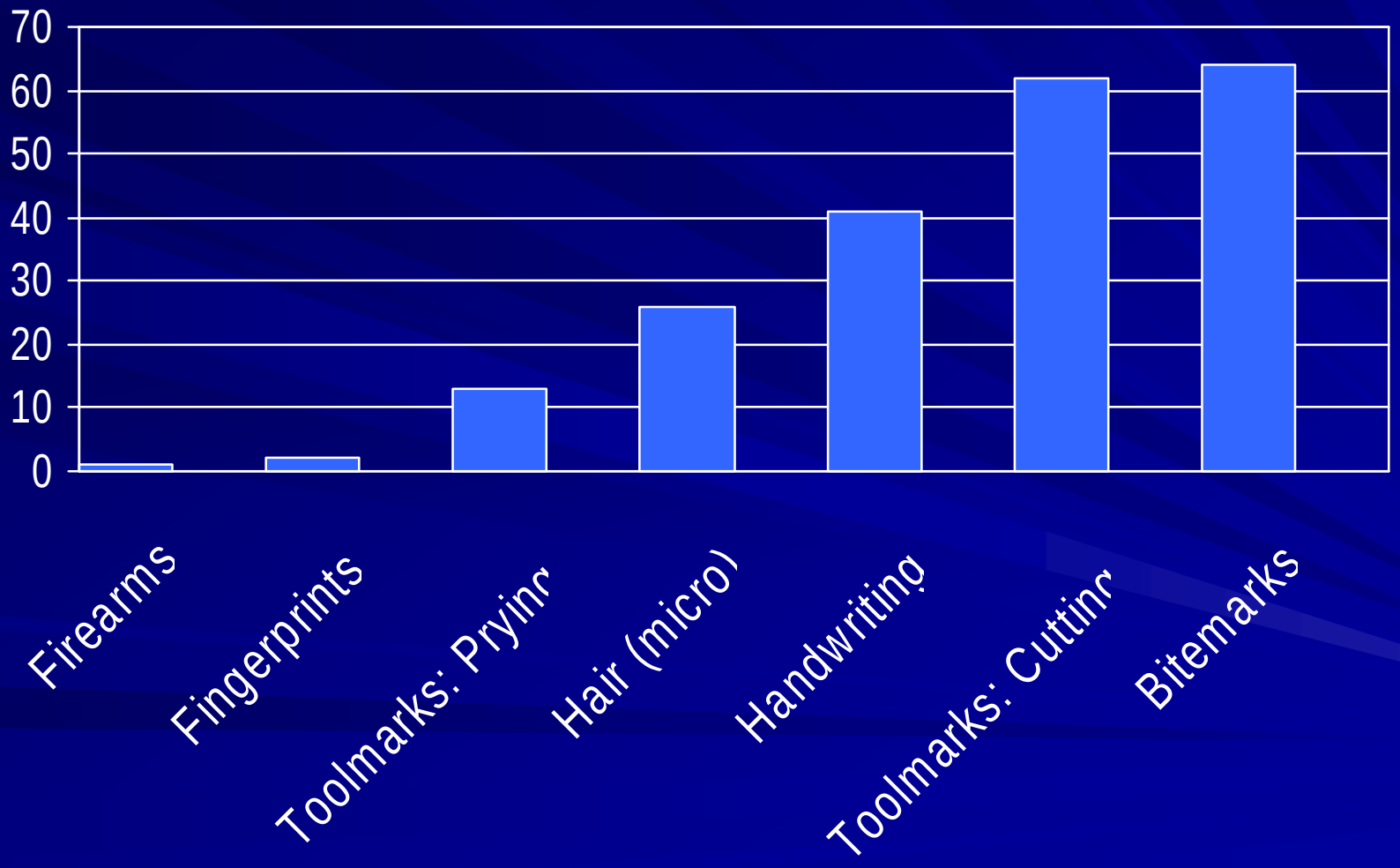
Determining task-specific error rates for forensic science processes is very important.

Error Rates (from various tests)



But if I personally could choose only one reform to undertake, it would not be to do research to tie down task specific error rates under test conditions, it would be to establish appropriate blind testing in actual practice.

Risinger, Saks, Thompson and Rosenthal,
Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion,
90 Cal. L. Rev. 1
(2002)

Any Information not necessary to the exercise of one's expertise will distort results

And the more that information engages emotions and desires, the stronger the distortion will be.

This is one of the best established and supported general propositions of modern cognitive psychology, and has given rise to a great improvement in the methodology of research and application in much of science.

Any process using a human as a perceptor, rater, or interpreter should be “as blind as possible for as long as possible”

--Robert Rosenthal

No area of forensic science can fulfill this fundamental responsibility without a formal consideration of what constitutes domain-relevant and domain-irrelevant information.

To date, no area has instituted such a study.

Why?

Think of the response of surgeons in the 1860's and 1870's to Joseph Lister's message about the need to control microbial infection during surgery.

The usual response by forensic practitioners is that their good faith and their training make them able to resist such distortions.

Thus the data and principles that have been shown universally applicable in every context in which they have been tested don’t apply to them.

Are Forensic Specialists immune?

Consider the case of Brandon Mayfield.

Merely Anecdotal?

How about this:

Procedural Bias in Forensic Examinations of Human Hair

Larry S. Miller

11 Law and Human Behavior
157
(1987)

56 hair identification tests were prepared.

Half the tests reflected the usual practice of presenting a known hair from a “suspect” and a single “questioned” hair from the crime scene, and asking if the two “matched”.
(a “show-up”)

The other half of the tests presented five “known” hairs from “suspects” to be compared to the “questioned” hair from the crime scene, and asked if the hair from the scene matched any of the suspects.
(a “line-up”)

In every test, the “crime scene” hair did not come from any of the “suspects,” though the hairs of all the “suspects” were selected to present characteristics not obviously dissimilar to the crime scene hair.

14 qualified examiners were given four tests each, two from each set of test designs

Erroneous declarations of “match” were found in 3.8% of the responses to the “line-up” condition, but in 30.4% of the responses to the “show-up” condition.

Still not convinced?

Visual hair comparison already known to be too unreliable to draw any general conclusion based on studying it?

Consider the Dror et al Study (2006 Forensic Sci. Int. 74-78)

Five experienced fingerprint examiners were asked by a colleague to evaluate the Mayfield prints after it was known that the FBI had misidentified them

In reality, they were given prints they themselves had found to match in actual cases

Four of the five now came to a different result.

One now said that the latent was too small and smudged to reach a conclusion

And three now concluded that the latent didn’t match the known, (when they had come to the opposite conclusion in the real case)

Think the n is too small?

Think the malleator is too unusual?

How about a replication using more normal context cueing.

Itiel Dror & David Charlton,

Why Experts Make Errors

56 J. Forensic Identification
600
(2006)

6 experienced fingerprint examiners were given eight sets of two prints each by their supervisor.

All of the print pairs given each examiner were from previous cases where that examiner had declared that there was a sufficient basis to declare a match (four each) or an exclusion (four each)

In addition, each of these cases had been rated as to difficulty by the examiner when originally performing the comparison.

In four of the test cases presented (two of previous “match” [one hard, one easy] and two of previous “exclusion,” [one hard, one easy], no context information was provided, merely a request for comparison

In the other four cases (similarly distributed), not uncommon context information was given (“suspect has confessed, etc”).

The test thus resulted in 48 decisions (6 examiners X 8 comparisons each)

Of those 48 decisions, 6 were inconsistent with the previously rendered decision in the actual case. (12.5%)

Two of the six examiners gave results completely consistent with their previous decisions. The other four did not.

Three of the four remaining examiners changed one decision each, and the other examiner changed three.

Four of the changes were in tests where context information was supplied, and two were in cases where no context information was supplied

Five of the switches were in cases rated as difficult, but the one switch in an easy case (from match to exclusion) was in a case containing context information suggesting exclusion.

These effects are not limited to the “forensic identification skill” areas.

DNA in mixed sample situations

Even Forensic Pathology

A Forensic Scientist is not a Detective!

(And should resist wanting to be one!)