FIU Steven J. Green School of International & Public Affairs
Department of Economics

Interrogation and Disclosure of Evidence

Jesse Bull

Florida International University

Department of Economics

Florida International University

Working Paper 2212

October 2022

11200 SW 8th Street, Miami, Florida 33199

https://economics.fiu.edu/

Interrogation and Disclosure of Evidence

Jesse Bull*

Revised September 2022

Abstract

In recent years there has been a shift to less confrontational interrogation methods. One argument in favor of these methods is that they reduce the scope for false confessions. This paper explores this idea from an information perspective with rational agents. Should a police officer choose to interrogate an accused person, the goal is a confession. The police officer bases her decision of whether to interrogate on both the realization of some evidence and on a private non-verifiable signal; both of these influence the outcome at trial, should the case proceed to trial. So there are potentially two channels of information to the accused: 1) that the police officer decided to interrogate, and 2) the presentation of evidence to the accused during interrogation. The accused chooses whether to confess by weighing his expected payoff from going to trial against that from confessing. Less confrontational interrogation methods close the second channel of information and can reduce false confessions while still inducing guilty to confess. It is helpful for the accused to be somewhat informed about the strength of the case against him, and this comes through being interrogated. However, it can be detrimental for the accused to be too well informed about the strength of the case against him.

1 Introduction

Viewers of television shows about police often see police officers aggressively question an accused person and early on present incriminating evidence in a confrontational manner. This is typically portrayed as inducing the guilty accused to confess. In North America most

*Florida International University. Email: bullj@fiu.edu. This paper began as a result of a related paper titled "Interrogation and Evidence Fabrication." Acknowledgements to be added.

of the aggressive interrogation techniques have their basis in what is known as the "Reid Technique." The most recent version of this technique is presented in Inbau et al. (2013).

Recently, in Canada, the UK, Denmark, and New Zealand, there has been a transition to a less manipulative and confrontational approach known as PEACE for Preparation and Planning, Engage and Explain, Account, Closure and Evaluate. See, for example Snook, Eastwood, Barron (2014) and Starr (2013). The approach is less manipulative and less confrontational. Instead of focusing on confronting the accused with incriminating evidence, the technique encourages an open minded approach to asking the accused for information. Bull (2018) states:

> The PEACE approach is based on a number of basic principles that were chosen by the police. These include:
>
> 1) "The purpose of investigative interviewing is to obtain accurate and reliable information from suspects, witnesses or victims in order to discover the truth about matters under investigation."
>
> 2) "Interviews should be approached with an open mind. Information obtained from the person who is being interviewed should always be tested against what the investigator already knows or what can reasonably be established."

Advocates of the PEACE approach focus on the scope for obtaining information from the accused and suggest that its use can reduce false confessions. Gudjonsson (2012) notes, "Furthermore, Gudjonsson and Pearse (2011) argue that the psychologically manipulative components of the Reid Technique make it more susceptible to inducing false confessions than the UK PEACE model."[1] Many of the criticisms of the Reid Technique focus on its confrontational approach, manipulation of the accused, and goal of obtaining a confession. It's felt that the non-confrontational nature of the PEACE approach is much less likely to pressure an accused into confessing.

The Reid Technique focuses on the interrogator first determining whether she believes the accused is guilty. Once she is convinced of the guilt of the accused, she is to use "The Reid Nine Steps of Interrogation©" as described in Inbau et al. (2013). These steps are:

1. Direct, Positive Confrontation
2. Theme Development
3. Handling Denials

[1] See Gudjonsson (2012) and Gudjonsson and Pearse (2008).

4. Overcoming Objections

5. Procurement and Retention of a Suspect's Attention

6. Handling the Suspect's Passive Mood

7. Presenting an Alternative Question

8. Having the Suspect Orally Relate Various Details of the Offense

9. Converting an Oral Confession into a Written Confession

As it suggests, the first step focuses on confrontation. While much of the motivation for the technique focuses on psychological aspects, three features stand out for the modeling at hand. First, police interrogate once they are adequately convinced of the guilt. The second is the focus on confrontation. of the accused. Some of the discussion regarding the steps suggests presenting incriminating evidence. Third, this interrogation method's goal is a confession. Throughout Inbau et al. (2013), there are discussions that relate the interrogation exercise to that of a salesperson. So while there is scope for obtaining information from the accused, the confession of someone who is believed to be guilty is a goal.[2]

The scope for false confessions is central to the comparison of these various interrogation methods. Much of the criticism of the Reid technique in the legal psychology literature highlights techniques aimed at inducing confession. Although it receives little attention in the legal psychology literature, the Reid technique is permissive of fabrication of evidence within legal limits, saying:

> While the courts have consistently upheld the interrogator's use of deceptive evidence ploys, the interrogator should exercise great caution in utilizing them. In general courts recognize the practical necessity in allowing such tactics so long as they do not result in involuntary or false confessions.

While this issue is not the main focus of this paper, Section 6 contains a brief discussion of how the model might extend to fabrication of evidence by police for interrogation, which is legally permissible in the U.S.

In addition to the differences regarding psychological manipulation, there are differences in the information disclosed to the accused. The analysis in this paper is aimed at better understanding the effects of these differences. This is done in the context of a game-theoretic

[2]The Reid team notes that it's possible to interrogate someone and later realize they are not guilty, and recommends not pursuing a confession in that instance.

model in which the accused, who knows his guilt or innocence, interacts with a police officer prior to choosing whether to confess for an anticipated lenient sentence, which could simply be due to avoiding the cost and stress of the trial process, or continue on to trial.[3] Before interacting with the accused, the police officer receives some information via the following two channels: 1) she observes whether some evidence against the accused exists, and 2) she receives a private unverifiable signal of the accused's guilt. Neither is observed by the accused and together are not sufficient to prove guilt with certainty. The police officer updates her belief that the accused is guilty, and then chooses whether interrogate. The idea is that she only interrogates if her posterior belief that the accused is guilty is high enough.

The potentially realized evidence could be along the line of a witness identifying the accused or some physical evidence placing the accused at the crime scene. I model the existence of this evidence as increasing the probability that the accused is found guilty should he proceed to trial; that probability depends on both the evidentiary outcome and the private, unverifiable signal.

I compare requiring non-confrontational interrogation to requiring confrontational interrogation. These methods are viewed as being required since the interrogation method used is typically specified by the police department, agency, etc., and some training is provided. For this comparison, I focus on whether the interrogation method is confrontational. There are other methods in addition to the PEACE approach that are non-confrontational, but it is currently the most widely used non-confrontational approach.

Given the focus on false confession in the comparison of these methods, I assume that society prefers for a guilty accused to confess and for an innocent accused to go to trial, and this forms the basis for my comparison. This is due to society's desire to avoid conviction of the innocent or non-conviction of the guilty. This assumption could be motivated by a quadratic loss function for society. In my model, confession leads to a conviction with certainty; going to trial leads to a conviction with a probability that depends on both the evidentiary outcome and the private unverifiable information. While a richer treatment of type 1 and type 2 errors might be attractive, the main result suggests that one does not need to come at the expense of the other. This model suggests that non-confrontational interrogation can be helpful for both inducing guilty to confess and innocent to not confess.

A key intuition is the following. When non-confrontational interrogation is required, the accused is uninformed about the realization of evidence. This effectively closes one channel of

[3]This is discussed further below. Being detained and interrogated is costly for the accused. Avoiding a prolonged trial process certainly represents a cost savings to the accused. Whether explicitly offered by the interrogating police officer, there is a sense in which the accused often anticipates some leniency.

information to the accused. However, the accused still knows that the police officer believes him to be guilty with a high enough probability that she interrogated, which leaves the other channel of information in place. So the accused forms an expected probability of being found guilty at trial and uses this, in combination with an expected reduction in cost/sentence from confessing, to decide whether to confess.

It turns out to be helpful for the accused to be uninformed about the evidentiary outcome. That is, it's helpful for the accused to have some information from being interrogated, but to not be too well informed by also observing the evidentiary outcome. If, instead, use of a confrontational interrogation method is required, the accused knows whether the evidence was realized, and he uses this information to form an expected probability of being found guilty at trial. This can prevent a punishment that induces the guilty accused from confessing when the evidence is not realized that does not also induce the innocent accused from confessing when the evidence is realized. In many of those situations, keeping the accused uninformed about the evidentiary outcome results in being able to induce the guilty to confess while not inducing the innocent to confess. This is because the accused knows that the police officer only interrogates when she has a high enough posterior belief that he is guilty, and uses this in forming his expected probability of being found guilty at trial. This leads to requiring use of non-confrontational interrogation weakly dominating requiring the use of confrontational interrogation

Related Literature

To my knowledge there has been little work, in the economics, and law and economics literatures, focusing on interrogation techniques. Bull (2012) considers interrogation and the fabrication of evidence. Ispano and Vida (2020) and Ispano and Vida (2021) study interrogation using cheap talk models. Their focus is on information disclosure through cheap statements. Although they do not focus on confessions or attempt to specifically model the widely used interrogation methods, they have some similar ideas related to police not revealing too much information early on. They also study commitment by the police. In my model, commitment comes from officers following police department protocol and not from a sense of one officer delegating tasks to another as Ipsano and Vida study.

Much of the literature on settlement has focused on contract settings and other issues. See, for example, Shavell (1989), Shavell (1993), and Rosenberg and Shavell (2006). Work on plea bargaining, such as Grossman and Katz (1983), has focused more on efficiency issues. Recently in the economics literature there has been considerable work addressing questions related to bail decisions, pretrial detention, and diversion. See, for example, Mueller-Smith

and Schnepel (2020), Arnold, Dobbie, and Yang (2018), and Dobbie, Goldin, and Yang (2018).

There is much recent work on evidence that is related. This includes Bull (2008a and 2008b), Bull and Watson (2007), Bull and Watson (2004), Sanchirico (1999, 2000, and 2001), Sanchirico and Triantis (2008). Sanchirico (2010) emphasizes the importance of incorporating fabrication into models of evidence. Bull and Watson (2019) study jury interpretation and the exclusion of evidence at trial in a setting with two channels of information. Although their model focuses on disclosure incentives and a different setting, it is related in there are two channels of information. Fluet and Lanzi (2018), in a different setting, also has two channels of information.

The law literature that has focused on the effects of Miranda rights on police interrogation has touched some upon police lying to suspects. Also the legal psychology literature has investigated issues related to interrogation. Some of these studies have addressed the circumstances that lead to false confessions. Magid (2001) suggests that all interrogation involves some degree of untruthfulness, and discusses limitations on some types of discussion. There is some debate as to whether the evidence on false confessions shows this to be a serious problem. Ofshe and Leo (1997) clearly suggest there are many examples of people who are incarcerated as the result of police-induced false confession. White (2001) suggests that the empirical evidence indicates that pernicious interrogation practices can lead to false confessions to an extent that societal concern is warranted. Further, White notes that the empirical evidence on police-induced false confessions studied by Leo and Ofshe (1998) "seems to suggest that such confessions are mostly likely to occur in high profile cases." This is because police are under greater pressure to solve the crime and devote a lot more time and resources to doing so than they are able to for lesser crimes.

Some studies, such as Weisselberg (2001), focus more on the details of how the Miranda rights are followed regarding questioning of suspects.[4] Kassin, et. al. (2007), based on police survey results, found that 81 percent of suspects waive their Miranda rights. They also reported, "The typical interrogation often, but not always, includes confronting the suspect with evidence of his or her guilt and appealing to his or her self-interests." Additionally, their respondents estimated "that 69.48% of guilty suspects provide a confession" and "23.30% of innocent suspects provide some form of confession." Overall, this line of research is very relevant for the current paper, and faces the problem of the lack of observability of guilt or innocence of the problem in using empirical studies to understand lying by police. The

[4]There is a law and economics literature on the right to silence and the 5th Amendment. See Seidman and Stein (2000), Mialon (2005), and Leshem (2010).

issues studied in these papers help to motivate the model presented here.

The rest of the paper is organized as follows. Section 2 presents the basic model. A numerical example is used to illustrate the basic model and its analysis in Section 3. Analysis of the basic model and the main result are presented in Section 4. Policy implications are discussed in Section 5. A brief discussion of how the main results might apply to evaluating the legally-protected practice of police fabricating some kinds of evidence for use in interrogation is contained in Section 6. Section 7 concludes. The Appendix contains proofs not contained in the text, further calculations for Example 1, and additional illustrative examples.

2 A Simple Model of Interrogation

This section describes the model of interaction between a police officer, denoted by P, and a person accused of a crime, who is denoted by A. I assume that A knows whether he is guilty (G) or innocent (I), but P does not. P observes whether evidence D is realized. This occurs probabilistically and I refer to this as the evidentiary outcome. A does not observe the evidentiary outcome directly. If P chooses to interrogate, the interrogation method in place determines whether P has a choice as to whether to represent to A that the event occurred.

In addition to observing this evidentiary outcome, P also observes a private, non-verifiable signal x. This conveys information about A's guilt to P, but P is unable to convey this to A. However, assume that the information conveyed by x becomes available in the trial process should the case go to trial. The idea is that although P cannot convey x to A during the interrogation, this information comes out at trial. As the trial process is not the focus here, I do not explicitly model it, but instead assume that the outcome is given probabilistically depending on the value of x and the evidentiary outcome.

Based on the realization of x and the evidentiary outcome, P chooses whether to interrogate A. Specifically, P forms an updated belief of A's guilt and if this posterior belief is high enough, P interrogates. This represents that P only interrogates if she thinks the accused is guilty. P choosing not to interrogate means the game ends. If P interrogates and a confrontational interrogation method is in place, P decides whether to represent to A that she has observed D. If D was not realized, announcing that she has observed D requires fabrication.

After being interrogated and observing P's choice, A chooses whether to confess. If A

does not confess, the case goes to trial, and the court finds A guilty with a probability that depends on the private information and the evidentiary outcome. Confessing corresponds to accepting a perceived reduced sentence. The sequence of interaction is represented in the timeline below.

1. A observes his guilt or innocence: $\theta \in \{G, I\}$. P does not observe θ.

2. P observes an evidentiary outcome and observes her private, non-verifiable signal $x \in X$. The evidentiary outcome is denoted by $e \in \{D, \emptyset\} = E$. The realization is according to distribution $f(\theta, e, x)$.

3. P updates for posterior belief A is guilty given by $b(e, x)$.

4. P decides whether to interrogate A. Interrogates when $b(e, x) \geq \underline{b} > r$. If P does not interrogate, the interaction ends. If P interrogates, we continue to 5 below.

5. If P interrogates and is required to use confrontational interrogation, she makes announcement $m \in \{d, n\}$. Under non-confrontational interrogation, P does not make such an announcement. A forms a belief about his probability of being found guilty if he goes to trial, which is given by $\overline{\rho}$. Also, A anticipates cost savings or leniency should he confess instead of going to trial.

6. A chooses whether to confess.

7. If the offer was rejected, they go to trial. The probability A is found guilty at trial is denoted by ρ, where, in general, $\rho : E \times X \to [0, 1]$.

It will sometimes be useful to write the probability of e and x conditional on θ, which is given by the standard conditional-probability formula:

$$f(e, x \mid \theta) \equiv \frac{f(\theta, e, x)}{f(\theta, E, X)}.$$

P's prior probability of G denoted by r is given by $f(G, E, X)$. Assume that P's decision of whether to interrogate is primarily influenced by the information she obtains from the realization of e and x so that $r < 1/2$.

I assume $f(D, X \mid G) > f(D, X \mid I)$, which implies $f(\emptyset, X \mid G) < f(\emptyset, X \mid I)$. So I say D is positive evidence of G and the absence of D, denoted by $\emptyset$, is negative evidence of I (Bull and Watson, 2004). So this allows for $f(\emptyset, x \mid G) > f(\emptyset, x \mid I)$, for some, but not all, x.

For notational simplicity, I assume a monotone likelihood ratio property (MLRP) with respect to x so that

$$\frac{f(e,x \mid G)}{f(e,x \mid I)} > \frac{f(e,x' \mid G)}{f(e,x' \mid I)},$$

for $x > x'$ and any $e \in E$.

Given (e,x), P forms a posterior belief that $\theta = G$ given by

$$b(e,x) = \frac{rf(e,x \mid G)}{rf(e,x \mid G) + [1-r]f(e,x \mid I)},$$

by using Bayes' Rule to update. Note that, due to the MLRP assumption, $b(e,x)$ is increasing in x. P interrogates when $b(e,x) \geq \underline{b} > r$, reflecting that P interrogates when she thinks it's likely enough that A is guilty. So, upon being interrogated, A knows that the realization of (e,x) was such that $b(e,x) > \underline{b}$. A forms a belief about ρ, the probability of being found guilty at trial, which takes into account that the realization of (e,x) was sufficient for P to interrogate.

For simplicity, assume $\rho = b(e,y)$, where $y \in Y$ denotes the signal received by the court. There are various assumptions that we can make on ρ, but reasonable ones would involve ρ being increasing in a belief by the court that A is guilty. How well informed the court is may differ from how well informed P is. This can be represented by either y being more informative, in a Blackwell sense, than x or vice versa. For $Y = X$, we have P and the court being equally informed. I assume this, primarily for simplicity, but also because the focus here is not on differences in how informed the court is relative to P. So $\rho(\theta, e, x) = b(e,x)$. However, it may be interesting to explore the court having a more accurate prior that is less than r.

Welfare

I've assumed that once P interrogates, she wishes to obtain a confession. This is consistent with the typical training police receive and how they are to make the decision to interrogate only when they believe the suspect is guilty and not admitting it. A payoff function for P that yields this is the following.

$$u_P = \begin{cases} 1 & \text{if A confesses} \\ 1/2 & \text{if A does not confess} \end{cases}.$$

The accused is simply concerned with his expected payoff.

I assume that society is concerned with both wrongful convictions and wrongful acquit-

tals. In this model, this translates into a desire for the guilty to confess and the innocent to go to trial. This ensures that the guilty is found guilty with probability 1, and the innocent has a better chance of being found not guilty. However, I acknowledge that this does not focus on the severity of the sentence. Additionally, we could specify society's preferences over type 1 and type 2 errors.

Two Channels of Information

When non-confrontational negotiation is used, A receives no information about e beyond that the realized (e, x) pair was sufficient to lead to interrogation. So we have the following expected value of ρ

$$\overline{\rho}(\theta) = \frac{\sum_{x \in X_D} f(D, x \mid \theta) b(D, x) + \sum_{x \in X_\emptyset} f(\emptyset, x \mid \theta) b(\emptyset, x)}{\sum_{x \in X_D} f(D, x \mid \theta) + \sum_{x \in X_\emptyset} f(\emptyset, x \mid \theta)},$$

where $X_D = \{x \in X \mid b(D, x) > \underline{b}\}$, and $X_\emptyset = \{x \in X \mid b(\emptyset, x) > \underline{b}\}$.

Assuming that fabrication or lying aren't permitted, when confrontational negotiation is used, A learns the realization of e so that he knows whether D was realized, in addition to knowing that (e, x) was sufficient for P to interrogate. So A's expected value of ρ depends on e, and we have

$$\overline{\rho}(\theta, D) = \frac{\sum_{x \in X_D} f(D, x \mid \theta) b(D, x)}{\sum_{x \in X_D} f(D, x \mid \theta)},$$

and

$$\overline{\rho}(\theta, \emptyset) = \frac{\sum_{x \in X_\emptyset} f(\emptyset, x \mid \theta) b(\emptyset, x)}{\sum_{x \in X_\emptyset} f(\emptyset, x \mid \theta)},$$

with X_D and $X_\emptyset$ as above.

Importantly, there are two channels of information for A. The first is that P choosing to interrogate implies the realization of a (e, x) pair such that $b(e, x) \geq \underline{b}$. So A knows that P believes it is likely enough that A is guilty that P interrogates. Regardless of the interrogation method, A knows this.

The second channel is the evidentiary outcome. When confrontational interrogation is used, A is informed of the evidentiary outcome. This may help A to have a better sense of the value of x that was realized. However, when non-confrontational interrogation is used, A is uninformed about e. In both cases, A forms an updated belief, conditional on being interrogated, that each (e, x) has been realized. These are used to calculate A's expected probability of being found guilty at trial.

A's Decision

Assume that, for A, confession leads to payoff R and losing at trial yields payoff H, with $H < R < 0$. I normalize the payoff of winning at trial to 0, and also simplify by ignoring the cost of going to trial. In reality R may be an anticipated reduced sentence. Police often present confessing as leading to leniency, which helps to motivate an accused to confess. This reduction is not modeled as a choice variable for P. [5] The scope for avoiding the cost of trial is another motivation.

When non-confrontational interrogation is used, A, who is assumed to be risk neutral for simplicity, confesses when $\overline{\rho}(\theta)H < R$. So A, of types G and I, respectively confesses when $\overline{\rho}(G)H < R$, and $\overline{\rho}(I)H < R$. Similarly, when confrontational interrogation is used, A of type θ confesses when $\overline{\rho}(\theta, e)H < R$.

3 Numerical Example

To get some intuition about the model and to see how non-confrontational interrogation can be helpful, let's consider a numerical example. Further calculations for this example are in the Appendix.

Example 1

Let $X = \{x_1, x_2\}$. Assume that $\underline{b} = \frac{2}{3}$.

Assume the joint probabilities are the following:

$$f(G, D, x_1) = \frac{1}{16}, f(G, D, x_2) = \frac{3}{32}, f(I, D, x_1) = \frac{1}{32}, f(I, D, x_2) = \frac{1}{32},$$

$$f(G, \emptyset, x_1) = \frac{1}{16}, f(G, \emptyset, x_2) = \frac{1}{8}, f(I, \emptyset, x_1) = \frac{17}{32}, f(I, \emptyset, x_2) = \frac{1}{16}.$$

These imply the prior $f(G, E, X) = \frac{11}{32} < \frac{21}{32} = f(I, E, X)$, and imply the following

[5] Even if the police imply a reduction, they typically would not have a lot of flexibility to set it. Often anticipated leniency can be a motivation for a suspect to confess. See, for example Magid (2001), and Ofshe and Leo (1997). Further, it has been suggested that the "minimization" technique, which involves the interrogator seeking to minimize the crime by providing scenarios that lend moral justification for the crime, can lead to the belief that leniency will follow even if leniency was not explicitly promised. See Kassin and McNall (1991). Inbau et al. (2013) discuss promises of leniency that are not considered coercion. These include promises to recommend a low bail or to say the suspect cooperated with investigators (p. 424). Importantly, they also state (p. 369), "As with threats, although to a lesser degree, a promise of leniency may have the effect of inducing an innocent person to confess." They also state, "To allow a suspect to believe that it may be beneficial if he tells the truth, the Reid Technique takes advantage of one of the fundamental principles of human nature..." In *United States v. Harris* the court stated, "Police are free to solicit confessions by offering to reduce the charges, so an offer of leniency was not coercion."

likelihood ratios:

$$\frac{f(D,x_1 \mid G)}{f(D,x_1 \mid I)} = \frac{42}{11}, \frac{f(D,x_2 \mid G)}{f(D,x_2 \mid I)} = \frac{63}{11},$$

$$\frac{f(\emptyset,x_1 \mid G)}{f(\emptyset,x_1 \mid I)} = \frac{42}{187}, \frac{f(\emptyset,x_2 \mid G)}{f(\emptyset,x_2 \mid I)} = \frac{84}{187}.$$

It's straightforward to verify that these satisfy the assumptions $f(D,x \mid G) > f(D,x \mid I)$, for all $x \in X$ and

$$\frac{f(e,x \mid G)}{f(e,x \mid I)} > \frac{f(e,x' \mid G)}{f(e,x' \mid I)},$$

for $x > x'$ and any $e \in E$.

We find the following posterior beliefs for P.

$$b(D,x_1) = \frac{2}{3}, b(D,x_2) = \frac{3}{4}, b(\emptyset,x_1) = \frac{2}{19}, b(\emptyset,x_2) = \frac{2}{3}.$$

Upon being interrogated, A forms an updated belief of ρ.

When non-confrontational interrogation is used, A receives no information about e. So he updates based on the values of e and x that would result in $b(e,x) \geq \underline{b}$ so that P chooses to interrogate.

Thus,

$$\overline{\rho}(\theta) = \frac{\sum_{x \geq x_D} f(D,x \mid \theta) b(D,x) + \sum_{x \geq x_\emptyset} f(\emptyset,x \mid \theta) b(\emptyset,x)}{\sum_{x \geq x_D} f(D,x \mid \theta) + \sum_{x \geq x_\emptyset} f(\emptyset,x \mid \theta)},$$

where x_D and $x_\emptyset$ denote the values of x for which $b(e,x) \geq \underline{b}$ for $e = D, \emptyset$, respectively.

So here, since $\underline{b} = 2/3$, $X_D = x_1, \{x_2,\}$ and $X_\emptyset = \{x_2\}$. That is, P interrogates when $(e,x) \in \{(D,x_1),(D,x_2),(\emptyset,x_2)\}$.

So we have

$$\overline{\rho}(G) = \frac{75}{108} \approx .694, \overline{\rho}(I) = \frac{11}{16} \approx .6875,$$

$$\overline{\rho}(G,D) = \frac{43}{60} \approx .7167, \overline{\rho}(I,D) = \frac{17}{24} \approx .70833,$$

and

$$\overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset) = \frac{2}{3}.$$

Note that $\overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset)$ since when D is not realized, A knows this and knows that $(\emptyset,x_2)$ has been realized. A's expectation of ρ is higher when he knows that D was realized.

When non-confrontational interrogation is used, A confesses when $\overline{\rho}(\theta)H < R$. So A, of types G and I, respectively confesses when $\overline{\rho}(G)H \approx .694H < R$ and $\overline{\rho}(I)H \approx .6875H < R$. Thus, for R such that $.694H < R < .6875H$, type G will confess but type I will not.

Similarly, when non-confrontational interrogation is used, A confesses when $\overline{\rho}(\theta, e)H < R$. Note that here $\overline{\rho}(G, D) = \frac{43}{60} \approx .7167 > \overline{\rho}(I, D) = \frac{17}{24} \approx .70833 > \overline{\rho}(G, \emptyset) = \frac{2}{3}$. So when fabrication is not allowed, it's not possible to induce G to confess when $\emptyset$ is realized and to also induce I to not confess when D is realized.

Non-confrontational interrogation results in A being uninformed of the evidentiary outcome. This allows G to be induced to confess while I chooses to go to trial. However, when A observes the evidentiary outcome, sequentially rational behavior can lead to an undesirable outcome. This is the main idea that is developed in the next section.

4 Main Results

Theorem 1 shows there is some ordering of the expect probabilities at trial $\overline{\rho}$. This is useful for describing and analyzing the different cases in extending the intuition of Example 1 more generally. Theorem 2 shows that it can be welfare improving for the accused to be not too well informed is welfare improving, which can be accomplished through use of non-confrontational interrogation. I address the practical limitations of this below.

Consider the ordering of $\overline{\rho}$, holding e constant.

Theorem 1: *The following hold: 1)* $\overline{\rho}(G, D) \geq \overline{\rho}(I, D)$*, 2)* $\overline{\rho}(G, \emptyset) \geq \overline{\rho}(I, \emptyset)$*, and 3)* $\overline{\rho}(G) \geq \overline{\rho}(I)$*. Further 1) and 3) hold strictly when* X_D *is non-singleton, and 2) holds strictly when* $X_\emptyset$ *is non-singleton.*

The issue when X_D or $X_\emptyset$ are singletons is there is only one possible (e, x) combination for each. So then the expected probability at trial is $b(e, x)$ for either type G or I. In the other cases, $\overline{\rho}(G, D) > \overline{\rho}(I, D)$ since the posterior beliefs $b(\theta, x)$ are increasing in x and type G has more relative weight on higher xs I.

For the consideration of non-confrontational interrogation to be interesting requires $\overline{\rho}(G, D) > \overline{\rho}(G, \emptyset)$. Otherwise, there wouldn't be sense in which the police could confront the accused with incriminating evidence. Although, I limit attention to this case for the main comparison, which is described below, the assumptions prior to this one do not rule out the opposite.[6]

[6] An assumption on the evidence environment that generates $\overline{\rho}(G, D) > \overline{\rho}(G, \emptyset)$. is not obvious. The assumption that $f(D, x \mid G) > f(D, x \mid I)$, for all $x \in X$, does not generate it. Adding the similar assumption for evidence not being realized of $f(\emptyset, x \mid G) < f(\emptyset, x \mid I)$, for all $x \in X$, is too strong. So I instead directly assume $\overline{\rho}(G, D) > \overline{\rho}(G, \emptyset)$.

However, these other evidence environments suggest a different modeling of police confronting the accused with incriminating evidence during interrogation. The Appendix contains some further examples.[7]

Theorem 2: *Non-confrontational interrogation weakly dominates confrontational interrogation. That is, non-confrontational interrogation is welfare improving, in that it induces the guilty to confess and the innocent to go to trial, in a many cases and does no worse in the others.*

The intuition for when non-confrontational interrogation does better is the following. When confrontational interrogation is used, A knows whether the evidence was realized. Additionally, A also knows that, given P's private signal and the evidentiary outcome, P's belief that A is guilty surpasses the bar for interrogation. However, with non-confrontational interrogation, A only knows, that P's belief of guilt was high enough that she chose to interrogate. When uninformed about the evidentiary outcome, the guilty A puts more probability weight on the evidence having been realized than does the innocent. However, when the guilty A sees that the evidence *has not* been realized, his expected payoff from not confessing is higher than when he is uninformed. Similarly, when the innocent A sees that the evidence *has* been realized, his expected payoff from confessing is higher This result illustrates the importance of the two channels of information and the restriction of the evidentiary outcome channel. In particular, since A knows the realized (e, x) was sufficiently high for P to interrogate, but is not informed of the evidentiary outcome.

Cases to Consider

As noted above, for there to be sense in which the police could confront the accused with incriminating evidence requires $\overline{\rho}(G, D) > \overline{\rho}(G, \emptyset)$, which is assumed. One may wonder whether it's possible to have $\overline{\rho}(I, D) < \overline{\rho}(I, \emptyset)$ when $\overline{\rho}(G, D) > \overline{\rho}(G, \emptyset)$. Theorem 1 shows that in this case the issue of which interrogation method to use only matters when $\overline{\rho}(G, \emptyset) = \overline{\rho}(I, \emptyset)$. In order for this to hold, requires the following ordering

$$(1) \qquad \overline{\rho}(G, D) > \overline{\rho}(G, \emptyset) \geq \overline{\rho}(I, \emptyset) > \overline{\rho}(I, D).$$

With $\overline{\rho}(G, \emptyset)$ strictly larger than $\overline{\rho}(I, \emptyset)$, it's possible to set R so that both types of the guilty accused confess and neither type of the innocent does. However, when $\overline{\rho}(G, D) >$

[7]These include cases in which A prefers the realization of D to that of $\emptyset$. Example 3 in the Appendix shows that it's possible to have both $\overline{\rho}(G, D) < \overline{\rho}(G, \emptyset)$ and $\overline{\rho}(I, D) < \overline{\rho}(I, \emptyset)$. Again, this does not fit with the idea of confronting A with incriminating evidence. It's also possible to have these in opposite directions. Example 2 in the Appendix provides an example of this, which fits with Lemma 1 below.

$\overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset) > \overline{\rho}(I,D)$, confrontational interrogation presents the tension of an R that induces the guilty to confess when $\emptyset$ is realized also induces the innocent to confess when D is realized. When non-confrontational interrogation is used, we can have R such that the guilty confesses but the innocent does not. This is because we have $\overline{\rho}(G) > \overline{\rho}(I)$.

Lemma 1: *When $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset) > \overline{\rho}(I,D)$, using confrontational interrogation can lead to the guilty confessing and the innocent not confessing. When $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset) > \overline{\rho}(I,\emptyset) > \overline{\rho}(I,D)$, using non-confrontational interrogation can both lead to the guilty confessing and the innocent not.*

Now let's consider the case where $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset)$ and $\overline{\rho}(I,D) > \overline{\rho}(I,\emptyset)$. Here, Theorem 1 implies that we have $\overline{\rho}(G,D) \geq \overline{\rho}(I,D)$ and $\overline{\rho}(G,\emptyset) \geq \overline{\rho}(I,\emptyset)$, with these holding strictly for non-singleton X_D and $X_\emptyset$, respectively. As noted previously, from a practical viewpoint, I would expect X to be quite rich and there to be multiple xs that in combination with both D and $\emptyset$ would lead to interrogation. So, despite using numerical examples with few xs, I view the singleton cases as more of a technical issue.

There are two possible orderings:

(2) $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset) \geq \overline{\rho}(I,D) > \overline{\rho}(I,\emptyset)$.

(3) $\overline{\rho}(G,D) \geq \overline{\rho}(I,D) > \overline{\rho}(G,\emptyset) \geq \overline{\rho}(I,\emptyset)$

For ordering (2), with $\overline{\rho}(G,\emptyset) > \overline{\rho}(I,D)$, use of confrontational or non-confrontational interrogation is not an issue (as in the previous case) because R can be set so that the guilty confesses and the innocent does not. However, when we have $\overline{\rho}(G,\emptyset) = \overline{\rho}(I,D)$, use of confrontational interrogation has the same tension as above in that of an R that induces the guilty to confess when $\emptyset$ is realized also induces the innocent to confess when D is realized. If, instead, when non-confrontational interrogation is required, we have $\overline{\rho}(G) > \overline{\rho}(I,)$, and the guilty can be induced to confess while the innocent is not.

Lemma 2: *When $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset) > \overline{\rho}(I,D) > \overline{\rho}(I,\emptyset)$, either method of interrogation can both lead to the guilty confessing and the innocent not confessing. When $\overline{\rho}(G,D) > \overline{\rho}(G,\emptyset) = \overline{\rho}(I,D) > \overline{\rho}(I,\emptyset)$, requiring non-confrontational interrogation can lead to the guilty confessing and the innocent not. However, requiring confrontational interrogation prevents inducing the guilty to confess when $\emptyset$ is realized and also inducing the innocent to not confess when D is realized.*

Next consider ordering (3). If either $\overline{\rho}(G,D) > \overline{\rho}(I,D)$ or $\overline{\rho}(G,\emptyset) > \overline{\rho}(I,\emptyset)$, then requiring non-confrontational interrogation does better by same method as in the other cases.

Since this will result in A being uninformed and we have $\overline{\rho}(G) > \overline{\rho}(D)$, there is an R that will induce the guilty to confess without inducing the innocent to do so.

However, if we have $\overline{\rho}(G, D) = \overline{\rho}(I, D) > \overline{\rho}(G, \emptyset) = \overline{\rho}(I, \emptyset)$, neither interrogation method being required does better.

Lemma 3: *When either* $\overline{\rho}(G, D) > \overline{\rho}(I, D) > \overline{\rho}(G, \emptyset) \geq \overline{\rho}(I, \emptyset)$ *or* $\overline{\rho}(G, D) \geq \overline{\rho}(I, D) > \overline{\rho}(G, \emptyset) > \overline{\rho}(I, \emptyset)$ *(or both) hold, we have* $\overline{\rho}(G) > \overline{\rho}(D)$ *and requiring non-confrontational interrogation can lead to the guilty confessing and the innocent not confessing. However, if* $\overline{\rho}(G, D) = \overline{\rho}(I, D) > \overline{\rho}(G, \emptyset) = \overline{\rho}(I, \emptyset)$*, requiring either interrogation method can lead to both the guilty confessing and the innocent not.*

This covers all of the cases where D is negative for A. In this setting, requiring non-confrontational interrogation is welfare improving in a many cases and does no worse in the others.

5 Policy Implications

The results suggest that a non-confrontational approach to interrogation can out perform a confrontational one in terms of reducing false confessions and inducing the guilty to confess. This is because a non-confrontational approach conveys some information to the accused about the evidence against him, but does not provide too precise of information.

This relies on two key features of how police officers behave: 1) they only interrogate when they are adequately convinced that the accused is guilty, and 2) they follow their department's interrogation protocol. Generally, these seem reasonable. However, as recent events show, there are situations where some police officers have not followed proper protocol. In what are hopefully very rare instances there have been entire departments that do not follow proper protocol. Anecdotal evidence suggest such cases are rare. However, it police officers detain and interrogate people without sufficient evidence to suggest the accused is guilty, this analysis would not hold. Gladwell (2019) contains a thorough discussion of these sort of issues related to the police department in Ferguson, Missouri.

Police officers following department protocol (as they are trained to) can be viewed as a way for police to commit to certain behavior. While it may be difficult for an individual officer to commit to not confront an accused person with incriminating evidence, it is possible for the department to do so. Based on conversations with police officers, it seems likely that the majority of officers follow the guidelines they receive in training. Of course, there is also the issue of ensuring that officers who interrogate have received adequate training.

6 A Related Interrogation Technique: Fabrication

As noted above, police lying during interrogation is legally protected in many states.[8] One particularly troubling example occurred in 2002 and was reported by the Los Angeles Times.[9] Jose "Peps" Ledesma, a member of the Vineland Boyz, a violent gang that controlled most of the illegal drug sales in its area, was accused of the murder of Christian Vargas. He was accused of shooting Vargas outside of 16-year old Martha Puebla's home.

Briefly, the major events of the murder and investigation occurred as follows. A teenage girl, who was a friend of Puebla, and Vargas parked in front of Puebla's home around 2 a.m. on November 27, 2003. The teenage girl, who was not named in the LA Times for her safety, went to Puebla's ground-floor window while Vargas sat in the car. While the girl was talking to Puebla, Vargas was shot and killed.

Ledesma was identified as a suspect. Ledesma and another Vineland Boyz member Mario Catalan were interrogated by Los Angeles Police Department detectives Martin Pinner and Juan Rodriguez. When he was interrogated, Ledesma did not call an attorney, and repeatedly denied involvement in the shooting. Pinner told Ledesma that they had multiple witnesses who would testify that he was the shooter. He then showed Ledesma a "six-pack," which is a binder of photos of possible suspects, with his photo circled,"M.P." written below it, along with "this is the guy who shot my friends boyfriend," and Martha's signature. It was a total fabrication. Martha had actually been unhelpful with the police throughout their investigation and when testifying at the preliminary hearing for Ledesma's murder trial.

Ledesma used a pay phone near his cell to call another Vineland Boyz member Javier Covarrubias, and discussed Martha, saying, "I need her to disappear." Several days after Martha Puebla testified at the preliminary hearing for Ledesma's murder trial, she was shot and killed, allegedly, by a member of the Vineland Boyz. Neither Martha nor her family had been made aware of her having potentially been put in harm's way by the interrogation techniques used by Pinner and Rodriguez.

In 2008, as part of a federal plea bargain to avoid the death penalty, Ledesma, Cavarrubias, and a Vineland Boyz member believed to be the shooter pleaded guilty to the murder of Puebla. However, before federal prosecutors and the LAPD had made progress on Puebla's murder case, Pinner and Rodriguez arrested Juan Catalan, Mario Catalan's brother, and interrogated him. Pinner and Rodriguez told Juan Catalan there were witnesses to him

[8] See, for example, Magid (2001) and Rubin and Bloomekatz (2008).

[9] See Rubin and Bloomekatz (2008) and the transcript of testimony by LAPD Detective Martin Pinner during the murder trial of Jose Ledesma.

shooting Puebla and showed him six-packs with his photo circled and witness signatures. These were all fabricated. The case against Juan Catalan was eventually dismissed when his attorney was able to acquire video showing Catalan at an LA Dodgers game at the time of the shooting. The "ruse" used by the LAPD detectives was legal. Further, state and federal courts in the U.S. have repeatedly upheld the right of police to lie to people they have in custody.

There is a sense in which the result here could apply to allowing police to fabricate evidence that is presented to the accused. We can consider the evidentiary outcome of D being fabricated. If the police officer always fabricates, the accused is uninformed about e. Recall that P only interrogates when she is convinced, based on her posterior belief, that A is guilty. In this case, it would be reasonable for P to want a confession.

Although extending Theorem 2 to the current practice of allowing police to lie or fabricate evidence during interrogation is possible, much caution is needed in practice. I suggest that it's not an appropriate extension for a variety of reasons. The analysis suggests that we, as a society, ought to be concerned with ensuring that those who are accused understand the institution when police are allowed to lie/fabricate. Anecdotal evidence suggests that many people are not aware that police are legally allowed to lie/fabricate during interrogation.[10] Certainly, it seems in the above case involving Jose "Peps" Ledesma the accused did not expect that police fabricated.

One possibility might be to require that the scope for police lying/fabricating be conveyed in a similar manner to "Miranda rights." However, it is well known that despite being informed of their right to silence and right to an attorney, many do not use these. So it may be that those being interrogated would not internalize this fact. It also seems there are reasons that police may not wish to inform the accused of this. This may make it less likely for an accused to believe police on other matters, and may lead to a distrust of police.

Additionally, the analysis of the model suggests limitations on when this result holds. It seems many police do not lie or fabricate when interrogating. In fact, the typical training discourages this. Further, it seems that many who become police officers value honesty. This may give them a very high cost of lying or fabricating.

However, some recent events suggest that not all police have these values.[11] While this may often be a small number, there are examples where this behavior is part of the

[10] In discussing this paper, it seemed many I spoke with were surprised to learn this.

[11] For example, Lamar Ferguson filed a lawsuit against the City of Minneapolis alleging, along with numerous other things, that an officer "falsely stated that there was a warrant out for" Ferguson's arrest. In 2017, the city paid $25,000 to settle. Andone, Silverman, and Alonso (2020) describes this.

culture. There can be societies or municipalities where police practice aggressive policing with everyone, or at least people of certain groups. It has been suggested that Ferguson, Missouri was such a place.[12] While there have been many studies suggesting that crime tends to be concentrated in extremely small geographic areas within cities known as crime "hot spots" and aggressive policing in those small areas is effective, aggressive policing everywhere is not desirable.[13] In areas where aggressive policing is used despite being unwarranted, there can be bias and interrogation despite a low posterior. This danger may also be a reason to not allow fabrication.

At a minimum the analysis here, and the complexity in practice, suggest a need for careful policies on police lying and fabrication. They also suggest a need for greater awareness for citizens, and more thorough training for police. I believe this should also include more clear guidelines for police.

7 Conclusion

This paper presented a simple model of interrogation in which a police officer potentially has the opportunity to convey information to the accused about whether the police have some evidence. The primary focus was on how requiring non-confrontational interrogation, which limits the information the accused receives, affects the behavior of the accused in choosing whether to confess. In the stylized setting, ideally, the guilty accused would confess and the innocent accused would not confess and go to trial.

Since, in this model, the police officer interrogates only if her posterior belief of guilt is high enough, there is a second information channel, which conveys some information about the police officer's private unverifiable information. Requiring use of a non-confrontational interrogation method closes the first channel of information and causes the accused to be uninformed about the evidentiary outcome, other than through the second information channel. This provides enough, but not too much information, to the accused. Theorem 2 shows that, in this setting, requiring non-confrontational interrogation weakly dominates requiring confrontational interrogation on a welfare basis.

The analysis presented has focused on the information aspects of stylized interrogation methods. While there are other compelling arguments for the use of non-confrontational interrogation methods, this provides a reason to favor non-confrontational methods that does not rely on cognitive limitations of the accused. Requiring the use of non-confrontational

[12] Gladwell (2019) contains a thorough discussion of this.

[13] See, for example, Sherman and Weisburd (1995) and Weisburd (2015).

interrogation can provide a way to avoid the misalignment of incentives of a police officer who thinks it's quite likely that the accused is guilty and those of society.

A Proofs

Theorem 1: *The following hold: 1)* $\overline{\rho}(G, D) \geq \overline{\rho}(I, D)$*, 2)* $\overline{\rho}(G, \emptyset) \geq \overline{\rho}(I, \emptyset)$*, and 3)* $\overline{\rho}(G) \geq \overline{\rho}(I)$*. Further 1) and 3) hold strictly when* X_D *is non-singleton, and 2) holds strictly when* $X_\emptyset$ *is non-singleton.*

Proof: Note that when X_D contains only a single x, $\overline{\rho}(G, D) = \overline{\rho}(I, D)$. This is because P interrogates when D and the highest x are realized, but no other xs. Similarly, when $X_\emptyset$ contains a single x, $\overline{\rho}(G, \emptyset) = \overline{\rho}(I, \emptyset)$.

Next consider the cases where the X_D and $X_\emptyset$ sets contain multiple xs.

First we show

$$\overline{\rho}(G, D) = \frac{\sum_{x \in X_D} f(D, x \mid G) b(D, x)}{\sum_{x \in X_D} f(D, x \mid G)} > \frac{\sum_{x \in X_D} f(D, x \mid I) b(D, x)}{\sum_{x \in X_D} f(D, x \mid I)} = \overline{\rho}(I, D).$$

Rewrite this as

$$\frac{\sum_{x \in X_D} f(D, x \mid G) b(D, x)}{\sum_{x \in X_D} f(D, x \mid I) b(D, x)} > \frac{\sum_{x \in X_D} f(D, x \mid G)}{\sum_{x \in X_D} f(D, x \mid I)}.$$

Since (D, x) for $x \in X_D$ results in $b(D, x) > r$, which is necessary for it to lead to interrogation, it must be that $f(D, x \mid G) > f(D, x \mid I)$ for all $x \in X_D$. We also have that $b(D, x)$ is increasing in x. Given these properties, the inequality holds.

Next we show, in a similar manner,

$$\overline{\rho}(G, \emptyset) = \frac{\sum_{x \in X_\emptyset} f(\emptyset, x \mid G) b(\emptyset, x)}{\sum_{x \in X_\emptyset} f(\emptyset, x \mid G)} > \frac{\sum_{x \in X_\emptyset} f(\emptyset, x \mid I) b(\emptyset, x)}{\sum_{x \in X_\emptyset} f(\emptyset, x \mid I)} = \overline{\rho}(I, \emptyset).$$

Rewrite this as

$$\frac{\sum_{x \in X_\emptyset} f(\emptyset, x \mid G) b(\emptyset, x)}{\sum_{x \in X_\emptyset} f(\emptyset, x \mid I) b(\emptyset, x)} > \frac{\sum_{x \in X_\emptyset} f(\emptyset, x \mid G)}{\sum_{x \in X_\emptyset} f(\emptyset, x \mid I)}.$$

Since $(\emptyset, x)$ for $x \in X_\emptyset$ results in $b(\emptyset, x) > r$, which is necessary for it to lead to interrogation, it must be that $f(\emptyset, x \mid G) > f(\emptyset, x \mid I)$ for all $x \in X_\emptyset$. We also have that $b(\emptyset, x)$ is increasing in x. Given these properties, the inequality holds.

Lastly, we show that

$$\overline{\rho}(G) = \frac{\sum_{x \in X_D} f(D, x \mid G) b(D, x) + \sum_{x \in X_\emptyset} f(\emptyset, x \mid G) b(\emptyset, x)}{\sum_{x \in X_D} f(D, x \mid G) + \sum_{x \in X_\emptyset} f(\emptyset, x \mid G)}$$

$$> \frac{\sum_{x\in X_D} f(D,x \mid I)b(D,x) + \sum_{x\in X_\emptyset} f(\emptyset,x \mid I)b(\emptyset,x)}{\sum_{x\in X_D} f(D,x \mid I) + \sum_{x\in X_\emptyset} f(\emptyset,x \mid I)} = \overline{\rho}(I).$$

Note that $\overline{\rho}(I,\emptyset)$ is a convex combination of $\overline{\rho}(G,D)$ and $\overline{\rho}(G,\emptyset)$, and $\overline{\rho}(I)$ is a convex combination of $\overline{\rho}(I,D)$ and $\overline{\rho}(I,\emptyset)$. So we have $\overline{\rho}(G) > \overline{\rho}(I)$. □

B Calculations for Example 1

Some details of Example 1 omitted from the text are below.

The joint probabilities imply the following conditional probabilities:

$$f(D,x_1 \mid G) = \frac{2}{11}, f(D,x_2 \mid G) = \frac{3}{11}, f(D,x_1 \mid I) = \frac{1}{21}, f(D,x_2 \mid I) = \frac{1}{21},$$

$$f(\emptyset,x_1 \mid G) = \frac{2}{11}, f(\emptyset,x_2 \mid G) = \frac{4}{11}, f(\emptyset,x_1 \mid I) = \frac{17}{21}, f(\emptyset,x_2 \mid I) = \frac{2}{21}.$$

These yield the following likelihood ratios:

$$\frac{f(D,x_1 \mid G)}{f(D,x_1 \mid I)} = \frac{42}{11}, \frac{f(D,x_2 \mid G)}{f(D,x_2 \mid I)} = \frac{63}{11}, \frac{f(\emptyset,x_1 \mid G)}{f(\emptyset,x_1 \mid I)} = \frac{42}{187}, \frac{f(\emptyset,x_2 \mid G)}{f(\emptyset,x_2 \mid I)} = \frac{84}{187}.$$

It's straightforward to verify that these satisfy the assumptions $f(D,x \mid G) > f(D,x \mid I)$, for all $x \in X$ and

$$\frac{f(e,x \mid G)}{f(e,x \mid I)} > \frac{f(e,x' \mid G)}{f(e,x' \mid I)},$$

for $x > x'$ and any $e \in E$.

Using Bayes' Rule to update. We find the following posterior beliefs for P.

$$b(D,x_1) = \frac{2}{3}, b(D,x_2) = \frac{3}{4}, b(\emptyset,x_1) = \frac{2}{19}, b(\emptyset,x_2) = \frac{2}{3}.$$

We have

$$\overline{\rho}(\theta) = \frac{\sum_{x\geq x_D} f(D,x \mid \theta)b(D,x) + \sum_{x\geq x_\emptyset} f(\emptyset,x \mid \theta)b(\emptyset,x)}{\sum_{x\geq x_D} f(D,x \mid \theta) + \sum_{x\geq x_\emptyset} f(\emptyset,x \mid \theta)},$$

For our example, assume that $\underline{b} = \frac{2}{3}$. So here $X_D = x_1, \{x_2,\}$ and $X_\emptyset = \{x_2\}$. That is, P interrogates when $(e,x) \in \{(D,x_1),(D,x_2),(\emptyset,x_2)\}$.

We have

$$\overline{\rho}(G) = \frac{\frac{2}{11}\frac{2}{3} + \frac{3}{11}\frac{3}{4} + \frac{4}{11}\frac{2}{3}}{\frac{2}{11} + \frac{3}{11} + \frac{4}{11}} = \frac{75}{108} \approx .694.$$

and

$$\overline{\rho}(I) = \frac{\frac{1}{21}\frac{2}{3} + \frac{1}{21}\frac{3}{4} + \frac{2}{21}\frac{2}{3}}{\frac{1}{21} + \frac{1}{21} + \frac{4}{21}} = \frac{11}{16} \approx .6875.$$

Next consider the case a confrontational method is used. Presume, as above, that P interrogates when $b(e,x) \geq \frac{2}{3}$. Here, A observes whether P has received D. So A knows whether $(e,x) \in \{(D,x_1),(D,x_2)\}$ or $(e,x) = (\emptyset, x_2)$. To denote A's type and whether the document is observed, write $\overline{\rho}(\theta, e)$. We have

$$\overline{\rho}(G,D) = \frac{\frac{2}{11}\frac{2}{3} + \frac{3}{11}\frac{2}{3}}{\frac{2}{11}+\frac{3}{11}} = \frac{43}{60} \approx .7167, \overline{\rho}(I,D) = \frac{\frac{1}{21}\frac{2}{3} + \frac{1}{21}\frac{3}{4}}{\frac{1}{21}+\frac{1}{21}} = \frac{17}{24} \approx .70833,$$

and

$$\overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset) = \frac{2}{3}.$$

Note that $\overline{\rho}(G,\emptyset) = \overline{\rho}(I,\emptyset)$ since when D is not realized, A knows this and knows that $(\emptyset, x_2)$ has been realized. A's expectation of ρ is higher when he knows that D was realized.

C When $\emptyset$ can Suggest Guilt

This Section presents two examples that are outside the class of environments studied in the main text. It's possible that for some realizations of x the evidentiary outcome $\emptyset$ is worse for A than is D.

I have assumed $f(D,x \mid G) > f(D,x \mid I)$, for all $x \in X$, but this does not prohibit, for some x (but not all) $f(\emptyset,x \mid G) > f(\emptyset,x \mid I)$. Clearly, this latter condition cannot hold for all xs or $f(G,X \mid G) < f(\emptyset,X \mid G)$, which is implied by $f(D,x \mid G) > f(D,x \mid I)$, for all $x \in X$, would not hold.

Example 2

It's possible to have the effect of D to be beneficial for one type of A and detrimental for the other type, in terms of expected probability of being found guilty at trial. The example below shows that's it possible to have $\overline{\rho}(G,D) < \overline{\rho}(G,\emptyset)$ but $\overline{\rho}(I,D) > \overline{\rho}(I,\emptyset)$.

Let $X = \{x_1,x_2,x_3\}$. Suppose $r = f(G,E,X) = \frac{15}{32}$, and assume $\underline{b} = .65$.

The joint probabilities are the following:

$$f(G,D,x_1) = \frac{1}{16}, f(G,D,x_2) = \frac{3}{32}, f(G,D,x_3) = \frac{3}{64},$$

$$f(I,D,x_1) = \frac{1}{32}, f(I,D,x_2) = \frac{1}{32}, f(I,D,x_3) = \frac{1}{32},$$

$$f(G,\emptyset,x_1) = \frac{1}{64}, f(G,\emptyset,x_2) = \frac{1}{16}, f(G,\emptyset,x_3) = \frac{5}{128},$$

$$f(I,\emptyset,x_1) = \frac{5}{16}, f(I,\emptyset,x_2) = \frac{1}{32}, f(I,\emptyset,x_3) = \frac{1}{32}.$$

These imply the following conditional probabilities:

$$f(D,x_1 \mid G) \approx 0.1359, f(D,x_2 \mid G) \approx 0.2034, f(D,x_3 \mid G) \approx 0.2373,$$

$$f(D, x_1 \mid I) \approx 0.0580, f(D, x_2 \mid I) \approx 0.0580, f(D, x_3 \mid I) \approx 0.0580,$$

$$f(\emptyset, x_1 \mid G) \approx 0.0339, f(\emptyset, x_2 \mid G) \approx 0.1356, f(\emptyset, x_3 \mid G) \approx 0.2542,$$

$$f(\emptyset, x_1 \mid I) \approx 0.7101, f(\emptyset, x_2 \mid I) \approx 0.0580, f(\emptyset, x_3 \mid I) \approx 0.0580.$$

These imply, using Bayes' Rule to update, the following posterior beliefs for P.

$$b(D, x_1) \approx 0.6667, b(D, x_2) = 0.75, b(D, x_3) \approx 0.7778,$$

$$b(\emptyset, x_1) \approx 0.0392, b(\emptyset, x_2) \approx 0.6667, b(\emptyset, x_3) \approx 0.7895.$$

These yield

$$\overline{\rho}(G) \approx .6926,$$
$$\overline{\rho}(I) \approx .6915,$$

$$\overline{\rho}(G, D) \approx .7418,$$
$$\overline{\rho}(I, D) \approx .7315,$$

and

$$\overline{\rho}(G, \emptyset) \approx 0.7468,$$

$$\overline{\rho}(I, \emptyset) \approx .7281.$$

Example 3

The following example shows that, generally, it's possible to have $\overline{\rho}(G, D) < \overline{\rho}(G, \emptyset)$ and $\overline{\rho}(I, D) < \overline{\rho}(I, \emptyset)$ even when we have $f(D, x \mid G) > f(D, x \mid I)$, for all x.

Let $X = \{x_1, x_2\}$.

The joint probabilities are the following:

$$f(G, D, x_1) = 0.1, f(G, D, x_2) = 0.05, f(I, D, x_1) = 0.05, f(I, D, x_2) = 0.1,$$

$$f(G, \emptyset, x_1) = 0.06, f(G, \emptyset, x_2) = 0.04, f(I, \emptyset, x_1) = 0.68, f(I, \emptyset, x_2) = 0.01.$$

These imply the prior $f(G, E, X) = 0.25 < 0.75 = f(I, E, X)$, and imply the following conditional probabilities:

$$f(D, x_1 \mid G) = 0.4, f(D, x_2 \mid G) = 0.2, f(D, x_1 \mid I) \approx 0.0667, f(D, x_2 \mid I) \approx 0.01333$$

$$f(\emptyset, x_1 \mid G) = 0.24, f(\emptyset, x_2 \mid G) = 0.16, f(\emptyset, x_1 \mid I) \approx 0.90667, f(\emptyset, x_2 \mid I) \approx 0.01333.$$

Using Bayes' Rule to update yields the following posterior beliefs for P.

$$b(D, x_1) = \frac{2}{3}, b(D, x_2) \approx 0.8333,$$

$$b(\emptyset, x_1) \approx 0.0811, b(\emptyset, x_2) \approx 0.80.$$

These yield the following expected probabilities of being found guilty at trial.

$$\overline{\rho}(G, D) \approx .7222, \overline{\rho}(I, D) \approx 0.6944$$

$$\overline{\rho}(G, \emptyset) = 0.8, \overline{\rho}(I, \emptyset) \approx 0.799\overline{9},$$

$$\overline{\rho}(G) \approx .7386, \overline{\rho}(I) \approx 0.7095$$

References

Andone, Dakin, Hollie Silverman, and Melissa Alonso. 2020. "The Minneapolis Police Officer Who Knelt on George Floyd's Neck Had 18 Previous Complaints Against Him, Police Department Says." *CNN*. May 31, 2020.

Arnold, D., W. Dobbie, and C.S. Yang. 2018."Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 1885–1932.

Baliga, Sandeep and Jeffrey Ely. 2016. "Torture and the commitment problem." *The Review of Economic Studies*, 1406–1439.

Bull, Jesse. 2008a. "Mechanism Design with Moderate Evidence Cost." *B.E. Journal of Theoretical Economics*, 8, article 15.

Bull, Jesse. 2008b. "Costly Evidence Production and the Limits of Verifiability,'" *B.E. Journal of Theoretical Economics* 8 article 18.

Bull, Jesse (2012), "Interrogation and Evidence Fabrication," working paper.

Bull, Jesse and Joel Watson. 2019. "Statistical Evidence and the Problem of Robust Litigation." *The RAND Journal of Economics*, 50, 4: 974–1003.

Bull, Jesse and Joel Watson. 2007. "Hard Evidence and Mechanism Design." *Games and Economic Behavior*, 58, 75–93.

Bull, Jesse and Joel Watson. 2004. "Evidence Disclosure and Verifiability." *Journal of Economic Theory*, 118: 1–31.

Bull, Ray. 1989. "Can Training Enhance Detection of Deception?" In J.C. Yuille (Ed.), *Credibility Assessment* (83–99). London: Kluwer Academic Publishers.

Daughety, Andrew and Jennifer Reinganum. 2020. "Reducing Unjust Convictions: Plea Bargaining, Trial, and Evidence Disclosure." *The Journal of Law, Economics, and Organization*, 36 (2), 378–414.

DePaulo, B.M., J.I. Stone and G.D. Lassiter. 1985. "Deceiving and Detecting Deceit." In B.R. Schlenker (Ed.) *the Self and Social Life* (323–370). New York: McGraw-Hill.

Dobbie, W., J. Goldin, and C.S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 108(2): 201–240.

Fluet, Claude and Thomas Lanzi. 2019. "Adversarial Persuasion with Cross-Examination," working paper.

Gladwell, Malcolm. 2019. *Talking to Strangers: What We Should Know about the People We Don't Know*. Little, Brown and Company.

Grossman, Gene M. and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare." *The American Economic Review*, 73, No. 4: 749–757.

Gudjonsson, Gisli H. 2012. "False Confessions and Correcting Injustices." *New England Law Review*, vol. 46, no. 4: 689–710.

Gudjonsson, Gisli H. and John Pearse. 2011. "Suspect Interviews and False Confessions." *Current Directions in Psychological Science*, 20(1): 33–37.

Inbau, Fred E., John E. Reid, Joseph P. Buckley, and Brian C. Jayne. 2013. *Criminal Interrogation and Confessions*, 5th Edition. Jones & Bartlett Learning.

Ispano, Alessandro and Péter Vida. 2020. "Custodial Interrogations." working paper.

Ispano, Alessandro and Péter Vida. 2021. "Designing Interrogations." working paper.

Kassin, Saul M., Sara C. Appleby, and Jennifer Torkildson Perillo. 2010. "Interviewing Suspects: Practice, Science and Future Directions." *Legal and Criminological Psychology.*

Kassin, Saul M., Richard A. Leo, Christian A. Meissner, Kimberley D. Richman, Lori H. Colwell, Amy-May Leach, and Dana La Fon. 2007 "Police Interviewing and Interrogation: A Self-Report Survey of Police Practices and Beliefs." *Law and Human Behavior*, 31: 381–400.

Kassin, Saul M. and Karlyn McNall. 1991. "Police Interrogations and Confessions: Communicating Promises and Threats by Pragmatic Implication." *Law and Human Behavior* 15: 233–251.

Leo, Rchard A. 2008. Police Interrogation and American Justice.

Leo, Richard A. and Richard Ofshe. 1998. "The Consequences of False Confessions: Deprivations of Liberty and Miscarriages of Justice in the Age of Psychological Interrogation." *Journal of Criminal Law and Criminology*, 88: 429.

Leshem, Shmuel. 2010. "The Benefits of a Right to Silence for the Innocent." *The RAND Journal of Economics*. 41, 2: 398-416.

Magid, Laurie. 2001. "Deceptive Police Interrogation Practices: How Far Is Too Far?" *Michigan Law Review*, 99:168–1210.

Mialon, Hugo M. 2005. "An Economic Theory of the Fifth Amendment." *The RAND Journal of Economics*, 36, 4: 833–848.

Meissner, Christian A. and Saul M. Kassin. 2002. "'He's Guilty!': Investigator Bias in Judgments of Truth and Deception." *Law and Human Behavior*, 26: 469–480.

Mueller-Smith, M. and K. Schnepel. 2020. "Diversion in the Criminal Justice System," working paper.

Ofshe, Richard J. and Richard A. Leo. 1997. "The Decision To Confess Falsely: Rational Choice and Irrational Action." *Denver University Law Review*, 74: 979–1122.

Rosenberg, David and Steven M. Shavell. 2006. "A Solution to the Problem of Nuisance Suits: The Option to Have the Court Bar Settlement." *International Review of Law and Economics*, 26: 42.

Rubin, Joel and Ari B. Bloomkatz. 2008. "Interrogation, then Revenge." *Los Angeles Times*. July 2, 2008.

Sanchirico, Chris. 1999. "Relying on the Information of Interested - and Potentially Dishonest - Parties." it American Law and Economics Review.

Sanchirico, Chris. 2000. "Games, Information, and Evidence Production: With Application to English Legal History." *American Law and Economics Review*, 2: 342–380.

Sanchirico, Chris. 2001. "Character Evidence and the Object of Trial." *Columbia Law Review*, 101: 1227.

Sanchirico, Chris. 2010. "Models of Evidence: Survey and Assessment." University of Pennsylvania Institute for Law & Economics Research Paper No. 10-28.

Sanchirico, Chris and George Triantis. 2005. "Evidentiary Arbitrage: the Fabrication of Evidence and the Verifiability of Contract Performance." *Journal of Law, Economics, and Organization*, 24:1.

Seidman, Daniel J., and Alex Stein. 2000. "The Right to Silence Helps the Innocent: A Game-Theoretic Analysis of the Fifth Amendment Privilege." *Harvard Law Review*, 114, 2: 430–510.

Shavell, Steven M. 1993. "Suit versus Settlement When Parties Seek Nonmonetary Judgments." *Journal of Legal Studies*, 22: 1.

Shavell, Steven M. 1989. "Sharing of Information Prior to Settlement or Litigation." *Rand Journal of Economics*, 20: 183–195.

Sherman, Lawrence W. and David Weisburd. 1995 "General Deterrent Effects of Police Patrol in Crime 'Hot Spots': a Randomized, Controlled, Trial." *Justice Quarterly*. 12, 4: 625–648.

Weisburd, David. 2015. "The Law of Crime Concentration and the Criminology of Place." *Criminology*. 53, 2: 133–157.

Weisselberg, Charles D. 2001. "In the Stationhouse after Dickerson." *Michigan Law Review*, 99: 1121–1167.

White, Welsh S. 2001. "Miranda's Failure to Restrain Pernicious Interrogation Practices." *Michigan Law Review*, 99: 1211–1247.