D55 Combining Human Analysis With Human or Computerized Analysis to Reduce Classification Errors

Barry Cushman, BS, American Polygraph Association, PO Box 8037, Chattanooga, TN 37414; and Frank Horvath, PhD, Ntl Acad for Credibility Assessment, 7540 Pickens Ave, Fort Jackson, SC 29207-6804*

After attending this presentation, attendees will have a better understanding of how forensic analysts might decrease classification errors by complementing human, manual analysis of data with computerized analysis or with a second analysis done blindly by another human analyst. While polygraph data are used here, the concept and application applies to other disciplines that are dependent on human analysis, e.g., fingerprints or tool mark impressions.

This presentation will impact the forensic science community by challenging forensic analysts to investigate how the pairing of analyses increases the accuracy of professional opinions in their respective disciplines.

Introduction: For some time, forensic analysts have recognized the value of having a second analyst review their work. Traditionally, reviewers have served in a quality-assurance capacity and how their conclusions impact error rates is not well known. This study examined how two analyses could be combined to reduce errors.

Method: In this study, polygraph charts from 88 single-issue field polygraph examinations were independently analyzed by 10 human evaluators seeking proficiency certification. In each of these examinations, "ground truth" (i.e., the certainty of the examinee's veracity) had been confirmed by criteria typically used in published studies: 48 deceptive and 40 truthful. Each of the evaluators blind-analyzed the data in each case and made one of three classifications: truthful, deceptive, or inconclusive (unable to decide). An "inconclusive" occurs when the data are either insufficient for proper analysis or when the analyst believes a truthful or deceptive opinion is equally probable. Inconclusives are intended to provide a way to avoid errors and are generally resolved with re-testing and are therefore not counted as errors.

The polygraph data in each case were re-analyzed separately using two commercially available and regularly used computer algorithms for the analysis of polygraph data: White Star 2 and OSS-3. Both algorithms were run from within an updated version of the polygraph software that created the original data (Axciton Computerized Polygraph, version 9.2.0). Each reports the probability of truthfulness or deception. These computerized classifications were individually paired with the human evaluators' classifications. After pairing, the decision was reclassified to "inconclusive" when the two decisions disagreed (i.e., one concluded truthful; the other, deceptive). If the human evaluator did not make a decision of truthfulness or deception (i.e., the classification was inconclusive), the algorithm's classification was used. Otherwise, the human evaluator's classification did not change. Each of the 10 human evaluators' 88 decisions was, in this way, reclassified twice; first by pairing them with the White Star and then with OSS-3 analysis.

The proportion correct and erroneous were totaled for each examiner for each method, i.e., the original (human) classifications and those based on the pairing with each of the two computer algorithms as described above. The data were then analyzed to assess whether there were any differences between the human evaluators' original classifications and the two reclassifications.

Results: Pairing of the human and computer algorithm outcomes resulted in statistically significant decreases in the proportion of errors for truthful or deceptive examines, or both, with both algorithms.

Specifically, the original classifications for the 10 evaluators averaged 87% correct and 13 % errors (excluding inconclusive classifications). The OSS-3 and White Star 2 analyses resulted in 82% correct and 18% errors and 90% correct and 10% errors, respectively. The overall proportions of correct and erroneous OSS-3 and White Star classifications were not significantly different from the averages of the human evaluators. When paired with OSS-3, human classifications averaged 89% correct and 11% errors. When paired with White Star, classifications averaged 95% correct and 5% errors. More simply, errors decreased an average of 21% when paired with OSS-3 and 62% with White Star.

Discussion: This study supports the idea that "two heads are better than one," whether a human evaluator's decision is paired with another (blind) human evaluator or a computer analysis, when the goal is to reduce decision errors. While human analysts were not paired with human analysts, the algorithms appear to represent two average human evaluators, and the results provide support for the premise that when two agree in their conclusions, the likelihood their analyses will yield fewer errors increases.

Blind Analysis, Computer Analysis, Reducing Errors

 ** Presenting Author*