A Comparison of Similarity Scores Between Bullet Casings: Forensic Analysts Versus an Algorithm

Maria Cuellar
Carnegie Mellon University

Follow this and additional works at: https://lib.dr.iastate.edu/csafe_conf

Recommended Citation

This Presentation is brought to you for free and open access by the Center for Statistics and Applications in Forensic Evidence at Iowa State University Digital Repository. It has been accepted for inclusion in CSAFE Presentations and Proceedings by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
A Comparison of Similarity Scores Between Bullet Casings: Forensic Analysts Versus an Algorithm

Comments
Posted with permission of CSAFE.

This presentation is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/csafe_conf/25
SIMILARITY SCORES OF BULLET CASINGS

Maria Cuellar, PhD
Criminology Department
July 31, 2018
Outline

- Motivation
- What we’d like to do
- What we’ve done so far
Motivation

• 2004 Madrid bombings
• Partial print inside bag
• FBI finds 20 candidates, five examiners choose one: Brandon Mayfield’s
• Spanish national police says no, FBI apologizes
• Dror experiment
• Theory is: Contextual bias

Itiel E. Dror, David Charlton, Ailsa E. Peron, “Contextual information renders experts vulnerable to making erroneous identifications.” (2005)
Contextual bias

- Different types
- Want to find evidence of this in firearm identification
- Reference materials bias

Dror 2017.
Originally: Look for evidence of reference bias

- *Test for reference bias in firearm identification*: Designed experiment
- Do individuals make worse decisions (higher error rate) when shown reference materials than when they are not?
  - ...
- What is the error rate without the reference bias?
- *Foundational validity* of firearm identification
  - ...
- We do something a little different: Can individuals distinguish between matches and non-matches, even in *the difficult cases*?
Firearm identification

• Firearm identification is the task of determining whether two bullets or cartridges were fired by the same firearm.

• Most is done subjectively (human).

• There is a shift to perform comparisons objectively (algorithm).
  • The National Integrated Ballistic Information Network (NIBIN) is a national database of digital images of spent bullets and cartridge cases that were found at crime scenes or test-fired from confiscated weapons.
  • Examiners use NIBIN to select candidates for comparison.
  • NIBIN algorithm is not available for use by researchers, so for our experiment we use the algorithm created by Tai (2017).
Firearm mechanism

CBC News report on firearm identification technology

Maria Cuellar, PhD
Images used for cartridge analysis

Head stamp

Primer

Firing pin impression

Head stamp

Firing pin impression

Breech marks

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org
Data

108 images, 12 guns (9 images per gun)

NBIDE (NIST Ballistics Imaging Database Evaluation)
Split into four groups using Tai’s algorithm

Non-matches

Distributions of the non-match pairs and the match pairs (Figure 12 in Tai, 2017).

Matches

Similarity score distribution split up into four groups
Our test...so far

- Overlap in Tai’s results, also these are most difficult cases
- Can individuals distinguish between matches and non-matches, even in the difficult cases?
  - Ask: How similar are two breech faces?
  - Select the easy and difficult cases by using Tai’s algorithm.
- Novices select degree of similarity in scale of 0-100 (no inconclusive).
Examples of images from four groups

(a) High similarity match.  (b) Low similarity match.  (c) High similarity nonmatch.  (d) Low similarity nonmatch.

Known matches  Known non-matches
Survey design

• **Two surveys** — Two conditions (LM, HN) and four conditions (HM, LM, HN, LN) and .

• **Software** — Survey in Qualtrics software, Respondents in Amazon Mechanical Turk, analysis in R.

• **Lots of checks** — Initial questions don’t count, randomization in order, black screen between questions, ask same questions twice to check consistency.
Preliminary results

High-similarity matches

Low-similarity matches

High-similarity non-matches

Low-similarity non-matches

Densities of responses by group, for four-group experiment

Research funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) - forensicstats.org
Preliminary results

Low-similarity matches

High-similarity non-matches

Densities of responses by group, for two-group experiment
Discussion

- Novices seem to distinguish between matches and non-matches even in difficult cases.
- How many of the answers were “wrong”? By one definition 30%, by another 10%.
- Next time want to ask for
  - Match vs. nonmatch (error rate?),
  - Similarity score (distributions?),
  - Confidence level (less confidence in difficult cases?).
- Try it with experts.
- ...Then we can start to find evidence of contextual evidence.
QUESTIONS?
Preliminary results

• T-test assumes data are drawn from normal distribution and the data (images and people) are independent.
  • Used a test (Kolmogorov—Smirnov) that does not assume parametric form or independence. Results hold.
  • Average standard deviation per individual is consistent through sets (between 19 and 24).
• Multiple comparisons are under-conservative.
  • Use Bonferroni correction. Results hold.

• Limitations.
  • This analysis has a small sample size (pairs of images, respondents), we did not record respondents’ confidence in their responses, we only had access to one algorithm, we have not yet surveyed experts.
Similarity algorithm

- **NIST 2007** — created an algorithm to provide a similarity score for a pair of breech face images (not publicly available).
- **Roth et al. 2015** — created algorithm that was the author’s “best guess” about what the NIST one did.
- **Xiao Hui Tai 2017** — received the Roth algorithm and modified it in several ways.
  - Procedure: 1) Automatically select breech face marks, 2) level the image, 3) remove the circular symmetry, 4) perform outlier removal and filtering, 5) estimate similarity by maximizing correlation through translations and rotations, and 6) compute the probability of obtaining a higher score by chance.
Participants: Four-group survey

- **Number of participants** — 120
- **Gender** — 49 were female and 71 were male.
- **Age** — 36 and 40 years old, on average.
- **Income** — $48,600 USD per year (with a standard deviation of $3,550 USD).
- **Educational level** — Mode being a 4-year college degree.
- **Duration** — 18.5 minutes with a standard deviation of 9.14 minutes.
- **Visual impairment** — 77 had none, 43 had an impairment with correction.
- **Training as firearm examiner** — None.
- **Gun ownership/use/training** — 88 did not own a gun, 25 did, and 7 would rather not say. 7 used it very infrequently, 12 used it once per month, and 6 used it once per year. 30 had gun training, 71 did not, and 4 would rather not say.
- **Payment** — $3 USD for participating in the survey and a bonus of $0.50 for finishing it.
Participants: Two-group survey

- Number of participants — 120
- Gender — 50 were female and 70 were male.
- Age — 31 and 35 years old, on average.
- Income — $42,600 USD per year (with a standard deviation of $5,850 USD).
- Educational level — Mode being a 4-year college degree.
- Duration — 11.71 minutes with a standard deviation of 6.50 minutes.
- Visual impairment — 83 had none, 35 had an impairment with correction.
- Training as firearm examiner — None.
- Gun ownership/use/training — 92 did not own a gun, 26 did, and 2 would rather not say. 6 used it very infrequently, 14 used it once per month, and 6 used it once per year. 31 had gun training, 88 did not, and 1 would rather not say.
- Payment — $3 USD for participating in the survey and a bonus of $0.50 for finishing it.