Should Fingerprint Examiners Make More Erroneous Identifications?

Willa Mannering, Macgregor Vogelsang
Faculty Mentor: Dr. Tom Busey, Department of Psychological and Brain Sciences *Indiana University*

ABSTRACT

This study was conducted as a response to the concerns about the consequences of latent fingerprint examinations. The goal is to determine if society's moral values align with the current bias towards erroneous exclusion decisions over erroneous identification decisions found in latent print examinations. Subjects of this experiment were asked to manipulate a web-based visualization that reflects the tradeoffs between putting guilty people in jail and keeping innocent people out of jail. The results of the experiment were analyzed to determine the similarities and differences between the opinions of fingerprint examiners and the opinions of students and members of the general public. In practice, examiners adopt more conservative decision criteria, because they could lose their job if they put an innocent person in jail. According to the results of this study, examiners seem to have a much more liberal exclusion criterion than they actually do in casework, and the public seems willing to tolerate a higher amount of erroneous identifications in exchange for a lower erroneous exclusion rate based on their average criteria placement in the visualization. The results of this study will help examiners align their responses to those of society, and help all citizens understand the tradeoffs that can occur with shifting decision criteria. If the results of the study indicate the need to shift the decision criteria to put more criminals in jail, additional safeguards may be necessary to guard against innocent people going to jail. Thus this dataset represents a rich framework for measuring, interpreting, and responding to the values and beliefs of what constitutes a just and moral society.

KEYWORDS: fingerprint decisions, taboo tradeoff, signal detection theory, society, criminal justice

INTRODUCTION

This study was conducted as a response to the concerns about the accuracy of latent fingerprint examinations. During normal casework, fingerprint examiners compare latent fingerprints that come from a crime scene and typically have a fair amount of noise, and exemplar prints which are taken in a controlled environment and are typically higher quality than the latent prints. There are two types of fingerprint pairs an examiner can compare: mated fingerprints and non-mated fingerprints. Mated fingerprints are pairs of latent and exemplar fingerprints that actually came from the same finger, and non-mated fingerprints are pairs of latent and exemplar fingerprints that actually came from different fingers. The status of a pair of fingerprints is rarely known for certain outside of an experimental context. Instead, the examiner must make a conclusion that represents his or her opinion of the status of the prints. To do this, an examiner conducts an analysis and comparison of the two prints to make one of three decisions about a pair of fingerprints: identification, exclusion, or inconclusive. An identification decision means that the examiner believes there is enough perceived detail in agreement between two fingerprints to say the fingerprints came from the same finger. An exclusion decision means that the examiner believes there is either not enough detail in agreement or that there are sufficient details in disagreement between the two fingerprints to say they did not come from the same finger. An inconclusive decision means the examiner believes there is not sufficient detail in agreement or disagreement to make an identification or exclusion decision.

In 2011, Ulery, Hicklin, Buscaglia, and Roberts measured the accuracy and reliability of latent fingerprint examiner's decisions in a study where 169 latent print examiners each compared around 100 pairs of latent and exemplar fingerprints from a pool of 744 pairs. Five examiners made erroneous identification decisions for an overall erroneous identification rate of 0.1%. Eighty-five percent of examiners made at least one erroneous exclusion decision for an overall erroneous exclusion rate of 7.5%. Further, 31.1% of the total mated fingerprints were classified as inconclusive and 11.1% of the non-mated fingerprints were classified as inconclusive (Ulery, Hicklin, Buscaglia, & Roberts, 2011).

This study brought to attention the current error bias in latent fingerprint examinations where examiners are more likely to make an erroneous exclusion decision—i.e. concluding that the fingerprints do not match when in reality they do—over an erroneous identification decision—i.e. concluding that the fingerprints do match when in reality they do not. This bias may be present for several reasons. It is possible that society has placed more importance on making sure innocent people are not put in jail, which would pressure the examiners into making more erroneous exclusion decisions over erroneous identification decisions. There is also the possibility that examiners fall back on inconclusive decisions in order to prevent making career-ending errors. Of course, if examiners say inconclusive all the time then no crimes will be solved. This leads to a set of tradeoffs that occur when examiners must decide where to place their decision criteria, which then determines how much evidence is required before making an exclusion or identification conclusion. However, it is up to the individual examiner where to place their decision criteria, which could be a function of several different and possibly competing factors. For example, in addition to considering the amount of perceived detail in agreement between the two prints, the examiner might consider the rarity of this information, the likelihood of the detective bringing a mated print, the costs to society of various errors, and the personal consequences to the examiner of these errors. While the error bias found by Ulery et al. reflects how examiners translate their own moral values and personal tradeoffs into the decision criteria, it may not accurately reflect the values that the whole of society holds. This is compounded by the fact that the consequences of changing the decision criteria will typically have both positive and negative outcomes. In the case of making an identification decision, if an examiner requires less detail in agreement before making an identification, they could potentially contribute information that will increase the number of criminals in jail, but it would also increase the number of innocent people in jail. Moving the decision criterion in the opposite direction would help keep more innocent people out of jail, but could also let more criminals free. Because this tradeoff involves negative consequences regardless of the direction, it is known as a Taboo Tradeoff (Fiske & Tetlock, 1997).

A taboo tradeoff is defined by Alan Fiske and Philip Tetlock as a, "mental comparison or social transaction that violates deeply-held normative intuitions about the integrity, even sanctity, of certain forms of relationships and of the moral-political values that derive from those relationships" (Fiske & Tetlock, 1997, p. 256). Relational theory proposes four different models that support social relationships and decisions: communal sharing, authority ranking, equality matching, and market pricing. Each model has implementation rules decided by different cultures that provide guidance for how to compare actions, values, objects, and relationships within those models. However, there is no central model that determines how to make choices between the four relational models. This means that when it is necessary to weigh alternatives and choose between one of the four, there is no clear-cut way to make the necessary tradeoffs.

Fiske and Tetlock propose that a tradeoff is considered taboo when the entities being compared do not belong to the same relational models. For example, while the market-pricing model provides a way to think of socially meaningful relationships such as prices, rent, or wages, the communal sharing model provides a framework to think about relationships that are considered shared, such as shared goods or romantic relationships. Now imagine that someone asks you to assign a market pricing relation to an entity that is normally considered something that belongs to the communal sharing model. The task could be something like, "How much money would you pay to breathe x-amount of air for a week?" or "How much money is your marriage worth?" Both of these questions seem very strange or even offensive. This is the idea of a taboo tradeoff. We believe that asking people to explicitly state their values when it comes to incarceration and exoneration is a taboo tradeoff and might cause discomfort among participants, because the prospect of putting an innocent person in jail violates the principles of freedom that form the basis of our society, yet sometimes it may be necessary to ensure an overall functioning society that places limits on crime. However, there is no easy solution, because it depends on the importance that one places on avoiding unjust incarceration, as well as maintaining justice for those affected by crime. Thus, an individual's solution to the taboo tradeoff reflects his or her own attempt to resolve the tradeoff in a way that optimizes the outcomes that fit his or her personal values.

The goal of the present experiment is to measure the values of the different outcomes of latent print examinations when subjects are given a graphical representation of this taboo tradeoff. We might find that the general public is less concerned with innocent people being put in jail and may not tolerate the large percentage of inconclusive decisions found by Ulery et al. To do this, we created a web-based visualization that was based on the findings in the Ulery et al. study that subjects were told to manipulate depending on their own personal values. To generate this visualization, we analyzed the Ulery et al. data using a model known as signal detection theory (Macmillan & Creelman, 2004) as described next, which allows us to quantify the exact nature of the taboo tradeoff.

Modeling the Taboo Tradeoff with Signal Detection Theory

To understand the importance of modeling the tradeoffs that can occur in forensic decision making, consider the following example from the related field of TSA baggage screening. As bags are scanned in the x-ray machine, the operator is compiling evidence that the bag contains a suspicious object. If so, the operator will ask to hand-search the bag, which causes a slow-down of the line and takes more personnel. However, if the suspicious object is allowed to pass, it could be used to hijack a plane. In addition, there are various factors that affect the decision to pull a bag for additional screening, such as the possibility of an imminent threat. In this case, the operator would adjust their criteria so that bags with even remotely suspicious items would get pulled for additional screening. This would result in many more false alarms and angry passengers, but it might be justified by the circumstances. All of these factors illustrate that the operator's choice of decision criteria (in this case how suspicious an item looks before pulling the bag for additional screening) affects both the probability of stopping a terrorist attack as well as the number of bags that are screened that turn out to be fine.

The costs of each outcome in the above example are fairly clear—we can calculate the additional cost of time and personnel of a false alarm, and we can estimate the cost of a terrorist attack. Note, however, that there is no way for the operator to simultaneously reduce the false alarm rate and reduce the chance of a terrorist attack. This is only possible through some additional scanning technology that specifically identifies suspicious objects either through some new form of x-ray or digital image processing.

When a similar analysis is applied to fingerprint examinations, we have additional constraints. First, the cost of a false alarm is much larger: an innocent person might go to jail, the real suspect may commit more crimes, and the examiner may be disciplined or fired if the error is discovered. Second, although a missed identification is not as bad as a terrorist attack, it still may lead to a guilty person going free and committing more crimes. The same tradeoff still exists as with the TSA screener—the examiner may want to adjust their criteria to make more identifications, which will put more criminals and more innocent people in jail. However, this is not necessarily a 1:1 ratio. Depending on the exact nature of the tradeoff, we might find that for every one innocent person in jail, we put an additional 50 criminals in jail. We as a society might feel that this is a justifiable tradeoff, because the only way to never have an innocent person in jail is to never put anyone in jail. The problem is that the tradeoff is quite complex and the ratio will differ for different decision criteria. We used the data from Ulery et al (2011) to construct a model using signal detection theory that allows us to present this tradeoff to our subjects as

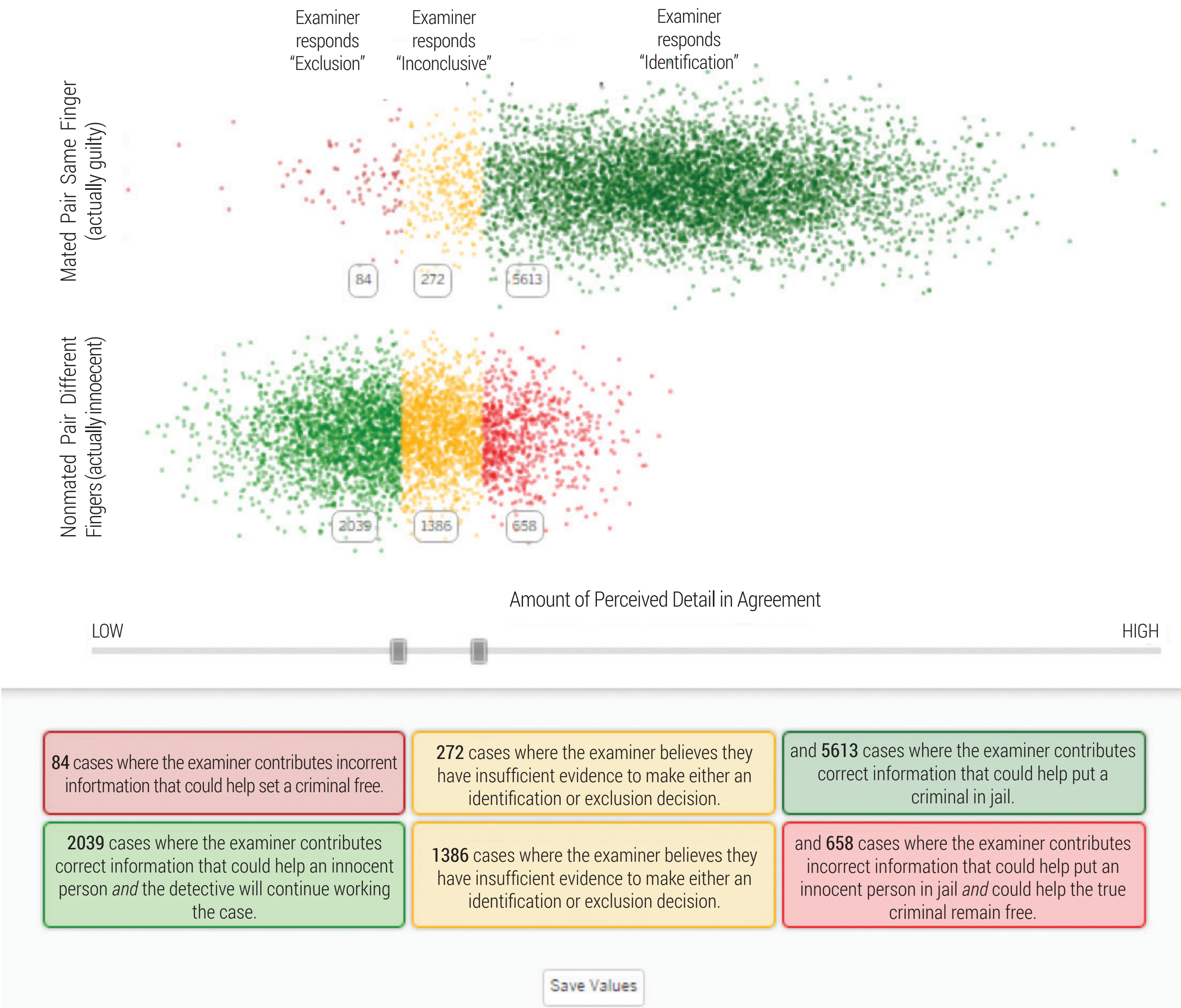


Figure 1: High nMated Web-based Model (Part 1)

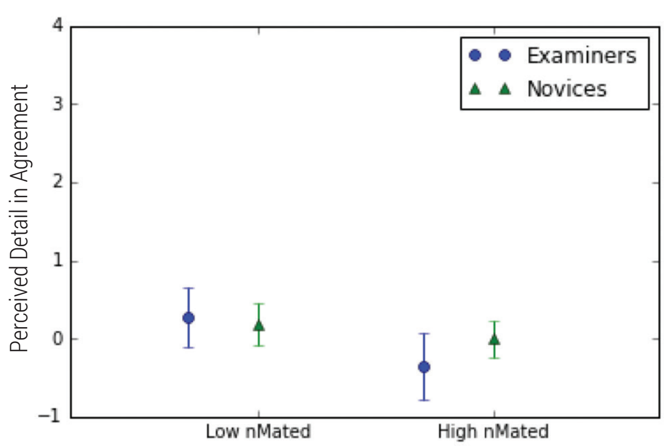
Figure 1: High nMated Web-based Model (Part 2)

a graphical representation.

In order to estimate the tradeoffs that occur for the different outcomes (i.e. innocent people in jail vs. guilty people in jail) as an examiner adopts different decision criteria, we constructed a mathematical representation of the underlying distributions of mated and non-mated fingerprint pairs. We assume that a fingerprint comparison results in an *amount of perceived detail in agreement* which creates an evidence axis that examiners use to make decisions. Higher values along this unidimensional evidence axis are more likely to produce an identification decision and lower values are more likely to produce an exclusion decision.

In reality, an image pair is either mated or non-mated. The goal of the examination is to determine which conclusion is best supported by the evidence. Signal detection theory allows us to estimate the mated fingerprint distribution and the non-mated fingerprint distribution using Gaussian curves, where the non-mated distribution is fixed with a mean of zero and a standard deviation of 1.0. This sets the scale of the evidence axis. We then adjust four free parameters: the location of the mated distribution, the standard deviation of the mated distribution, and the two decision criteria (one that separates exclusion from inconclusive responses and the other that separates inconclusive from identification responses). These parameters are fitted using minimization procedures (maximum likelihood estimation) to find parameter values such that the predicted proportions of different responses are as close as possible to the obtained proportions of responses. So, by using signal detection theory, we created a mathematical model that accurately predicts the proportions of examiner's decisions based on the actual

Figure 2



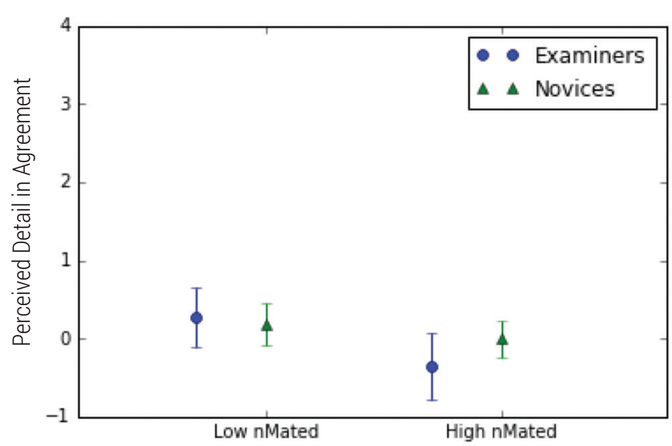


Figure 3: Exclusion Criterion Placement. All error bars in this figure and following figures are a 95% confidence interval based on the standard error of the mean multiplied by 1.96.

Figure 4: Identification Criterion Placement

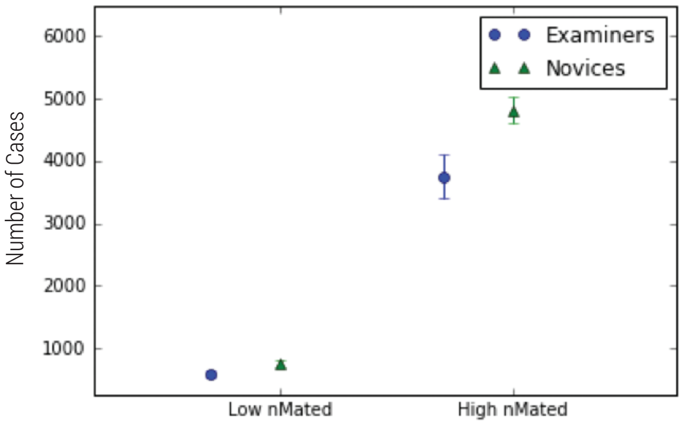


Figure 5: Average number of Criminals in Jail

proportions of decisions obtained from the Ulery et al. data. Table 1 illustrates the response proportions from the Ulery et al. data, and Table 2 illustrates the predicted proportions from signal detection theory. The best fitting parameters are: mated mean=3.42, mated standard deviation=1.54, exclusion criterion=1.21, and identification criterion=2.97. This means that the mated distribution is slightly more spread out than the non-mated distribution, and that examiners have adopted an extremely conservative decision criterion for the identification criterion, given that it is almost 3 standard deviations away from the center of the non-mated distribution.

The parameters that were freely estimated allow us to build a complete model of the tradeoffs that occur at various decision criteria. For example, if examiners were to adjust the identification criterion to the left, say adopting a value of 2.5 instead of 2.97 along the evidence axis, we know that both the number of criminals in jail would increase and the number of innocent people in jail would increase. However, they would not increase by the same amount, or even proportionately as shown by the predictions of our model. Instead, we can determine the amounts that each would increase by asking how much more area under the non-mated and mated distributions falls to the right of the new location of the identification criterion. Figure 1 illustrates this using a graphical interface, and shows the data for two different decision criteria locations along with the consequences for each criterion.

The ability to directly compute the consequences of different decision criteria allows us to quantify the taboo tradeoff in such a way that we can explore the values expressed by different participants. For example, if a subject is uncomfortable with a certain number of innocent people in jail, they can shift the decision criteria to higher values along the evidence axis. However, this will simultaneously affect the number of criminals in jail, which will drop by an amount determined by the mathematical model. The visualization provides both a graphical representation of the two distributions and immediate feedback for the consequences of different decision criterion choices.

METHODOLOGY

Participants

A total of 222 subjects participated in this experiment. The subjects were split into two types: examiners and novices. There were 147 novice subjects who were undergraduate students attending Indiana University and were recruited from the Psychological and Brain Sciences Course Credit Subject Pool. These subjects were tested in a lab to ensure that they understood the background and importance of this task before participating and received course credit as compensation. The 75 examiner subjects that participated in this study were recruited from numerous conferences. These subjects accessed the experiment through a web-link and performed

Table 2: Proportion predictions made by the signal detection theory model

Pair Type	Exclusion	Inconclusive	Individualization	Total Mates or non-mates
Mates (matches)	0.075	0.311	0.614	1
Non-mates (non-matches)	0.887	0.111	0.001	1

Pair Type	Exclusion	Inconclusive	Individualization	Total Mates or non-mates
Mates (matches)	0.075	0.311	0.614	1
Non-mates (non-matches)	0.887	0.111	0.001	1

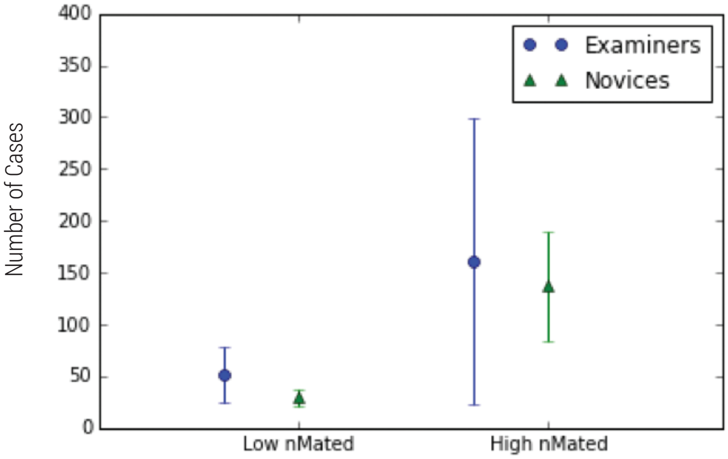


Figure 6: Average number of Criminals Set Free

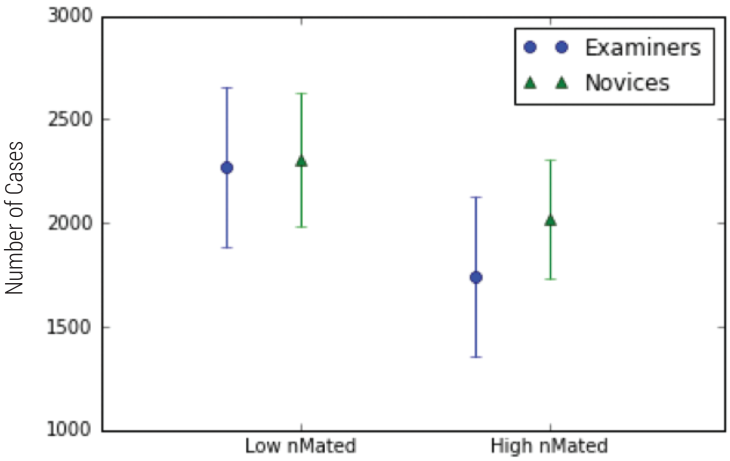


Figure 7: Average number of Innocents in Jail

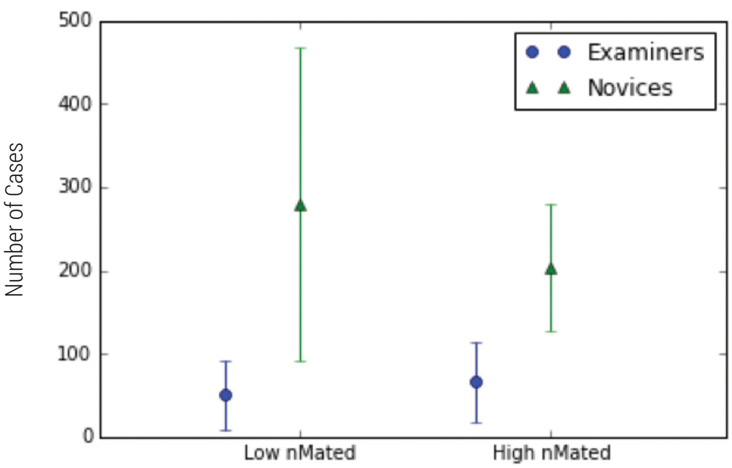


Figure 8: Average number of Innocents Set Free

the experiment on their own. All subjects watched a 6 minute instructional video before being directed to manipulate the web-based visualization (Figure 1). After saving their exclusion criterion and identification criterion placements, subjects were asked to fill out demographic data. Subjects that failed to fill out demographic data were discarded. The experiment took

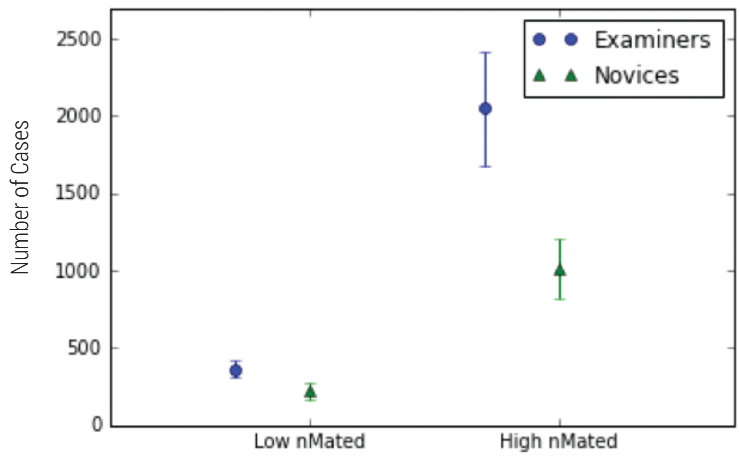


Figure 9: Average number of Inconclusive Decisions for Mated Pairs

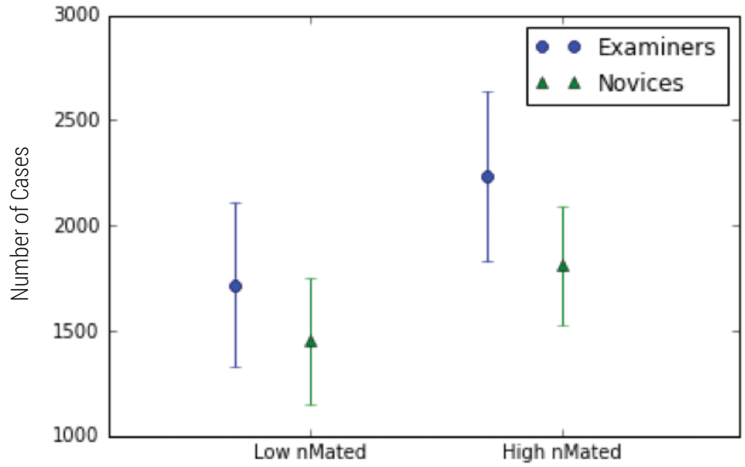


Figure 10: Average number of Inconclusive Decisions for Non-Mated Pairs

approximately 15 minutes to finish.

Procedure

Optimizations will include the improvements in the feed The subjects were randomly assigned one of two web-based visualizations to manipulate. The only difference between the two visualizations was the number of mated fingerprints (nMated) represented in the top cloud: either the High nMated condition where nMated=5969 or the Low nMated condition where nMated=1000. The High nMated condition reflects the distributions found in the Ulery et al. data while the Low nMated condition was created to determine how great of an effect the actual numbers of cases shown in the visualization affected criteria placement. Figure 2 shows the model with the Low nMated condition. The top cloud in the figure represents pairs of mated fingerprints and the bottom cloud represents pairs of non-mated fingerprints. The x-axis represents the amount of perceived detail in agreement between the two fingerprints. The y-axis is an arbitrary axis that serves to separate the mated and non-mated groups. The first slider represents the exclusion criterion; everything to the left of this slider is an exclusion decision and everything to the right is an inconclusive decision. The second slider represents the identification criterion; everything to the right of this slider is an identification decision and everything to the left is an inconclusive decision. The boxes below the clouds explain the outcomes of placing the sliders in a specific position and are color coordinated to indicate good outcomes (green), bad outcomes (red), and inconclusive outcomes (yellow). The colors of the clouds match the colors of the boxes in the same positions; for example, the green cloud in the upper right hand corner corresponds to the outcomes in the green box in the upper right hand corner. As the sliders are moved, the number of cases in each box changes proportionally. The subjects were instructed to carefully read all of the outcomes in each box and move the sliders to a position where they were comfortable with the outcomes in the boxes. After the subjects were comfortable with the position of the sliders, they clicked the "Save Values" button below the boxes and proceeded to fill out demographic data.

RESULTS

Table 3 shows the mean criteria placement and standard deviation for the exclusion criterion for both the High nMated and Low nMated conditions for both subject types. Running a two-tailed, unpaired t-test comparing the exclusion criterion placement of High nMated examiners and High nMated novices shows that there is not significant difference at $p<0.05$ between exclusion criterion placement of the examiners (mean=-0.36) and novices (mean=-0.01) in the High nMated condition ($t=-1.46$, $p=0.15$). The results of a two-tailed, unpaired t-test comparing the exclusion criterion placement of Low nMated examiners and Low nMated novices shows that there is not a significant difference at $p<0.05$ between exclusion criterion placement of the examiners (mean=0.27) and novices (mean=0.18) in the Low nMated condition (t0.35, $p=0.73$).

Table 4 shows the mean criteria placement and standard deviation for the identification criterion for both the High nMated and Low nMated conditions for both subject types. The novice subjects were broken down into two groups for this criterion: Random Novice and Fixed Novice. The Fixed Novice group was created because significant data collection occurred with a link for the web-based model that only provided the High nMated condition. The Random Novice group consists of the novices that were randomly assigned the High nMated condition. There is no significant difference between the mean identification criterion for each novice group ($t=0.0184$, $p=0.9854$). Because there was not a significant difference between the Fixed Novice group and the Random Novice group, the Fixed Novice data was not used in the creation of any other graphs or tables. The results of a two-tailed, unpaired t-test comparing the identification criterion placement of the High nMated examiners and the High nMated random novices shows that there is a significant difference at $p<0.01$ between the mean identification criterion placement of the examiners (mean=2.84) and the novices (mean=1.87) in the High nMated condition ($t=4.97$, $p=4.22e-6$). The results of a two-tailed, unpaired t-test comparing the identification criterion placement of the Low nMated examiners and the Low nMated novices shows that there is a significant difference at $p<0.01$ between the mean identification criterion placement of the examiners (mean=3.03) and the novices (mean=2.17, $t=3.92$, $p=1.91e-4$).

Figure 3 shows the average exclusion criterion placement for

Table 3: Exclusion Criterion Placement

Condition Type	Subject Type	Mean Criteria Placement	Standard Deviation	Sample Size
High nMated	Examiner	-0.36	1.29	35
	Novice	-0.01	0.77	41
Low nMated	Examiner	0.27	1.20	38
	Novice	0.18	0.84	40

Table 4: Identification Criterion Placement

Condition Type	Subject Type	Mean Criteria Placement	Standard Deviation	Sample Size
High nMated	Examiner	2.84	0.80	35
	Fixed Novice	1.87	1.26	68
	Random Novice	1.87	0.87	41
Low nMated	Examiner	3.03	0.82	38
	Novice	2.17	1.06	40

Figure 11: Average Examiner Placement of Exclusion and Identification Criteria according to the "Black Box" Study

Figure 12: Average Examiner Placement of Exclusion and Identification Criterion according to this study

both examiners and novices. The y-axis represents the perceived detail in agreement between two fingerprints and the x-axis represents the High nMated and Low nMated conditions. The values on the y-axis are equal to the standard deviation of the non-mated distribution, which is fixed at 1.0.

Figure 4 shows the average identification criterion placement for both examiners and novices. The y-axis represents the perceived detail in agreement between two fingerprints and the x-axis represents the High nMated and Low nMated conditions. The values on the y-axis are equal to the standard deviation of the non-mated distribution, which is fixed at 1.0.

Figures 5-10 illustrate the different outcomes of the cases that result from moving the decision criteria to different locations. Figure 5 shows the average amount of "Criminals in Jail" indicated by both examiners and novices for the High nMated and Low nMated conditions. Figure 6 shows the average amount of "Criminals Set Free" indicated by both examiners and novices for the High nMated and Low nMated conditions. Figure 7 shows the average amount of "Innocents in Jail" indicated by both examiners and novices for the High nMated and Low nMated conditions. Figure 8 shows the average amount of "Innocents Set Free" indicated by both examiners and novices for the High nMated and Low nMated conditions. Figure 9 shows the average amount of "Inconclusive Decisions for Mated Pairs" indicated by both examiners and novices for the High nMated and Low nMated conditions. Figure 10 shows the average amount of "Inconclusive Decisions for Non-Mated Pairs" indicated by both examiners and novices for the High nMated and Low nMated conditions.

Figure 11 shows the average placement of the identification and exclusion criteria of examiners found by Ulery et al. represented using the High nMated web-based visualization. Figure 12 comparatively shows the same distributions that were used in Figure 10 but instead uses the mean placement of the identification and exclusion criteria of the examiner subjects in this study. Figure 13 shows the same distribution used in both Figure 11 and Figure 10 but instead uses the average placement of the identification and exclusion criteria that the novice subjects indicated in this study.

DISCUSSION

The results of this study seem to support the idea that the general public shows less of a bias toward erroneous exclusion decisions than examiners and are less tolerant of a large amount of inconclusive decisions. According to the results shown in Figure 3, there is a significant difference between the novice and examiner placement of the identification criterion. As shown in Table 2, the mean placement of the identification criterion by the novice subjects is lower than the examiner subjects for both the High nMated and Low nMated conditions. This suggests that the novice subjects are more willing to accept an erroneous identification decision in exchange for fewer inconclusive decisions.

This finding can also be seen looking at the different placements of the exclusion and identification criteria in Figures 11, 12, and 13. Figure 11 represents the average criteria placement of the examiners found in the Ulery et al. study. In this figure, the identification criterion is located at 2.97 along the evidence axis and the exclusion criterion is located at 1.21. The placement of these criteria represents how examiners are actually classifying prints during casework. Figure 12 however represents the examiner's average criteria placement found in this study. It is interesting to note that while the identification criterion is relatively close to the placement in Figure 11, the exclusion criterion is placed significantly lower than in Figure 11. This indicates that while examiners place their criteria in one place during casework, ideally examiners would rather increase the amount of inconclusive decisions in order to decrease the amount of erroneous exclusion decisions. On the other hand, by looking at Figure 13 we can see that the novice subjects placed their exclusion criterion relatively close to the examiner's exclusion criterion in Figure 12. However, while the novices and examiners had fairly similar placement of the exclusion criterion, the novice subjects placed their identification criterion significantly lower than the examiners. This indicates that novices would rather allow more erroneous identification decisions in exchange for fewer inconclusive decisions and are comfortable with having more erroneous exclusion decisions than examiners currently allow in casework.

The difference in criterion placement between the two groups may indicate that the decisions of examiners in latent print examinations do not accurately reflect the values of society. Currently, there is a bias towards erroneous exclusion decisions in examinations. It is seen as a much worse mistake to accidentally incarcerate an innocent person than to accidentally exonerate criminals, in fact if an examiner does commit an erroneous identification decision there is a possibility of losing his or her job. This pressure may be what is causing this erroneous exclusion bias. However, according to the results of this study, perhaps society would actually be more comfortable increasing the amount of erroneous identification decisions in exchange for less erroneous exclusion decisions.

AUTHOR INFORMATION

All correspondence should be sent to the first author: wmanneri@indiana.edu.

REFERENCES

Bellehumeur, C., Li, L., Sun, Q., & Gu, P. (2004). Modeling of Fiske, A. P. & Tetlock, P. E. (1997). Taboo Tradeoffs: Reactions to Transactions That Transgress the Spheres of Justice. Political Psychology, 18(2), 255-297.

Macmillan, N. A. & Creelman, C. D. (2004). Detection Theory: A User's Guide: Taylor & Francis.

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.