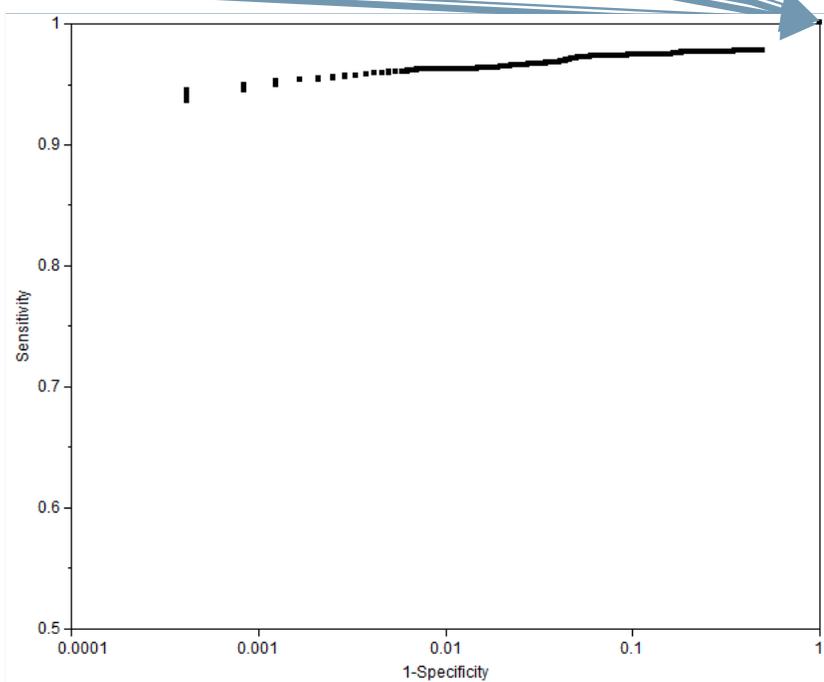
Hypothetical data



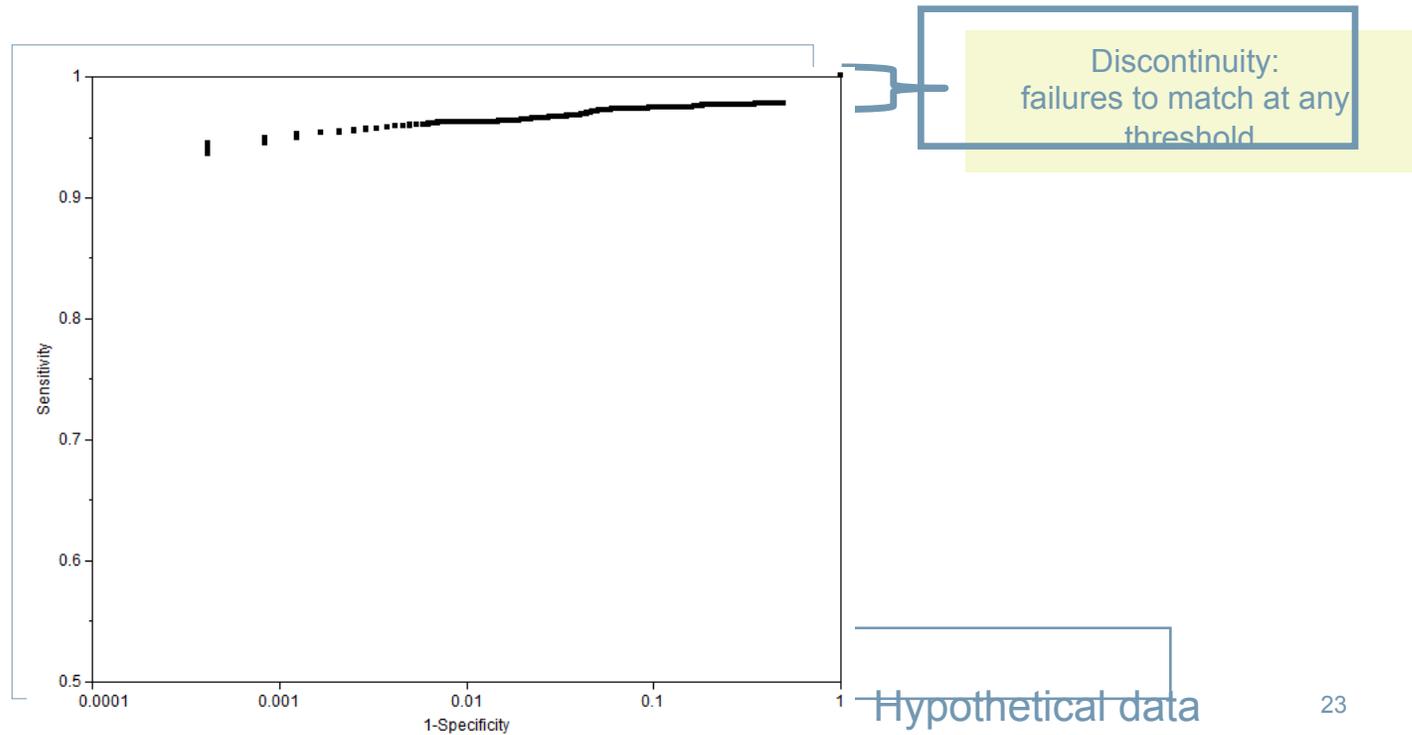
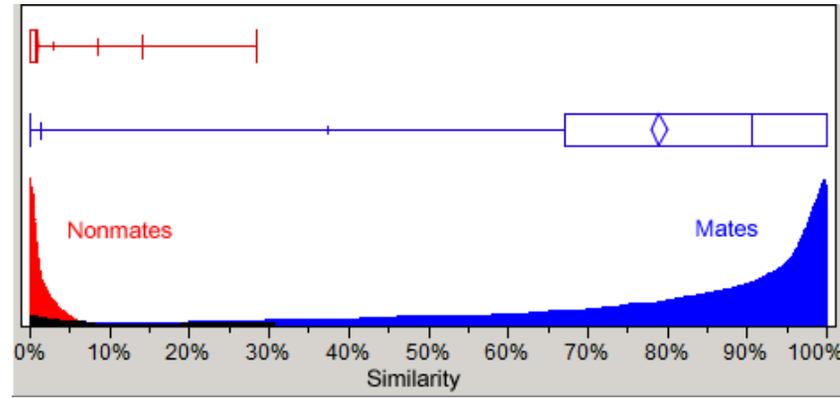
No similarity:  
Mate with poor quality

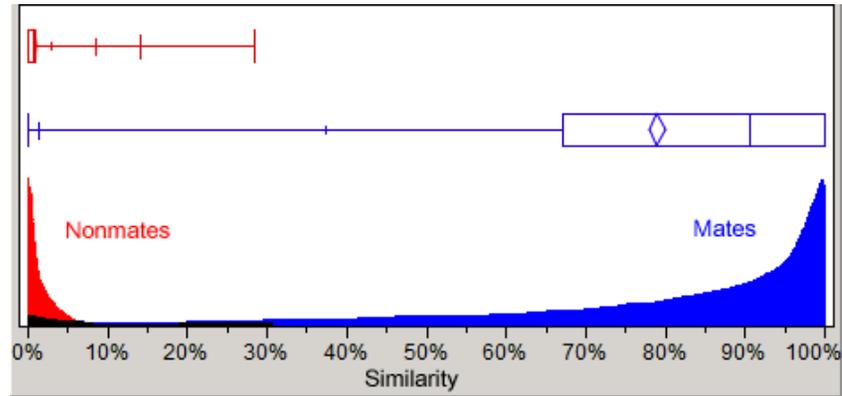
No similarity:  
Mate with no overlap

No similarity:  
Nonmate

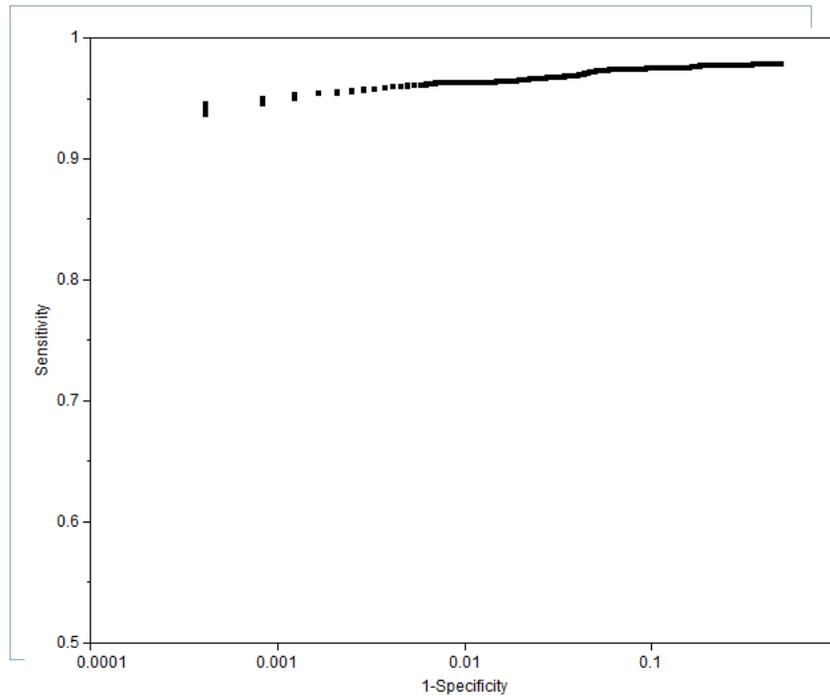


Hypothetical data

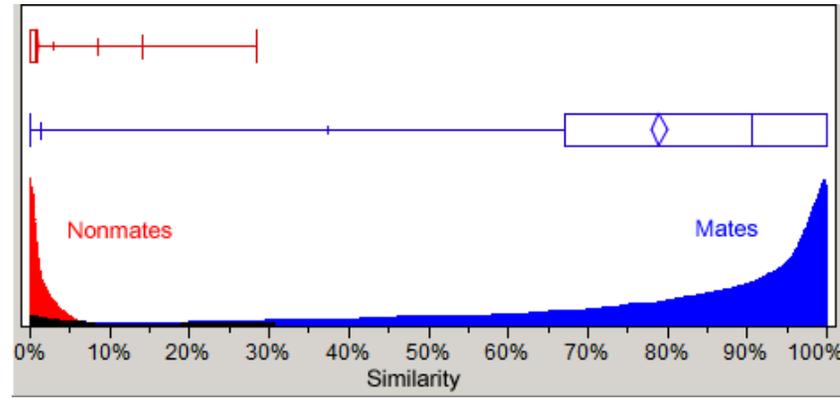




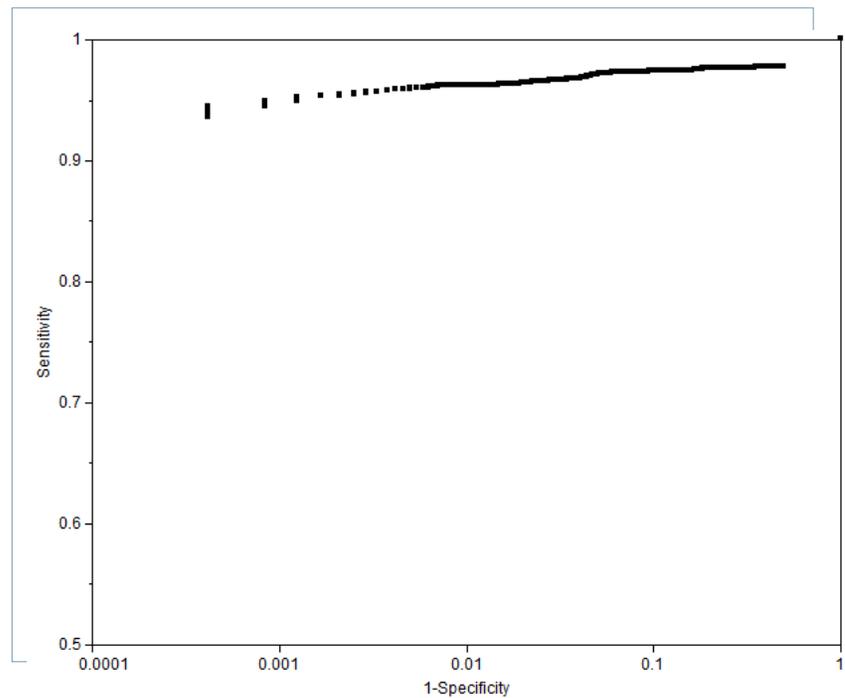
Moderate similarity:  
Allows empirical measurement of error rate tradeoffs



Hypothetical data



High similarity:  
 No empirical measurement of specificity (false matches)

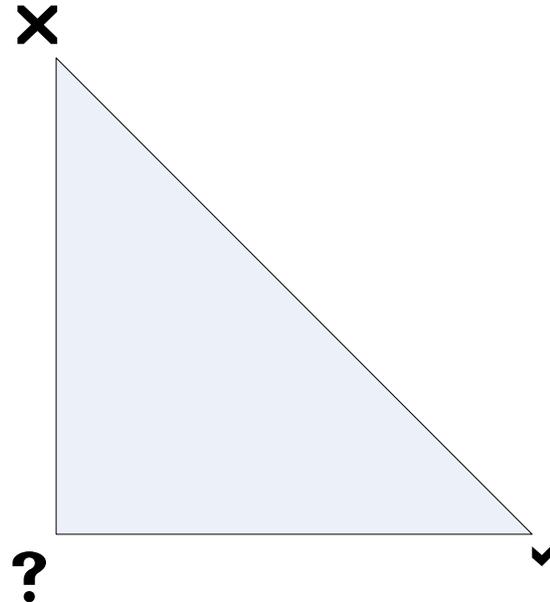
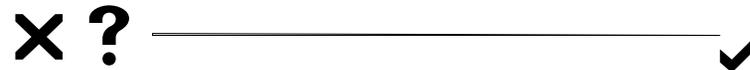


Hypothetical data

# Human examiner determinations

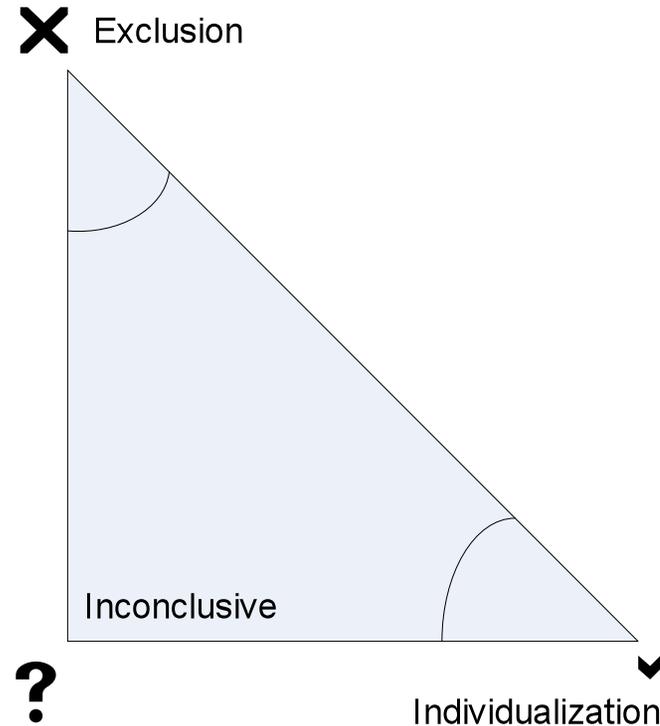
# Automated matchers & human examiners

- Automated matchers
  - Single function: similarity
  - Continuous distribution of values
  - 2 determinations:
    - Match
    - Failure to match (Nonmatch/ Inconclusive)
- Human examiners
  - 2 functions:
    - Similarity
    - Difference
  - 3 determinations:
    - Same source / Individualization
      - (highly similar)
    - Different source / Exclusion
      - (highly different)
    - Inconclusive

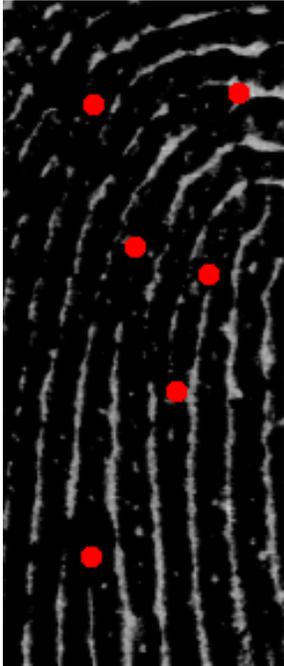


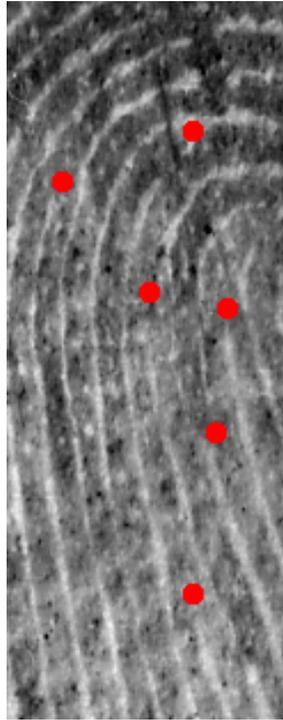
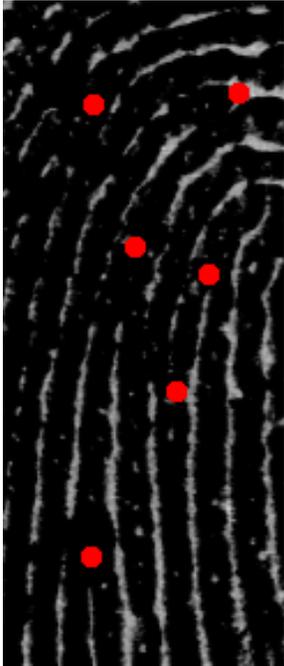
# Automated matchers & human examiners

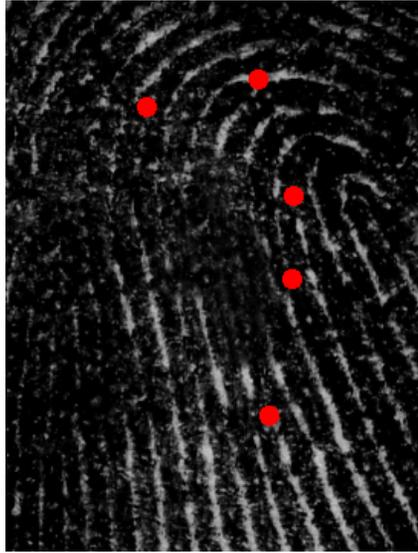
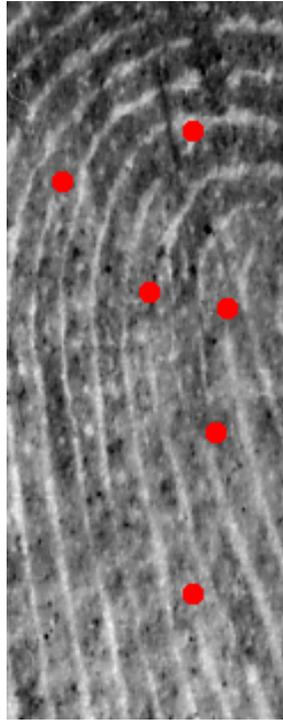
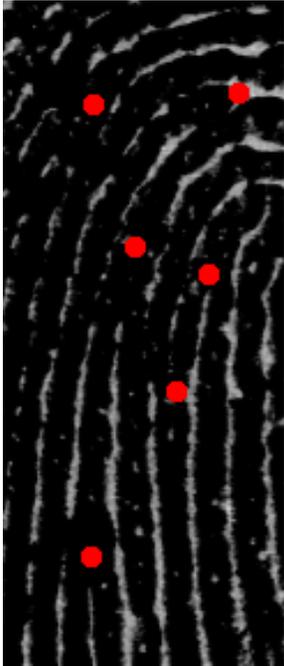
- Decision thresholds
  - Automated matchers
    - Continuous distribution of values allows requirements-based threshold adjustment
  - Human examiners
    - Responses are Boolean, not continuous, and are very deliberately cautious

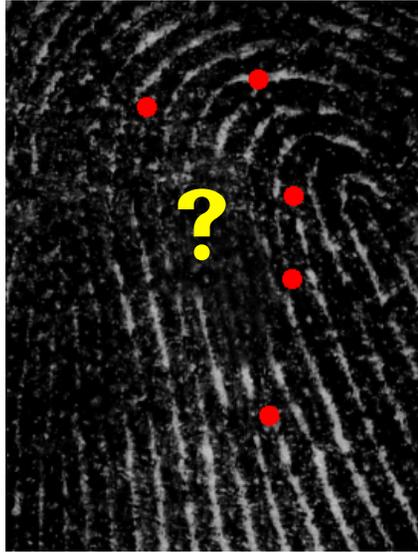
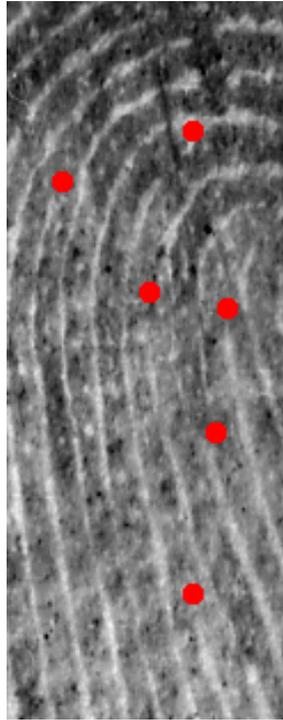
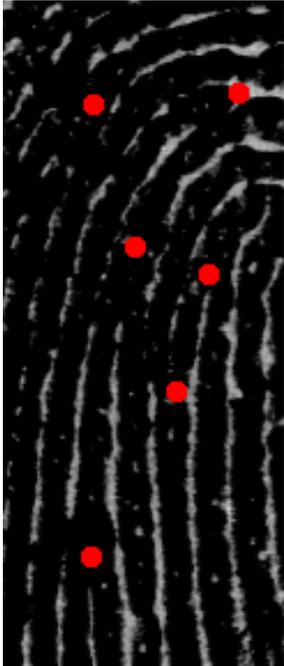


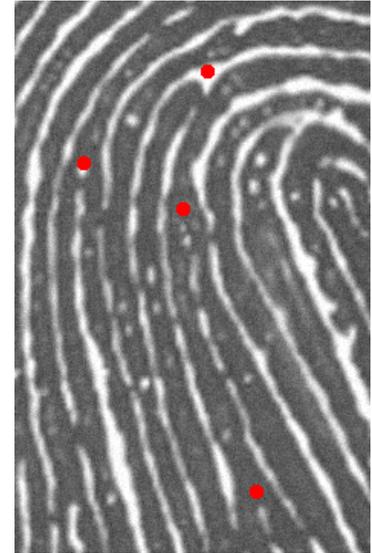
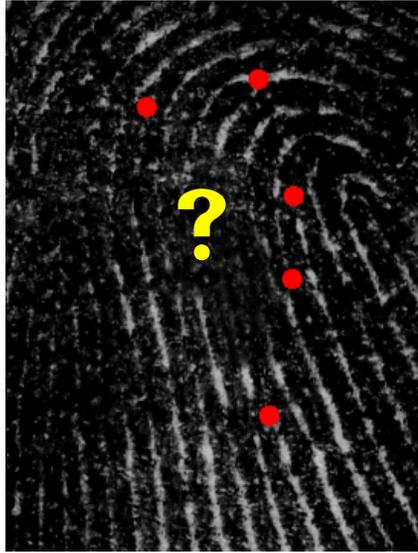
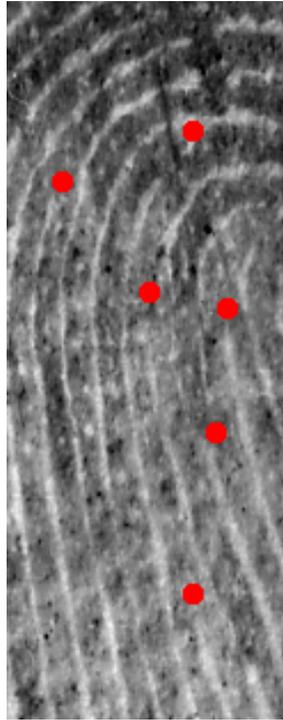
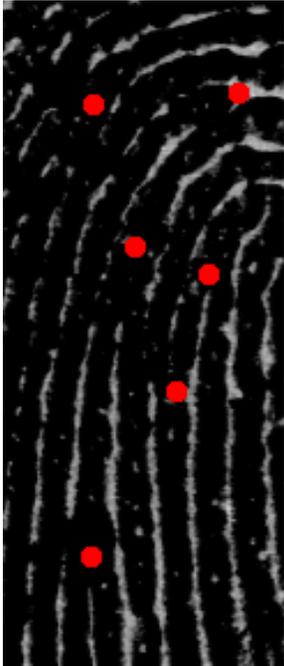
# Feature-level similarity, difference, and inconclusive assessments

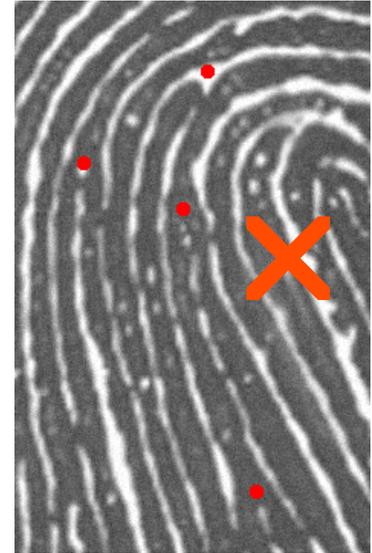
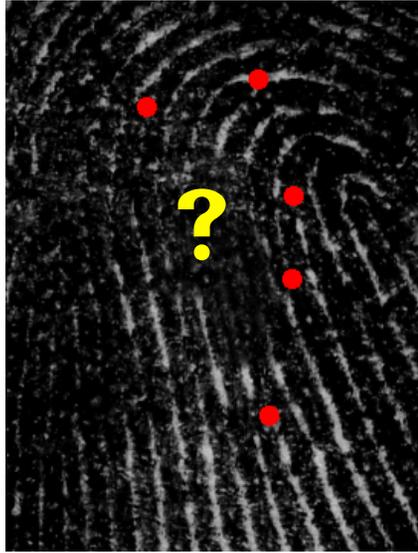
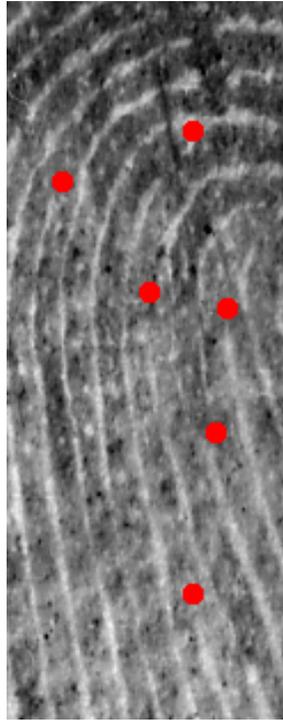
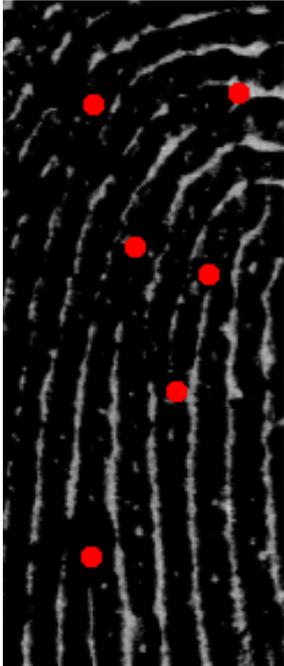


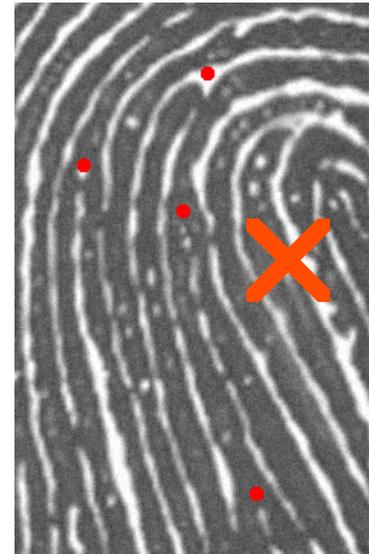
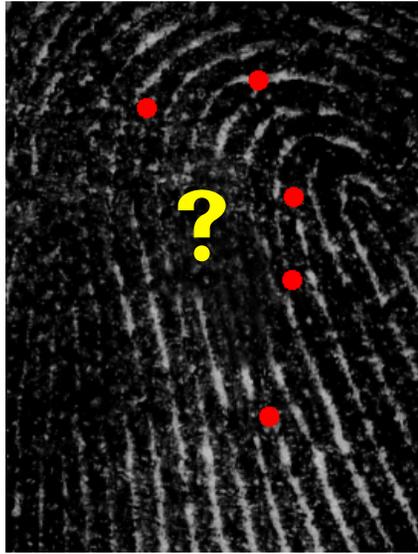
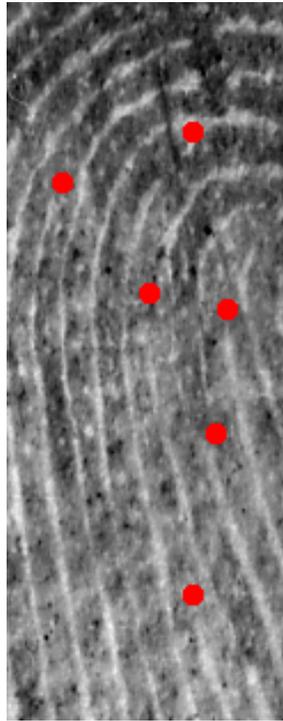
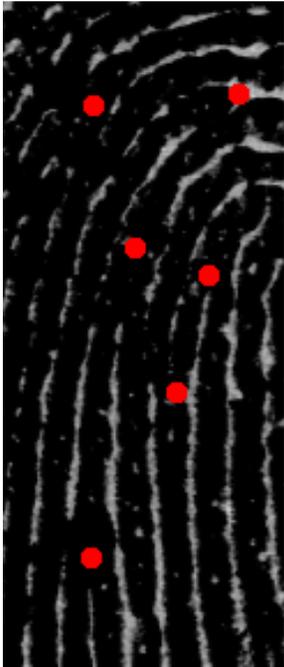




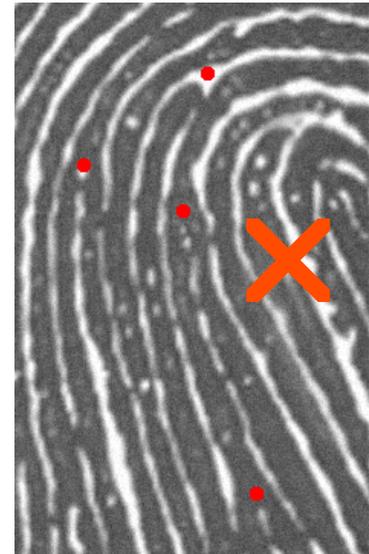
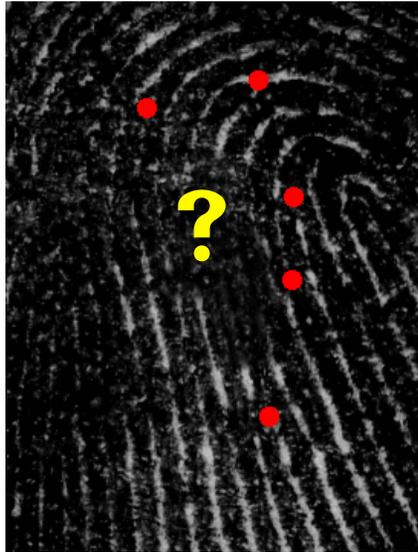
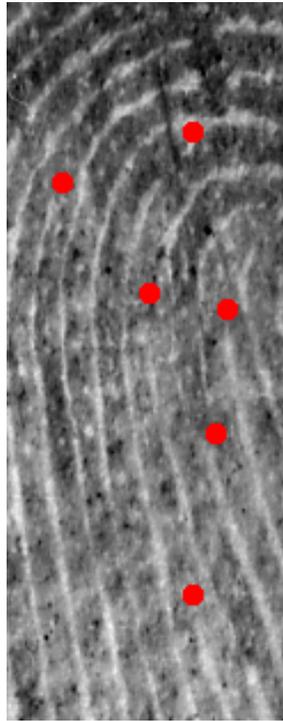
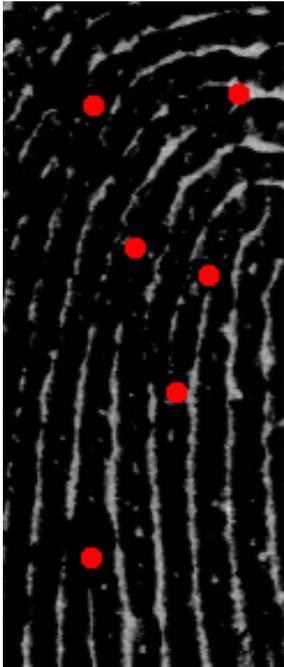








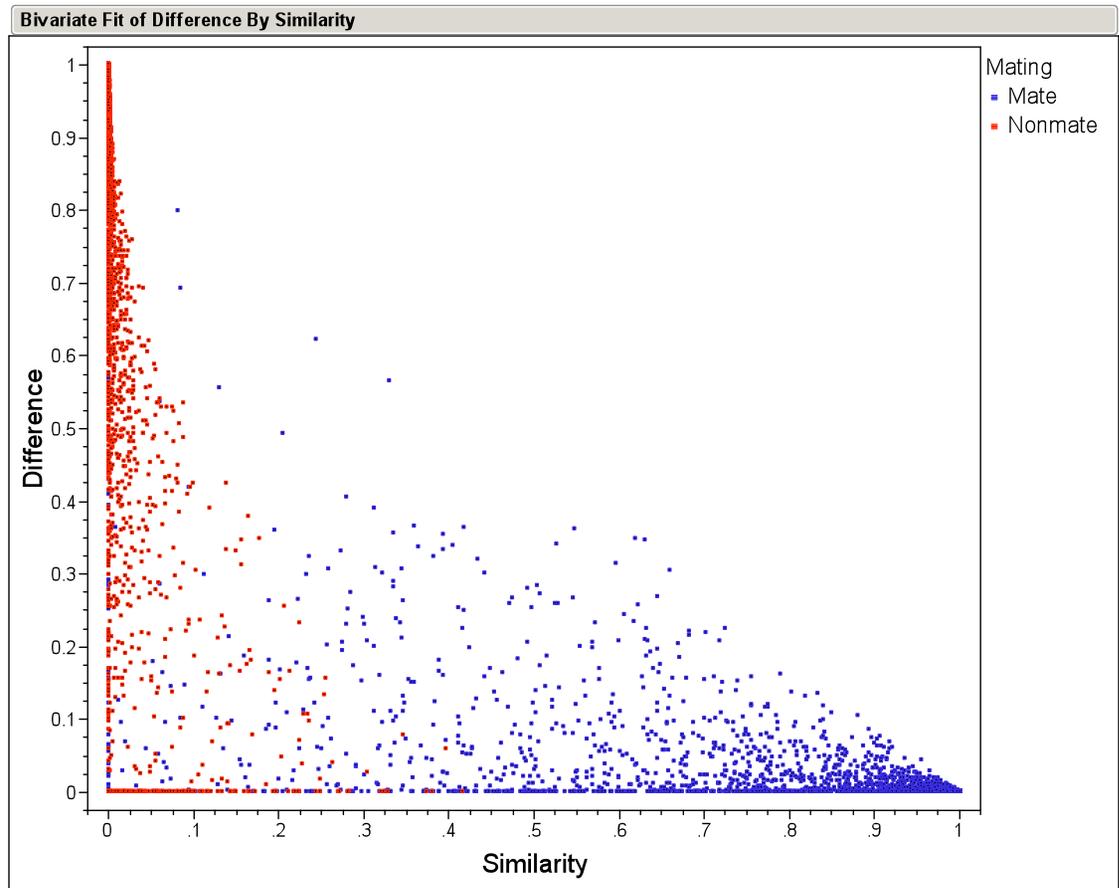
- Examiners assess each feature or grouping of features in a comparison in terms of similarity, difference, and inconclusive



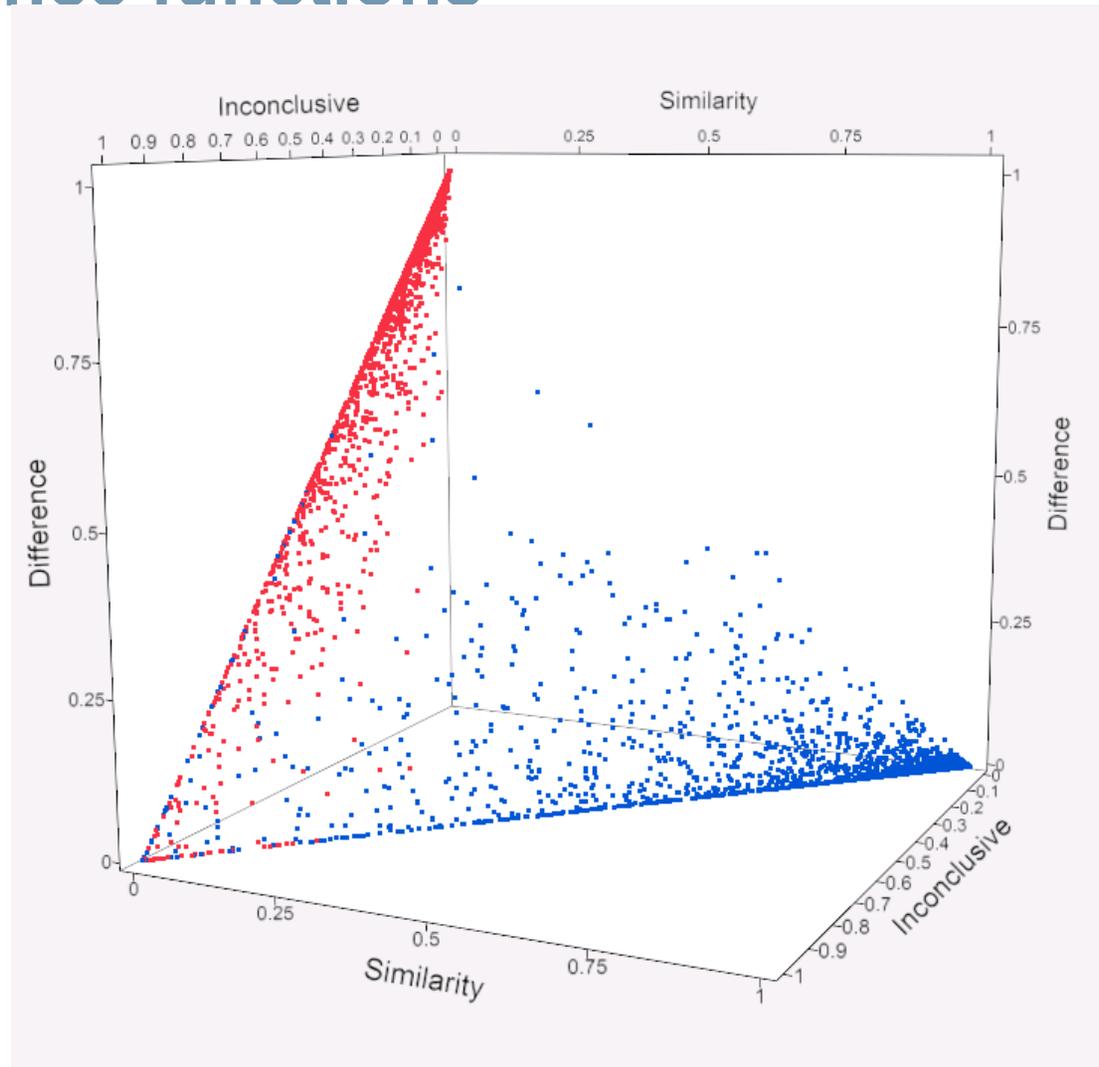
- Feature-level difference measures require
  - Accurate feature detection
  - and/or
  - Feature-level quality/confidence algorithms

# Hypothetical matchers

# Hypothetical matcher with both similarity and difference functions



# Hypothetical matcher with both similarity and difference functions



# Summary

- Difference measures are not the same as 1-similarity
- Automated matchers
  - Return a single continuous monotonic similarity value
  - An adjustable threshold can be applied
  - Decisions are match vs nonmatch (including exclusion and inconclusive)
- Human examiners
  - Return a tri-state decision
    - Same source (individualization)
    - Different source (exclusion)
    - Inconclusive
- These decisions are made at three levels
  - Overall
  - Feature groups
  - Individual features
- Automated matchers could in theory return continuous difference scores in addition to similarity scores
  - Potential for improved accuracy
  - Would allow additional functionality (exclusion)
  - Requires reliable feature extraction and/or reliable feature-level quality metrics
- In the absence of accurate difference measures
  - Similarity + Difference + Inconclusive = 1
  - An inconclusive measure could be defined as a measurement of applicability/quality

Austin Hicklin

[hicklin @ noblis.org](mailto:hicklin@noblis.org)

Brad Ulery

[ulery @ noblis.org](mailto:ulery@noblis.org)