Publics' Perceptions of

Machine Learning Based Risk Assessments

by

Anna Fine

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2021 by the
Graduate Supervisory Committee:

Nicholas Schweitzer
Jessica Salerno
Laura Smalarz

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

In the legal system, the prediction of a person's risk of committing a crime has mostly been based on expert judgment. However, newer techniques that employ machine learning (ML)—a type of artificial intelligence—are being implemented throughout the justice system. Yet, there is a lack of research on how the public perceives and uses machine learning risk assessments in legal settings. In two mock-trial vignette studies, the perception of ML-based risk assessments versus more traditional methods was assessed. Study 1 was a 2 (severity of crime: low, high) x 2 (risk assessment type: expert, machine learning) x 2 (risk outcome: low, high) between-subjects design. Participants expressed ethical concerns and discouraged the use of machine learning risk assessments in sentencing decisions, but punishment recommendations were not affected. Study 2 was a within-subjects design where participants were randomly assigned read through one of three crime scenarios (violent, white-collar, sex offense) and one of three risk assessment techniques (expert, checklist, machine learning). Consistent with Study 1, participants had ethical concerns and disagreed with the use of machine learning risk assessments in bail decisions, yet their own decisions and recommendations did not reflect these concerns. Overall, laypeople express skepticism toward these new methods, but do not appear to differentially rely on ML-based versus traditional risk assessments in their own judgments.

ACKNOWLEDGMENTS

I would like to thank my mentor and thesis committee chair, Dr. Nicholas Schweitzer, who has helped me grow into the researcher I am today. I am forever grateful for the endless support and encouragement throughout this process. I would also like to extend my gratitude to Dr. Jessica Salerno and Dr. Laura Smalarz for their mentorship and contributions to my thesis.

I am grateful for Arizona State University for giving me the resources I needed to complete my thesis. I would also like to thank the American Psychology and Law Society and the Arizona State University Graduate and Professional Student Association for their financial support of this project.

DEDICATION

I dedicate my thesis work to my friends who stayed by my side for countless hours until the very end, offering me words of encouragement and endless support. Thank you for helping me cross the finish line, I would not have made it without your love and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Incarceration rates in the United States are the highest in the world with 1.8 million people incarcerated at the end of 2020 (Kang-Brown, 2021). Not only does the United States have the most people behind bars, the number of inmates per 100,000 is also higher than any other country. According to the World Prison Brief, the United States has 655 inmates per 100,000 people of the population (Walmsley, 2018, p. 2). With incarceration rates being as high as they are, many researchers and legal scholars are pushing for the use of evidence-based practices to decrease incarceration rates.

To decrease the amount of people incarcerated, experts propose providing community based alternative options for offenders who do not pose a substantial risk to the public if released (Warren, 2007). This strategy can be used as a guideline to recognize offenders who are at a low risk to reoffend and minimize their sentence (Wolff, 2008; Warren, 2008; Monahan & Skeem, 2016; Bird et al., 2011). The justice system uses evidence-based practices such as expert risk assessments in many different areas, such as pre-trial, sentencing, probation, and parole hearings (Barabas et al., 2018; Corbett-Davies et al., 2017; Desmarais et al., 2020; Kehl, 2016; Scurich & Monahan, 2016; Scurich & Krauss, 2020; Skeem, Scurich, & Monahan, 2019). These forensic risk assessments typically involve evaluating and predicting the likelihood of antisocial behavior and future offending (Singh, 2012).

Many risk assessments used to predict recidivism fall into two categories: clinical and actuarial (Dawes et al., 1989). Clinical judgments are based on human decision-making processes and expert intuition. Actuarial methods use numeric ratings

that are weighted and combined into a total score which is referenced against a table which provides a risk outcome. Recently, actuarial methods have become more advanced and are using machine learning techniques (Barabas et al., 2018).

With the push for evidence-based practices into different trial areas, the general public's exposure to risk assessments is increasing. There is little research that explores public perceptions of the different types of risk assessments used in court. It is important to understand the public's perceptions on the implementation of machine learning risk assessments as future policies might rely on the public vote. For the sake of this paper, I am going to focus on arraignment and sentencing hearings. The current study will explore how the public perceives the use of machine learning risk assessments compared to other types of risk assessments.

Risk Assessments

Risk assessments examine factors (biological, psychological, and sociological) that either increase the likelihood of antisocial behavior (risk factors) or decrease the likelihood of antisocial behavior (protective factors). Further, the factors can be either *static* (will not change), *acutely dynamic* (malleable), or *stably dynamic* (malleable but most likely will not change) (Andrews and Bonta, 2010). Some common factors used in risk assessments include substance abuse, residential instability, and criminal thinking (Perry, 2013; Bonta et al., 1998; Andrews & Bonta, 1994; Gendreau et al., 1996; Baskin-Sommers et al., 2013).

Pretrial risk assessments are used in arraignment hearings to decrease the amount of people placed in jail when they are at a low risk for failing to appear in court and

committing a new crime before trial (Demarais et al., 2020). Pretrial risk assessments are used in many U.S. jurisdictions. One examination of 91 U.S. jurisdictions found that more than 66% used a pretrial risk assessment (Pretrial Justice Institute, 2019). Judges use these risk assessments to inform their decision-making in arraignment hearings, during which they determine if a defendant who pleads not guilty should be offered bail and decides how much (Harris, Gross, & Gumbs, 2019). Risk assessments are also used to inform sentencing. A review of risk assessments used in sentencing revealed that these assessments are used in multiple areas of sentencing (Monahan & Skeem, 2016). Judges use risk assessments on the front end of sentencing to help guide their decision for the appropriate punishment. They are also used on the back end of sentencing to shorten sentences and release inmates to community-based programs.

Clinicians' ability to accurately make predictions of risk has been called into question by many researchers (Monahan et al, 2001; Grove & Meehl, 1996; Harris, 2006). In 20 different areas (e.g., length of hospitalization and parole violations) humans have been shown to be inferior forecasters compared to actuarial tools (Meehl, 1954). Two meta-analyses looked at the accuracy of clinical and actuarial methods. A meta-analysis of 136 studies explored this in multiple domains (e.g., forensic assessment, admissions, medical diagnoses) and found that clinical judgments were 10 percent less accurate than actuarial methods (Grove et al., 2000). Another meta-analysis of 67 studies found that mental health practitioners' clinical judgment was 13 percent less accurate than actuarial assessments and 17 percent less accurate when predicting future violent or criminal behavior (Ægisdóttir et al., 2006).

Human judgments are motivated and influenced by cognitive biases (Gilovich, 1991; Kunda, 1994). When an expert makes a prediction of recidivism, these judgments are shaped by their experiences and biases. Actuarial assessments are generally more accurate because they decrease human bias in the prediction (Monahan et al, 2001; Grove & Meehl, 1996; Harris, 2006). However, some experts claim that their judgments are superior to others (Commons et al., 2012; Ehrlinger, Gilovich, & Ross, 2005) which is related to a phenomenon known as *bias blind spot* (Pronin, Lin, & Ross, 2002). A recent survey of forensic experts demonstrated this phenomenon as the forensic experts claimed that they perceive themselves as less biased than their colleagues (Neal & Brodsky, 2016). Experts are seemingly underestimating the impact of cognitive biases and motivated reasoning on their judgments.

Most experts rely on their previous experiences to guide their decision-making processes, which can make it difficult to pinpoint the exact reasoning or process that has produced their judgment. The expert is prompted to search their memory for relevant information, as intuition is nothing more than recognition (Simon, 1992). One technique to evaluate risk that is commonly used by clinicians is an unstructured interview, which has been shown to have low predictive validity (Devaul et al., 1987). Predictive validity is the ability for an assessment to accurately predict criminal behavior, which is one of the most important aspects for criminal risk assessments (Bonta, 2002). During an unstructured interview with a client, clinicians receive both diagnostic (i.e., useful to their judgment) and non-diagnostic (i.e., not useful to their judgment) information. This could be problematic if the clinician neglects the diagnostic information when there is the

non-diagnostic information present, which is known as the *dilution effect* (Dana et al., 2013).

The human brain has difficulty summarizing complex information and making judgments based on those summaries (Kahneman, 2011). To simplify complicated calculations, experts rely on heuristics and other short cuts (Nisbett & Ross, 1980). The human brain has a limited amount of cognitive processing it can handle and uses heuristics to reserve resources (Simon, 1983). This can be seen in parole rulings made by experienced judges. Danziger, Levav, and Avnaim-Pesso (2011) tracked judicial parole decisions for roughly a two-month period. The judges receive two meal breaks throughout their day which help replenish their mental resources. They found that the percentage of favorable rulings for the defendants drops moderately over time, then increases after their break. This shows how depletion of mental resources can cause even experienced experts to fall to heuristics. Heuristic processing can increase bias, which has been shown in both experts and novice decision makers (Kahneman, Slovic, & Tversky, 1982). Yet, despite the overwhelming evidence that human forecasters are flawed, people still assign more weight to advice from a human compared to advice from a computer (Önkal et al., 2009).

As predictive technology expands, the criminal justice system is using different subsets of artificial intelligence (AI) to better predict future criminal behavior. Machine learning is a form of data analysis that automatically creates models using immense amounts of data and can make predictions with high levels of uncertainty (Robert, 2014). These models are then used to make decisions. The algorithm learns and makes

adjustments to the model to give better predictions with the more data points fed into it. Despite the promise of this technology, scholars have warned that algorithmic risk assessments are far from perfect. Machine learning algorithm models are made up of data from past criminals. Larsen and colleagues (2016) analyzed COMPAS—an algorithmic risk assessment. They examined over 10,000 criminal defendants from Broward County, Florida and compared the predicted recidivism rates to the actual rate in which they reoffended over a two-year period. Black defendants were mistakenly marked as a high risk twice as much as white defendants (45% vs. 23%) and white defendants were classified as a low risk twice as much as black defendants. These trends are problematic and often erroneous. While algorithms are becoming a more popular method of crime prediction there is a strong push to increase predictive validity and decrease error rates.

Researchers are pursuing a new method of risk assessment, *neuroprediction* (Kiehl et al., 2018). Neuroprediction uses chronological age, demographics, social, and psychological characteristics, as well as a person's "brain age", which is an index of the volume and density of grey matter in the brain. A model was created using structural MRI of incarcerated males and then the model's ability to predict recidivism was tested with a longitudinal sample of male offenders. This research team found that they were able to more accurately predict recidivism using "brain age" compared to chronological age ($R^2 = 0.032$), as it considers individual differences in brain structure and activity over time, which influence decision making and risk taking (Kiehl et al., 2018). With an increase in predictive validity, algorithmic risk assessments may be more readily accepted by the public.

Perceptions of Artificial Intelligence

As the justice system increases the implementation of machine learning, understanding public perceptions and concerns are essential to smoothly integrate this technology. It is important to ensure that the public's expectations of artificial intelligence (AI) do not deviate from its capabilities. Laypeople are commonly unaware that forecasting models are probabilistic and not certain (Fast & Horvitz, 2016; Perry, 2013). Trust plays a large role in the disuse of an AI system, which occurs when a user neglects to use it for its purpose, which generally occurs after witnessing an error. Smoke detectors, for example, commonly set off false alarms and sometimes users will disable them out of frustration, even if the machine is not broken (Parasuraman & Riley, 1997).

To better understand the AI-human relationship, a basic understanding of trust will be discussed. Lee and See (2004), explain that trust consists of *ability* (i.e., trustee's performance quality), *integrity* (i.e., the overlap of values between the trustee and the truster, dependability), and *benevolence* (i.e., the extent the trustee's actions align with the goals of the truster). Lee and Moray (1992) describe three factors that influence trust for AI-human relationships: *performance* refers to the reliability and predictability of the AI, *process* refers to the extent to which the algorithm is appropriate for different situations and the ability to achieve the user's goals, and *purpose* is the degree to which the algorithm is being used for its programed intent.

There is a large difference in the way trust is formed in interpersonal relationships compared to an AI-human relationship (Lee & See, 2004). Interpersonal trust tends to start low and grows over time based on predictability, integrity, dependability, and

benevolence. However, a user's formation of trust in AI works in the opposite direction: Trust in AI starts high and is based on belief in the capabilities of the AI system. Once an error has been made, the user's trust decreases immensely (Lee & See, 2004).

Kramer and colleagues (2017) found that prior exposure to AI decision makers is a strong positive predictor of preference of an AI decision maker compared to a human decision maker. In other words, those who had prior experience with a computer making a decision for them, assuming it did not make an error, were more likely to prefer the computer compared to a human when given the choice. One explanation for this is the mere *exposure effect*, which suggests that continuous exposure of a stimulus can increase positive affect towards that stimulus (Zajonc, 2001). This might have a positive effect if judges are increasingly exposed to algorithmic risk assessments, considering that the accuracy is high and there is low chance for error.

Algorithm Aversion

Across a vast number of studies, statistical algorithms have been shown to outperform human decision-makers in making predictions under uncertainty (Dawes, 1979; Dawes et al., 1989; Grove et al., 2000; Silver, 2012; Dietvorst et al., 2015). This would suggest that the public should trust algorithms more than expert judgments. However, individuals are less confident in algorithms performing tasks and more likely to choose a human forecaster, even when algorithm accuracy is higher (Diab et al, 2011; Dietvorst et al., 2015; Eastwood et al., 2012). People are reluctant and even against using algorithms over human forecasters, which is known as *algorithm aversion* (Dietvorst et al., 2015).

It seems counterintuitive that people would choose less precise decision makers, and previous research has described several possible explanations for algorithm aversion. First, there is incentive to take advice from a human expert, especially when there is a perceived high risk or a negative consequence when making an error (Harvey & Fischer, 1997). This allows the advice seeker to share and diffuse some of the responsibility onto the human expert, instead of being solely responsible if an error were to occur (Bonacio & Dalal, 2006). Some have also argued that it is unethical to use algorithms to make consequential decisions regarding the lives of others, as it "turns individuals into numbers" (Dawes, 1979).

Some errors are perceived as more acceptable for humans to make but are perceived as intolerable for computers. For example, when a human driver makes a navigation error it is easily corrected and forgotten the next time that route is taken. However, if a GPS system makes a navigation mistake, it is much costlier, and the user is likely to lose confidence in the machine, making them hesitant to use it again (Dietvorst et al., 2015). These previous studies explain why people are averse to relying on algorithms. When exposed to algorithmic risk assessments, algorithm aversion may be exacerbated due to the consequential nature of using them in trial decisions. These studies suggest that people might be skeptical of machine learning for these reasons, despite believing that they are accurate.

Research Overview

In two studies, the current research examines laypeople's perceptions of the use of machine learning risk assessments compared to other types of risk assessments. The first

study was a 2 (severity of crime: low severity arson, high severity arson) x 2 (risk assessment technique: expert, machine learning) x 2 (risk assessment outcome: low, high) between-subjects design. Participants read a brief summary of a criminal court case involving an arson in which the defendant lit his ex-girlfriend's place of work on fire. The summary gave an overview of the case and statements from two witnesses. Participants were then given a description of the sentencing process, risk assessments, and read testimony from an expert who described the risk assessment. Finally, participants rated the defendant's perceived responsibility and risk assessment procedure, as well as gave punishment recommendations.

The second study used a within-subjects design in which participants were instructed to give their opinions on different bail procedures. Participants were randomly assigned to read three crime scenarios (violent, non-violent, sex offense) and three risk assessment procedures (machine learning, expert-based risk assessment, and automated checklist). First, participants were given one of the three crime scenarios paired with one of the three risk assessment procedures. Then, participants gave bail judgments and opinions on the procedure. These steps repeat until participants have read all three crime scenarios, risk assessments, and gave bail judgments and opinions on the procedure. Finally, participants rated how strongly they would recommend each risk assessment procedure if they were to be implemented in their city.

Research Questions and Hypotheses. In these studies, I wanted to examine people's perceptions of the use of machine learning risk assessments compared to other more traditional risk assessment methods. Research shows that people are hesitant to rely

on algorithms over experts and are concerned about their involvement in consequential decisions (Dietvorst et al., 2015). This has been tested in other domains (e.g., admissions decisions) however, at this time, there are few if any studies examining how people will react to them being used in the justice system. Therefore, these two studies examine people's opinions of the use of machine learning risk assessments in the sentencing phase (Study 1) and the bail phase (Study 2). For both Studies 1 and 2, I predicted that participants would generally distrust the use of machine learning risk assessments in both Sentencing (Study 1) and Arraignment (Study 2) and would have more negative perceptions towards the machine learning risk assessment compared to the expert.

Study 1 Method

Participants

Using Prolific Academic, I obtained a sample of 387 US residents. I removed 6 people for taking the survey too quickly (less than 200 seconds, which was the quickest 5%) and 20 people for not passing our manipulation checks. Therefore, I ended with a participant sample of 361 US residents (see table 1 for demographics).

Table 1

Participant Demographics

N	361
Mean age	33.03 (10.58)
Range	18-76
% Female	52.9%
% Hispanic / Latino / Central / South American	6.1%
% White / Caucasion	69.4%
% Black / African American	11.5%
% Middle East / North African	0.5%
% Asian / Pacific Islander	10.1%
% Other	2.5%
At least a bachelor's degree	49.4%
Moderate Political Views	15.4%
Liberal Political Views	64.7%
Conservative Political Views	19.9%

Materials

Instructions. Prior to reading the trial summary, participants receive instructions that the defendant has already been found guilty by jury and that they will be reading a summary of the trial, then be moving on to the sentencing/punishment phase.

Guilt Trial Summary. Participants read a summary of a crime where the defendant went to his ex-girlfriend's place of work, got into an argument, was asked to leave, and then stormed off the premises. In the low severity condition, he came back

after the business was closed and lit the building on fire, causing property damage. In the high severity condition, the defendant came back during business hours and lit the building on fire, causing bodily injuries to people in the building in addition to property damage. Participants then read prosecution and defense opening statements which summarized their arguments. For the prosecution, participants read a description of the defendant's criminal history. The defense described that the defendant had a difficult upbringing and also included witnesses that spoke to his character.

Sentencing Trial. The participants were told that they were going to read a brief description of the risk assessment conducted by neuroscientist Dr. Pavy, who used either his expert judgment or a machine learning algorithm to form an opinion of risk. In the machine learning condition, Dr. Pavy took brain scans of the defendant and entered them into a piece of specialized computer software. The brain scans were analyzed by the machine learning algorithm using a complex set of mathematical equations based on analyses of hundreds of criminal offenders. Dr. Pavy came to either a low or high-risk outcome, which came from his judgment (expert condition) or from the machine learning risk assessment (machine learning condition). In the expert condition, instead of using a computer program, Dr. Pavy used his expertise of evaluating hundreds of criminal offenders and his intuition to help make his decision of either low or high risk.

Attention and Manipulation Checks. To ensure the quality of the data, participants were asked about the crime that was committed (multiple choice format: murder, assault, armed robbery, arson) and how old the defendant was (multiple choice format: 55, 18, 38, 28). To ensure the risk assessment manipulation was successful,

participants were asked about the risk assessment method they saw (multiple choice: Dr. Pavy's expertise, machine learning algorithm, blood tests, Big Five personality test) and the risk outcome given to the defendant (multiple choice: low risk, medium risk, high risk, inconclusive).

Measures. After reading the trial summary, participants were asked to give ratings of responsibility, which were shown in a random order: the extent to which the defendant was in control of his actions (9-point Likert scale item "Not at all in control" to "Fully in control"), the likelihood that the defendant is going to re-offend (9-point Likert scale item "Will almost certainly NOT reoffend" to "Will almost certainly reoffend"), the future dangerousness of the individual (9-point Likert item from "Not at all dangerous" to "Extremely dangerous"), the severity of the crime (9-point Likert item from "Not at all severe" to "Extremely severe"). Then participants gave ratings of the risk assessment procedure used, which were shown in a random order: believability (9-point Likert item from "Certainly do not believe" to "Certainly believe"), accuracy (9-point Likert item from "Not at all accurate" to "Perfectly accurate"), the extent in which it matched the participant's belief (9-point Likert item from "Did not match at all" to "Completely matched"), ethicality (9-point Likert item from "Certainly NOT ethical" to "Certainly ethical"), and the extent to which that particular procedure should be used (9-point Likert item from "Certainly should NOT use" to "Certainly should use"). Further, they were asked for sentencing and punishment recommendations: the extent to which the defendant should be punished (9-point Likert item from "To the MINIMUM extent allowable" to "To the MAXIMUM extent allowable").

Procedure

Participants were randomly assigned to one of 8 conditions in a 2(crime severity: low v high) x 2(evaluation type: ML v expert) x 2(risk outcome: low v high) between-subjects design. Participants first read a summary of a crime that varied on whether the people were injured during the arson (crime severity manipulation). Then participants read a summary of the guilt trial along with opening statements. Next, during the sentencing phase, participants read the neuroscientific risk assessment. In all conditions, the expert conducted an fMRI of the defendant's brain and then examined the brain structure to predict future recidivism. Participants read one of two evaluation types for the risk assessment: Either the expert examined the fMRI images and made the decision using their previous subjective experience or the fMRI data was entered into a computer and the ML algorithm made the decision (evaluation type manipulation). Further, in the ML condition, an explanation of ML and how it works was given. Participants were then asked to provide punishment and sentencing recommendations, perceptions of the risk assessment including, ethicality and if they should be used, likelihood of defendant recidivism, dangerousness, controllability of the defendant's actions and demographic measures.

Study 1 Results

Frequency analyses were conducted to ensure that our conditions were evenly distributed (see table 2).

Table 2.

Participant distribution in each condition

Condition	N	Total %
Severity of Crime		
Low	181	50.1%
High	180	49.9%
Risk Assessment Technique		
Expert	174	48.2%
Machine Learning	187	51.8%
Risk Outcome		
Low	177	49.0%
High	184	51.8%

Note: All the conditions are evenly distributed.

Manipulation Checks. The severity manipulation was successful. The main effect for severity on the perceived severity of the crime was significant, with the perceived severity of the crime being higher in the high severity condition (M = 7.31, SD = 1.27) compared to the low severity condition (M = 6.87, SD = 1.36), F (1, 352) = 10.053, p = .002, ηp^2 = 0.028. The main effect for risk outcome on the defendant's perceived recidivism was significant, with the defendant being perceived as a higher risk of reoffending when he received a high-risk outcome (M = 7.05, SD = 1.29) compared to a low-risk outcome (M = 6.31, SD = 1.70), F (1, 353) = 21.921, p < .001, ηp^2 = 0.058.

Judgments of Defendant Responsibility

Controllability. A 2x2x2 univariate ANOVA examined the effect of the severity of the crime (low vs high), risk assessment type (ML vs expert) and risk outcome (low vs high) on participant perceptions of to what extent the defendant had control over their actions. The three-way interaction was not significant, F (1, 352) = 0.605, p = .437, ηp^2 = 0.002. However, there was a significant two-way interaction between the risk assessment technique and the risk outcome on the defendant's perceived controllability (see figure 1 below), F (1, 352) = 7.168, p = .01, ηp^2 = 0.020. Simple effects tests revealed that participants who read the expert-based risk assessment believed the defendant was in control of his actions more when the expert deemed the defendant as a low risk (M = 7.53, SD = 1.94) compared to a high risk (M = 6.65, SD = 2.35), F (1, 171) = 7.047, p = .009, η^2 = 0.040. However, participants' perceptions of the defendant's control over his actions were not affected by the machine learning's risk outcome, F (1, 185) = 1.332, p = .250, η^2 = 0.007. The interactions between severity and risk outcome, F (1, 352) = 2.055, p = .153, ηp^2 = 0.006 and severity and risk assessment technique, F (1, 352) = 0.911, p = .341, ηp^2 = 0.003 were not significant. Further, the main effects of risk outcome, F (1, 352) = 1.440, p = .231, ηp^2 = 0.004, risk assessment technique, F (1, 352) = 0.050, p = .824, $\eta p^2 < 0.001$, and severity, F (1, 352) = 0.100, p = .752, $\eta p^2 < 0.001$, were not significant.

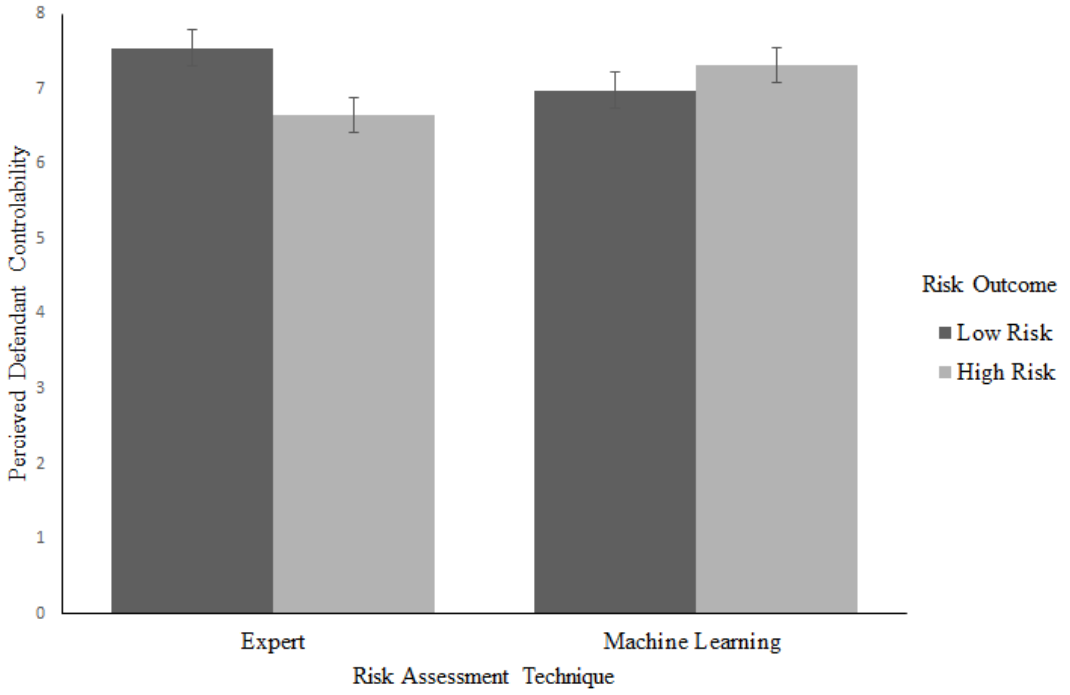

Figure 1. The effect of risk assessment technique and risk outcome on the defendant's perceived controllability.
Note. Higher levels of perceived defendant controllability mean that the participant believed that the defendant had high levels of control over his actions at the time of the crime.

Recidivism. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level outcome (low vs high) on the defendant's perceived recidivism. The three-way interaction was not significant, F (1, 353) = 0.011, p = .917, $\eta p^2 < 0.001$. Further, the two-way interactions between risk assessment technique and risk outcome, F (1, 353) = 1.358, p = .245, $\eta p^2 = 0.004$, severity and risk outcome, F (1, 353) = 0.011, p = .541, $\eta p^2 = 0.001$, and severity and risk technique, F (1, 353) = 0.043, p = .837, $\eta p^2 < 0.001$, were all not significant. The main effects of risk assessment technique, F (1, 352) = 0.077, p = .781, $\eta p^2 < 0.001$, and

severity, F (1, 352) < 0.001, p = .997, ηp^2 < 0.001, were not significant. The main effect for risk outcome is reported above in the manipulation checks section.

Dangerousness. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on the perceived dangerousness of the defendant. The three-way interaction was not significant, F (1, 352) = 0.185, p = .667, ηp^2 = 0.001. There is a significant two-way interaction between the severity of the crime and risk assessment type on the defendant's perceived dangerousness (see figure 2 below), F (1, 352) = 3.958, p < .05, ηp^2 = 0.011. Simple main effects analysis showed that when the crime severity was high, participants were affected by risk assessment type such that participants who read about the expert based risk assessment rated the defendant as more dangerous (M = 7.12, SD = 1.36) compared to participants in the machine learning condition (M = 6.59, SD = 1.53), F (1, 177) = 4.788, p < .05, η_p^2 = 0.027. However, when there were no differences in the defendant's perceived level of dangerousness when the severity of the crime was low, F (1, 177) = 0.412, p = .522, η^2 = 0.002. Further, the two-way interactions between risk assessment technique and risk outcome, F (1, 352) = 0.236, p = .627, ηp^2 = 0.001, and severity and risk outcome, F (1, 352) = 1.403, p = .237, ηp^2 = 0.004, were not significant. Unsurprisingly, the main effect for severity on the defendant's perceived future dangerousness was significant, with the defendant being perceived as more dangerous in the high severity condition (M = 6.86, SD = 1.47) compared to the low severity condition (M = 6.46, SD = 1.52), F (1, 352) = 7.466, p < .01, ηp^2 = 0.021. However, the main

effects of risk assessment technique, F (1, 352) = 1.151, p = .284, ηp^2 = 0.003, and risk outcome, F (1, 352) = 3.072, p = .081, ηp^2 < 0.01, were not significant.

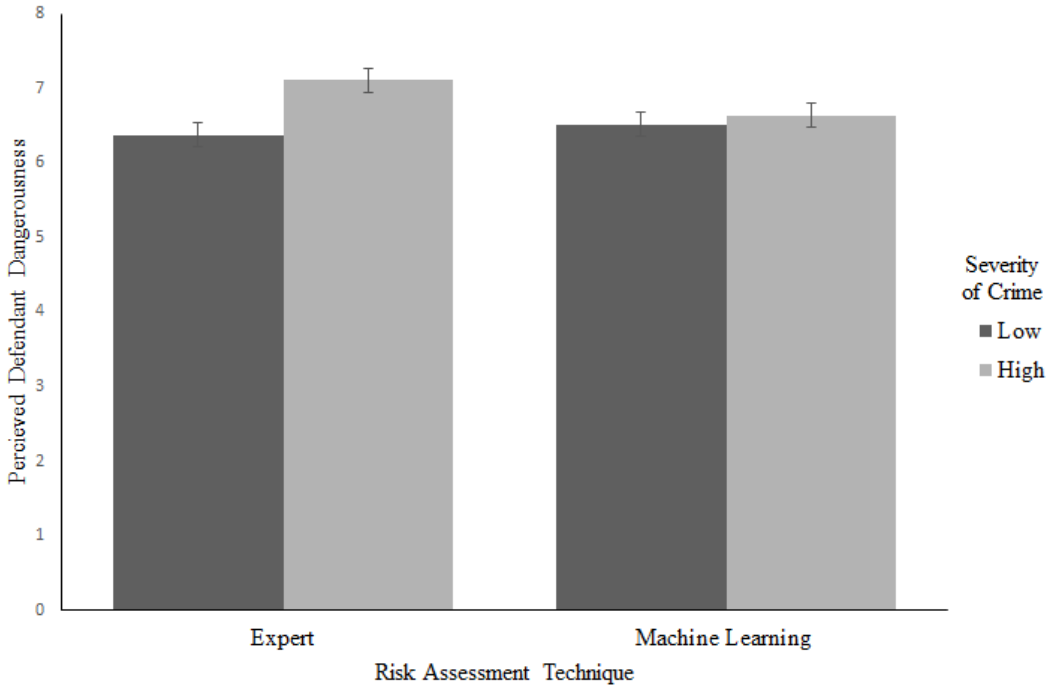

Figure 2. The effect of severity of crime and risk assessment technique on the defendant's perceived dangerousness.
Note. Higher levels of perceived defendant dangerousness mean that the participant believes that the defendant has higher levels of future dangerousness.

Ratings of Risk Assessment

Believability. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on the perceived believability of the risk assessment. The three-way interaction was not significant, F (1, 351) = 0.958, p = .328, ηp^2 = 0.003. The two-way interactions between risk assessment technique and risk outcome, F (1, 351) = 0.708, p = .401, ηp^2 = 0.002, severity and risk outcome, F (1, 351) = 0.049, p = .826, ηp^2 < 0.001, and severity and risk

technique, F (1, 351) = 0.006, p = .940, $\eta p^2 < 0.001$, were not significant. The main effect for risk outcome was significant such that, participants perceived the risk assessment as more believable when the defendant was labeled as a high risk (M = 5.54, SD = 2.08) than when he was labeled as a low risk (M = 3.64, SD = 2.07), F (1, 351) = 73.887, $p <$ .001, $\eta p^2 = 0.174$. However, the main effects of risk assessment technique, F (1, 351) = 0.243, p = .622, $\eta p^2 = 0.001$, and severity, F (1, 351) = 0.238, p = .626, $\eta p^2 = 0.001$, were not significant.

Accuracy. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on the perceived accuracy of the risk assessment. The three-way interaction was not significant, F (1, 343) = 0.467, p = .495, $\eta p^2 = 0.001$. The two-way interactions between risk assessment technique and risk outcome, F (1, 343) = 0.161, p = .688, $\eta p^2 < 0.001$, severity and risk outcome, F (1, 343) = 0.001, p = .975, $\eta p^2 < 0.001$, and severity and risk technique, F (1, 343) = 0.095, p = .758, $\eta p^2 < 0.001$, were not significant. The main effect for risk outcome was significant such that, participants perceived the risk assessment as more accurate when the defendant was labeled as a high risk (M = 5.37, SD = 1.92) than when he was labeled as a low risk (M = 3.69, SD = 1.85), F (1, 343) = 66.466, $p < .001$, $\eta p^2 = 0.162$. However, the main effects of risk assessment technique, F (1, 343) = 0.626, p = .429, $\eta p^2 = 0.002$, and severity, F (1, 343) < 0.001, p = 1.00, $\eta p^2 < 0.001$, were not significant.

Matched Belief. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on

whether the risk assessment outcome matched their belief. The three-way interaction was not significant, F (1, 351) = 0.321, p = .571, ηp^2 = 0.001. The two-way interactions between risk assessment technique and risk outcome, F (1, 351) = 1.962, p = .162, ηp^2 = 0.006, severity and risk outcome, F (1, 351) = 0.654, p = .419, ηp^2 = 0.002, and severity and risk technique, F (1, 351) = 0.400, p = .528, ηp^2 = 0.001, were not significant. The main effect for risk outcome was significant such that, participants felt the risk outcome matched their belief more when the risk outcome was high (M = 6.71, SD = 1.79) than when he was labeled as a low risk (M = 4.92, SD = 2.60), F (1, 351) = 351.044, p < .001, ηp^2 = 0.500. However, the main effects of risk assessment technique, F (1, 351) = 0.041, p = .840, ηp^2 < 0.001, and severity, F (1, 351) = 1.176, p = .279, ηp^2 = 0.003, ηp^2 = 0.003, were not significant.

Ethicality. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on perceived ethicality of the risk assessment. The three-way interaction was not significant, F (1, 349) = 2.139, p = .144, ηp^2 = 0.006. The two-way interactions between risk assessment technique and risk outcome, F (1, 349) = 3.090, p = .080, ηp^2 = 0.009, severity and risk outcome, F (1, 349) = 0.226, p = .635, ηp^2 = 0.001, and severity and risk technique, F (1, 349) = 0.349, p = .555, ηp^2 = 0.001, were not significant. The main effect for risk assessment technique was significant such that, participants who read about the machine learning technique (M = 3.75, SD = 2.17) perceived the risk assessment as less ethical compared to those who read about the expert-based risk assessment (M = 4.48, SD = 2.39), F (1, 349) = 9.352, p = .002, ηp^2 = 0.026. However, the main effects of risk

assessment outcome, F (1, 349) = 2.989, p = .085, ηp^2 = 0.008, and severity, F (1, 349) = 0.134, p = .714, ηp^2 < 0.001, were not significant.

Use of Risk Assessment. A 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on whether the risk assessment should be used in sentencing. The three-way interaction was not significant, F (1, 352) = 1.881, p = .171, ηp^2 = 0.005. The two-way interactions between risk assessment technique and risk outcome, F (1, 352) = 2.274, p = .132, ηp^2 = 0.006, severity and risk outcome, F (1, 352) = 0.285, p = .594, ηp^2 = 0.001, and severity and risk technique, F (1, 352) = 0.765, p = .382, ηp^2 = 0.002, were not significant. The main effect for risk assessment technique was significant such that, participants who read about the machine learning technique (M = 3.75, SD = 2.25) were less likely to advocate for the use of the risk assessment in sentencing compared to those who read about the expert-based risk assessment (M = 4.24, SD = 2.29), F (1, 352) = 9.352, p = .002, ηp^2 = 0.026. However, the main effects of risk assessment outcome, F (1, 352) = 1.436, p = .232, ηp^2 = 0.004, and severity, F (1, 352) = 0.168, p = .682, ηp^2 < 0.001, were not significant.

Punishment Recommendations

Finally, a 2x2x2 ANOVA examined the effect of severity of crime (low vs high), risk assessment type (ML vs expert) and risk level conclusion (low vs high) on participants' punishment recommendations. The three-way interaction was not significant, F (1, 353) = 1.816, p = .179, ηp^2 = 0.005. The two-way interactions between

risk assessment technique and risk outcome, F (1, 353) = 0.222, p = .638, ηp^2 = 0.001, severity and risk outcome, F (1, 353) = 0.919, p = .338, ηp^2 = 0.003, and severity and risk technique, F (1, 353) = 0.151, p = .698, ηp^2 < 0.001, were not significant. The main effect for severity was significant such that participants who read the high severity crime (M = 6.83, SD = 1.47) gave higher punishment recommendations compared to those who read the lower severity crime (M = 6.12, SD = 1.82), F (1, 353) = 16.767, p < .001, ηp^2 = 0.045. However, the main effects of risk assessment outcome, F (1, 353) = 1.032, p = .310, ηp^2 = 0.003, and risk technique, F (1, 353) = 0.055, p = .814, ηp^2 < 0.001, were not significant.

Study 1 Discussion

In Study 1, I wanted to examine people's perceptions of the use of machine learning risk assessments in the sentencing compared to other risk assessment methods. The manipulations for severity of the crime and risk outcome were successful. The high severity crime increased perceptions of crime severity, dangerousness, and punishment recommendations. Also, when the defendant was deemed a high risk, perceived recidivism risk rates increased.

Some of the participants' perceptions of the defendant were affected by the expert-based risk assessment but not the machine learning risk assessment. This was seen with perceived controllability, a risk factor for recidivism. When presented with the expert-based risk assessment, participants were sensitive to the risk outcome given by the expert such that the defendant’s perceived controllability at the time of the crime was higher when the expert deemed them as a low risk compared to when he was deemed as a

high risk. Participants were also sensitive to the risk assessment when the severity of the crime was high. They gave higher ratings of dangerousness when they read the expert risk assessment compared to the machine learning risk assessment. It also seems as though participants felt the defendant was a high risk and didn't believe the risk assessment, found it less accurate, and disagreed with the outcome when the defendant was labeled as a low risk.

Previous research exploring algorithm aversion has suggested that people are hesitant to rely on algorithms when making consequential decisions (Dietvorst et al., 2015). Regardless of the severity of the crime and the risk outcome, participants expressed concern about the use of machine learning risk assessments and questioned their ethicality. Interestingly, their ethical preferences did not seem to have any effect on their judgements as there were no differences in sentencing when the risk assessment was informed by an expert versus a machine learning algorithm. If courts are implementing machine learning risk assessments in different trial phases, it is important to see if the effects from study 1 are replicated in a bail trial setting. Therefore, in study 2 I examine people's perceptions of the use of this risk assessment method in a pretrial scenario to see if these effects are present in a different trial phase.

Study 2 Method

Study 2 examined if the machine learning risk assessment method would negatively affect participant bail judgments and gauged participants' perceptions of these different procedures compared to other risk assessment techniques. This study used a within-subjects design where participants read about three separate crimes, which were

used to counterbalance the study, and were randomly paired with the three risk assessment procedures (our main manipulation). Given that people questioned the ethicality and felt that machine learning risk assessment should not be used in sentencing, I was interested to see if they felt the same way about the use of these risk assessments in bail hearings.

Participants

Using Prolific Academic, I obtained a sample of 452 US residents and removed 5 participants for failing our two quality control attention checks. Therefore, I ended with a sample size of 447 US residents (see table 3 for demographics). I needed a sample of 272 individuals, which is based off an a priori power analysis conducted using G*Power3 with the effect size f for ethics ratings for the risk assessment from our previous study (effect size f = 0.12), $\alpha = .05$, to have a power of .80.

Table 3

Participant Demographics

N	447
Mean age	32.59
Range	18-72
% Female	50.2%
% Non-White	30.6%
At least a bachelor's degree	60.0%
Moderate Political Views	15.6%
Liberal Political Views	53.7%
Conservative Political Views	18.7%

Design. After being recruited from Prolific Academic, participants took a Qualtrics survey where they read that a judge was to decide how to hold bail hearings and gathered the public's opinion. They were given a brief description of the purpose of the risk assessments and told that they were all reviewed by a committee and deemed 90% accurate. Next, they were randomly assigned to see one of three crimes: violent crime, white collar crime, and a sex offense. Then they were randomly assigned to view one of three risk assessment procedures: human, checklist/algorithm, or machine learning.

Materials

Bail Instruction. Prior to reading the case scenarios, participants read about the bail process. The summary included a description of arrest, booking, and then the bail hearing. Then they were told that they were going to be asked for their opinions on different risk assessment procedures used in the bail process.

Crime Scenarios. To give a range of crime severity, three types of crimes were chosen: violent, non-violent, and a sex offense. The violent crime was a description of an arson, where the defendant was accused of lighting a building on fire during the day when people were inside. For the non-violent crime, the defendant was accused of embezzling millions of dollars into a foreign account. For the sex offense, the defendant was accused of public indecency for revealing genitals to the public.

Risk Assessments. Participants read three types of risk assessments methods: expert forensic psychologist, standardized checklist, and machine learning algorithm. In

the expert risk assessment procedure, the defendant is interviewed by a forensic psychologist who uses their expert judgment and experience to evaluate their level of risk. The checklist is standardized and has items that are weighted based on known factors that affect the likelihood of a defendant showing up to their trial date. Each factor has a weighted score, which is added up to get the total risk score. Finally, the machine learning risk assessment describes a computer program that uses machine learning techniques--a set of complex mathematical equations that predict the level of risk based on evaluations of hundreds of criminal defendants. Once the risk factors are entered into the machine, the computer software uses an algorithm that is based on information from hundreds of criminal defendants. The algorithm considers all risk factors and compares them to previous defendants and prints out a risk score of how likely the defendant will return to his trial.

Measures. Participants were asked for their bail judgment (No bail necessary, Bail - with $ amount, and No bail allowed - sent to jail). Then they were asked about their perceptions of the defendant if they were released in terms of dangerousness (9-point Likert item from “Not at all Dangerous” to “Extremely Dangerous”) and the likelihood that they will show up to court using a sliding scale from 0 to 100. Further, they are asked to rate the risk assessment procedure used in that scenario in terms of ethicality (9-point Likert item from “Not at all Ethical” to “Extremely Ethical”), Accuracy (9-point Likert item from “Not at all Accurate” to “Extremely Accurate”), and if it should be used (9-point Likert item from “Definitely should NOT be used” to “Definitely should be used”). After reading all three scenarios, participants are asked to rate how strongly they

would recommend each risk assessment procedure if their state was choosing between using the three assessment tools for bail.

Attention Checks. To ensure the quality of the data, I included two attention checks. One attention check asked participants to leave it blank and the other asked them to select the third of five response options.

Procedure. After being recruited from Prolific Academic, participants took a Qualtrics survey where they read that a judge was to decide how to hold bail hearings and gathered the public's opinion. Next, they were randomly assigned to see one of three crimes: violent crime, white collar crime, and a sex offense. Then they were randomly assigned to view one of three risk assessment procedures: human, checklist/algorithm, or machine learning. Because participants in Study 1 seemed to inherently believe that all defendants are high risk, the participants in this study were told that the assessment classified the defendant as a "generally low risk" so as to make any impact of the assessment more apparent in the data. Finally, participants were asked to give their bail recommendations and opinions on the risk assessment method used.

Coded Variables. To analyze the data, I coded variables using syntax that are the order in which they saw each crime scenario (Scenario 1-3) and risk assessment (Risk 1-3). Further, scenario and risk order variables were also coded. These variables were used to ensure that there were no order effects. Finally, to measure the effect of the risk assessment procedure on judgments, I took the main 7 DVs and created variables for each risk assessment procedure. For the three bail amount variables for each risk assessment, I

did a 90% winsorization and created three new variables to replace the lower and upper extreme values (Wicklin, 2017). This means that the bail amounts below the 5th percentile would be replaced with the 5th percentile and the amounts above the 95th percentile would be replaced with the 95th percentile.

Study 2 Results

To examine the effect risk assessment procedure on bail judgments, I ran a mixed effects ANOVA, with the within-subjects factor having three levels. The three levels are the risk assessment type (expert, checklist, machine learning). I ran this 7 times using the main DVs: bail, bail amount, show up, danger, ethical, accurate, and should use. Additionally, I included crime scenario order as a between-subjects factor to ensure the order in which they saw the crime scenarios did not matter. I reported the Greenhouse-Geisser correction, which corrects for sphericity (Armstrong, 2017) (see table 4).

Table 4.

Effect of Risk Assessment Procedure on Bail Judgments and Risk Assessment Ratings

Measure	Mean Expert (SD)	Mean Checklist (SD)	Mean ML (SD)	Test
Bail	1.94 (0.61)	1.90 (0.56)	1.92 (0.59)	F (2, 872) = 0.440, p = .643, η_p^2 = 0.001
Bail Amount	26109 (62299.84)[b]	16323 (27275.88)[a]	29349 (66384.77)[b]	F (1.9, 296) = 3.31, p = .041, η_p^2 = 0.021
Show up	56.40 (26.96)	56.94 (26.18)	56.80 (27.58)	F (2, 862) = 0.071, p = .931, η_p^2 < 0.001
Danger	4.21 (2.35)[a]	4.57 (2.41)[b]	4.56 (2.45)[b]	F (2, 876) = 3.46, p = .032, η_p^2 = 0.008
Ethical	5.60 (2.26)[b]	5.48 (2.13)[b]	5.20 (2.21)[a]	F (1.92, 836) = 8.29, p = .0003, η_p^2 = 0.019
Accurate	5.39 (1.93)	5.45 (1.97)	5.28 (2.05)	F (2, 869) = 1.64, p = .194, η_p^2 = 0.004
Should use	5.53 (2.16)[a]	5.35 (2.13)[a]	5.13 (2.25)[b]	F (1.97, 864) = 7.67, p = .0005, η_p^2 = 0.017
Recommend	72.13 (21.55)[a]	63.68 (23.11)[b]	53.94 (28.17)[c]	F (1.87, 799) = 77.67, p = 3.1678E-30, η_p^2 = 0.154

When exploring bail amounts the main effect was significant such that when participants read about the machine learning risk assessment (M = 29349.38, SD = 5431.79), they suggested significantly higher bail amounts than when they read the checklist risk assessment (M = 16323.43, SD = 2271.61), but not when compared to the expert-based risk assessment (M = 26109.38, SD = 5147.69). Participants rated the defendant as less dangerous when they read the expert-based risk assessment (M = 4.21,

SD = 2.35) compared to the checklist (M = 4.57, SD = 2.41) and machine learning risk assessment (M = 4.56, SD = 2.45). When asked about the ethicality of each risk assessment procedure, they rated the expert (M = 5.60, SD = 2.26) as the highest, followed by the checklist (M = 5.45, SD = 2.13), and finally the machine learning risk assessment (M = 5.20, SD = 2.21). When asked if each risk assessment procedure should be used, they rated the expert (M = 5.53, SD = 2.16) as the highest, followed by the checklist (M = 5.35, SD = 2.13), and finally the machine learning risk assessment (M = 5.13, SD = 2.25). Finally, when asked the extent to which they would recommend the different procedures, participants rated experts the highest (M = 72.13, SD = 21.55), followed the checklist (M = 63.68, SD = 23.11), and finally the machine learning risk assessment (M = 53.94, SD = 28.17).

The design was not fully crossed, as there were the three key risk assessments which were counterbalanced by the three crime scenarios. To look at more of a standard study design where I can separate the effects of scenario from assessment, I conducted a series of exploratory 3 (crime scenario: arson, embezzlement, indecency) x 3 (risk assessment type: expert, checklist, machine learning) univariate ANOVAs on the first of the three conditions presented to the participants in order to examine if any of the effects of the risk assessments differed by crime scenario. In the exploratory analysis to test whether any of the effects of the assessment differed by crime scenario, the only test that was significant was participants' perceptions of the ethicality of the risk assessment.

Bail. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on bail recommendations was not significant, F (4, 437) = 1.784, p

= .131, ηp^2 = 0.016. The main effects for risk assessment, F (2, 437) = 0.387, p = .679, ηp^2 = 0.002 and crime scenario, F (2, 437) = 1.067, p = .679, ηp^2 = 0.005 were also not significant.

Bail Amount. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on the suggested bail amount was not significant, F (4, 285) = 0.561, p = .692, ηp^2 = 0.008. The main effects for risk assessment, F (2, 285) = 0.548, p = .579, ηp^2 = 0.004. There was a significant effect for a crime scenario such that participants suggested higher bail amounts for the embezzlement crime (M = 189905, SD = 52483) compared to the arson (M = 17197, SD = 47067) and the indecency crime (M = 6518, SD = 43365), F (2, 285) = 4.237, $p < .05$, ηp^2 = 0.029.

Show up. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on the defendant's perceived likelihood of showing up was significant, F (4, 436) = 3.114, $p < .05$, ηp^2 = 0.028. However, the simple main effects were not significant. The main effects for risk assessment, F (2, 436) = 1.413, p = .245, ηp^2 = 0.006 and crime scenario, F (2, 436) = 1.199, p = .302, ηp^2 = 0.005 were not significant.

Dangerousness. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on the defendant's perceived dangerousness was not significant, F (4, 438) = 0.889, p = .470, ηp^2 = 0.008. The main effect for risk assessment was not significant, F (2, 438) = 0.145, p = .865, ηp^2 = 0.001. However, the main effect for crime scenario was significant such that the defendant who committed embezzlement (M = 2.57, SD = 1.76) was rated as less dangerous than the defendants who committed

arson (M = 4.58, SD = 2.19) and the indecency crime (M = 4.34, SD = 2.02), F (2, 438) = 40.020, $p < .001$, $\eta p^2 = 0.155$.

Ethical. For wave 1, the two-way interaction between crime scenario and risk assessment procedure was not significant, F (4, 436) = 0.983, $p = .416$, $\eta p^2 = 0.009$. The main effect for the crime scenario was not significant, F (2, 436) = 1.476, $p = .230$, $\eta p^2 = 0.007$. The main effect for risk assessment was significant such that the machine learning risk assessment (M = 5.09, SD = 0.17) was rated as less ethical for use in bail decisions compared to expert (M = 5.79, SD = 0.16) and checklist (M = 5.55, SD = 0.21), F (2, 436) = 4.57, $p < .05$, $\eta p^2 = 0.019$.

Accuracy. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on the risk assessment's perceived accuracy was not significant, F (4, 437) = 1.749, $p = .138$, $\eta p^2 = 0.016$. The main effects for risk assessment, F (2, 437) = 1.422, $p = .242$, $\eta p^2 = 0.006$ and crime scenario, F (2, 437) = 1.539, $p = .216$, $\eta p^2 = 0.007$ were also not significant.

Should Use. For wave 1, the two-way interaction between crime scenario and risk assessment procedure on whether the risk assessment should be used to inform bail decisions was not significant, F (4, 438) = 0.487, $p = .745$, $\eta p^2 = 0.004$. The main effects for risk assessment, F (2, 438) = 1.445, $p = .237$, $\eta p^2 = 0.007$. The main effect for crime scenario was significant such that, participants were more likely to say that the risk assessment should be used for the indecency crime (M = 5.57, SD = 0.17) compared to the arson (M = 4.94, SD = 0.17) and the embezzlement crimes (M = 5.21, SD = 0.19).

Study 2 Discussion

Study 2 explored how people perceived the use of machine learning risk assessments in a bail hearing and how these perceptions might affect their bail judgments. Participants' perceptions of the defendant were affected by which risk assessment they read. When they read the expert-based risk assessment, the defendant was rated as less dangerous compared to the checklist and the machine learning risk assessment. Given that all three risk assessments settled on a “low risk” conclusion, it is possible that participants did not believe the machine learning or checklist risk assessment. Additionally, participants suggested bail amounts were higher for machine learning compared to the checklist risk assessment but not the expert risk assessment.

Consistent with Study 1, I found that people questioned the ethicality of the machine learning risk assessments and believed that they should not be used in arraignment hearings. Further, when asked to rate how strongly they would recommend each risk assessment procedure if their city was choosing between using the three assessment tools for bail, participants rated the machine learning risk assessment lowest. Overall, people had consistently negative perceptions of machine learning risk assessments and were not accepting of its use.

In regard to the exploratory analysis to test whether any of the effects of the assessment differed by crime scenario, there was only one significant interaction between crime scenario and risk assessment procedure on the defendant’s perceived likelihood of showing up. However, when I ran post-hoc analyses, none of the simple main effects were significant, so we will not interpret this effect. We also found similar results to the

main analyses which indicated that participants questioned the ethicality of the risk assessment and were hesitant about its use within arraignment hearings. Further, the embezzlement crime received the highest bail amounts compared to arson and the indecency crime. This is unsurprising given that the defendant had embezzled money into a foreign account which increases the likelihood that the defendant might flee to another country.

General Discussion

In both studies, there are some notable results regarding the participants' perceptions of machine learning risk assessments in sentencing and arraignment hearings. Consistently through both studies people had expressed ethical concerns and disagreed with the use of machine learning risk assessments in both sentencing and bail hearings. In Study 1, participants questioned the ethicality and agreed that machine learning risk assessments should not be used in Sentencing. In Study 2 participants recommended against the implementation of machine learning risk assessments in arraignment hearings within their own city compared to the checklist and expert risk assessments.

Participants' negative perceptions of the machine learning risk assessment seemed to influence their judgments of defendant responsibility. In Study 1, participants who read the expert-based risk assessment believed the defendant was in control of his actions more when the expert deemed the defendant as a low risk compared to a high risk. However, participants' perceptions of the defendant's control over his actions were not affected by the machine learning's risk outcome. This indicates that regardless of the risk outcome given by risk assessment, participants had already made a judgment about the

defendant. However, they were willing to change it when an expert said so but were not when the machine learning risk assessment differed from their judgment. They completely ignored the risk outcome given by the risk assessment and made their own judgment. It is unsurprising that participants have such negative perceptions towards the machine learning risk assessment, and they are most likely experiencing algorithm aversion (Dietvorst et al., 2015). Given that the courts are already implementing these types of risk assessments, future research might explore interventions on how to increase positive perceptions of machine learning risk assessments. It is suggested that exposure to artificial intelligence forecasters increases the likelihood of using them (Kramer et al., 2017). Therefore, increasing exposure to machine learning risk assessments might be the best place to start.

In study 2 participant's suggested bail amount was affected by the risk assessment type. When participants read the machine learning risk assessment, regardless of the crime, they suggested higher bail amounts compared to the checklist risk assessment but not the expert risk assessment. It is possible that people distrusted the machine learning risk assessment compared to the checklist; however, what is unclear is that the expert risk assessment was not significantly lower than the machine learning risk assessment, which would have had a much more explainable effect. It is possible that the effect I found is a fluke and might have been created by extraneous factors that cannot be explained. The risk assessment measures the future dangerousness and likelihood the defendant will show up to their court date and not the severity of the crime, therefore it should not be affecting the suggested bail amount.

There were some notable factors that might have distracted the participants from our risk assessment manipulations in both studies. In Study 1, participants read a brief description of the risk assessment conducted by neuroscientist Dr. Pavy, who used either his expert judgment or a machine learning algorithm to form an opinion of risk. In both conditions, Dr. Pavy took a fMRI brain scan of the defendant's brain and then entered it into a piece of specialized computer software. Given that the expert was involved up until this point, participants might have believed that the expert had more say in the risk assessment outcome than we originally intended. In Study 2, there was a judge involved in all three risk assessment procedures, as the judge was presiding over the arraignment hearings. Given that the expert was always involved, participants' perceptions of the risk assessment method might have been more positive than they might actually be if there was no expert involved at all. While this might not be currently realistic, it is possible with the expansion of technology that there might be less human involvement. Future research might try to fully remove human involvement for the machine learning risk assessment to fully test the effects of algorithm aversion in a court setting.

To increase ecological validity, it would be beneficial to get a judicial sample, as judges are the ones who would be using the risk assessments to decide bail and sentencing. It would be interesting to see how they perceive the use of machine learning risk assessments and if it affects their judgments in different trial phases. Further, future research might explore how people evaluate machine learning risk assessments and how much value they place on different factors such as ethicality compared to accuracy.

Conclusion

The United States has the highest incarceration rates in the world and experts are advocating for the use of risk assessments in multiple trial phases (Warren, 2007). This strategy will help lower incarceration rates by minimizing sentences and offering community-based alternatives for low-risk offenders. Given that machine learning risk assessments are currently being implemented into courts, it is essential to understand peoples' perceptions of them and if they are averse to machine learning risk assessments, then the next step is to better understand how to mend that relationship. Machine learning risk assessments can be used to help the courts operate more efficiently and reduce incarceration rates. These risk assessments can better predict which offenders are low risk to offer community-based alternatives and lesser sentences. This research aimed to fill the gap in the literature, as there are very few studies exploring the publics' perceptions of the use of machine learning risk assessments in different trial phases.

Previous research on algorithm aversion suggests that people prefer human forecasters over algorithms. The current research found corroborating evidence of algorithm aversion. Participants consistently had negative perceptions of machine learning risk assessments, discouraged their use in both sentencing and arraignment hearings, and did not like the idea of them being implemented in their own city. This is also consistent with a poll conducted by Pew Research Center (2018), which found that a majority of Americans found it unacceptable to use algorithms in criminal risk assessments for parole decisions. Policy makers should keep this in mind as they implement these risk assessments around the US in trial decisions.

References

Andrews, D. A., & Bonta, J. (2006). *The Psychology Of Criminal Conduct*. Routledge.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Providence, NJ: Matthew Bender & Company, Inc.

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016) "Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks," *Propublica*. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Armstrong R. A. (2017). Recommendations for analysis of repeated-measures designs: testing and correcting for sphericity and use of manova and mixed model analysis. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 37(5), 585–593. https://doi.org/10.1111/opo.12399

Barabas, C., Dinakar, K., Ito, J., Virza, M., & Zittrain, J. (2017). Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. *Proceedings of Machine Learning Research, 81*, 1–15. http://arxiv.org/abs/1712.08238

Baskin-Sommers, A. R., Baskin, D. R., Sommers, I. B., & Newman, J. P. (2013). The intersectionality of sex, race, and psychopathology in predicting violent crimes. *Criminal Justice and Behavior,* 40(10), 1068–1091. https://doi.org/10.1177/0093854813485412

Basu, C., & Singhal, M. (2016, March). Trust dynamics in human autonomous vehicle interaction: a review of trust models. In *2016 Aaai Spring Symposium Series*.

Bird, S. M., Goldacre, B., & Strang, J. (2011). We should push for evidence-based sentencing in criminal justice. *BMJ : British Medical Journal (Online)*, *342*(February), 1–2. https://doi.org/10.1136/bmj.d612

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. *Organizational Behavior And Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Bonta, J. (2002). Offender Risk Assessment: Guidelines for Selection and Use. *Criminal Justice and Behavior*, *29*(4), 355–379. https://doi.org/10.1177/0093854802029004002

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: a meta-analysis. *Psychological Bulletin*, *123*(2), 123.

Branham, L. S. (2012). Follow the Leader: The Advisability and Propriety of Considering Cost and Recidivism Data at Sentencing. *Federal Sentencing Reporter, 24*(3), 169-171. doi:10.1525/fsr.2012.24.3.169

Commons, M. L., Miller, P. M., Li, E. Y., & Gutheil, T. G. (2012). Forensic experts' perceptions of expert bias. *International Journal of Law and Psychiatry, 35*(5–6), 362–371. https://doi.org/10.1016/j.ijlp.2012.09.016

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In *Proceedings of The 23rd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining* (Pp. 797-806).

Warren, R. (2007). Evidence-Based Practice to Reduce Recidivism: Implications for State Judiciaries. Washington, DC: National Institute of Corrections.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. Judgment and Decision Making, 8(5), 512–520. Retrieved from http://journal.sjdm.org/12/121130a/jdm121130a.pdf

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892. https://doi.org/10.1073/pnas.1018033108

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582. http://dx.doi.org/10.1037/0003-066x.34.7.571

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668–1674. http://dx.doi.org/10.1126/science.2648573

Desmarais, S. L., Zottola, S. A., Duhart Clarke, S. E., & Lowder, E. M. (2020). Predictive Validity of Pretrial Risk Assessments: A Systematic Review of the Literature. *Criminal Justice and Behavior*, 1–23. https://doi.org/10.1177/0093854820932959

Devaul, R. A., Jervey, F., Chappell, J. A., Caver, P., Short, B., Keefe, S. O., & Briggs, M. (1987). Medical School Performance of Initially Rejected Students Psychological profile. Library, 257(1), 47–51.

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in us and non‐us samples. *International Journal of Selection and Assessment*, *19*(2), 209-216.

Dietvorst, B. J., Simmons, J., & Massey, C. (2014). Understanding algorithm aversion: forecasters erroneously avoid algorithms after seeing them err. In *Academy of Management Proceedings* (Vol. 2014, No. 1, P. 12227). Briarcliff Manor, Ny 10510: Academy of Management.

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: attitudes toward decision‐making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458-468

Eastwood, J., & Luther, K. (2016). What you should want from your professional: The impact of educational information on people's attitudes toward simple actuarial tools. *Professional Psychology: Research and Practice*, *47*(6), 402–412. https://doi.org/10.1037/pro0000111

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., … Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. The Counseling Psychologist, 34(3), 341–382. https://doi.org/10.1177/0011000005285875

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin, 31(*5), 680–692. https://doi.org/10.1177/0146167204271570

Gendreau, P., Little, T., & Goggin, C. (1996). A meta‐analysis of the predictors of adult offender recidivism: what works! *Criminology*, *34*(4), 575-608

Gilovich, T. (1991). *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life.* New York: Free Press.

González-Tapia, M. I., & Obsuth, I. (2015). "Bad genes" & criminal responsibility. *International Journal of Law and Psychiatry*, *39*, 60–71. https://doi.org/10.1016/j.ijlp.2015.01.022

Green, B. (2020). The false promise of risk assessments: Epistemic reform and the limits of fairness. FAT* 2020 - *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 594–606. https://doi.org/10.1145/3351095.3372869

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment*, 12, 19–30. http://dx.doi.org/10.1037/1040-3590.12.1.19

Grove, WM., & Meehl, RE. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law*, 2,293-323.

Harris, P. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgment. *Federal Probation*, 70(2), 8-14.

Harris, H. M., Goss, J., & Gumbs, A. (2019). Pretrial risk assessment in California. *Public Policy Institute of California.* https://www. ppic. org/publication/pretrial-risk-assessment-in-california.

Harvey, N., & Fischer, I. (1997). Taking advice: accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133. https://doi.org/10.1006/obhd.1997.2697

Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *IEEE Intelligent Systems*, *28*(1), 84–88. https://doi.org/10.1109/MIS.2013.24

Hyatt, J., Bergstrom, M., & Chanenson, S. (2011). Follow the Evidence: Integrate Risk Assessment into Sentencing. *Federal Sentencing Reporter, 23*(4), 266-268. doi:10.1525/fsr.2011.23.4.266

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., P. Slavic and A. Tversky, 1982. *Judgments under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press.

Kajita, M., & Kajita, S. (2017). Crime prediction by data-driven green's function method. *Arxiv Preprint Arxiv:1704.00240*.

Kang-Brown, J. (2021). People in Jail and Prison in 2020. January, 1–2. https://www.vera.org/people-in-jail-and-prison-in-2020

Kehl, D., Guo, P., & Kessler, S. (2016). Algorithms in the criminal justice system: assessing the use of risk assessments in sentencing. *Responsive Communities*. Retrieved from https://cyber.harvard.edu/ publications/2017/07/algorithms.

Kiehl, K. A., Anderson, N. E., Aharoni, E., Maurer, J. M., Harenski, K. A., Rao, V., … Steele, V. R. (2018). Age of gray matters: neuroprediction of recidivism. *Neuroimage: Clinical*, *19*(June), 813–823. https://doi.org/10.1016/j.nicl.2018.05.036

Kramer, M. F., Schaich Borg, J., Conitzer, V., & Sinnott-Armstrong, W. (2018, December). When do people want AI to make decisions? In *Proceedings Of The 2018 Aaai/Acm Conference On Ai, Ethics, and Society* (Pp. 204-209).

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and alloca- tion of function in human machine systems. *Ergonomics*, 22, 671–691.

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Mayer, R., Davis, J., & Schoorman, F. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review, 20*(3), 709-734. Retrieved from http://www.jstor.org.ezproxy1.lib.asu.edu/stable/258792

Meehl, P. E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of The Literature.* Minneapolis, Mn: University of Minnesota Press.

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual review of clinical psychology*, *12*, 489-513.

Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Clark Robbins, P., Mulvey, E.P, Roth, L.H., Grisso, T., & Banks, S. (2001). Rethinking risk assessment. New York: Oxford University Press.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783-792.

Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *22*(1), 58–76. https://doi.org/10.1037/law0000077

Nisbett, R. E., & Ross, L. (1980). *Human inference : strategies and shortcomings of social judgment.* Prentice-Hall.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., And Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*, 390–409. https://doi.org/10.1002/bdm.637

Parasuraman, R., & Riley, V. (1997). Humans and Automation : Use , Misuse , Disuse , Abuse. *Human Factors*, *39*(2), 230–253.

Perry, W. L. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand Corporation.

Peters, T. W. & Warren, R. K. (2006). Getting Smarter about Sentencing: NCSC's Sentencing Reform Survey. National Center for State Courts.

Pretrial Justice Institute. 2019. Scan of Pretrial Practices. *Pretrial Justice Institute* (2019). https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=24bb2bc4-84ed-7324-929c-d0637db43c9a&forceDialog=0

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*(3), 369-381.

Robert, C. (2014). Machine learning, a probabilistic perspective.

Rozin, P. (2005). The Meaning of "Natural": Process More Important Than Content. Psychological Science, 16(8), 652–658.

Silver, N. (2012). *The Signal and The Noise: Why so many predictions fail–but some don't*. New York, Ny: Penguin Press.

Singh, J. P. (2012). Handbook of juvenile forensic psychology and psychiatry. In E. Grigorenko (Ed.), *Handbook of juvenile forensic psychology and psychiatry* (pp. 215–225). New York: Springer.

Simon, H. (1983). Reason in human affairs (Harry Camp lectures at Stanford University; 1982). Stanford, Calif.: Stanford University Press.

Simon, H. A. (1992). What Is an Explanation of Behavior? Psychological Science, 3(3), 150–161. https://doi.org/10.1111/j.1467-9280.1992.tb00017.x

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. Stanford Law Review, 66, 803–872

Stetler, D. A., Davis, C., Leavitt, K., Schriger, I., Benson, K., Bhakta, S., … Bortolato, M. (2014). Association of low-activity MAOA allelic variants with violent crime in incarcerated offenders. *Journal Of Psychiatric Research*, *58*, 69–75. https://doi.org/10.1016/j.jpsychires.2014.07.006

Walmsley, R. (2018). *World prison population list* (12th ed.). Institute for Criminal Policy Research.

Warren, R. K. (2008). The Most Promising Way Forward: Incorporating Evidence-Based Practice into State Sentencing and Corrections Policies, *20*(5), 322–325. https://doi.org/10.1525/fsr.2008.20.5.322.322

Wicklin, R. (2017, February 8). Winsorization: The good, the bad, and the ugly. SAS: Analytics, Artificial Intelligence and Data Management. https://blogs.sas.com/content/iml/2017/02/08/winsorization-good-bad-and-ugly.html

Willis, J., & Todorov, A. (2006). Making up your mind after 100-ms Exposure to Face. *Psychological Science, 17*(7), 592–598.

Wolff, M. A. (2008). Evidence-Based Judicial Discretion: Promoting Public Safety through State Sentencing Reform. *New York University Law Review*, *83*(5), 1389–1419.

Wolfgang, M, Figlio, RM, Tracy, PE, Singer, SJ (1985). National Survey of Crime Severity (NCJ-96017). Washington, DC: Government Printing Office.

World Prison Brief, Institute for Criminal Policy Research (2017). Highest to Lowest – Prison Population Rate. http://www.prisonstudies.org/highest-to-lowest/prison_population_rate?field_region_taxonomy_tid=All

Xu, A., & Dudek, G. (2015, March). Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *Proceedings of The Tenth Annual*

Acm/Ieee International Conference on Human-Robot Interaction (Pp. 221-228). Acm.

Zajonc, R. B. (2001). Mere Exposure: a gateway to the subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228. https://doi.org/10.1111/1467-8721.00154

APPENDIX A

MATERIALS FROM STUDY 1

Machine Learning v. Expert Risk Neuroprediction
Start of Block: Intro/Consent
Q1 Greetings!
We are researchers at Arizona State University in the United States, and we would like to invite you to participate in short survey on how jurors evaluate evidence in a legal case. This will involve reading a short description of a crime, and answering additional questions about evidence you will be instructed to examine. You will not be able to refer back to the trial summary once you're finished with that page.
Your participation in this survey is voluntary. You have the right not to answer any questions, and to stop participating at any time. If you choose not to participate or to withdraw from the study at any time, there will be no penalty.
If you decide to participate, we expect the survey to take you 8-10 minutes. Please note that there is a 30 minute total time limit. You will be compensated $1.00 for completing our survey. Although there may be no other direct benefits to you, the possible benefits of your participation in the research include the opportunity to be involved in and learn about research. There are no foreseeable risks or discomforts to your participation. You must be 18 or older to participate in the study.
All information obtained in this study is strictly confidential and your responses will be anonymous. The anonymous data are stored on a password protected computer hard disk in a secure location so that only the study investigator may access it. The results of this research may be used in reports, presentations, and publications, but the researchers will not identify you.
If you have any questions concerning this study, please contact the research team / study investigator via email at laclab@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.
Please click the "NEXT" button to proceed with the survey.
Note: Please DO NOT use the "back" button on your browser, as it will invalidate your responses.
End of Block: Intro/Consent
Start of Block: Captcha
Q2 For quality control purposes, please complete this CAPTCHA. Sorry, we know these are annoying.
End of Block: Captcha
Start of Block: Case Intro
State v Adams_LS **State v. Adams**
You are about to read a summary of a court case in which the defendant, 28 year-old David Adams was charged with First-Degree Arson for setting fire to a building in which his ex-girlfriend worked.
In this case, we already know the outcome of the trial -- Mr. Adams was found guilty by a jury. What we would like your thoughts on is how Mr. Adams should be punished.
We will start by giving you a brief summary of what happened at the trial, and then we will move on to the sentencing / punishment phase of the trial. Please read all of the materials carefully as we will be asking for your honest impressions and judgments. Also

note that you will not be able to refer back to the trial summary as you complete our questions, so please take your time reading.
End of Block: Case Intro
Start of Block: Trial - Low Severity
Low Severity **Case Background**
On March 18, 2018 Mr. Adams drove to his ex-girlfriend Lisa's place of work. The two were seen screaming and arguing in the parking lot. After several minutes, one of Lisa's coworkers intervened, and asked Mr. Adams to leave. Mr. Adams quickly fled, but 30 minutes later, after the store was closed and all employees had left, Mr. Adams returned and dumped lighter fluid around the building and set it on fire. The fire was extinguished by the fire department, but caused nearly $10,000 worth of damage.
A witness called the police and gave a description of Mr. Adams's vehicle and license plate number. After searching the area, the police were able to find the vehicle and locate Mr. Adams's place of residence, where he was then arrested.

Page Break

OS: Pros **Trial Evidence** During the trial, the prosecution and defense both made opening statements and produced serveral witnesses.
The prosecution specifically noted that Mr. Adams had an existing criminal history. He started having trouble with the law a couple years after high school. At the age of 20, he was charged for property damages for breaking his neighbors window with a rock. About a year later, Mr. Adams was convicted for vandalism of a local grocery store. In addition, the defendant's neighbor testified that Mr. Adams frequently held loud parties at his house and on multiple occasions had the police show up due to fights breaking out. Finally, the prosecution showed surveillance video from a nearby building that showed Mr. Adams lighting the fire at his ex-girlfriend's place of employment.
The defense then explained that, while Mr. Adams has some criminal history, he has had a difficult upbringing. He grew up in poverty and his parents died while he was very young. One of Mr. Adams's coworkers testified that Mr. Adams was just a "normal guy" and that the alleged behavior was not at all the sort of thing that he would expect. He went on to exaplain that Mr. Adams likes to spend his time surfing at the beach--he wakes up early on the days that he works so that he can go surfing before he has to drive to work. Mr. Adams also spends a lot of time playing the guitar and will usually spend his weekends playing music with other local musicians. He has always been nice those he works with and is just like any normal 28 year old.
End of Block: Trial - Low Severity
Start of Block: Trial - High Severity
Q66 **Case Background**
On March 18, 2018 Mr. Adams drove to his ex-girlfriend Lisa's place of work. The two were seen screaming and arguing in the parking lot. After several minutes, one of Lisa's co-workers intervened, and asked Mr. Adams to leave. Mr. Adams quickly fled, but returned 30 minutes later and dumped lighter fluid around the building and set it on fire. The fire was extinguished by the fire department, but caused nearly $125,000 in damage,

and several smoke inhalation injuries and minor burns to the employees working in the building.
A witness called the police and gave a description of Mr. Adams's vehicle and license plate number. After searching the area, the police were able to find the vehicle and locate Mr. Adams's place of residence, where he was then arrested.

Page Break

Q67 **Trial Evidence**
During the trial, the prosecution and defense both made opening statements and produced several witnesses.
The prosecution specifically noted that Mr. Adams had an existing criminal history. He started having trouble with the law a couple years after high school. At the age of 20, he was charged for property damages for breaking his neighbors window with a rock. About a year later, Mr. Adams was convicted for vandalism of a local grocery store. In addition, the defendant's neighbor testified that Mr. Adams frequently held loud parties at his house and on multiple occasions had the police show up due to fights breaking out. Finally, the prosecution showed surveillance video from a nearby building that showed Mr. Adams lighting the fire at his ex-girlfriend's place of employment.
The defense then explained that, while Mr. Adams has some criminal history, he has had a difficult upbringing. He grew up in poverty and his parents died while he was very young. One of Mr. Adams's coworkers testified that Mr. Adams was just a "normal guy" and that the alleged behavior was not at all the sort of thing that he would expect. He went on to exaplain that Mr. Adams likes to spend his time surfing at the beach--he wakes up early on the days that he works so that he can go surfing before he has to drive to work. Mr. Adams also spends a lot of time playing the guitar and will usually spend his weekends playing music with other local musicians. He has always been nice those he works with and is just like any normal 28 year old.
End of Block: Trial - High Severity
Start of Block: Sentencing - Intro
Q213 **Sentencing**
Mr. Adams was found guilty at his trial and is now awaiting a decision on sentencing for first degree arson. We'd like you to read this brief description of the arguments presented to the court during the sentencing hearing. Again, please read it carefully as we will be asking for your impressions and judgments. You will not be able to refer back to the information after you move on.

Page Break

Q214 **Sentencing Risk Assessment**
A common event that occurs within the sentencing phase is to categorize the defendant’s risk level for engaging in future violent behavior. This categorization of risk is then used to help guide the sentencing decision. This decision has very important implications for both the defendant and the general public. We will now describe the details of this risk assessment procedure and outcome.

End of Block: Sentencing - Intro
Start of Block: Machine Learning Risk Assessment Low
ML Risk **Risk Assessment**
To assess the likelihood that the defendant would pose a risk to the public in the future, Dr. Pavy, a neuroscientist, took a fMRI scan of Mr. Adams's brain and entered it into a piece of specialized computer software. The computer then "looks" at the brain scan data and analyzes it using machine learning—a form of artificial intelligence. Because we know that certain brain structures, like the ones that are associated with impulsivity and aggression, are related to the likelihood of someone engaging in criminal behavior, the computer software is able to use the data from the defendant's brain to make such a prediction. Using a set of complex mathematical equations that were based on analyses of hundreds of other criminal offenders, the algorithm in the machine learning software automatically predicts the amount of risk a person poses, and simply classifies them as "low," "medium," or "high" risk.

Page Break

ML Low **Risk Assessment Outcome**
At the sentencing hearing, Dr. Pavy testified as to the results of the assessment, stating that: "The machine learning risk assessment is completed within a matter of hours after entering the defendant's fMRI data. The software provides a simple readout of its prediction, and I am here to read the results of the machine learning risk assessment. In this case, the computer program identified Mr. Adams as someone of low risk to reoffend in the future."

Page Break

ML Low Closing **Closing Statements**
Prosecution - "The conclusion found by the machine learning risk assessment, does not matter. We can not see into the future and we have no idea if Mr. Adams will reoffend again or not. Even if the computer concluded he is a low risk, Mr. Adams has committed a crime and deserves to be punished for his actions and his sentence should be maximized."
Defense - "The machine learning risk assessment concluded, based on a statistical algorithm, that my client is a low risk. This means he is at low risk to commit the crime again. I agree with the decision made by this computer program and believe Mr. Adams's sentence should be minimized."
End of Block: Machine Learning Risk Assessment Low
Start of Block: Machine Learning Risk Assessment High
ML Risk **Risk Assessment**
To assess the likelihood that the defendant would pose a risk to the public in the future, Dr. Pavy, a neuroscientist, conducted a fMRI scan of Mr. Adams's brain and entered it into a piece of specialized computer software. The computer then "looks" at the brain scan data and analyzes it using machine learning—a form of artificial intelligence. Because we know that certain brain structures, like the ones that are associated with

impulsivity and aggression, are related to the likelihood of someone engaging in criminal behavior, the computer software is able to use the data from the defendant's brain to make such a prediction. Using a set of complex mathematical equations that were based on analyses of hundreds of other criminal offenders, the algorithm in the machine learning software automatically predicts the amount of risk a person poses, and simply classifies them as "low," "medium," or "high" risk.

Page Break

ML High **Risk Assessment Outcome**

At the sentencing hearing, Dr. Pavy testified as to the results of the assessment, stating that: “The machine learning risk assessment is completed within a matter of hours after entering the defendant's fMRI data. The software provides a simple readout of its prediction, and I am here to read the results of the machine learning risk assessment. In this case, the computer program identified Mr. Adams as someone of high risk to reoffend in the future.”

Page Break

ML High Closing **Closing Statements**

Prosecution - "Mr. Adams was deemed as a high risk according to the machine learning risk assessment, conducted by the computer program. He needs to be punished to the maximum extent. It is likely he will reoffend and severely punishing him will discourage him from doing so."

Defense - "The machine learning risk assessment has concluded, based on the statistical algorithm, that my client is a high risk. It is unethical to punish my client for something that hasn’t occurred yet. My client deserves to be punished for the crime that was committed and nothing more."

End of Block: Machine Learning Risk Assessment High

Start of Block: Expert Risk Assessment Low

Expert Risk **Risk Assessment**

To assess the likelihood that the defendant would pose a risk to the public in the future, Dr. Pavy, a neuroscientist, conducted a fMRI scan of Mr. Adams’s brain. Dr. Pavy then looks at the brain scans and evalutes them based on his expertise. Because we know that certain brain structures, like the ones that are associated with impulsivity and aggression, are related to the likelihood of someone engaging in criminal behavior, Dr. Pavy is able to use the brain scan images to make such a prediction. Using his experience evaluating hundreds of other criminal offenders, Dr. Pavy can predict the amount of risk a person poses and classify them as "low," "medium," or "high" risk.

Page Break

Expert Low **Risk Assessment Outcome**

In the sentencing hearing, Dr. Pavy testified as to the results of the assessment, stating that “The risk assessment is completed after a couple of hours, as I take time to examine the fMRI. I look at the patterns and structure of the brain and then estimate the

defendant's risk based on other defendant's I've examined. I am here to report the results of the risk assessment I conducted that is based on my expertise. I identified Mr. Adams as someone of low risk to reoffend when released."

Page Break

Expert Low Closing **Closing Statements**

Prosecution - "The conclusion found by Dr. Pavy's risk assessment does not matter. We can not see into the future and we have no idea if Mr. Adams will reoffend again or not. Even if Dr. Pavy concluded that he is a low risk, Mr. Adams has committed a crime and deserves to be punished for his actions and his sentence should be maximized."

Defense - "Dr. Pavy who conducted the risk assessment has concluded, based on his expertise, that my client is a low risk. This means he is at low risk to commit the crime again. I agree with Dr. Pavy's risk assessment and believe Mr. Adams's sentence should be minimized."

End of Block: Expert Risk Assessment Low

Start of Block: Expert Risk Assessment High

Expert Risk **Risk Assessment**

To assess the likelihood that the defendant would pose a risk to the public in the future, Dr. Pavy, a neuroscientist, conducted a fMRI scan of Mr. Adams's brain. Dr. Pavy then looks at the brain scans and evalutes them based on his expertise. Because we know that certain brain structures, like the ones that are associated with impulsivity and aggression, are related to the likelihood of someone engaging in criminal behavior, Dr. Pavy is able to use the brain scan images to make such a prediction. Using his experience evaluating hundreds of other criminal offenders, Dr. Pavy can predict the amount of risk a person poses and classify them as "low," "medium," or "high" risk.

Page Break

Expert High **Risk Assessment Outcome** In the sentencing hearing, Dr. Pavy testified as to the results of the assessment, stating that "The risk assessment is completed after a couple of hours, as I take time to examine the fMRI. I look at the patterns and structure of the brain and then estimate the defendant's risk based on other defendant's I've examined. I am here to report the results of the risk assessment I conducted that is based on my expertise. I identified Mr. Adams as someone of high risk to reoffend when released."

Page Break

Expert High Closing **Closing Statements**

Prosecution - "Mr. Adams was deemed as a high risk according to Dr. Pavy, based on his experience. He needs to be punished to the maximum extent. It is likely he will reoffend and severely punishing him will discourage him from doing so."

Defense - "Dr. Pavy who conducted the risk assessment has concluded, based on his expertise, that my client is a high risk. It is unethical to punish my client for something that hasn't occurred yet. My client deserves to be punished for the crime that was committed and nothing more."

End of Block: Expert Risk Assessment High
Start of Block: Responsibility
Resp
Thank you for reading the trial summary. Now we would like you to answer the following questions about your impressions of the defendant, Mr. Adams. There are no right or wrong answers here; we are just looking for your gut reaction.

Control On a scale of 1 - 9, please indicate to what extent you feel that the defendant was in **control** of his actions at the time of the assault.

Reoffnd On a scale of 1 - 9, please indicate to what extent you feel that the defendant is to **re-offend**, or to commit a crime again in the future.

Danger On a scale of 1 - 9, please indicate how **dangerous** of a person the defendant is.

Severe On a scale of 1 - 9, please indicate how **severe** the crime committed was.

End of Block: Responsibility
Start of Block: Risk Assessment Questions
Risk

Next, we would like you to think about the risk assessment described by Dr. Pavy in the sentencing hearing. Please answer the following questions about what you think about that assessment.

Believe On a scale of 1 - 9, please indicate how much you **believe** the risk outcome given by the risk assessment.

Accurate On a scale of 1 - 9, please indicate **how accurate** you believe this risk assessment is.

Matched_Belief On a scale of 1 - 9, please indicate how well the risk assessment **matched your own belief** in the defendant's future risk?

Ethical On a scale of 1 - 9, please indicate how **ethical** is it to use this type of risk assessment.

Should_be_Used On a scale of 1 - 9, please indicate if you think risk assessments in general **should be used** to inform punishments for criminals?

Odds As you read in the risk assessment, criminal defendants are classified as either Low, Medium, or High risk for re-offending. In your estimation, what do you feel the odds are that a defendant in each category will actually re-offend (commit another crime sometime in the future)?

Odds (0-100%) that a **LOW RISK** defendant will re-offend ()	
Odds (0-100%) that a **MEDIUM RISK** defendant will re-offend ()	
Odds (0-100%) that a **HIGH RISK** defendant will re-offend ()	

End of Block: Risk Assessment Questions
Start of Block: Punishment
Punish **Now we would like to get your opinions about what should happen to Mr. Adams. Please answer the following questions related to punishment. There are no right or wrong answers here--we are just looking for your gut reaction.**
Severe_Punish First, without getting into exact amounts, how severely do you believe Mr. Adams should be punished?

- o To the **MINIMUM** extent allowable (1)
- o (2)
- o (3)
- o (4)
- o (5)
- o (6)
- o (7)
- o (8)
- o To the **MAXIMUM** extent allowable (9)

Sentence The defendant in this case was convicted of First-Degree Arson. In most states the punishment (in terms of time in prison) ranges from just a few months to a several years. If you were to recommend a sentence for this defendant, how long would the defendant's sentence be? (Note: if you don't believe the defendant should be jailed at all, you can enter 0)

- o Years (1) _____
- o Months (2) _____

End of Block: Punishment
Start of Block: Manipulation checks
Manip
Next, we'd like you to answer the following questions about the criminal case that you read. YOU WILL BE PAID REGARDLESS OF HOW YOU RESPOND TO THESE ITEMS. However, we would simply like to know how well you remember these items. Please do your best and click NEXT when you're finished.
Manip_Crime Mr. Adams was found guilty of what crime?

- o Murder (1)
- o Assault (2)
- o Armed robbery (3)
- o Arson (4)

Manip_Age Mr. Adams was how old?

- o 55 (1)
- o 18 (2)

o 38 (3)
o 28 (4)

Manip_Blank For quality control, please leave this question blank.

_____

Manip_Risk The risk assessment was based on what technique?
o Dr. Pavy's expertise (1)
o Machine learning algorithm (2)
o Blood tests (4)
o The big five psychological test (6)

Manip_Outcome What was the outcome of the Risk Assessment?
o Mr. Adams was LOW risk (1)
o Mr. Adams was MEDIUM risk (2)
o Mr. Adams was HIGH risk (4)
o The Risk assessment was inconclusive (6)

End of Block: Manipulation checks

Start of Block: Demos

Demo

Finally, we have some basic questions about yourself.

Q22 What is your gender?
o Male (1)
o Female (2)
o Other / Prefer not to say (3)

Q23 What is your age?

_____

Q24 Which ethnicity do you most identify with?
o Hispanic / Latino/ Central/South American (1)
o White / Caucasian (2)
o Black / African American (3)
o Middle East / North African (4)
o Asian / Pacific Islander (5)
o Other (6) _____

Q25 In which state do you currently reside?
▼ Alabama (1) ... I do not reside in the United States (53)

Q26 What is your highest level of education?
▼ Less than High School (1) ... Doctoral Degree (7)

Q27 Generally speaking, which of the following most closely describes your political views?
▼ Very Conservative (1) ... Very Liberal (7)

Q66 Generally speaking, how familiar would you say you are with each of the following concepts?

	Not familiar at all (51)	Slightly familiar (52)	Moderately familiar (53)	Very familiar (54)	Extremely familiar (55)

Artificial Intelligence (4)	o	o	o	o	o
Machine Learning (5)	o	o	o	o	o
Risk Assessment (6)	o	o	o	o	o
Criminal Court Trials (7)	o	o	o	o	o
Mathematics and Statistics in General (8)	o	o	o	o	o

APPENDIX B

STUDY 2 MATERIALS

Machine Learning v. Exp (Bail)

Start of Block: Intro/Consent

Q1 Greetings!

We are researchers at Arizona State University, and we would like to invite you to participate in short survey on how people evaluate evidence in a legal case. This will involve reading a short description of a criminal case, and answering additional questions about evidence you will be instructed to examine.

Your participation in this survey is voluntary. You have the right not to answer any questions, and to stop participating at any time. If you choose not to participate or to withdraw from the study at any time, there will be no penalty.

If you decide to participate, we expect the survey to take you about 5-10 minutes. Please note that there is a 30 minute total time limit. You will be compensated $1.50 for completing our survey. Although there may be no other direct benefits to you, the possible benefits of your participation in the research include the opportunity to be involved in and learn about research. There are no foreseeable risks or discomforts to your participation. You must be 18 or older to participate in the study.

All information obtained in this study is strictly confidential and your responses will be anonymous. The anonymous data are stored on a password protected computer hard disk in a secure location so that only the study investigator may access it. The results of this research may be used in reports, presentations, and publications, but the researchers will not identify you.

If you have any questions concerning this study, please contact the research team / study investigator Nick Schweitzer via email at laclab@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

Please click the "NEXT" button to proceed with the survey.

Note: Please DO NOT use the "back" button on your browser, as it will invalidate your

responses.

End of Block: Intro/Consent

Start of Block: Captcha

Q2 Before we begin, please complete this CAPTCHA. Sorry, we know these are annoying.

End of Block: Captcha

Start of Block: Case Intro

Instructions
Opinions on Bail Procedures: Some Background When a person in the US is arrested and accused of committing a crime, they must be "arraigned." During the arraignment, the accused person will enter a plea of either *guilty* or *not-guilty*. If the accused pleads not-guilty, a judge will determine whether that accused person is eligible for **bail**. This is what we are interested in studying today.

The purpose of **bail** is to ensure that the accused individual will be present for the subsequent trial or legal proceedings against him or her (which may take months or even years). Bail is set as an amount of money that the accused must pay (either in cash, equity, or bonds) to the court as a guarantee that he or she will return to stand trial. Once the trial process is complete, any cash or equity paid by the accused is returned. But, if the accused flees or does not show up for the trial, they may forfeit the bail.

The amount of bail is to be set in proportion to an accused's risk of fleeing / not showing up for trial. This might also include not requiring any bail (if a person is considered low-risk), or deciding that no amount of money is sufficient (if they are very high risk). If an accused cannot pay the bail, they will be held in prison until trial, which is why the 8th Amendment of the US Constitution explicitly forbids "excessive bail."

Click NEXT to continue.

Page Break

Q109 Recently there have been many methods tested for how to accurately decide how much of a flight risk someone is, and, thus, what their bail amount should be. These "pre-trial risk assessments" are based on a wide variety of different methods. We are interested in hearing your opinions on these risk assessment procedures.

On the subsequent pages, you will read through three different summaries of criminal cases along with the bail procedure used by the judge in each case. You will be asked about your opinions about how bail should be set in the case.

These summaries are based on actual cases, and in all situations had defendants with no previous criminal records (and, therefore, more difficult to predict whether they will show up for trial). We are only providing a very brief amount of information about each case. We know that you don't have enough to fully judge them--we are only interested in your initial impressions.

When you are ready to read the first case summary, click NEXT to begin.

End of Block: Case Intro

Start of Block: Scenarios

Case Creepy **Case Summary:**
Defendant C.J. has been arrested and is awaiting trial. He was accused of traveling to five different grocery stores over a one-week period wearing an oversized face mask, and displaying his genitals to several women inside of the stores. The defendant was arrested after the police received multiple complaints and descriptions from customers at the stores. Police claim that security footage of the perpetrator removing his mask outside one of the stores shows the defendant is the perpetrator. The defendant does not have any prior offenses and claims that he is innocent and not the individual pictured in the security footage.

Case Nonviolent **Case Summary:**
Defendant A.M. was a manager for a chain of local restaurants who had access to purchasing and payroll accounts. While cleaning his office, his assistant found emails that indicated that he was embezzling money from his work. He was arrested after the police received a warrant and searched his office, where they found 10 years worth of financial statements which showed $2.5 million dollars was embezzled into a foreign account. The name on the financial statements did not match the defendant, however. The defendant claims that he was framed and has not committed a crime before.

Page Break

Case Violent **Case Summary:** Defendant J.G. was accused of arson and assault after an altercation with an ex-girlfriend. The defendant was seen seen screaming and arguing

with the victim in the parking lot. About an hour later, the victim saw a hooded figure dump lighter fluid on her car and workplace entrance and light them on fire. The victim called the police and said the perpetrator sped off in a black SUV. After searching the area, the police found the defendant in a park near the scene wearing a hooded sweatshirt. The defendant claims that his argument wiht his ex-girlfriend was over and he had no reason to light a fire, nor did he drive a black SUV.

Expert In this particular jurisdiction, the arraignment proceedings use a bail risk assessment procedure in which each defendant is evaluated by an expert forensic psychologist who specializes in bail proceedings. This expert interviews the defendant, and, based on known suggest that indicate that someone might flee or miss their trial (along with the expert's past history with defendants), makes a recommendation to the judge as to the defendant's level of flight risk. The judge will then assign higher bail to individuals that the expert determines are a flight risk.

Checklist The arraignment hearings in this jurisdiction use a bail risk assessment procedure called the Bail Risk Checklist. Court personnel will gather information from outside sources on a set of over 30 risk factors which are entered into a checklist. The checklist then gives point values based on how each factor predicts a defendant showing up to their trial date. The more points each defendant receives, the more risk they are considered to have. The judge the uses the risk score when assigning bail and increases the bail amount as the score increases.

ML In this jurisdiction’s arraignment hearings, the courts use a bail risk assessment procedure called the Advanced Risk Evaluation Tool. This procedure involves an automated computer program that uses artificial intelligence to predict a defendant's level of risk. The computer uses past data to "learn" what makes people more or less likely to jump bail, and then analyzes each defendant's file to create a risk store. The judge will assign bail based on the outcome of the computer's assessment.

Instructions 1 **We would like to get your initial opinions about the case and procedure you just read. There are no right or wrong answers here--we are just looking for your gut reaction.**

Instructions 2 **We would like to get your initial opinions about the case and procedure you just read. There are no right or wrong answers here--we are just looking for your gut reaction.**

Instructions 3 **We would like to get your initial opinions about the case and procedure you just read. There are no right or wrong answers here--we are just looking for your gut reaction.**

Bail 1
To help us better understand your reactions to this case, we would like you to imagine that each of the defendants you read about were determined to be a "generally low risk" by the assessment procedure. (This is the most common risk assessment outcome for first-time offenders.)

Further, for the purposes of setting bail amounts, we will assume that the defendants all earn about $3000 per month after taxes.

Given all of this information, if you were the judge, how would you set the bail in this case?

- o No Bail Required ($0) (1)
- o Require Bail of: (Type in your recommended bail amount below in numbers only) (14) _____
- o Remand to Jail (No Bail Allowed) (15)

Bail 2
To help us better understand your reactions to this case, we would like you to imagine that each of the defendants you read about were determined to be a "generally low risk" by the assessment procedure. (This is the most common risk assessment outcome for first-time offenders.)

Further, for the purposes of setting bail amounts, we will assume that the defendants all earn about $3000 per month after taxes.

Given all of this information, if you were the judge, how would you set the bail in this case?

- o No Bail Required ($0) (1)
- o Require Bail of: (Type in your recommended bail amount below in numbers only) (14) _____
- o Remand to Jail (No Bail Allowed) (15)

Page Break

Bail 3

To help us better understand your reactions to this case, we would like you to imagine that each of the defendants you read about were determined to be a "generally low risk" by the assessment procedure. (This is the most common risk assessment outcome for first-time offenders.)

Further, for the purposes of setting bail amounts, we will assume that the defendants all earn about $3000 per month after taxes.

Given all of this information, if you were the judge, how would you set the bail in this case?

- o No Bail Required ($0) (1)
- o Require Bail of: (Type in your recommended bail amount below in numbers only) (14) _____
- o Remand to Jail (No Bail Allowed) (15)

Showup 1 If, hypothetically, the judge released the accused without any bail, what to you think the odds (%) are that the accused in this case would show up for his trial?

Would definitely NOT show up	Would definitely show up

Released on bail? ()	

Showup 2 If, hypothetically, the judge released the accused without any bail, what to you think the odds (%) are that the accused in this case would show up for his trial?

Would definitely NOT show up	Would definitely show up

Released on bail? ()	

Showup 3 If, hypothetically, the judge released the accused without any bail, what to you think the odds (%) are that the accused in this case would show up for his trial?

Would definitely NOT show up	Would definitely show up

Released on bail? ()	

Danger 1 If the accused in this case was released on bail, how much of a danger do you think he would pose to the community while awaiting trial?

Danger 2 If the accused in this case was released on bail, how much of a danger do you think he would pose to the community while awaiting trial?

Danger 3 If the accused in this case was released on bail, how much of a danger do you think he would pose to the community while awaiting trial?

Ethical 1 On a scale of 1 - 9, please indicate how ethical you think it is this to use this bail risk assessment procedure for the defendant?

Ethical 2 On a scale of 1 - 9, please indicate how ethical you think it is this to use this bail risk assessment procedure for the defendant?

Ethical 3 On a scale of 1 - 9, please indicate how ethical you think it is this to use this bail risk assessment procedure for the defendant?

Accurate 1 On a scale of 1 - 9, please indicate how accurate you think this bail risk assessment procedure is?

Accurate 2 On a scale of 1 - 9, please indicate how accurate you think this bail risk assessment procedure is?

Accurate 3 On a scale of 1 - 9, please indicate how accurate you think this bail risk assessment procedure is?

Should use 1 On a scale of 1 - 9, please indicate if you think this specific type of risk assessment should be used to inform bail decisions?

Should use 2 On a scale of 1 - 9, please indicate if you think this specific type of risk assessment should be used to inform bail decisions?

Should use 3 On a scale of 1 - 9, please indicate if you think this specific type of risk assessment should be used to inform bail decisions?

End of Block: Scenarios

Start of Block: Overall Procedure Comparison

Q346 Now that you have read about these different types of bail procedures: If your home state was choosing between using these three assessment types for setting bail, how strongly would you recommend each of these procedures?

	Would NOT Recommend	Would STRONGLY Recommend
Forensic Expert Interview ()		
Bail Risk Checklist ()		
Artificial Intelligence Tool ()		

Manip_Crime For quality purposes, please pick the third option.
- o 5 (1)
- o 1 (2)
- o 6 (3)
- o 7 (4)

End of Block: Overall Procedure Comparison

Start of Block: Demos

Demo **Finally, we have some basic questions about yourself.**

Q22 What is your gender?
- o Male (1)
- o Female (2)
- o Other / Prefer not to say (3)

Q23 What is your age?

_____

Q24 Which ethnicity do you most identify with?
- o Hispanic / Latino/ Central/South American (1)
- o White / Caucasian (2)
- o Black / African American (3)
- o Middle East / North African (4)

o Asian / Pacific Islander (5)
o Other (6) _____

Q25 In which state do you currently reside?
▼ Alabama (1) ... I do not reside in the United States (53)

Q26 What is your highest level of education?
▼ Less than High School (1) ... Doctoral Degree (7)

Manip_Blank For quality control, do not write anything in the space below.
_____

Q27 Generally speaking, which of the following most closely describes your political views?
▼ Very Conservative (1) ... Very Liberal (7)

BailJump If you had to guess, what percentage of criminal defendants in the US who are released on bail flee and/or fail to show up for their eventual trial?

Q66 Generally speaking, how familiar would you say you are with each of the following concepts?

	Not familiar at all (51)	Slightly familiar (52)	Moderately familiar (53)	Very familiar (54)	Extremely familiar (55)
Artificial Intelligence (4)	o	o	o	o	o
Machine Learning (5)	o	o	o	o	o
Risk Assessment (6)	o	o	o	o	o

Criminal Court Trials (7)	o	o	o	o	o
Mathematics and Statistics in General (8)	o	o	o	o	o

Q44 Thank you. Click NEXT to finish.

End of Block: Demos

APPENDIX C

IRB

ASU Knowledge Enterprise Development

EXEMPTION GRANTED

Nicholas Schweitzer
Social and Behavioral Sciences, School of
-
njs@asu.edu

Dear Nicholas Schweitzer:

On 1/28/2014 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Neuroscience in the Courtroom
Investigator:	Nicholas Schweitzer
IRB ID:	STUDY00000579
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	None

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 1/28/2014.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc:
Alisha Meschkow