Examining the Feasibility and Effectiveness of Fingerprint Lineups using Fingerprint Matching Techniques

by

Jerry Huang

Advisors:
Dr. Adele Quigley-McBride
Dr. David Banks

Committee Members:
Dr. David Banks
Prof. Brandon Garrett
Dr. Alexander Hartemink

Table of Contents

Acknowledgements

I would like to thank my advisors Dr. Quigley-McBride and Dr. Banks as well as my committee members, Professor Garrett and Dr. Hartemink, for their guidance, support, and individual expertise in each of the fields I explored in this project. In particular, I cannot be more thankful to have Dr. Quigley-McBride as my advisor as she introduced me to this exciting project and helped guide me through many of the challenges that I encountered time and time again. I would also like to thank Dr. Eldridge for her expertise in forensic analysis and for taking the time to evaluate the fingerprint lineups that we generated.

I would also like to thank my parents Xin and Wei Huang for continually to encourage my interest in technology, as well as my friends, notably Harry Xie, who has listened to my rants and ideas about my research countless times.

Abstract

In the past 15 years, researchers from many disciplines have come together to systematically investigate how cognitive bias that affect the conclusions reached by forensic analysts and develop methods to counteract cognitive bias in forensic analyses. The current project focuses on fingerprint analyses—a pattern recognition technique which relies on human perception and judgment. Cognitive bias in fingerprint analysis can arise from analysts having 1) information that would normally be withheld during a traditional, blind scientific experiment and 2) the responsibility of making impactful *categorical* decisions (e.g., same source, different source, inconclusive) that might result in the arrest or release of a suspect under time constraints.

In this project, we examined the feasibility and effectiveness of fingerprint lineups. Quigley-McBride and Wells (2018) demonstrated that fingerprint lineups could reduce or eliminate biases in fingerprint analyses. This is because comparing a lineup of fingerprints to the crime print rather than a single fingerprint mimics the blinding that occurs in a traditional experiment. When examining fingerprint lineups, analysts cannot use irrelevant or potentially biasing information to determine if the suspect's print is a match to the print from the crime—they first need to determine *which* fingerprint belongs to the suspects. Several computer science algorithms were tested that could compute the level of similarity between two fingerprints. The algorithm selected for use was able to distinguish between fingerprints that were or were not from the same source with an impressive ~85% accuracy when using non-pristine fingerprints.

Using this algorithm, fingerprint lineups were generated using databases maintained by NIST, and an expert fingerprint analyst was asked to evaluate the lineups. The analyst's feedback indicates that the lineups had the intended effect on the decision-making—they had to carefully examine and pay attention to details in each fingerprint to perform the task. Unfortunately, our

results also demonstrate how difficult it is to generate lineups that are hard enough to challenge fingerprint examiners. We concluded that a very large database of fingerprints is necessary to produce a lineup of fingerprints similar enough that an expert examiner would need to carefully consider the confirming and disconfirming evidence in each fingerprint to make a conclusion.

Definitions

As this project spans the fields of computer science, statistics, and forensic science, this definitions page is included to provide additional context for readers.

Friction ridge pattern: Refers to ridges that exist on the tip of a person's finger, which form the fingerprint marking when that fingertip comes into contact with a surface.

Latent print/fingerprint: A fingerprint that has been lifted from a surface, rather than rolled on a ten-print ink card or some other intentional collection method.

> *Print*: In this thesis, "print" and "fingerprint" are used interchangeably.
>
> *Pristine print*: Fingerprints that show the complete or almost complete friction ridge pattern and are free of distortions.
>
> *Non-pristine print*: Fingerprints that suffer from information loss and/or distortions.
>
> *Crime scene fingerprints*: Fingerprints found at the scenes of crimes. Typically, these fingerprints are also considered non-pristine prints.
>
> *Matching versus non-matching prints*: Matching prints are two fingerprints that originate from the same source/finger. Non-matching prints are prints that may be similar but are ultimately not derived from the same source/finger.
>
> *Suspect fingerprint*: After a fingerprint is found at a crime scene, investigators will try and find a set of fingerprints that match the crime scene fingerprint by querying a fingerprint database. The set of prints returned from the database (which may or may not include true matches) is a set of suspect fingerprints. Alternatively, a suspect fingerprint could be obtained from someone who has become a suspect using other evidence or investigative methods (e.g., eyewitness identification or recall, CCTV footage, motive).

Fingerprint lineup: Designed to mirror traditional police lineups, fingerprint lineups include one print from a suspect and several other prints (usually four to five) from people known-to be innocent, but who have fingerprints that look like the suspect's fingerprint. The fingerprint examiner's job is to determine whether there is a fingerprint that "matches" the fingerprint collected from a crime scene within the lineup.

> *Fillers*: Inked fingerprints that are from people who are known to be innocent (e.g., people who are deceased or were incarcerated when the crime occurred), but exhibit similar features to the suspect fingerprint. The suspect fingerprint is embedded among several of these "filler" examples to test the perception of the analyst.
>
> *Fair lineup*: A fair lineup is a lineup where filler fingerprints closely resemble the suspect's fingerprint so that an examiner must spend time carefully looking at each fingerprint to complete the task.
>
> *Biased lineup*: A biased lineup is a lineup where the suspect fingerprint stands out from the fillers. An extreme example would be if the suspect fingerprint was a whorl, but all the filler fingerprints were loops.

Minutiae: Points in a fingerprint where ridge lines end or split. See the below definitions for "bifurcations" and "ridge endings".

Bifurcations: Points in a fingerprint where friction ridge lines split into two or more lines. Also sometimes called “forks”.

Ridge endings: Points in a fingerprint where friction ridge lines end. Also sometimes called “terminations”.

Graphs (vertices and edges): Fingerprints can be modeled as graphs, which are composed of points (vertices) and lines that connect these points (edges). Typically, in fingerprint matching algorithms, vertices are placed wherever minutiae exist.

Introduction

In 2009, the National Research Council (NRC) called into question the reliability and accuracy of forensic science disciplines after completing a comprehensive review of the research, policies, and practices currently in place in the USA. The NRC found that almost all forensic techniques lacked the basic features of scientific endeavors, such as information about error rates, procedures that safeguard against error and bias, and standardized criteria, rules, and procedures within each forensic discipline. In response to this critical report, researchers and practitioners have sought ways to establish error rates, standardize the approaches currently being used in the United States, and create procedures that minimize problematic forensic decision-making practices and the negative consequences of those practices, notably including wrongful convictions.

One of the main issues that arose from the 2009 NRC report was the impact of cognitive biases on forensic decisions. Over the past two decades, a large body of literature has already been written, demonstrating the presence of contextual bias effects in many forensic domains. The term "cognitive bias" refers to a class of cognitive phenomena in which the way people collect, perceive, interpret, and evaluate information is influenced by the decision-maker's pre-existing beliefs, expectations, motivations, and/or the situational context in which the decision is being made. In the case of forensic science, cognitive bias stems from analysts having different mindsets, having access to different information, or completing their work in different contexts. In such circumstances, two highly qualified, experienced people analyzing the same evidence might come to different conclusions (Kassin, Dror, & Kukucka, 2013). The possibility that these cognitive biases could have an influence on forensic analyses is real—forensic analysts frequently have access to information that would not be available in blind scientific experiments. Sometimes analysts are asked to approach their analyses in particular ways or communicate the

information in informal ways (in addition to the formal report). For example, analysts are often contacted to answer specific questions from police investigators, asked to complete an analysis quickly so the suspect can be arrested, or instructed to provide evidence that the prosecution or defense can use at trial. Research has shown that these small differences in mindset can significantly affect how analysts evaluate evidence, resulting in bias in favor of a particular conclusion (Spellman, Eldridge, & Bieber, 2021). Normally, biases of this kind are controlled using standardized and blind procedures, and the error attributable to chance and biases can be accurately quantified and expressed using statistics. However, such procedures and practices are uncommon in the forensic science disciplines.

There are many underlying sources of bias and error in fingerprint analysis (Cole, 2005). For instance, media (notably in the form of fiction or news) has encouraged people without training in scientific reasoning and methodology, such as jurors, judges, attorneys, defendants, and the public, to believe in the credibility of fingerprint evidence (Cole & Dioso-Villa, 2009). Although error rate statistics among fingerprint examiners is more readily available now (Thompson, Tangen, & McCarthy, 2014; Ulery et al., 2011; Ulery et al., 2012; Ulery et al., 2014), the people who make important decisions in criminal cases (police, attorneys, jurors, and judges) tend not to question the validity and credibility of fingerprint analyses and conclusions, leading to a higher risk of wrongful conviction when erroneous fingerprint evidence is used in a case (Edmond, 2022). Even in the absence of other corroborating evidence against the defendant or in the presence of exonerating evidence, people tend to over rely on what the fingerprint evidence suggests is true and might discount information that suggests otherwise.

Potential solutions for eliminating bias

Now that there are ample studies demonstrating the existence of cognitive bias and error in fingerprint evidence, researchers have begun to explore solutions. One popular solution

involves the use of "information management protocols," which were inspired by other domains where erroneous decisions can have serious consequences, such as medical procedures (Kohn et al., 2000). In the context of forensic science, this type of procedure is typically referred to as *linear-sequential unmasking* (*LSU or LSU-Expanded;* Dror & Kukucka, 2021). Although LSU protocols are the most well-known solution, and perhaps the simplest to implement (Quigley-McBride et al., 2022), there are other effective solutions.

One solution that *eliminates* the biasing effects of additional, suggestive information is a lineup procedure (Quigley-McBride & Wells, 2018; Quigley-McBride, 2020). This method involves finding other fingerprints that are very similar to the crime print and the suspect print but are known to come from someone innocent (e.g., the police chief, someone with a solid alibi, or someone incarcerated at the time). The suspect's fingerprint is then embedded among these similar-looking fingerprints from people who, even if the analyst incriminates them, would never be prosecuted for the crime. Ultimately, when presented with this procedure, extra information incriminating the suspect might make the analyst more inclined to make a "same source" conclusion, but that information cannot affect decisions about a lineup of fingerprints. The analyst will still need to determine *which print* came from the suspect, which the additional, analysis-irrelevant information cannot help with. Identifying whether one of the prints belongs to the same person who left the fingerprint at the crime scene is, by nature, a data-driven task requiring the analyst to compare the friction ridge patterns and minutiae in each fingerprint.

Police have traditionally used lineup procedures when asking an eyewitness to determine if the police suspect is the person the eyewitness saw commit the crime. If the eyewitness's memory is good enough, he or she should be able to distinguish the culprit from a group of similar-looking people in a controlled setting. Even though fingerprint analysis is a perceptual task rather than a memory task, this logic also applies to fingerprints. A fingerprint lineup

changes the task from "are these two fingerprints from the same source?" to "is *one* of these fingerprints from the same source as the one found at the crime scene and, if so, which one?" Importantly, other solutions to forensic cognitive bias merely reduce bias or reduce opportunities for bias to occur, but evidence lineups have the potential to *eliminate* bias and errors of the type that lead to wrongful convictions—false positives. In addition, if the analyst picks one of the fingerprints from known innocent people, then they have made an incorrect determination, so in this way each lineup can serve as an on-the-job proficiency test.

The primary issue associated with the implementation of the evidence lineup procedure in practice is limits on time, funding, and personnel. Not only does the task itself take longer because there are more fingerprints to examine, but also the lineups themselves need to be created. A fair lineup that effectively "hides" the suspect fingerprint among similar others is difficult to create, especially if the decision-maker is an expert in that comparison task. For the lineup to work, the fingerprints need to be similar enough that an experienced analyst would need to slow down and actually consider the possibility that none of the fingerprints are from the same person as the crime fingerprint. This would require either another analyst to create the lineup, a form of software that can create the lineups, or both.

Given that laboratories are already dealing with significant backlogs, it is not feasible to expect them to set aside the time of some analysts to create lineups for their colleagues. However, it is not unusual for algorithms and software to be used in a forensic laboratory setting. In fact, fingerprint analysts already make use of a software called "Automated Fingerprint Identification System" (AFIS), which is used to scan fingerprints from crime scenes and find other candidates that are in the database that have a similar configuration of features or minutiae. This software was first developed in the 1980s and is primarily used to find candidates in cases where there is no suspect identified via other investigatory means. Although there are issues

associated with bias and accuracy when using such software (Dror et al., 2012), there are also benefits that come from offloading some cognitive tasks to technology (referred to as distributed cognition; Dror & Mnookin, 2010). In a similar way, the task of creating lineups of similar looking prints could be offloaded to algorithms to increase both the practicality and efficiency of fingerprint lineups in the field.

Thus, the current study sought to test out an algorithm created to produce lineups that are difficult enough that examiners will need to perform the data-driven task of determining whether one of the fingerprints in the lineup matches the crime print. Eventually, once the algorithm was perfected and could produce lineups that were difficult enough, we planned to ask expert examiners to review the lineups (some of which would contain a match and some of which would not) and make decisions about those lineups as well as answer follow up questions about their experience doing this task. To create the algorithm, several different approaches were tested for the purpose of creating fingerprint lineups to find an optimal approach and determine whether the available fingerprint databases were sufficient for the task of creating lineups. We hope this tool can be used in the future by researchers and by laboratories to create evidence lineups for use in casework—particularly for cases or analyses that are prone to bias and with errors that result in serious consequences, such as verification analyses of another examiner's conclusions.

Method

To create an algorithm that would be able to generate high quality fingerprint lineups, the first step was to create an algorithm that could accurately distinguish between matching and non-matching fingerprints. Each algorithm that was experimented with was inspired by commonly used machine learning techniques and/or fingerprint matching papers, and each algorithm was tested using both pristine and non-pristine fingerprint databases as suspect prints are commonly non-pristine. Afterwards, the fingerprint generation algorithm would use the algorithm mentioned above to generate fingerprint lineups based on similarity scores assigned to pairs of fingerprints. It is important that each filler print in a lineup has a similarity score that is high enough such that a fingerprint examiner must carefully examine the print, but not too high as to confuse the fingerprint examiner. Once fingerprint lineups were generated, fingerprint analysts would be asked to complete surveys about the difficulty and quality of the lineups and report their conclusion for each lineup presented.

Fingerprint Databases

At first, we used the fingerprint data set used in the Quigley-McBride and Wells (2018) experiment. This database includes a total of 735 labeled, non-pristine fingerprints which served as the benchmark for evaluating the accuracy of all fingerprint-matching algorithms which were written. When it became necessary to use pristine fingerprints for the lineups, we made use of a publicly available set of fingerprints maintained by the National Institute of Standards and Technology (NIST) called Special Database 302A, which features just over 1,900 labeled fingerprints. This was the database used to generate fingerprint lineups. A few other data sets with pristine prints and distorted prints were experimented with, but it was concluded early in the

testing process that these databases also had too few samples to produce lineups that were similar enough that forensic fingerprint examiners found them at least somewhat challenging.

Results

Developing a Fingerprint Matching Algorithm

The first step in building an algorithm that could generate fingerprint lineups was to create an algorithm that could accurately and efficiently differentiate between matching and non-matching fingerprints. Initially, effort was put into exploring whether traditional machine learning methods (e.g., neural networks, SVMs, etc.) could be used for this purpose, but the results quickly showed that the accuracy was too low. As a result, we pivoted to using algorithms that abstract the details of each fingerprint into graphs and compare the similarity of the generated graphs to determine whether a given pair of fingerprints was a match or non-matching pair. The graph generation process involved creating individual vertices for each minutia (most notably, ridge endings and bifurcations) and then adding edges between each pair of vertices. It is important to note that the length and angle between each edge is also stored. Finally, to effectively assign a "similarity score" to a pair of graphs, the algorithm goes through these graphs and tries to match as many edges as possible. In other words, the algorithm identifies minutiae (bifurcations and terminations) in the fingerprint and then maps their location and arrangement. It is the extent to which the arrangement of the minutiae, as well as the type and number of minutiae, are similar that is used to compute a similarity score and distinguish fingerprints.

Implementation

Initially, we wrote a fingerprint matching algorithm that mimicked the approach a 2014 paper by Akinyokun, Alese, and Iwasokun. The approach in this paper proposes to first identify the core of the fingerprint and then constructs a graph by storing the relative distances and angles of each individual minutia. Unfortunately, the accuracy of this algorithm (called Fingerprint Matcher v.1) was insufficient when working with non-pristine fingerprints, which would be necessary in real cases. The algorithm performed poorly on fingerprints that had distortions or information loss – especially in situations where the core was affected by these imperfections – resulting in accuracy levels of around 65% when adjusting for different thresholds to account for the precision and recall of the algorithm.

After encountering this roadblock, we discovered an open-source fingerprint matching algorithm, called SourceAFIS that utilized a different technique for constructing these graphs. Rather than connecting each minutia to a finger's core point, SourceAFIS used a probabilistic method of trying to match corresponding edges that exist between any pair of minutiae. So, when looking at two matching fingerprints from the same source, any pair of corresponding minutiae should produce edges which have a similar proportional length as well as very similar angles between the edge and other edges in their graphs. Thus, when computing the level of similarity between any two fingerprints, the probability of finding a matching pair of edges should be significantly higher for matching prints compared to that of non-matching prints. Using this new open-source algorithm, a new algorithm was written and tested (called Fingerprint Matcher v.2) and the performance metrics are reported in Figure 1 and 2 below.

	True Positive	True Negative
Predicted Positive	170	0
Predicted Negative	615	785

Measure	Value	Derivations
Sensitivity	0.2166	TPR = TP / (TP + FN)
Specificity	1.0000	SPC = TN / (FP + TN)
Precision	1.0000	PPV = TP / (TP + FP)
Negative Predictive Value	0.5607	NPV = TN / (TN + FN)
False Positive Rate	0.0000	FPR = FP / (FP + TN)
False Discovery Rate	0.0000	FDR = FP / (FP + TP)
False Negative Rate	0.7834	FNR = FN / (FN + TP)
Accuracy	0.6083	ACC = (TP + TN) / (P + N)
F1 Score	0.3560	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.3485	TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

Figure 1. Performance data from the Fingerprint Matcher v.2 algorithm when drawing on fingerprints from a database of *non-pristine* fingerprints.

	True Positive	True Negative
Predicted Positive	440	0
Predicted Negative	8	448

Measure	Value	Derivations
Sensitivity	0.9821	TPR = TP / (TP + FN)
Specificity	1.0000	SPC = TN / (FP + TN)
Precision	1.0000	PPV = TP / (TP + FP)
Negative Predictive Value	0.9825	NPV = TN / (TN + FN)
False Positive Rate	0.0000	FPR = FP / (FP + TN)
False Discovery Rate	0.0000	FDR = FP / (FP + TP)
False Negative Rate	0.0179	FNR = FN / (FN + TP)
Accuracy	0.9911	ACC = (TP + TN) / (P + N)
F1 Score	0.9910	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.9823	TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

Figure 2. Performance data from the Fingerprint Matcher v.2 algorithm when drawing on fingerprints from a database of *pristine* fingerprints.

These results were produced by generating an equal number of matching and non-matching pairs of fingerprints and running the algorithm on each of them. However, when reviewing these performance metrics, it is quite clear that the fingerprint matching algorithm still struggled to operate on non-pristine fingerprint images, with an accuracy of around 60-65% (see row labelled "Accuracy" in Figure 1). In contrast, when given pristine fingerprints, the algorithm's accuracy was close to 100% (see row labelled "Accuracy" in Figure 2; 99.11%).

At this point, we spent more time diving into why a large fraction (~27%) of matching fingerprints were being classified as non-matching. After an analysis of the prints that were being misclassified, we realized that many of the misclassified pairs of images had different scaling factors being used (i.e. the prints had the same graph but the edges were stretched by a constant factor), and this was not something that had been accounted for in the open-source code. To remedy this issue, we added a few pre-processing scripts and modified the similarity scoring metrics to overcome this issue (Fingerprint Matcher v.3). After running the same test as above, the performance metrics were obtained and examined again for non-pristine fingerprints (refer to Table 1). Now that the accuracy range was between 80 to 86% for non-pristine fingerprints (and 99%+ accuracy on pristine prints), we determined that this was sufficient to begin process of creating and developing the fingerprint lineup procedure.

Table 1.

Performance data from the Fingerprint Matcher v.3 algorithm when drawing on fingerprints from a database of non-pristine fingerprints.

Threshold	**TP**	**TN**	**P**	**N**	**(TP+TN)/(P+N)**	**TP/P**	**TN/N**
30	672	618	785	785	82.17%	85.61%	78.73%

35	630	725	785	785	86.31%	80.25%	92.36%
40	591	758	785	785	85.92%	75.29%	96.56%
45	567	774	785	785	85.41%	72.23%	98.60%
50	538	781	785	785	84.01%	68.54%	99.49%
55	521	783	785	785	83.06%	66.37%	99.75%
60	498	783	785	785	81.59%	63.44%	99.75%
65	471	784	785	785	79.94%	60.00%	99.87%

Fingerprint Lineup Generation

A significant challenge associated with real cases is that the crime scene fingerprints are imperfect and will commonly have distortions and information loss. For this reason, it was important for us to develop an algorithm that was able to deal with non-pristine fingerprints. Fingerprint lineups in the real world should help demonstrate whether a fingerprint examiner has been able to distinguish a suspect's fingerprint from other similar fingerprints by comparing them to a given a crime scene fingerprint. In the real world, it is almost always impossible to know whether a "same source" conclusion is actually accurate – a suspect is a suspect because the police are not sure if they are the culprit. However, it is possible to know something about the analyst's accuracy—a filler selection in the real world is always inaccurate.

In experiments, though, we know which fingerprints are from the same source and which are not, so we can systematically manipulate that variable and obtain full accuracy rates. Using the background knowledge from Dr. Quigley-McBride and her suggested readings, we began to

create a set of fingerprint lineups using the algorithm that we created. We started by creating a set of 5 fingerprint lineups that contained the following features:

- Five lineups with four to five prints in each lineup.
- Two to three lineups in which there is a single fingerprint that matches the crime scene fingerprint embedded among other fingerprints that share similar features to the matching fingerprint.
- Two to three lineups in which there is no fingerprint from the same source as the crime scene fingerprint, but the lineup contains four to five fingerprints that all contain features that are similar to those found in the crime print.

If the lineups are fair and well-constructed, the fingerprint examiner should be able to determine when one of the fingerprints matches the crime print if it is presented in the lineup, and should be able to correctly reject lineups in which no matching suspect fingerprint is present.

The fingerprint lineup generation process involves finding a subset of non-matching prints that are close enough in similarity to the suspect fingerprint such that a fingerprint examiner would have to carefully analyze them to distinguish them. With a working fingerprint matching algorithm that has a scoring metric which can evaluate how similar two fingerprints are (similarity score), we began the process of creating lineups.

In the eyewitness identification, researchers and police must ensure that the lineups are not too similar—that the eyewitness is not presented with a lineup of identical twins—because a lineup should not confuse eyewitnesses, merely test their memory (Fitzgerald, Oriet, & Price, 2015). We kept this in mind too as the logic also applies to fingerprint examiners. In fact, there exists one study where AFIS was used to generate candidate prints that could be used to create a lineup for experts (Kukucka, et al., 2020). The fingerprints generated by AFIS were so similar to

the crime print that the fingerprint designed to be the non-matching decoy actually looked more similar to the crime print than did the fingerprint that was the true match, evidenced by more false positive decisions in the non-matching condition than there were hits in the matching condition.

Using the NIST special database 302a, the algorithm was run to first rank each fingerprint in terms of similarity. Those with too high of a similarity would theoretically be removed as they may pose a challenge too hard for human examiners, and then the top 5 similar fingerprints would then be used in a set of lineups.

Pilot Testing the Fingerprint Lineups.

As the current study aimed to assess the feasibility and effectiveness of computer-generated fingerprint lineups *for expert fingerprint examiners*, the original goal was to generate approximately 10 fingerprint lineups and send a random subset of five of these lineups to expert fingerprint analysts who could evaluate their efficacy and difficulty they thought the lineups were and how much time they took to examine the lineups. However, before launching a full study, Dr. Quigley-McBride helped connect me with Dr. Heidi Eldridge, a researcher and forensic fingerprint analyst at RTI International.

We asked Dr. Eldridge to examine the fingerprint lineups included in Appendix A, generated using the similarity scores from Fingerprint Matcher v.3 applied to the SD 302A database. Specifically, we wanted to know: 1) how long these would take analysts to complete so we could plan our full study to be an appropriate length, 2) difficult enough for an expert analyst such that they would need to slow down and look closely at the minutiae in more than one of the fingerprints in each lineup, and 3) if there were any other observations she made that might help

us to improve our approach. We emailed her the lineups, some of which contained a fingerprint from the same source as the crime scene fingerprint and some of which did not, alongside these questions and instructions. She was told take as long as she needed to get us feedback.

Overall, Dr. Eldridge concluded the lineups were far too easy for an expert fingerprint examiner as each lineup took between 10 seconds to around 1 minute to complete. Although to a novice's eye the lineups appeared to include four to six very similar fingerprint patterns, an expert's capabilities require a much higher level of similarity before the task is difficult enough to be effective for reducing bias. In addition, experts are practiced at examining non-pristine fingerprints, so crime scene fingerprints that we thought suffered from significant information loss were not so alarming to Dr. Eldridge. Included in Table 2 are some excerpts from our email conversation with Dr. Eldridge's regarding her thought process when she was completing the lineup tasks.

As the comments in Table 2 show, the task was far too easy. She was able to eliminate fingerprints easily, without even looking closely at the minutiae in some cases. Thus, our fillers were not sufficiently similar to create a lineup that could test examiner's ability to discriminate between matching and non-matching fingerprints without the help of contextual information. We concluded that our main limitation was our database size—SD 302A only has approximately 2000 fingerprints. Given the amount of variability in minutiae type and arrangement, we would need a much larger database of fingerprints to achieve lineups with fillers similar enough to test expert examiners. A full discussion of the implications of this conclusion can be found in the discussion section. For this reason, though, we chose to end the project at this point temporarily and seek out a larger database. Larger databases are not publicly available and so it will be time

consuming to obtain one, but we are in the process of obtaining one from the FBI that is approximately 20,000 fingerprints.

Table 2.

Feedback on the Fingerprint Lineups Generated (Appendix A) from Dr. Eldridge.

Lineup	Dr. Eldridge's Comments
Figure A1 – Contains a print from the same source as the crime scene print.	*Ease of Task:* "The 'latent' is quite clear. It is very easy to identify the lower right impression."
	*Quality of Fillers: "*The upper right and lower left impressions are easy to almost instantly exclude." And "The upper left impression is more of a challenge because it has some creases and distortion in the areas I want to use to compare. Nevertheless, I can exclude it relatively easily."
	Time Taken to Complete Task: Less than 2 minutes
Figure A2 - Contains a print from the same source as the crime scene print.	*Ease of Task:* "The latent is very clear. The matching exemplar is very clear."
	Quality of Fillers: "The other 3 exemplars are reasonably clear, but more importantly, they are entirely different sub-classifications of pattern type, so they can be excluded more or less instantly as soon as you look at them."
	Time Taken to Complete Task: Less than 10 seconds
Figure A3 – Contains a print from the same source as the crime scene print.	*Ease of Task:* "The latent is very clear and has a highly diagnostic feature above the core that made it easy to search."
	Quality of Fillers, Quote 1: "The non-mated exemplars have some distortion, which would slow you down a bit in comparison if you hadn't already found it, but it was easy to find." *Quality of Fillers, Quote 2:* "As far as whether the other 5 made good distractors or not – the ridge count (number of ridges between the core and the delta) is too low on the top 3. It is better on the bottom 2, but the one on the right is a different sub-class. So, I'd say only the bottom middle one has even the potential to be a close non-match and it has a distinctive feature just to the left of the core that is clearly not in the latent, so it would also be eliminated pretty quickly. I'd say that this would be a pretty easy trial to exclude, even if I hadn't made the ID so quickly."
	Time Taken to Complete Task: Under 30 seconds
Figure A4 – Does not contain a print from the same source as the crime scene print.	*Ease of Task:* "Very easy"
	*Quality of Fillers: "*All 4 exemplars are left slant loops and the latent is a right slant loop, so they would be instantly excluded without any thought really.*"*
	Time Taken to Complete Task: 9 seconds
Figure A5 – Does not contain a print from the same source as the crime scene.	*Ease of Task: "*The top two were more difficult than the bottom two…for the top two, I had to at least look closely. But it was still pretty easy."
	*Quality of Fillers, Quote 1: "*The bottom left is completely the wrong pattern sub-type. The bottom right has a very distinctive core that is clearly and instantly recognizable as not the same as the latent." *Quality of Fillers, Quote 2:* [For the top two] "the one on the left had a too large ridge count, but I checked minutiae just to be sure. The one on the top right was the biggest challenge because the ridge count was about right and the area above the core was cut off, so I had to take a moment to find an area that was definitely present in both impressions that I could compare."
	Time taken to complete task: Less than 30 seconds.

Discussion

Although the fingerprint lineups generated from the publicly available database maintained by NIST (SD 302A database) were too easy to test fingerprint examiners, this series of experiments led to the creation of a new fingerprint-matching algorithm as well as a potential, real world application for that algorithm beyond simply computing similarity scores or ranking candidate fingerprints (as AFIS does). The utility of lineups has been well established, both in eyewitness identification research and field work as well as with fingerprint lineups (Kukucka et al., 2020; Quigley-McBride & Wells, 2018; Quigley-McBride, 2020). Lineups successfully reduce the number of forensically-relevant errors, or false positives, in memory-based eyewitness identification tasks and in perception-based fingerprint comparison tasks.

But creating a lineup is not easy. There is a wealth of research on how to create effective, fair lineups with faces (e.g., see the Quigley-McBride and Wells [2021] Chapter on Eyewitness Identification Research Methods), and the same care needs to be put into the creation of fingerprint lineups. This includes accounting for the perceptual expertise that fingerprint examiners have gained as a result of training and regular practice (Thompson & Tangen, 2014) and the fact that fingerprint analyses are *perceptual* tasks rather than *memory* tasks. That said, even a very large database is not sufficient. Kukucka and colleagues (2020), for instance, created lineups that were too similar using the AFIS software and associated databases. Lineups that are too similar are problematic because the goal is not to trick the participants—the goal is to improve their decisions and shift most errors onto fingerprints that will not result in a wrongful conviction (see Fitzgerald et al. [2015] for a discussion of face lineups that are too similar).

The databases associated with AFIS are extremely large and continuing to grow. Thus, this database size is probably too large to produce effective lineups if the lineup creator simply

selects the top candidates for the lineup. Moreover, research shows that using very large databases of photographs to create eyewitness identification procedures produces lineups that are too similar to be effective (Bergold & Heaton, 2018). In this study, they tested databases that were 5000 photos, 25000 photos, and 125000 photos in size. Although it seems intuitive to think the largest database would be best, the largest database produced the most errors because the task became too difficult—even participants with a very good memory of the crime or culprit would not perform well when presented with these lineups.

Similarly, studies have been run with the AFIS software too. These show that the size of the database (AFIS has ten-print cards from 6.964 million people) means that a large number of extremely close non-matches—fingerprints that are known not to match the crime scene fingerprint but share a very high number of features with the crime scene print—are returned by the software (Li et al., 2021). With a database on the scale that AFIS is, it is likely that experimental studies would need to be run to determine the optimal similarity score range for filler fingerprints so that the lineups are not *too* similar (as was the case in Kukucka et al., 2020), but *do* create lineups that are a challenge for expert examiners.

In an attempt to create lineups that were more difficult, we tried to distort parts of fingerprints within the lineup, however we realized that this approach would lead to experiments that would not resemble how fingerprint lineups would function in practice. After examining all of NIST's publicly available databases and the parameters of the fingerprints within them, we realized that there were no issues with the fingerprints or the algorithm. Rather, it was the database size (around 1,900 images) that was the main factor constraining the overall difficulty of the generated lineups. After sorting the fingerprints by their overall friction ridge patterns (e.g., double loops, tented arches, or whorls), we realized that the algorithm had far fewer

samples to select from when trying to find potential candidates for a lineup. Thus, we consider the creation of the algorithm to be a success. It was able to distinguish between non-pristine fingerprints with relatively high accuracy, which is very promising. In addition, although too easy for an expert examiner, the lineups that we generated did result in the desired type of decision-making process. That is, Dr. Eldridge had to examine and pay attention to details in each fingerprint to perform the task, considering both confirming and disconfirming evidence in each fingerprint to make a conclusion. We believe the next step for obtaining more concrete statistical results on the efficacy of fingerprint lineups is to obtain a much larger database. Due to the complicated and time-consuming nature of obtaining permission to access larger forensic databases, we have not yet been able to run another trial with a new database of fingerprints.

Overall, this project has simultaneously demonstrated both the feasibility and difficulty of using fingerprint lineups (and potentially lineups with other kinds of evidence) in forensic science. Through the collaboration of Dr. Quigley-McBride, Dr. Banks, Dr. Eldridge, and myself, we were able to leverage fingerprint-matching algorithms to create a software program that can efficiently generate fingerprint lineups that might help to reduce bias in forensic analyses with further testing. Although we have not been able to create lineups that are similar enough yet and test them with real analysts, we believe that this would be an effective tool with a larger database and some studies to establish a similarity score range that is appropriate for forensic experts analyzing fingerprint lineups. In addition, the interdisciplinary nature of this project shows that new ground can be covered—both creatively, practically, and theoretically—when scholars with different backgrounds come together. This project would not have been possible without the knowledge from experts in the fields of computer science, statistics, psychology, and forensic science.

References

Akinyokun, O. C., Alese, B. K., & Iwasokun, G. B. (2014). Fingerprint Matching Using Spatial Characteristics. *Proceedings of the World Congress on Engineering, I.*

Cole, S.A. (2005). More than Zero: Accounting for Error in Latent Fingerprint Identification. *Journal of Criminal Law and Criminology, 95*(3). Available at SSRN: https://ssrn.com/abstract=1025772

Cole, S. A., & Dioso-Villa, R. (2009). Investigating the "CSI Effect" Effect: Media and Litigation Crisis in Criminal Law. *Stanford Law Review*, *61*(6), 1335–1373. http://www.jstor.org/stable/40379713

Dror, I.E., & Mnookin, J.L. (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk*, 9(1), p. 47–67. https://doi.org/10.1093/lpr/mgp031

Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., & Walajtys, J. (2012). The impact of human-technology cooperation and distributed cognition in forensic science: biasing effects of AFIS contextual information on human experts. *Journal of Forensic Sciences*, *57*(2), 343–352. https://doi.org/10.1111/j.1556-4029.2011.02013.x

Dror, I.E., & Kukucka, J. (2021). Linear Sequential Unmasking–Expanded (LSU-E): a general approach for improving decision making as well as minimizing noise and bias. *Forensic Science International: Synergy, 3*, e100161. https://doi.org/10.1016/j.fsisyn.2021.100161

Edmond, G. (2022). Latent justice? A review of adversarial challenges to fingerprint evidence. *Science & Justice*, *62*(1), 21-29. https://doi.org/10.1016/j.scijus.2021.10.006

Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect filler similarity in eyewitness lineups: A literature review and a novel methodology. *Law and Human Behavior*, *39*(1), 62–74. https://doi.org/10.1037/lhb0000095

Institute of Medicine (US). Linda T. Kohn, *et al.* (Eds.), Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System, National Academies Press (US) (2000). https://doi.org/10.17226/9728

Kukucka, J., Dror, I.E., Yu, M., Hall, L., & Morgan, R.M. (2020). The impact of evidence lineups on fingerprint expert decisions. *Applied Cognitive Psychology, 34*(5), 1143-1153. https://doi.org/10.1002/acp.3703

NIST Special Database 302. NIST. (2021, November 2). Retrieved April 14, 2022, from https://www.nist.gov/itl/iad/image-group/nist-special-database-302

National Research Council (2009). Strengthening Forensic Science in the United States: A Path Forward. Available at: https://www.ojp.gov/pdffiles1/nij/grants/228091.pdf

Quigley-McBride, A., Dror, I.E., Roy, T., Garrett, B.L., & Kukucka, J. (2022). A practical tool for information management in forensic decisions: Using Linear Sequential Unmasking-Expanded (LSU-E) in casework. *Forensic Science International: Synergy, 4,* 100216. https://doi.org/10.1016/j.fsisyn.2022.100216

Spellman, B. A., Eldridge, H., & Bieber, P. (2021). Challenges to reasoning in forensic science decisions. *Forensic Science International: Synergy*, 100200. https://doi.org/10.1016/j.fsisyn.2021.100200

Thompson M.B., & Tangen J.M. (2014) The Nature of Expertise in Fingerprint Matching: Experts Can Do a Lot with a Little. *PLoS ONE, 9*(12): e114759. https://doi.org/10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior*, *38*(1), 84–93. https://doi.org/10.1037/lhb0000051

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. *PloS ONE, 7*(3), e32800. https://doi.org/10.1371/journal.pone.0032800

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7733-7738. https://doi.org/10.1073/pnas.1018707108

Ulery, B. T., Hicklin, R. A., Roberts, M. A., & Buscaglia, J. (2014). Measuring what latent fingerprint examiners consider sufficient information for individualization determinations. *PloS ONE, 9*(11), e110179. https://doi.org/10.1371/journal.pone.0110179

Vazan, Robert (2022). SourceAFIS [Computer software]. Retrieved from https://sourceafis.machinezoo.com/

APPENDIX A. Fingerprint Lineups Generated Using Fingerprint Matcher v3.

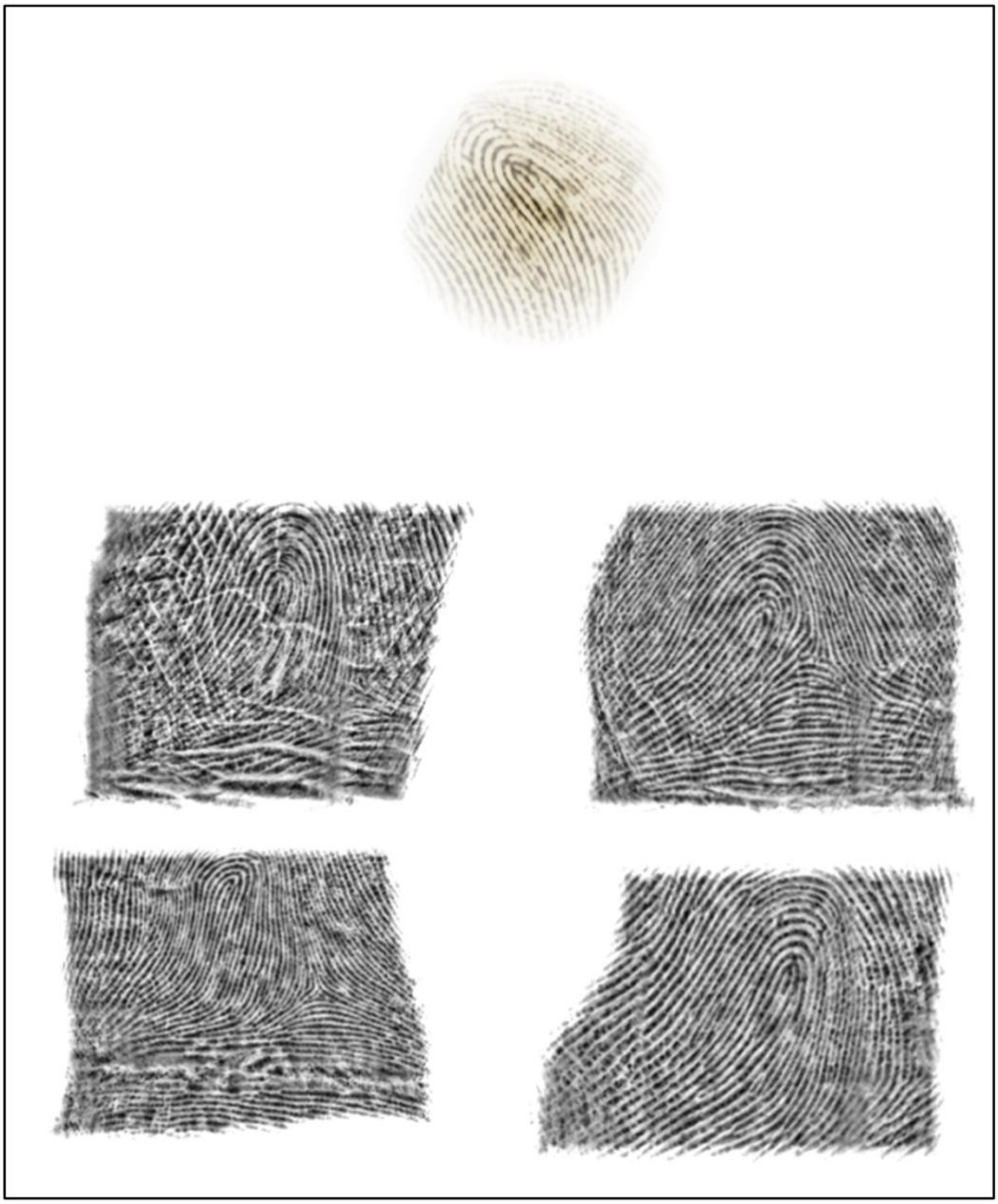

Figure A1. A fingerprint lineup generated using Fingerprint Matcher v.3 that *does* contain a fingerprint from the same source as the crime print (top middle).

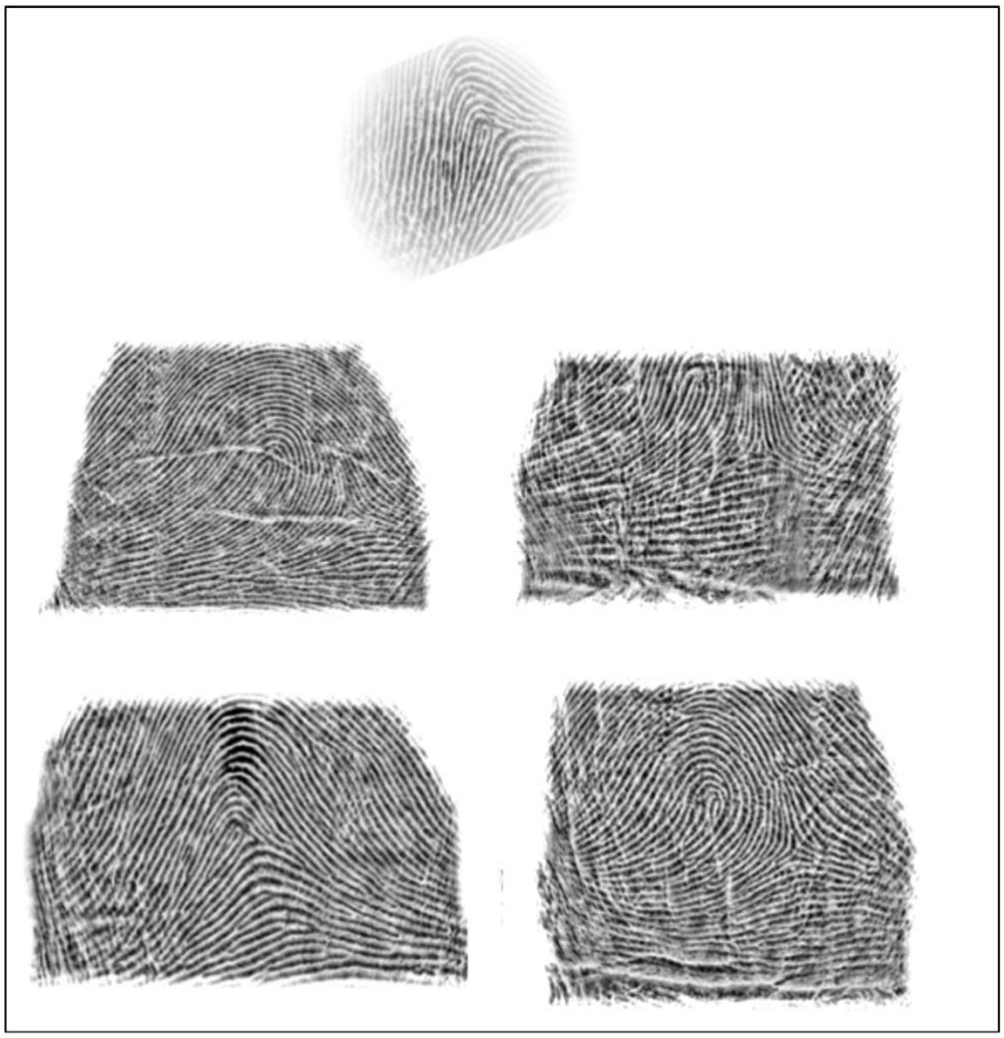

Figure A2. A fingerprint lineup generated using Fingerprint Matcher v.3 that *does* contain a fingerprint from the same source as the crime print (top middle).

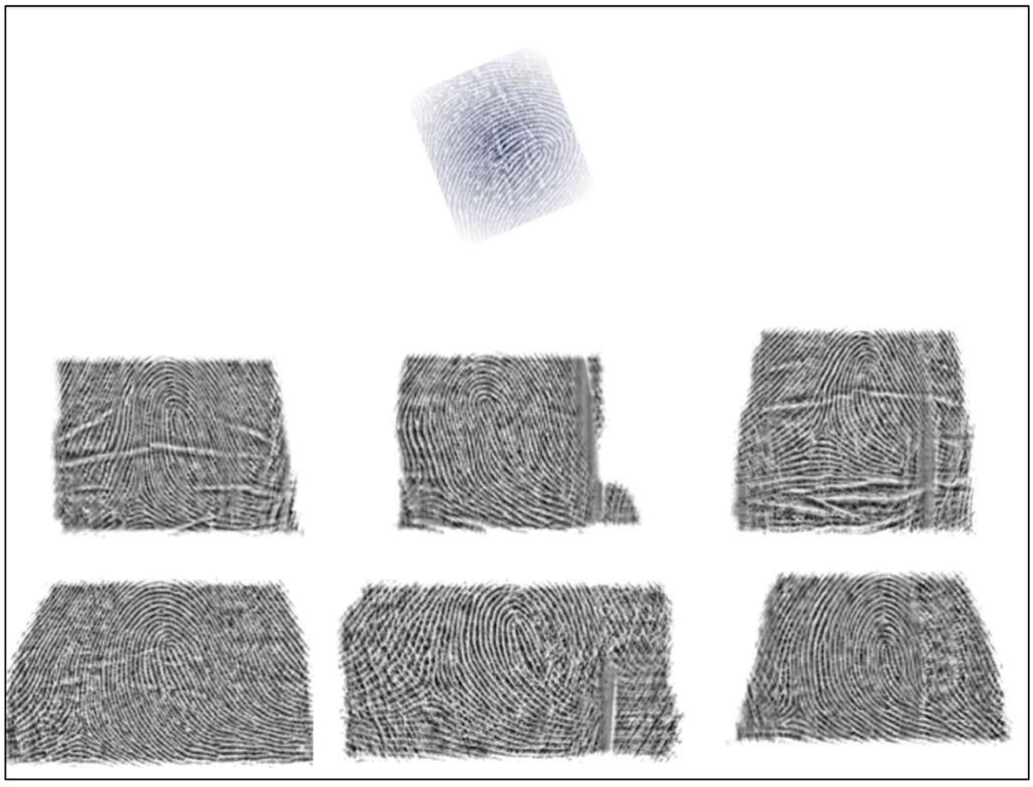

Figure A3. A fingerprint lineup generated using Fingerprint Matcher v.3 that *does* contain a fingerprint from the same source as the crime print (top middle).

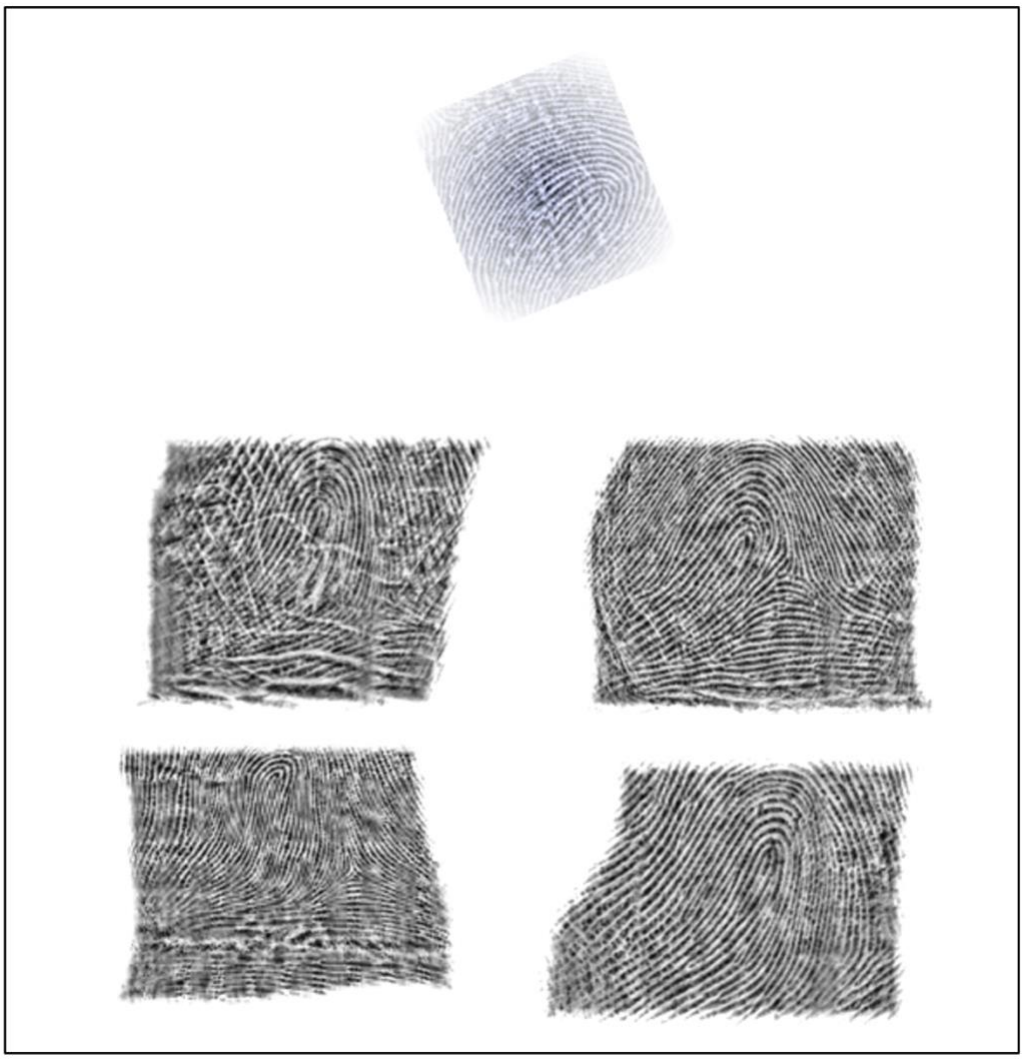

Figure A4. A fingerprint lineup generated using Fingerprint Matcher v.3 that *does* ***not*** contain a fingerprint from the same source as the crime print (top middle).

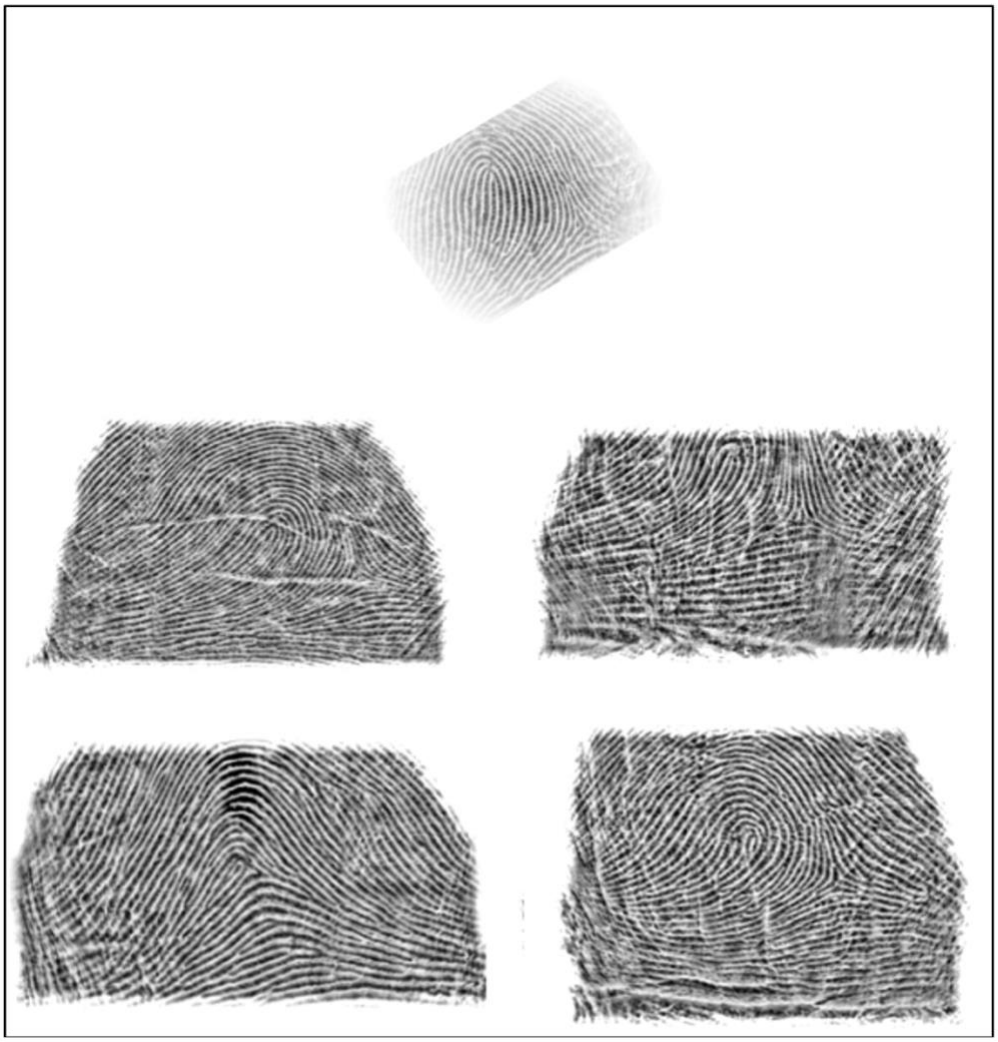

Figure A5. A fingerprint lineup generated using Fingerprint Matcher v.3 that *does* ***not*** contain a fingerprint from the same source as the crime print (top middle).