ICAO Ninth Symposium and Exhibition on MRTDs, Biometrics and Border Security 22-24 October 2013

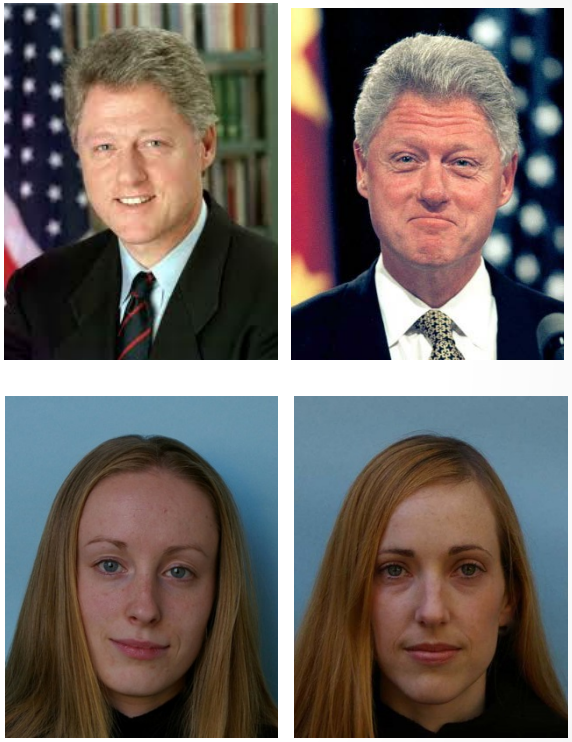
Challenges in facial image verification: Psychological aspects of issuing and checking biometric travel documents

Richard Kemp

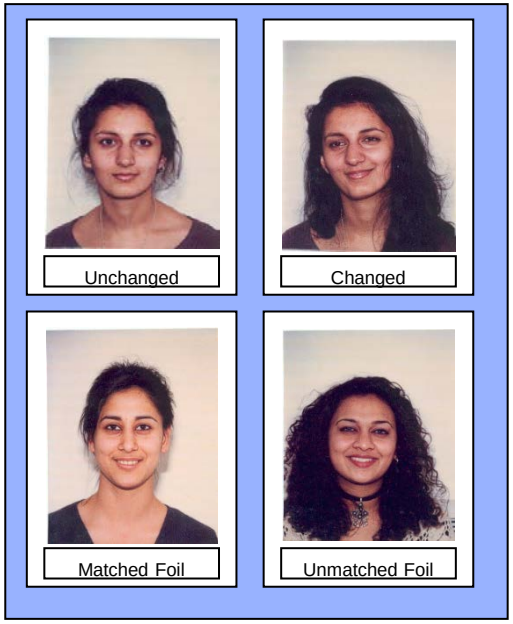
School of Psychology, University of New South Wales, Australia

Why Psychology?

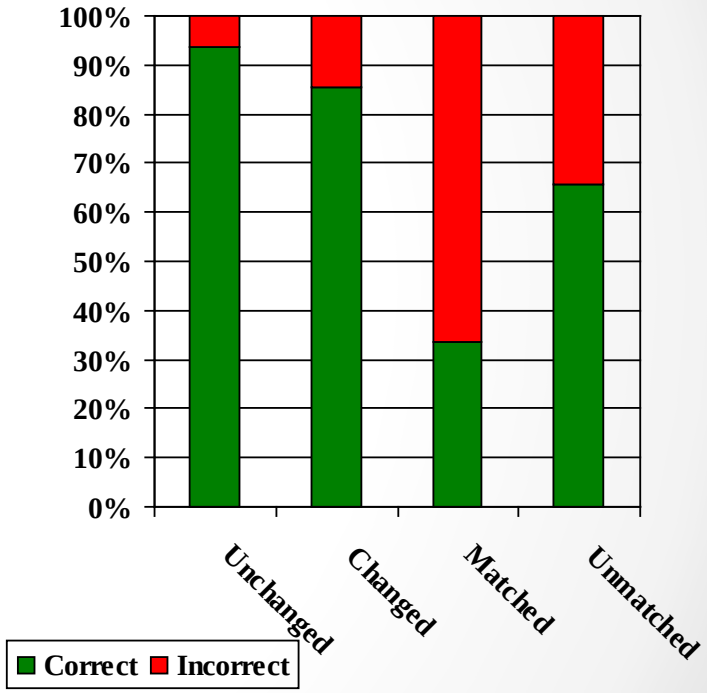
- I am an experimental Psychologist – I study human perception and memory, in particular face recognition
- Travel Documents which include a photograph (facial biometric) are validated by comparing the photograph to the person presenting the document.
- This is Unfamiliar Face Matching
- Psychological research shows that this is a difficult and error prone task.



Unfamiliar Face matching

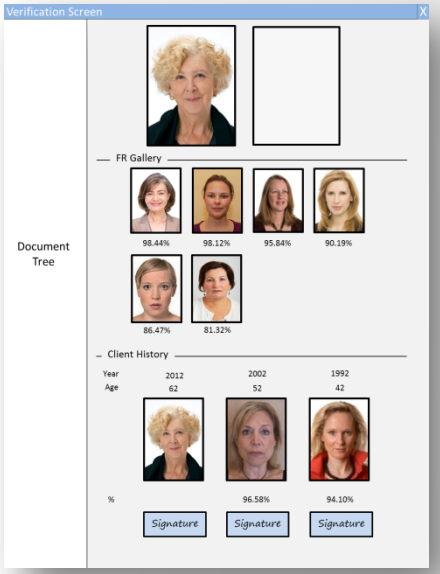


Kemp et al, 1997



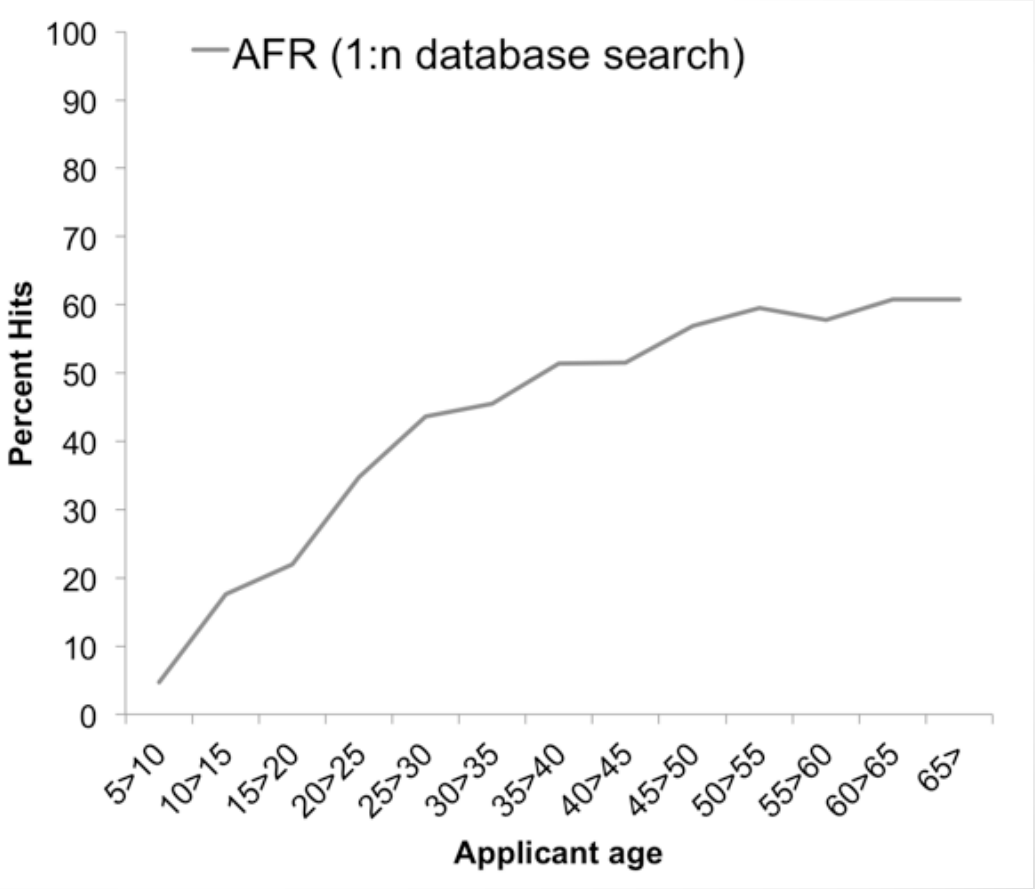
A problem not solved by AFR

- This is a problem if we rely on human operators to validate facial biometrics –either at the border or as part of the passport issuance process
- Automatic Face Recognition (AFR) systems don't eliminate the problem – we still rely on humans to make final decisions



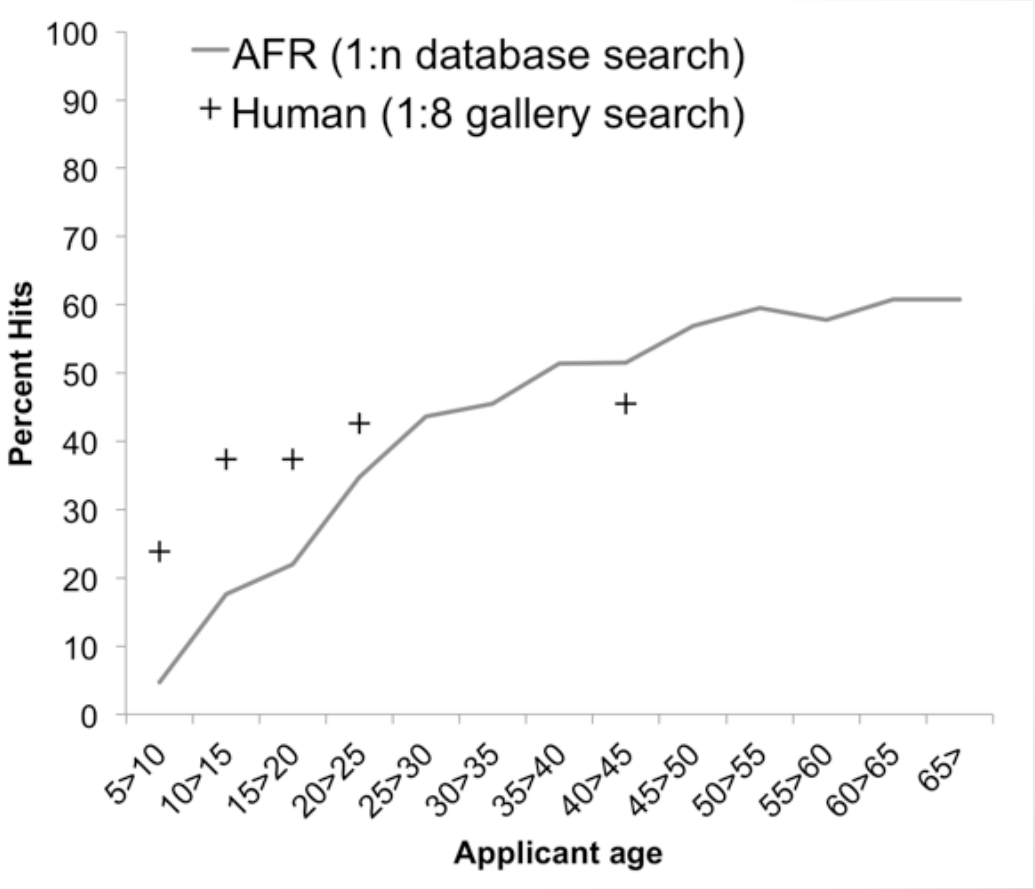
A problem not solved by AFR

- For one particular AFR system
- The percentage of occasions that a client history image is returned in the top 8 gallery positions



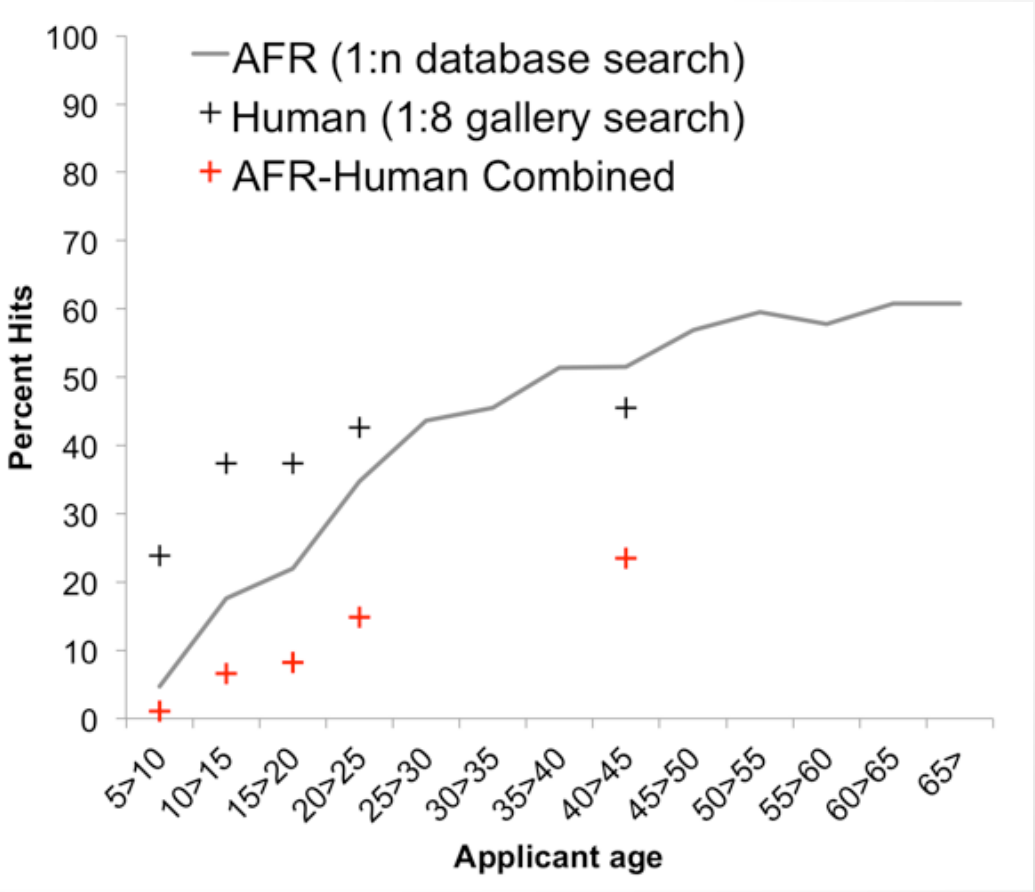
A problem not solved by AFR

- For one particular AFR system
- The percentage of occasions that a client history image is returned in the top 8 gallery positions
- Accuracy of Human operators in a 1-to-8 gallery matching task



A problem not solved by AFR

- For one particular AFR system
- The percentage of occasions that a client history image is returned in the top 8 gallery positions
- Accuracy of Human operators in a 1-to-8 gallery matching task
- Estimated accuracy of human operators searching the gallery returned by the AFR

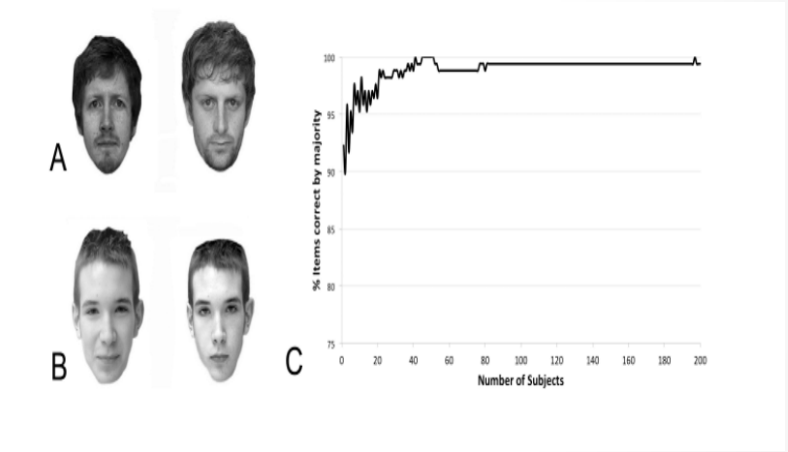


Improving performance

- The human operator is a major determinant of the fraud detection accuracy of the total system
- How can we improve Unfamiliar Face Matching performance in human operators?
- Four approaches:
  - Selection
  - Training
  - Workflow
  - Image format

Selection

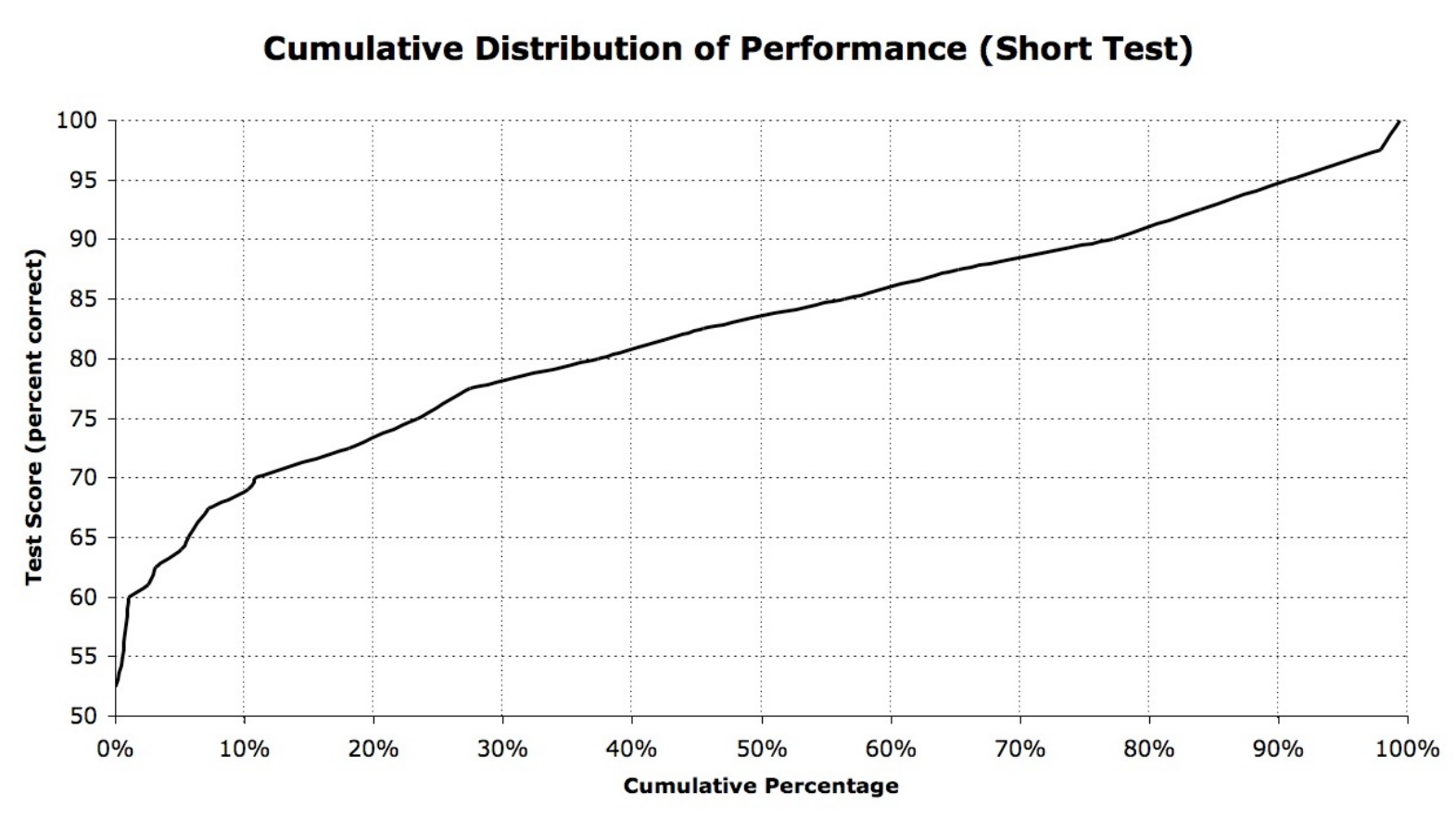
- There are stable individual differences in unfamiliar face matching performance
- There is a very wide range in performance
- We can measure performance using the Glasgow Face Matching Test (GFMT)



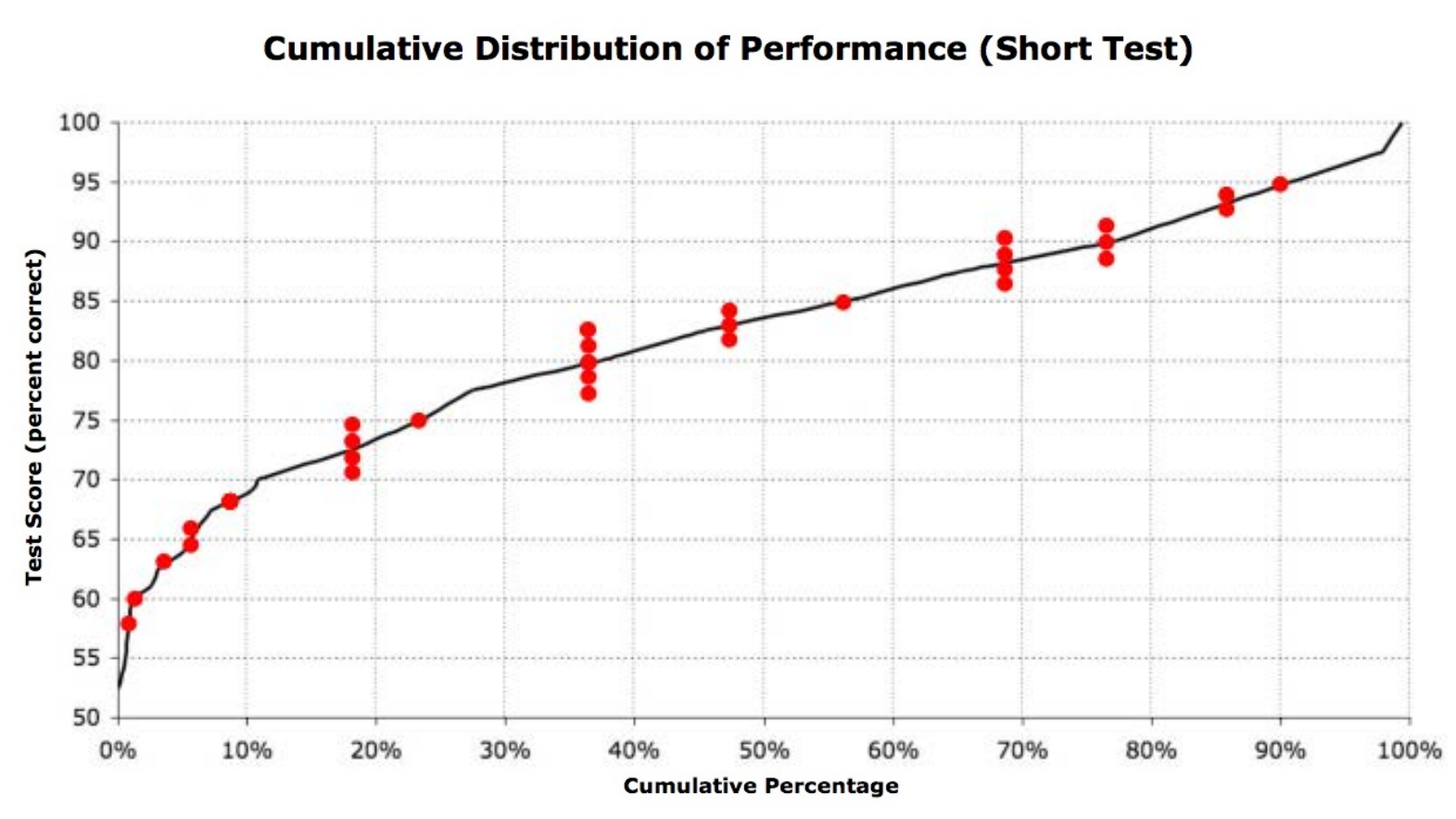
Glasgow Face Matching Task

Burton, White & McNeill (2010). *Behavior Research Methods*, 42, 286-291.

Glasgow Face Matching Test

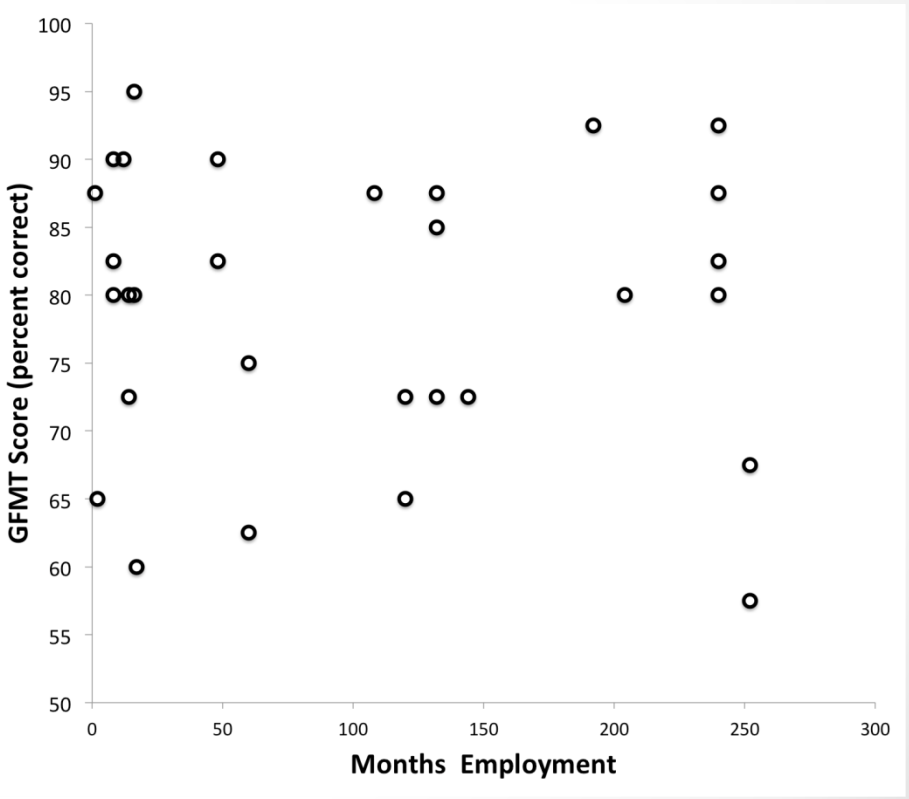


Glasgow Face Matching Test



Glasgow Face Matching Test

- Staff who validate facial biometrics as part of their job perform no better or worse than the general population on the GFMT
- Experience in the job is not correlated with performance on the GFMT



Training

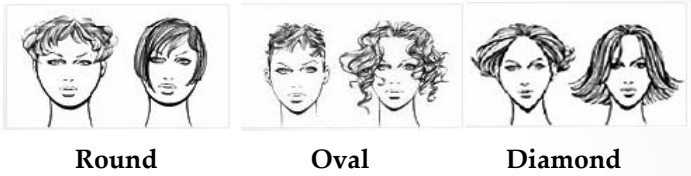
- Several training programs are in use around the world: No evidence available regarding effectiveness
- Common components include:
  - **Face Shape Strategy**
  - **Facial Anatomy training**
  - **Photographic training**

- We are evaluating these approaches

Testing the Face Shape Strategy

The Principle:
Faces are categorized into a number of shapes.

Faces with different shapes are likely to be of different individuals.



Method:
Participants trained to use strategy then shown 100 photos to categorize (5 images of each of 20 identities – one image twice).

Face matching tested before and after training

Testing the Face Shape Strategy

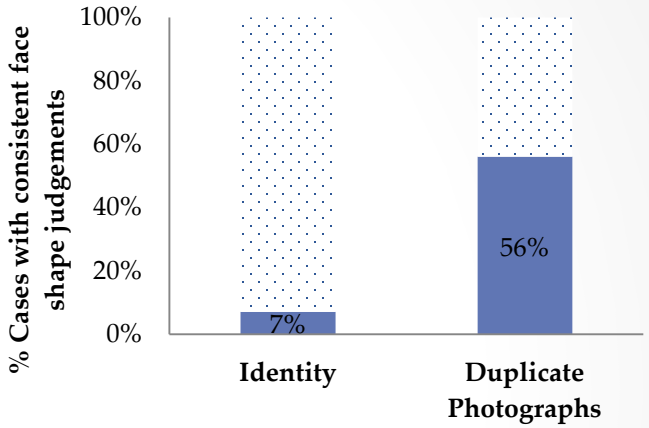
Results:
Participants couldn't consistently apply the face shape strategy.

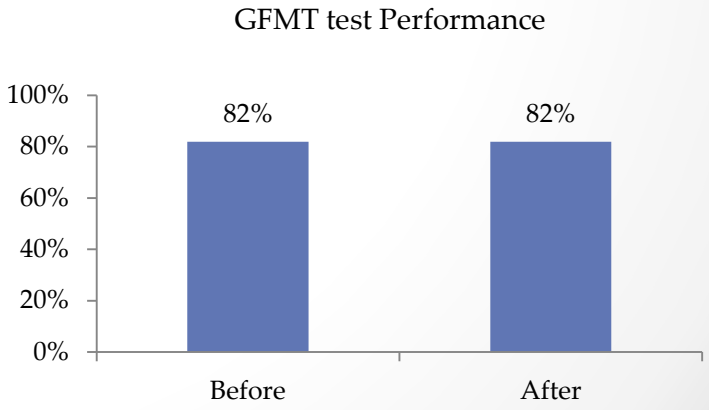
Only in 7% of cases were all photos of the same person classified as having the same shape.

Identical photos were given different classifications in 44% of cases.

Training didn't improve GFMT

Conclusions:
No evidence that the Face Shape Strategy is useful (when used in isolation).





Other training approaches being tested

Face Anatomy

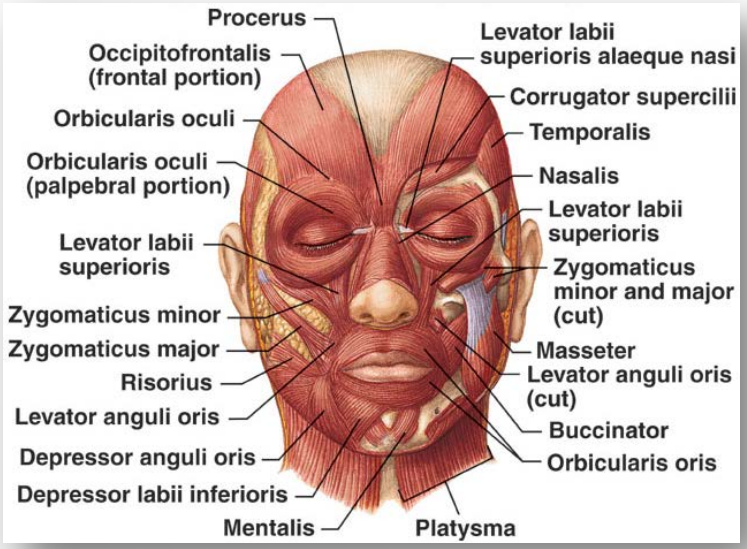


Image Capture

Training: Feedback

- One problem with Facial Biometric validation is that operators rarely receive feedback on the accuracy of their decisions.
- Psychological research shows feedback is critical to improving performance in a range of tasks.
- What happens if we provide feedback training?

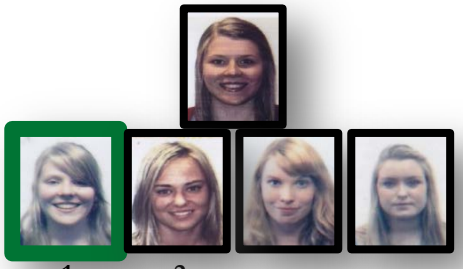
Feedback Training

Procedure:
1-to-8 array task. Real passport images*.
50% target present.

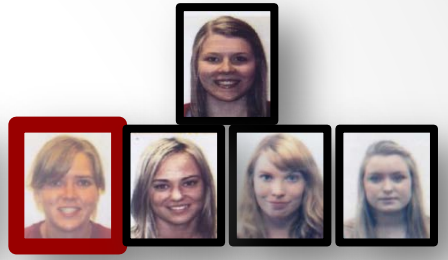
Feedback group received feedback after each trial.

Completed 400 trials on day 1, then another 400 one week later.

* Images shown here are not real passport photos



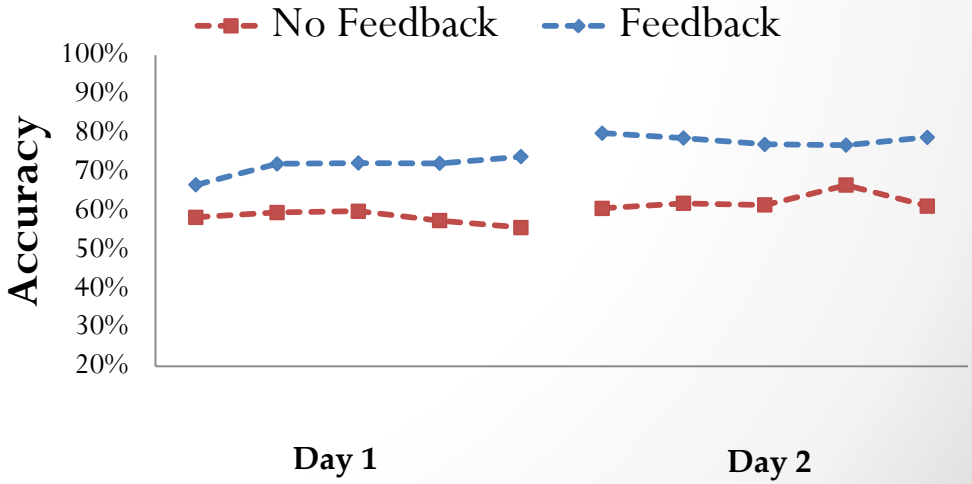
Correct!
You found the target.



Incorrect!
The target is absent.

Results:
Matching performance rapidly improves with feedback and improvement is sustained over time.

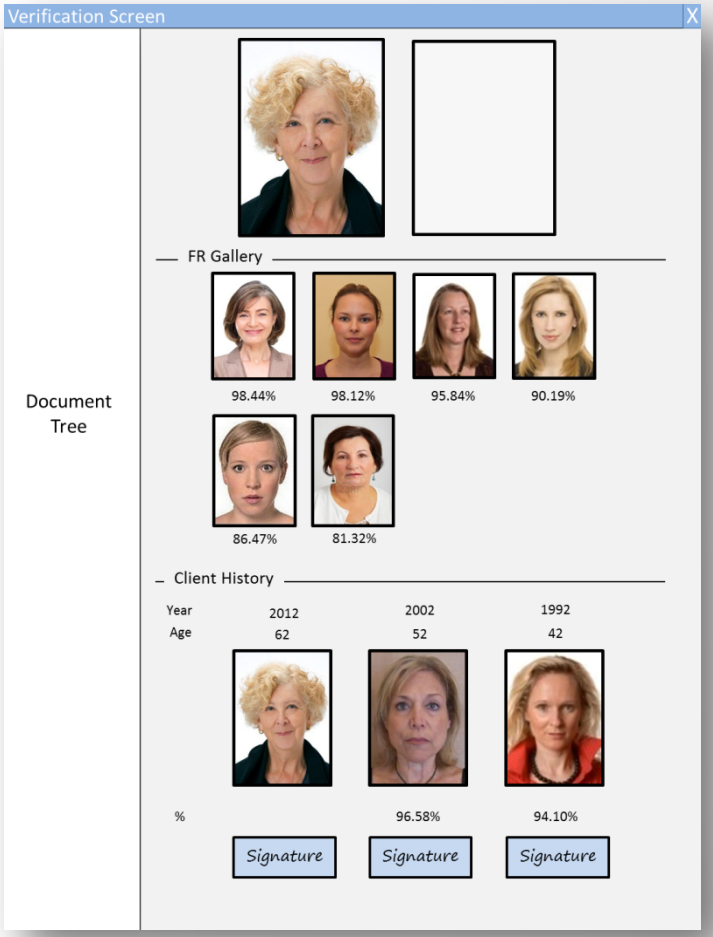
Improvement is for Target Absent trials only (i.e. improves detection of imposters but not for detection of duplicates).



Conclusions:
Feedback improves performance in Target Absent trials.
Latest results suggest effect is preserved after removal of feedback

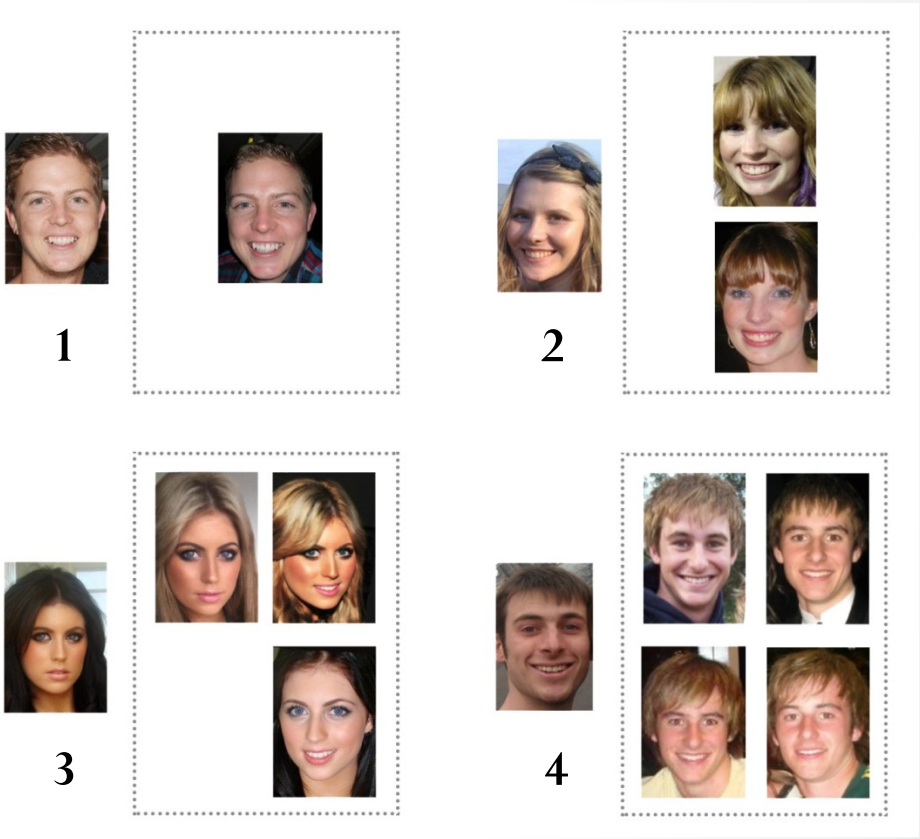
Workflow

- Familiarisation using client history images
- Aggregating responses



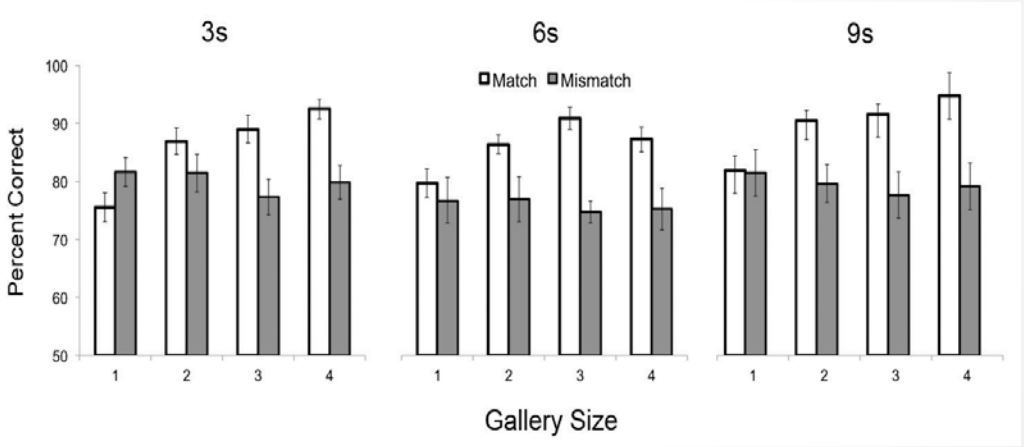
Familiarisation

- We know Familiar face matching is much easier than unfamiliar face matching
- So can we use client history images to "familiarise" ourselves with the target ?
- Does performance increase with additional history images



Familiarisation

- Familiarisation with client history images improves performance
- But this improvement is only seen for Match trials
- That is familarisation leads to better detection of duplicates, but doesn't help in imposter detection

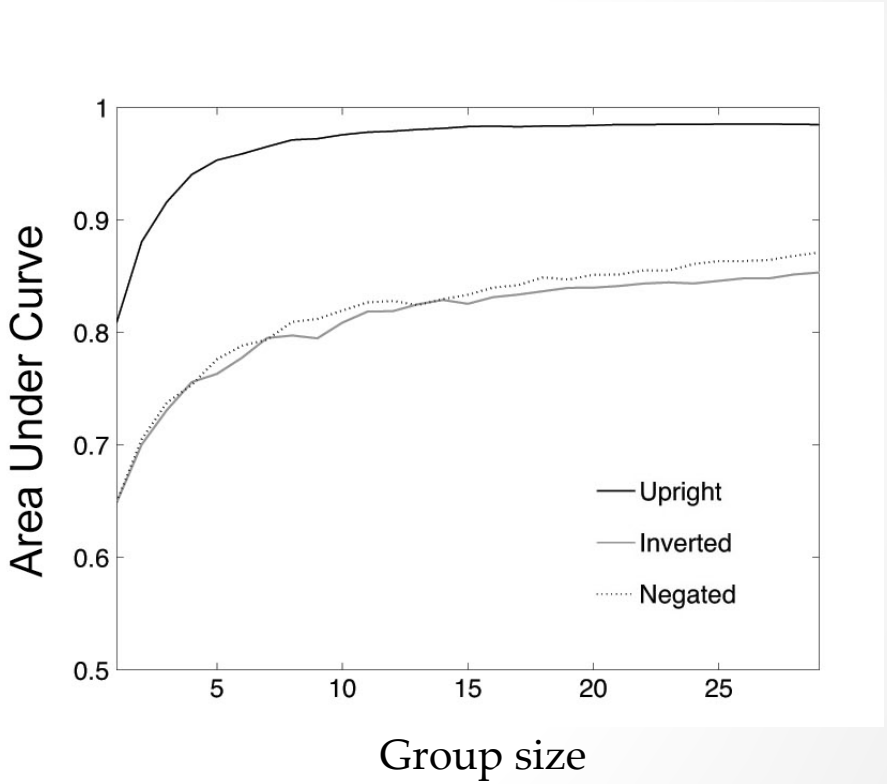


Response Aggregation

- Can we improve performance by aggregating the independent judgments of several operators?
- "Wisdom of the Crowd"

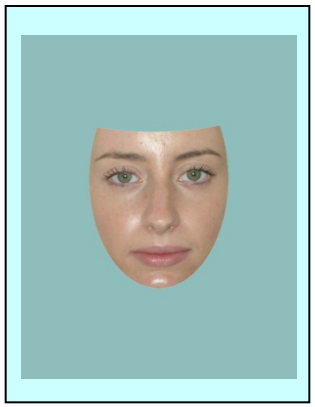
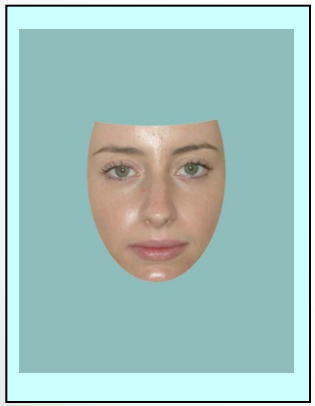
Response Aggregation

- Asked "How likely is it that these images are of the same person?"
- Random sampling to create 300 permutations of group for each crowd size
- Average across group to produce group response



Changing image format

- Can we change the format of the facial biometric to improve matching accuracy?
- One difference between familiar and unfamiliar face matching is that unfamiliar face matching is more dependent on external features of the face
- These also change over time so are unreliable cues to identity
- Prediction: matching may be improved if the external features of the face are matched.

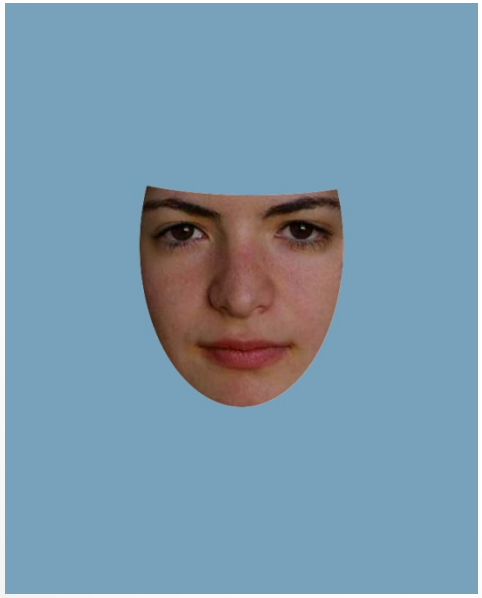



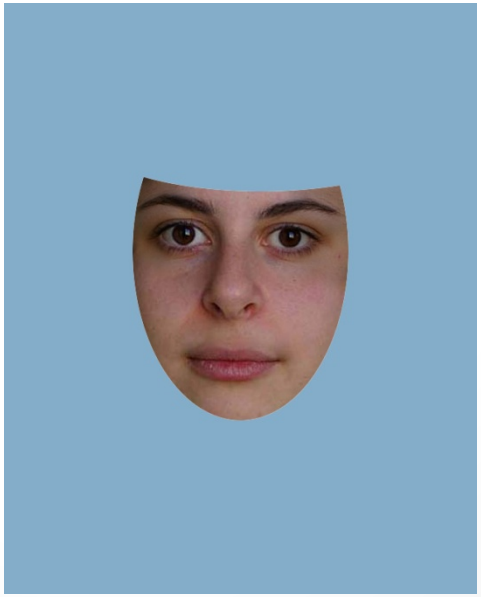
Presentation format

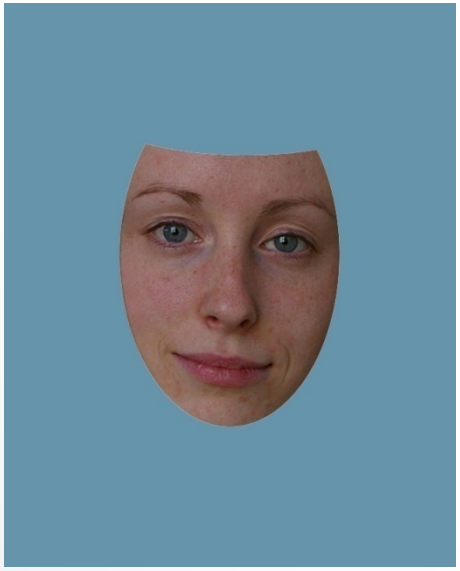
Whole-Whole

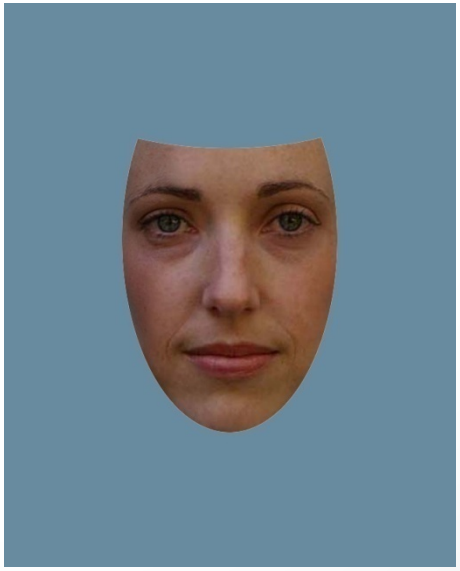
Whole-Internal

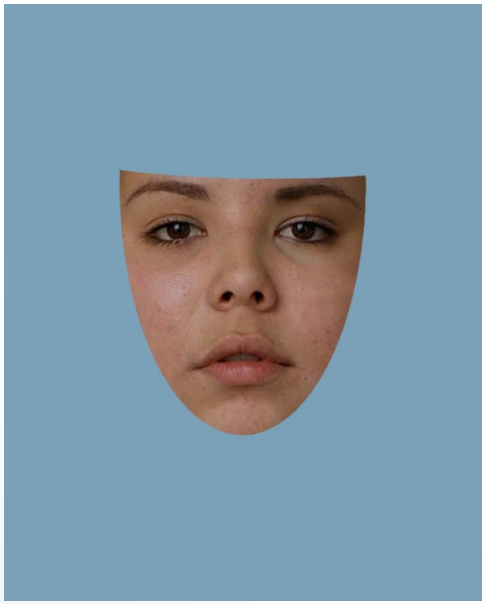
Internal-Internal

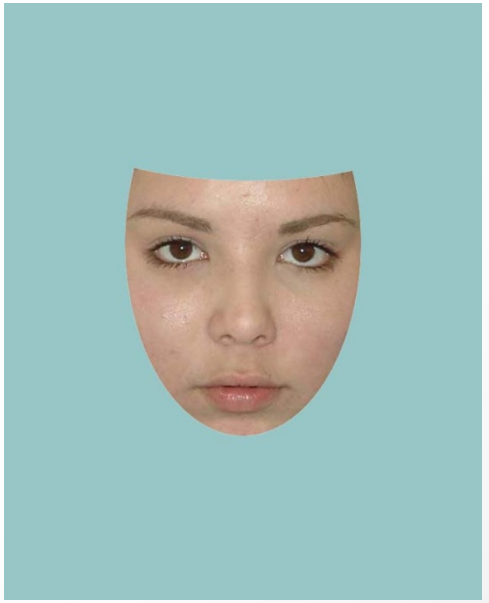


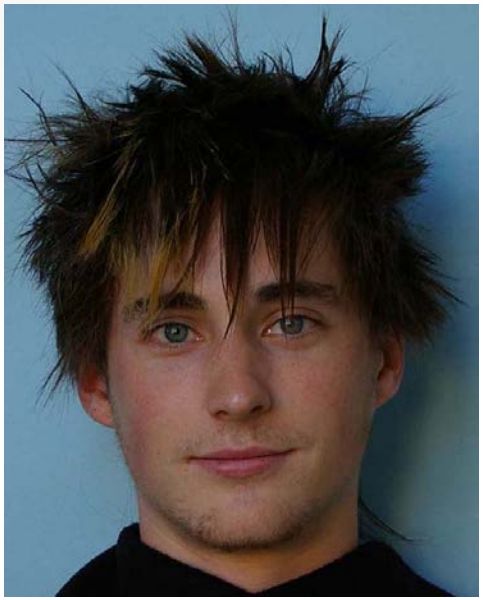


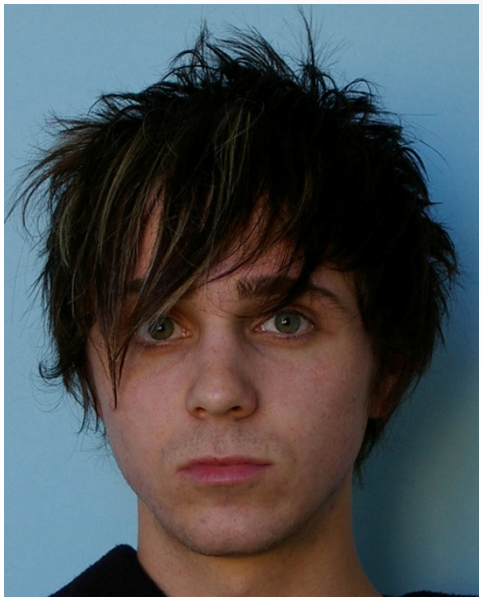


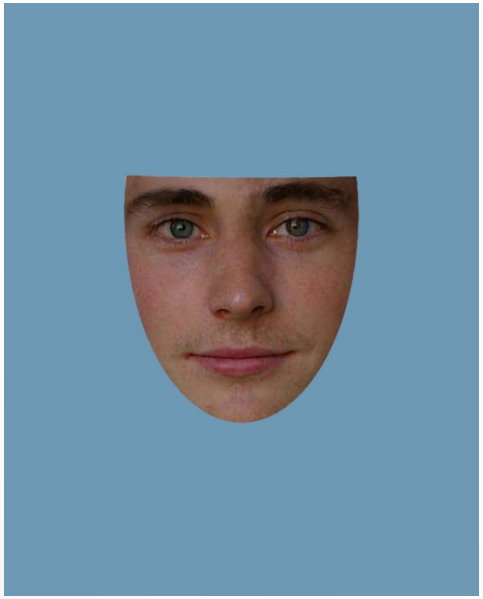


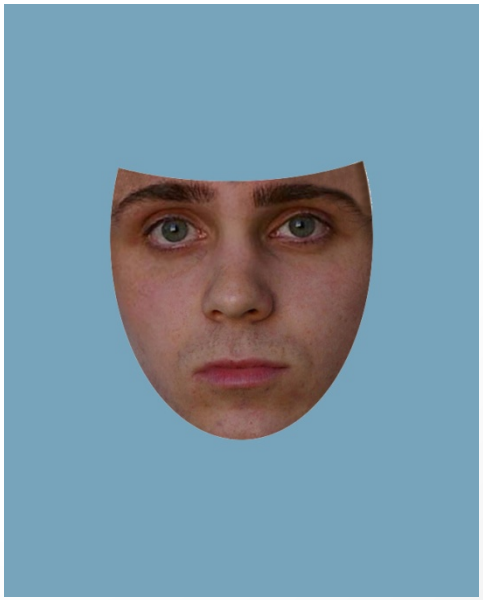






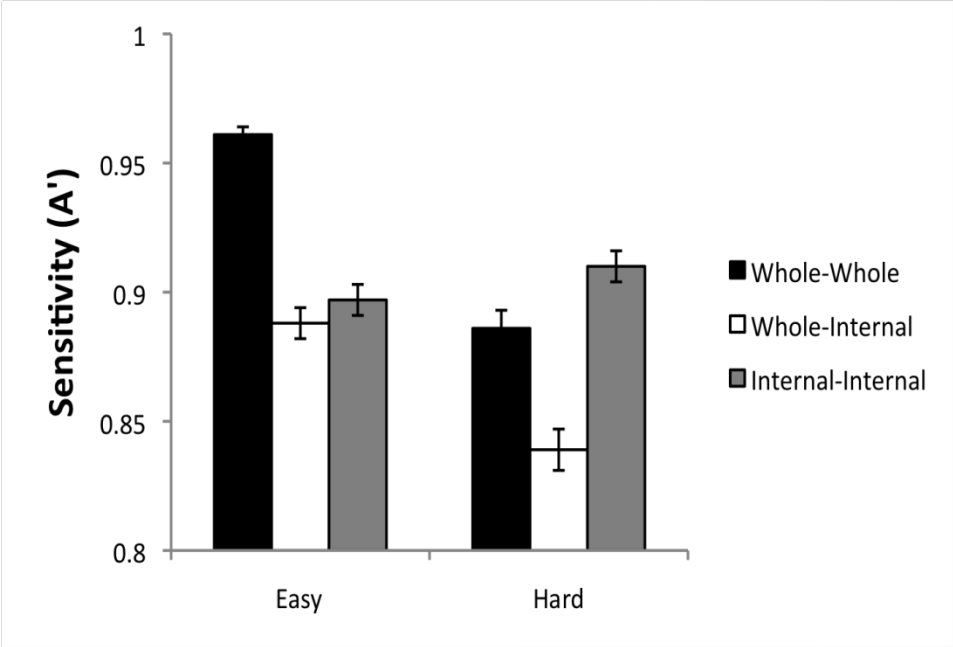






Changing image format

- For hard trials where appearance has changed over time or where the foil is well matched to target, masking external features significantly improves matching performance relative to the unmasked trials



Conclusions & Future research

- There are lots of opportunities to make small improvements to performance
- Added together these changes may represent a very significant improvement in the ability of human operators to detect fraud
- We can "tune" the performance of the humans
- We need to consider both the human and AFR components of the total system
- We should aim to maximize the performance of the total system through a better understanding relative strengths of AFR and human components

Acknowledgements

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA | FACULTY OF SCIENCE

- David White,
- Alice Towler
- Nadia Menon
- Melissa She
- James Dunn
- Manuela Tan
- Ruth Elijah
- Donna Li

Australian Government
Department of Foreign Affairs and Trade

- Michael Matheson
- Shashi Samprathi

Australian Government
Australian Research Council

THANK YOU

Contact Information:

Richard Kemp
University of New South Wales, Sydney

E-mail: richard.kemp@unsw.edu.au