Complexity, Level of Association and Strength of Fingerprint Conclusions

By Michele Triplett [1]

Abstract

False convictions and false incarcerations have pushed the topic of forensic errors into the national spot light. Friction ridge comparisons (referred to as fingerprints for the remainder of this paper) are very accurate but errors have occurred. The strength of any conclusion needs to be indicated since criminal proceedings rely heavily on this type of information. The following paper discusses a possible explanation for errors and offers a more accurate and transparent approach for arriving at and reporting results. The proposed approach labels the complexity and demonstrable level of association found between two impressions which allow others to more accurately discern the strength of a conclusion.

Keywords: cold cases, fingerprint comparison, false convictions

[1] Michele Triplett is the Forensic Operation Manager for the King County Regional AFIS Program in Seattle, WA. She is a Certified Latent Print Examiner and holds a BS in Mathematics and Statistical Analysis. She has been employed in the friction ridge identification discipline since 1991 and is actively involved in several committees, organizations and educational events.

Standard Conclusions

Historically, fingerprint conclusions have been reported in a categorical fashion, such as 'the impression has been identified to John Doe'. Reporting conclusions in this manner has made conclusions sound conclusive, when in reality they may be strongly supported with visual data, marginally supported with visual data, or lack visual data that can be successfully demonstrated to others (i.e., simply the beliefs of the practitioners stating the conclusion). In order to determine the strength of the conclusion, the basis behind the conclusion needs to be assessed. Conclusions have been reported categorically as a means of simplifying a very intricate process that was based on a large number of non-quantifiable variables. No statistical model has been able to express the strength of conclusions despite on-going and previous efforts dating back to the late 1800s.

Criterion of Inclusions

Sufficiency to establish an identification is commonly based on either a practitioner's own tolerance level or non-validated administrative point standards set by an agency. Even without a validated sufficiency threshold, past conclusions have seemed fairly reliable; opposing conclusions and errors appeared virtually nonexistent. With the advent of the internet, information sharing has become easier and the variation in practitioners' conclusions has become increasingly more apparent, conclusions are not as definitive as once claimed (*Jackson v. Florida*, 2015; Stacey, 2005; Possley, 2015).

Evaluating Correctness of Conclusions

The lack of a clearly defined criterion for arriving at conclusions makes it difficult to evaluate practitioner's conclusions; without a standard, there is no means of judging correctness. This is extremely concerning when people's liberties and lives are on the line. Currently, the only way to assess a conclusion is to ask for another practitioner's opinion; which is mistakenly viewed as a measure of accuracy. Repeating a conclusion is simply measuring whether or not the conclusion is acceptable to another practitioner; it is not establishing absolute truth.

Establishing Error Rates

In the last decade, millions of dollars have been spent on error rate studies. These studies have assessed the accuracy of practitioner conclusions when comparing manufactured impressions to ground truth conclusions. The studies did not compare the error rates of different methods for arriving at conclusions. The studies indicate that the error rate is low but perhaps higher than previously assumed. Some studies assessed the repeatability of supporting data but they have not evaluated the acceptability of the support behind the conclusions (e.g., an accurate conclusion arrived at illogically would have been determined to be correct for the purposes of the research).

In casework, the ground truth is never known; casework conclusions are labeled as errors when others disagree with the conclusion. Since the research studies are assessing a measurement that does not apply to casework, the results of these studies may not accurately represent the error rate for casework. More importantly, the significant question to attorneys, judges, and the person identified as depositing a fingerprint at a crime scene is not how often experts make errors, rather which conclusions, and which methods, are at a higher risk of error? In order to reduce error rates and strengthen forensic conclusions, improved research would compare the error rates of different methods in order to show which technique produces the best results.

Paradigm Shift

The time has come where it is now essential to establish standards for arriving at conclusions and clear articulation of the strength of subsequent conclusions. Doing so will improve conclusions and give the ability to measure the correctness of conclusions. Instead of oversimplifying conclusions as categorical variables (identification or exclusion), it is more appropriate to present decisions on a continuum that expresses the complexity of a comparison (e.g., *Basic*, *Advanced*, *Complex*) and the demonstrable level of association (such as: overwhelming, marginal, or none). The complexity of a comparison is important because it determines the extent of testing required to ensure the interpretation and amount of data hold up under a critical review. The results of the testing establish the acceptable level of association, which indicates the strength of a conclusion (e.g., a complex comparison does not indicate that a conclusion is weak, it indicates that addition quality assurance measures are required to establish a strong conclusion). It is possible to assess the complexity of an impression in isolation of a comparison; however, the complexity may change during a comparison, making a pre-comparison assessment of an impression unnecessary.

Measuring information with words instead of numbers may seem unusual however; this is common in disciplines that are unquantifiable. For instance, hospitals rate the condition of patients on a wording scale (critical, severe, good, fair, etc.). The words chosen are not simply at the discretion of the doctor, there are criteria for each category so that every doctor rates patients the same. For bone fractures, doctors do not simply report that a leg is broken; they rate the severity of the fracture

in words (compound, hairline, etc.). Again, there are specific definitions for each description to ensure fractures are rated the same. Additionally, doctors use a 4 stage scale when reporting the severity of a cancer diagnosis; with specific definitions for each rating. The pattern evidence disciplines can and should follow suit and report more than a conclusion. Adding information that indicates the strength of the association found would benefit all interested parties.

Scientific Criterion: Data and Testing Over-Confidence

The primary question asked regarding fingerprint comparisons is how much information is enough to establish an identification. As stated above, the answer is not a quantifiable number, however, the accepted criterion used by other non-quantifiable comparative sciences (i.e., based on analytical reasoning) fits well within the realm of fingerprint comparisons. The criterion is to ensure conclusions have sufficient justification within established fundamental principles, to hold up against strong scrutiny. This is often times referred to as general consensus, although the term general consensus can be misconstrued as meaning that the majority of people would arrive at the same conclusion. General consensus is better defined as the conclusion has been debated until all doubt has been resolved. Resulting conclusions may be referred to as inferences that are supported by data. The strength of an inference is determined by assessing whether the support behind the inference satisfies any doubts presented by others.

Conclusion

Conclusions based on specific criterion and vetted against rigorous scrutiny will preempt errors and make conclusions more trustworthy than conclusions based on personal thresholds and confidence levels. Clear thresholds also make it possible to judge the acceptable level of association used to support a conclusion; which helps assess the risk of error for each conclusion (example to follow). Measuring acceptance or rejection based on a criterion is a far more informative approach than judging conclusions based on the beliefs of other individuals. Ultimately, utilizing the following method will provide stronger conclusions and allow others to assess the strength of conclusions.

Simplicity/Complexity Scale (*Basic, Advanced, Complex*)

The following rankings are intentionally minimized into three groups for simplicity. The number of rankings could be expanded but has been found to be unnecessary because the minor differences of opinion that may occur are insignificant to the end result. The criterion listed for each ranking are based on the prevailing views, i.e. tenprint comparisons are considered Basic, latent comparisons are considered Advanced, and comparison based on highly ambiguous or minimal data are considered Complex. Comparisons between listed rankings can be labeled as semi-advanced or semi-complex.

Those using this method must be trained in fingerprint comparisons in order to determine the region and orientation of impressions. Users must be trained in scientific protocols in order to understand concepts such as the amount of adequate testing required. For example, scientific conclusions are never based on one piece of data, such as excluding a person as the source based on the pattern type alone. Plausible conclusions must be tested before arriving at a well-supported conclusion. The testing required for each ranking is based on standard testing requirements for non-quantifiable comparative sciences (ensuring the conclusion holds up to rigorous scrutiny). Demonstrating the basis behind a conclusion is required upon any request.

The determination that the conditions for each ranking are met is not at the discretion of

the practitioner. Whether or not the conditions are met must hold up under rigorous scrutiny.

Many who attempt to rank comparisons in this manner quickly find that this is no different from how they have assessed images in the past. The main difference is that the weight of a conclusion is put in the data, not in the practitioner's beliefs or abilities, which protects against over interpretation and errors.

I. *If* There is sufficient data to establish, not presume, the region and orientation; and;

The data being interpreted consists of clear Galton points, spatial relationship, and intervening ridges; and

The correlation of data would easily be interpreted by others; and

The amount of information is large, not all data needs to be assessed or utilized (such as the majority of tenprint to tenprint comparisons)

Then **The comparison is considered *Basic***

Testing (such as consultation, corroboration, supporting documentation, or testing against strong scrutiny) is not scientifically required for this simplistic of conclusions, a practitioner can determine if the data used and the conclusion will meet the criteria (ID: Holds up to strong scrutiny, Exclusion: region, orientation and a clear target group of minutia).

A review of the conclusion is not necessary but may be set by agency policy.

Examples: Standard tenprint comparisons, comparisons with dissimilarities/discrepancies may be considered *Basic* when the area with the discrepancy is not needed to perform a comparison and arrive at a conclusion (the appearance of differences/discrepancies may exist but the reason unknown. Differences/discrepancies do not necessarily indicate a comparison overall is advanced, complex, or that an identification is not warranted).

Latent print comparisons where the region and orientation are known and the features are very clear and large (more data than necessary) are considered *Basic*.

II. *If* There is insufficient data to establish, not presume, the region and orientation (making the search more difficult); or

Ancillary features (scars, creases, incipient ridges) are being interpreted; or

The interpretation of data has slight ambiguity (may not initially be interpreted the same by others); however, the interpretation of data can easily be demonstrated to the satisfaction of others

Then **The comparison is considered *Advanced***

Testing (such as consultation, corroboration, supporting documentation, or testing against strong scrutiny) is optional but recommended (since the interpretation of data can easily be demonstrated); a practitioner can determine if the data used and the conclusion will meet the criteria (ID: Holds up to strong scrutiny, Inconclusive: No consistency found, Exclusion: region, orientation and a clear target group of minutia or multiple target areas if ambiguity is present).

A review of the conclusion is not necessary but may be set by agency policy to ensure appropriate testing.

Examples: Standard latent comparisons, known impressions deposited with extreme deposition pressure, twisting or smearing, complete tonal reversals, the use of creases, or relying on mostly ancillary features.

III. *If* The interpretation of data (Galton or ancillary features) has predominant ambiguity (the interpretation of data is questionable making it difficult to demonstrate to the satisfaction of others); or

The correlation of data is extremely limited (making it necessary to use rarity, ridge shapes, edges, pores, or features in simultaneous impressions)

Then **The comparison is considered *Complex***

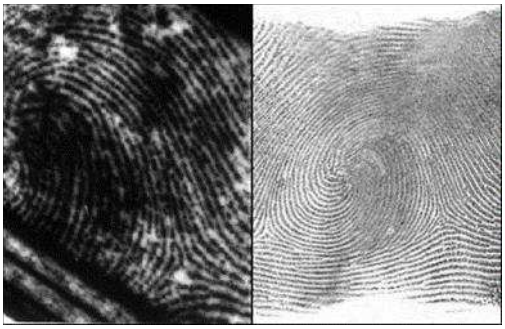
Testing (such as consultation, corroboration, supporting documentation, or testing against strong scrutiny to establish a consensus conclusion) is required to arrive at a conclusion that is well supported and tested under intense scrutiny.

A review is essential to ensure the appropriate amount of testing was performed.

Examples: Tonal shifts, relying on highly ambiguous data (SCRO, Mayfield, Daoud). *Note*: Complexity is distinguished from difficulty in that difficulty level is based on a person's ability while complexity is based on the data in the impressions (either the unknown or known).

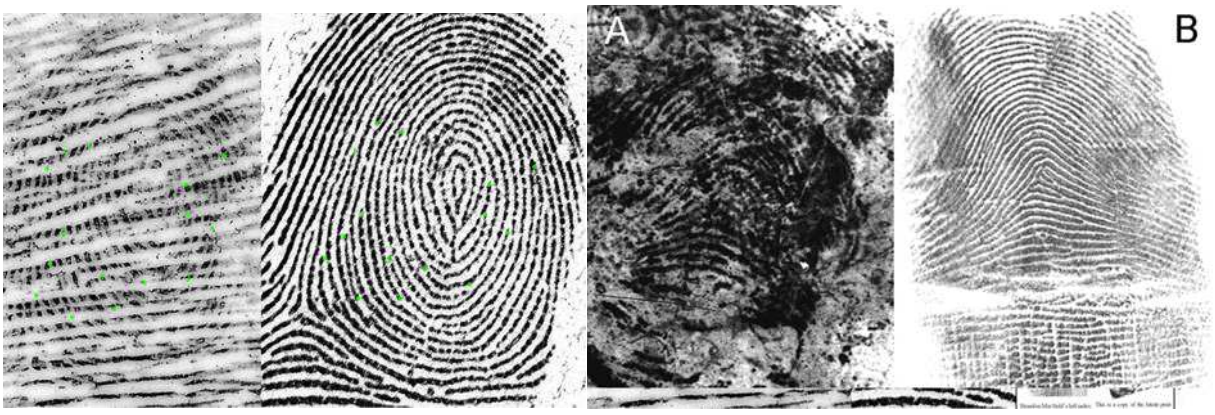
Below are examples of Basic Comparisons. The region and orientation are easily determined. The conclusion can be determined with the use of clear Galton points, their spatial relationship and the number of intervening ridges. The amount of information is abundant and not all data needs to be utilized (CLPEX.com fig's 36, 32, 40) (CLPEX, 2015). Conclusions from Basic Comparisons are very reliable.

The following example may be at the high end of Basic or Semi-Advanced. The region and orientation can be presumed. The conclusion can be determined with the use of Galton points, their spatial relationship and the number of intervening ridges. The amount of information is large and not all data needs to be assessed (FBI fingerprint image).



The comparison below may be considered Advanced since the region and orientation are not standard. However, the features within the image are clear and plentiful (CLPEX.com fig 68) (CLPEX, 2015).

The comparisons below fall into the category of Complex because the features within the unknown impressions are ambiguous, the interpretation of data may not be successfully demonstrated to others (CLPEX.com fig 95, Mayfield fingerprint comparison) (CLPEX, 2015; Saks & Koehler, 2005). Testing the interpretation of data for acceptability is essential to establish the appropriate conclusion.



The complexity of a comparison is based on the amount of ambiguity. The acceptable level of association is based on demonstrability and/or testing performed, which in turn determines the strength of a conclusion. The chart below can be used as a quick reference guide.

SIMPLICITY/COMPLEXITY SCALE: QUICK REFERENCE CHART

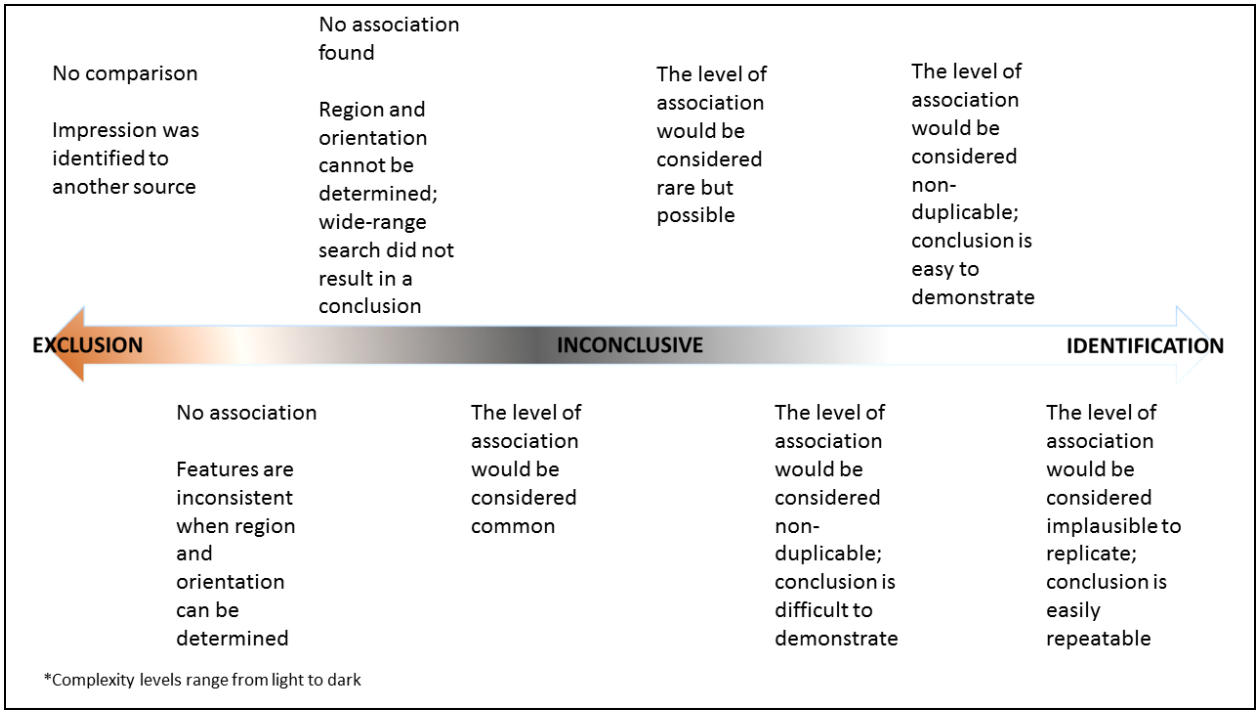
BASIC
Region and orientation can be determined. Use of Galton features, spatial relationship and intervening ridges (assessment of other features is not needed). Not all data needs to be assessed. Testing against scrutiny not required.

ADVANCED
Region and orientation may be questionable. Use of additional features (scars, creases, incipient ridges) or additional aspects (clarity, slight ambiguity of features). Not all data needs to be assessed. The use of ancillary features may be considered at the high end of Advanced. Testing against scrutiny recommended.

COMPLEX
Galton features are predominantly ambiguous. May include the use of edges, pores, or simultaneous impressions due to limited correlation of Galton features. Testing against scrutiny required.

Level of Association Continuum

It may seem reasonable to assume that erroneous identifications are more likely to occur as the level of association decreases (close non-matches; the gray ranges in the level of association continuum); however, this is not the case. Research into past identification errors demonstrates that misinterpretation of ambiguous data and reliance on reproducibility as the test for acceptability are the primary causes of errors. Past errors were found and acceptable associations established by testing the conclusion to ensure the interpretation of data holds up against strong scrutiny, ensuring the basis for the conclusion can be demonstrated to the satisfaction of others (i.e., general consensus) (Stacey, 2005; CBS Interactive, 2012).



Some agencies state the number of Galton points as an attempt at providing a weight to their conclusion. Stating a number of Galton points can be very misleading because it implies a weight that may not actually exist. A high correlation does not establish the strength of a conclusion because the assessment of those points may be a misinterpreted, as seen with the Mayfield error and the Dandridge error (Possley, 2015). The level of association is only meaningful if it can be successfully demonstrated to others, as required by the standard for non-quantifiable sciences.

Articulation of Conclusions

There is a lot of information that results from utilizing this method and therefore a variety of ways to articulate the information. For instance:

The comparison was considered:

a) basic
b) semi-advanced
c) advanced
d) semi-complex
e) complex

Testing performed:

a) none
b) tested against strong scrutiny for acceptable interpretation of data

The level of association is:

a) impression associated to another person, exclusion to this subject by deduction

b) overwhelming inconsistency, exclusion to this subject

c) features too broad to determine specific search area, no consistency found after a wide-range search

d) the level of association is limited or marginal, an amount of consistency seen in others

e) the level of association is high or considerable, not expected in others but plausible (may be referred to as an investigative lead or a person of interest)

f) the level of association is persuasive, difficult to demonstrate but considered implausible to replicate

g) the level of association is compelling, easy to demonstrable, and considered implausible to replicate

h) the level of association is overwhelming, easily repeatable by other experts, and considered implausible to replicate

Specific conclusion in casework could be articulated as one of the following:

"The comparison is *Basic*. The level of association is overwhelming and easily repeatable by others."

Or, "The comparison is *Advanced*. The level of association is *compelling*, easy to demonstrate, and considered implausible to replicate."

Or, "The comparison is *Complex*. *Testing against strong scrutiny* determined the level of association to be persuasive and considered implausible to replicate."

Conclusions presented with this type of information demonstrate to others that the practitioner relied on criteria and demonstrable data to protect against over-interpretation and to ensure conclusions are as solid as humanly possible. This method can also be beneficial to re-assess conclusions arrived at using a different method. The level of complexity, the degree of testing performed, and the level of association will establish the strength behind any conclusion.

The well-known 2004 FBI erroneous identification to Brandon Mayfield can be assessed under this method. Under this method, the identification to Mayfield would have been labeled complex since many of the associations used were ambiguous. A complex rating indicates that testing against rigorous scrutiny is essential. Rigorous scrutiny was not performed by the FBI since the culture at that time discouraged disagreement among examiners (Stacey, 2005). The Spanish experts ap-

proached their review in a more critical fashion by questioning the interpretation of data and the conclusion. If rigorous testing against scrutiny had occurred within the FBI or by the external practitioner reviewing the comparison for Mayfield, then the conclusion would have been labeled 'e' at best. If it had been known that it was a complex comparison and rigorous testing against scrutiny had not been performed, yet a conclusion of 'f', 'g' or 'h' was being reported, then others would clearly see the red flags in this case. Other past errors can be tested against this system as well. Each would show that an identification would not have held up under this standard.

References

CBS Interactive. (2012, November 4). Lana Canen freed over bad fingerprint evidence after 8 years in prison for Indiana murder. *CBS News*. Retrieved from http://cbsn.ws/1MItPhY

CLPEX. (2015). *Complete latent print examination*. Retrieved from http://www.clpex.com.

Federal Bureau of Investigation. (2013, July 9). *Latent hit of the year award: Fingerprint tool helps solve 1999 murder*. Retrieved from http://1.usa.gov/1WYhuaD

Kim Jackson v. State of Florida, SC13-2090 (Fla. 2015). Retrieved from http://fla.st/1H1VJVK.

Possley, M. (2015, October 12). Beniah Alton Dandridge. *The National Registry of Exonerations*. Retrieved from http://bit.ly/1Mor8AH.

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309*(5736). DOI: 10.1126/science.1111565

Stacey, R. B. (2005). Report on the erroneous fingerprint individualization in the Madrid train bombing case. *FBI Forensic Science Communications*. Retrieved from http://1.usa.gov/1LfhKv3.

Correspondence concerning this article should be addressed to:

Michele Triplett, 516 3rd Ave, Mail Stop: KCC-SO-0100, Seattle, WA 98104, Email: s.triplett@comcast.net

How to cite this article:

Triplett, M. (2016). Complexity, level of association and strength of fingerprint conclusions. *Journal of Cold Case Review, 2*(1), 6-15.