Lie Detection: A Strategic Analysis of the Verifiability Approach

Konstantinos Ioannidis, *University of Amsterdam*, Theo Offerman, *University of Amsterdam*, Randolph Sloof, *University of Amsterdam*

Send correspondence to: Konstantinos Ioannidis, CREED, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WV Amsterdam, the Netherlands; Tel: +31205259111 and Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, the Netherlands; Tel: +31205984581;
E-mail: ioannidis.a.konstantinos@gmail.com.

The Verifiability Approach is a lie detection method based on the insight that truth-tellers provide precise details whereas liars sometimes remain vague to avoid being exposed. We provide a game-theoretic foundation for the strategic effect that underlies this approach. We consider a speaker who wants to be acquitted and an investigator who prefers to find out the truth. The investigator can verify the speaker's statement at some cost; verification gets more reliable the more details are provided. If, after a falsified statement, the investigator convicts, an additional penalty is imposed. Constructing precise but false statements is assumed to be cognitively costly. We derive all equilibria and thereby the conditions under which the investigator can infer valuable information from the speaker's statement at face value. If cognitive costs are not prohibitively high, these require that liars are deterred from making false precise statements if always verified. Strategic information revelation by the speaker and verification by the investigator then necessarily work in tandem in a partially pooling equilibrium. Improvements in reliability result in more valuable information via the statements per se, whereas larger lying costs or a harsher penalty do not once the deterrence condition for the existence of this equilibrium is met. (*JEL*: C72, D01, D82, K14)

We gratefully acknowledge the highly valuable and constructive suggestions from the editor Albert Choi and two anonymous referees that improved the exposition considerably. We would also like to thank participants at the 2018 CeDEx-CBESS-CREED

American Law and Economics Review
doi:10.1093/aler/ahac005
Advance Access Publication on July 6, 2022

1. Introduction

After decades of research on lying detection, psychologists have recently made a breakthrough in revealing who is lying. The early literature focused on the idea that liars can be identified by facial microexpressions of emotions and other unintentional behaviors. In two meta-analyses, DePaulo et al. (2003) and Bond Jr and DePaulo (2006) showed that nonverbal cues of lying are weak and unreliable. A typical finding is that approximately 54% of examiners' judgments are correct, only slightly better than chance (50%). One important reason why nonverbal cues are unreliable is that liars try to mimic the expressions of truth tellers when they become aware of which cues are used by investigators. For example, Ekman et al. (1988) have shown that truth tellers often smile when they express genuine positive feelings and that liars mimic them by also smiling. The challenge that examiners then face is that they have to distinguish between fake and genuine smiles.

The breakthrough involves recent methods of lie detection which focus on the content of what is being said. In the Verifiability Approach (VA), the examiner judges a statement based on the presence and frequency of verifiable details. VA exploits a dilemma that liars face. Liars have an incentive to include verifiable details in their statement, because detailed accounts are more likely to be believed (Bell and Loftus, 1989). At the same time, presenting specific details is risky because it makes it easier for the examiner to check a statement (Nahari et al., 2014a). Truth-tellers typically do not have this dilemma and can reveal as many verifiable details as possible. The relative frequency of verifiable details in a statement may then become an informative signal of its truth. Using VA, examiners' judgments are correct in approximately 70% of the cases (Vrij, 2018).[1] Moreover, in contrast to the nonverbal cues, the accuracy of VA is enhanced when interviewees are made aware of it. Doing so results in truth tellers adding more verifiable

meeting in Norwich, U.K., the 2019 Annual Conference of the European Association of Law and Economics in Tel Aviv, Israel and the 2019 Annual Conference of the European Association of Psychology and Law in Santiago de Compostela, Spain for their insightful comments on the article.

1. Vrij (2018) provides an elaborate discussion of the state-of-the-art methods in lying detection. Besides VA, he discusses six prominent methods; see the next section for a brief overview. Among all these methods, VA stands out because of its success and the ease with which it is implemented.

details to their statement than liars do (Harvey et al., 2017; Nahari et al., 2014a).

In this article, we provide a game-theoretic foundation for the strategic effect that underlies VA and explore the potential interaction among its main drivers. Our analysis takes into account the cognitive costs of fabricating precise but false statements, the higher reliability of verifying detailed (rather than vague) statements, as well as the potential use of penalties for "obstruction of the investigation process" that VA may allow for. The main focus is on how these different elements jointly affect the strategic trade-off liars face and contribute to precise statements becoming an informative signal per se (even without being actually verified).

Our model considers a speaker who wants to convince an investigator that he is innocent and an investigator who pursues the truth. Applications of this type of strategic interaction abound. A mother may want to find out if her son is using drugs; a parole officer is interested to know if an offender lives up to the agreement made; an airport officer wants to find out if a passenger is carrying dangerous items; an insurance company wants to find out whether a claim was rightly made; an employer interviews an applicant (and potentially verifies references) to learn whether he has been thorough and truthful in drafting his CV; a judge questions a suspect to assess whether he is guilty. Throughout the article, we use labels that correspond to the judge-suspect example for ease of illustration. A suspect is privately informed about whether he is guilty or innocent. The judge has already collected some evidence that furnishes a prior belief about whether the suspect is guilty. The suspect is asked to make a statement about what happened. He either makes a precise statement that includes verifiable and distinctive details, or a vague statement. Providing a false precise statement is assumed to be cognitively costly. After listening to the suspect, the judge can decide to reach a verdict immediately or to check the statement at some cost. Checking a precise statement gives a more reliable signal than checking a vague statement does. If the judge convicts the suspect after his statement was checked and falsified, an additional obstruction of justice penalty is imposed on the suspect (Decker, 2004). The suspect always wants to be acquitted whereas the judge wants to reach a correct verdict. Moreover, she (weakly) prefers to wrongly acquit a guilty suspect over wrongly convicting an innocent one.

We derive all perfect Bayesian equilibria of the game and identify the conditions under which either full or partial information revelation occurs in equilibrium (and when this information is truly beneficial). A separating equilibrium exists only when the cognitive costs of lying are prohibitively high, such that guilty types always refrain from making a precise statement. The research on VA has identified interviewing techniques that can increase the cognitive load of lying.[2] Moreover, experimental evidence shows that deception can sometimes be (probabilistically) detected even without possibilities for ex post verification and consequences for lying (Jupe et al., 2017). Nevertheless, full separation may be hard to achieve in actual practice given the high stakes liars typically have in hiding the truth. In that case, lower (but positive) cognitive lying costs may still enable a partial pooling equilibrium in which the guilty suspect mixes between being precise and remaining vague (and an innocent suspect always makes a precise statement). However, in order for this information to be truly valuable to the judge and increase her expected payoffs, the possibility of actual verification then plays a key role. In particular, valuable information transmission then necessarily requires that the guilty type is deterred away from always lying if the judge would *always* verify a precise statement. If this "deterrence-by-verification condition" is not met, equilibria may still exist in which information is revealed via either the statement or the investigation, but the judge never gains in terms of expected payoffs relative to reaching a verdict immediately based on the prior belief.

Larger cognitive lying costs, a higher reliability of verification and a higher obstruction penalty all contribute to meeting the deterrence-by-verification condition. If indeed met, a partially pooling equilibrium exists in which the guilty suspect mixes between being precise and remaining vague and the judge only now and then verifies a precise statement (with a vague statement leading to immediate conviction). The judge then effectively has two sources of information complementing each other: the strategic behavior of the suspect (i.e., the statement per se) and the outcome of the

2. For example, asking surprise questions or requesting a narrative in reverse chronological order have been shown to be successful in Vrij et al. (2007) and Sorochinski et al. (2014). Other methods such as the Sheffield Lie Test exploit the fact that lying takes time and that response times can be used to distinguish truth tellers from liars (Suchotzki et al., 2017).

occasional verification. Within this equilibrium, increasing either the cognitive lying costs or the obstruction penalty further does not increase the provision of valuable information. What the judge can learn from the suspect's statement per se remains unaffected, because guilty suspects keep on lying with the same frequency. And the amount of valuable information obtained via verification is actually reduced, because higher lying costs or a higher obstruction penalty induces the judge to investigate less. Improvements in the verification technology that make it more reliable continue to have a beneficial impact; however, because a higher accuracy does induce the guilty type to lie less often. Interestingly, although all else equal the improved accuracy would by itself have led to more valuable information obtained via verification as well, in equilibrium it actually leads to less. The main drivers here are that precise statements are made less often by the guilty type and (therefore) also verified less often by the judge. Hence, if the verification technology becomes more accurate, the additional benefits that come with it are purely due to the deterrence effect of the potential verification. Finally, a decrease in the investigation costs has similar effects on the amount and source of valuable information obtained in the partial pooling equilibrium as an improved reliability has; driven by the deterrence effect again more information is obtained from the strategic behavior of the suspect itself and less from the actual verification of messages.

The overall upshot of our analysis is that especially improvements in the accuracy of the verification technology are beneficial. Even if the guilty type is willing to always lie, such improvements make actual verification that may occur in a pooling (on precise) equilibrium more cost effective. And as soon as the threshold that deters the guilty type from always lying is met, such improvements enlarge the deterrence effect. As a result, the guilty type reveals more information via the statement per se and the actual verification process itself actually yields less valuable information. Higher lying costs or a higher obstruction penalty play a supporting role in meeting the relevant deterrence threshold. But once this threshold is met, they do not facilitate further valuable information transmission.

We extend our analysis in two ways. First, we allow the suspect to confess to receive a penalty reduction. In that case, the guilty type no longer provides a vague statement in equilibrium and mixes between mimicking the innocent type by providing a precise statement and confessing instead.

The equilibrium analysis is essentially equivalent as for the baseline model, with the single difference that the lying costs should now be enlarged with the opportunity costs of not taking the opportunity to confess. A penalty reduction after confession thus complements the cognitive lying costs and the obstruction penalty in facilitating more informative equilibria and in fact represent two sides of the same coin.[3] That is, to reduce mimicking of the innocent type one can either make it more costly via the lying costs or the obstruction penalty, or less attractive via the penalty reduction. Second, we also consider the case in which the suspect has a "right to silence." In that case silence cannot be held against the suspect, effectively restricting the judge's choice of action in case the suspect refrains from making a precise statement. Such a right to silence may alter, but does not eliminate, strategic information revelation by the suspect and thus neither its complementary role to direct verification of statements.[4] Moreover, for an intermediate range of prior beliefs the lying costs and the obstruction penalty no longer play a supportive role, reinforcing that especially a higher reliability is advantageous.

The remainder of this article is organized as follows. Section 2 briefly discusses various lie detection methods that have received attention in the psychology literature and the verifiability approach in particular. It also discusses how we account for the key features of this approach in our theoretical analysis. Section 3 presents the setup of our baseline model. In Section 4, we derive the set of perfect Bayesian equilibria. Additionally, we discuss how the amount of (valuable) information revelation and the effective reliance on the different information sources varies with the characteristics of the verification technology. In Section 5, we consider two extensions of the baseline setup: the possibility of plea bargaining and accounting for a right to silence. Here, we also discuss the connection with earlier game-theoretic analyses of the latter two aspects within the law and economics literature. Section 6 summarizes the article and concludes.

3. As such, our article relates to earlier game-theoretic analyses of plea bargaining; see Grossman and Katz (1983), Reinganum (1988), Baker and Mezzetti (2001), Bjerk (2007), Kim (2010), and Tsur (2017). When discussing plea bargaining in Subsection 5.1, we make this connection (as well as the differences) with our model more precise.

4. The right to silence has been analyzed from a game-theoretic perspective by Seidmann and Stein (2000), Seidmann (2005), Mialon (2005), and Leshem (2010). In Section 5.2, we discuss the insights from these studies within the context of our model.

2. Lie Detection and the Verifiability Approach

The origins of deception detection research can be traced back to Zuckerman et al. (1981) who categorized emotion, arousal, control, and cognitive processing as four different cues to deception. Various methods were developed over the years which were based on the first three of these cues. The methods focused on nonverbal behavior, compared levels of arousal between liars and truth-tellers and did not intervene in the information gathering process. In a meta-analysis, DePaulo et al. (2003) showed that those methods were not reliable as the observed behaviors showed no direct links to deception. According to Vrij (2019), to overcome those issues, modern research in deception detection has made three major shifts. Modern methods (1) focus on the content of a statement, (2) take into account the cognitive process behind lying, and (3) have developed interview protocols to optimize the information gathering process. Some of these are already admissible as evidence in courts in countries like the United States, Germany and the Netherlands (Vrij, 2008).

Vrij (2018) provides an elaborate discussion of the state-of-the-art methods in deception detection. He compares the seven most prominent methods in terms of how ready they are to be applied in judicial systems. The list of methods includes Criteria Based Content Analysis (CBCA), Reality Monitoring (RM), Scientific Content Analysis (SCAN), Cognitive Credibility Assessment (CCA), Strategic Use of Evidence (SUE), the Verifiability Approach (VA), and Assessment Criteria Indicative of Deception (ACID). He does so on the basis of 14 criteria, which can be grouped into two sets: academic, such as whether the method has been tested and whether it has been subjected to peer review, as well as procedural, such as whether it is easy to use and whether it provides an information gathering protocol. Five of those criteria, also known as the Daubert standard, are the minimal requirements for scientific evidence to be admissible in US courts (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993).[5] Of the seven methods, only three abide by the Daubert standard, namely RM, ACID, and VA. However, ACID is not easy to incorporate in interviews and RM does not provide a

5. The full list of criteria for admissibility of scientific evidence in US courts is: (i) Has the technique been tested in actual field conditions (and not just in a laboratory)?; (ii) Has the technique been subject to peer review and publication?; (iii) What is the known

within-subject measure of truthfulness. Hence, our article models VA as the investigation mechanism available to the judge.

As the name suggests, VA is based on the verifiability of details. A detail is considered verifiable if it describes an activity experienced with an identifiable person or witnessed by an identifiable person or recorded through technology (Nahari et al., 2014a). Based on the finding that lying is cognitively more demanding (Vrij et al., 2017), there exist interviewing techniques which aim to magnify the cognitive task for liars. On the one hand, the interviewer asks the interviewee to include as many details as possible. On the other hand, the interviewee would like to avoid mentioning details that can easily be checked by the interviewer. Balancing those orthogonal incentives, one would expect a liar to provide many nonverifiable details in a statement. The ratio of verifiable over nonverifiable details is a within-subject measure of the probability that a statement is true or fabricated. Additional benefits of VA are the fact that it is robust to countermeasures (Nahari et al., 2014b) and that VA scoring could be computer-automatized as suggested by Kleinberg et al. (2016).

To capture the essential element of VA within a simple model of strategic information transmission, we extend an otherwise standard sender-receiver game with an (bilaterally) endogenous verification technology. On the disclosing side, the sender can choose between various statements that differ directly in their costs and indirectly in their degree of verifiability. More precise statements in principle allow for a more reliable investigation than less precise—that is more "vague"—statements do. At the same time, coming up with a precise false statement is cognitively costly. On the receiving side, the receiver subsequently decides whether to indeed investigate the actual statement made (at same costs) or not. Lying—that is fabricating a precise but false statement as to mislead the receiver—is clearly possible. However, with cognitive costs and with potential verification, the sender's statement is not pure cheap talk. From a modelling perspective our article thus fits within the broader theoretical literature on strategic communication with either intrinsically costly (cf. Kartik, 2009), or detectable deceit (cf.

---

or potential rate of error?; (iv) Do standards exist for the control of the technique's operation?; and (v) Has the technique been generally accepted within the relevant scientific community?

Holm, 2010; Dziuda and Salas, 2018; Balbuzanov, 2019; Ispano and Vida, 2021). Our setup differs, among other things, in verification being costly and at the receiver's discretion.

Our article is motivated by the above discussed findings from psychology that lying is cognitively more demanding. We model this psychological component both explicitly as a direct lying cost, as well as via the implied reliability of the investigation technology and its relationship to the sender's statement and underlying type. Glazer and Rubinstein (2012) provide a model of strategic persuasion in which a speaker is boundedly rational in the sense that she uses the truth as an anchor for cheating. In particular, when fabricating a false set of answers to a given questionnaire, the speaker starts from the truth and tries to modify her answers to satisfy the listener's pre-set acceptance conditions. Modifications are limited to adapting the consequence of originally violated "if-then" conditions.[6] As a result, truth-tellers always get their way, whereas liars—given their truth anchor and modification limits—may not be able to satisfy the listener's "codex." Similar to other recent papers on detectable deceit (Dziuda and Salas, 2018; Balbuzanov, 2019), we simply incorporate this element directly by allowing for (endogenous) variation in lie detection probabilities. The main focus is then on the implications for the amount of strategic information revelation that results.

Key within VA is the *verifiability of distinctive details*. We label statements that contain many such verifiable details as "precise" and statements with none or only a very few of these as "vague." This labeling does not

6. Glazer and Rubinstein (2012) provide experimental evidence that supports these two key ingredients (truth anchor and the specific type of modifications considered) of bounded rationality. In Glazer and Rubinstein (2014), they study a related setup in which the speaker does not know the exact set of acceptance conditions, but can make inferences about these from the observed earlier acceptance decisions of the listener. Here bounded rationality is captured by limitations on the complexity of the regularities that the speaker can detect in the acceptance data. By making the questionnaire sufficiently complex, the listener can almost completely eliminate successful cheating by liars. A different microfoundation for using complex interview protocols rooted in bounded rationality is provided by Jehiel (2021). He analyses multi-round cheap talk communication assuming liars have more limited memory than truth-tellers have. The liar's fear of issuing inconsistent statements over time can then be exploited to facilitate information revelation.

necessarily coincide with using precise or vague language though.[7] In linguistics, a term is considered vague if it exhibits borderline cases. For instance, there are no clear cut bounds on the number of grains that define a "heap" of sand (O'Connor, 2014). Within our setting, clear statements that might nevertheless be hard to verify (like "I was home alone sleeping in my bed") are considered vague. Similarly so are essentially empty statements that are (almost) tautologically true, like "I was on planet earth." (Note that such statements are also not cognitively demanding to fabricate.) Key difference between a precise and a vague statement in our setup is the larger extent to which the former provides a convincing alibi when verified as well as reason for serious suspicion if falsified, that is, its distinctiveness.

3. Baseline Model

Although the strategic interaction that we model arguably matches various real life applications (cf. Section 1), for concreteness we describe it in terms of the interaction between a suspect (speaker) and a judge (investigator). Assume a crime has been committed and a suspect (he) is being questioned. The judge (she) can use the statement of the suspect to update her beliefs on his innocence. She can do so immediately or after conducting a costly investigation that can, with some commonly known error probability, verify or falsify the statement. The suspect wants to be acquitted and the judge wants to reach a correct verdict, viz. to acquit innocent suspects and to convict guilty ones. Additionally, the judge prefers to acquit a guilty suspect over convicting an innocent one.

We note up front that our conceptualization of the interaction between a suspect and a judge is based on a number of simplifications. In real life, a suspect can get arrested by the police, provide a statement and a prosecutor may decide whether to file charges and bring the case to court or not. If she

7. In cheap talk experiments, messages such as "The true value of a variable belongs to set S" are labeled precise if S is a singleton and vague otherwise. Using this definition, vague language has been shown to increase efficiency in experiments involving public good games with hidden value (Serra-Garcia et al., 2011) and coordination games with multiple equilibria (Agranov and Schotter, 2012). Without the possibility to verify messages before taking an action, that is, when messages are pure cheap talk, both type of messages would be considered vague in our setup.

does so, the suspect becomes a defendant and may provide additional testimony during the trial. All evidence is examined by a judge and/or a jury and once a verdict is reached, the judge imposes a penalty or not. In our reduced form model we have condensed the timing, the actors and the type of information provided. We use "judge" as label for a representative of the judicial system with the understanding that in practice some actions described in the model might be taken by prosecutors or the jury.[8] Essentially, our model assumes that at some point during the entire judicial process, the suspect will be asked to provide some information. The untruthfulness of this information is assumed to have consequences for the sentence the suspect may be facing, if he gets convicted.

Our model corresponds to a sender–receiver game, where the sender is the suspect and the receiver is the judge. The suspect knows his own type (T), that is whether he is innocent ($T = I$) or guilty ($T = G$). The type of the suspect is unknown to the judge, but she holds a commonly known prior belief of $b = Pr(T = I)$ that the suspect is innocent. These prior beliefs can be interpreted as the evidence collected by the judge before questioning the suspect, so that in principle she can convict (or acquit) without requesting a statement.

The suspect can choose between two actions. He can choose to answer all the questions,[9] which results in a precise statement ($S = P$), or he can choose to refrain from providing clear and distinctive answers, which results in a vague statement ($S = V$).[10] Providing a precise, but false statement is cognitively costly. After seeing the statement, the judge must reach a verdict to acquit (A) or convict (C) the suspect. This decision can be taken either before or after having investigated (I) the statement made.

8. Assuming a unitary actor for the judicial system is an arguably reasonable simplification to the extent that the various actors within the judiciary share the same preferences and information. We briefly return to this in Section 5.1 where we discuss the possibility of plea bargaining and relate our strategic setup to existing models of plea bargaining in the literature.

9. An implicit assumption in the model is that when answering questions, an innocent suspect tells the truth whereas a guilty suspect lies. Allowing both of them to choose whether to answer truthfully or not is a possible extension for future research.

10. In a previous version of this article, we considered the extension towards allowing the suspect to reveal an arbitrary number of verifiable details. All the main insights of the baseline model presented here remain valid in this richer model.

The investigation mechanism works as follows. If the judge decides to investigate statement $S \in \{V, P\}$, the investigation mechanism provides an outcome that has a probability of r_S of being correct (which means verified for the statement of the innocent type and falsified for the statement of the guilty type) and a probability of $1 - r_S$ of being wrong (which means falsified for the statement of the innocent type and verified for the statement of the guilty type). Parameters r_V and r_P thus reflect the reliability of investigating the various statements. We assume that the investigation mechanism has at least some informational value, in the sense that it gets the judge closer to the truth. This assumption translates to both probabilities r_V and r_P being larger than $\frac{1}{2}$.[11] Aligned with the psychology literature on content-based deception detection methods, we also assume that the differences in content between the statement of the innocent and the guilty type will be more pronounced in a more detailed statement (Harvey et al., 2017). With a precise statement, the judge then gets better clues exactly what to look for, allowing her to steer her investigation in a more promising direction. As a result, investigating a precise statement is more likely to produce a correct outcome than investigating a vague one, that is, we assume that investigation probabilities satisfy $\frac{1}{2} < r_V < r_P < 1$.

Preferences of the two suspect types depend on both the statement they provide and on the decision of the judge. To capture that fabricating a detailed lie might be cognitively costly to the suspect, we assume that the guilty type suffers a direct lying cost equal to $\lambda_P \geq 0$ when providing a precise but false statement. Making a vague statement does not entail any cognitive costs, however, and neither does telling the truth in full detail via a precise statement for the innocent type. The choices of the judge affect the payoffs of both suspect types in the following way. Both suspect types get a payoff of 1 if they get acquitted. If they get convicted, they receive a lower payoff which depends on the amount of evidence that resulted in their conviction. If they get convicted on the basis of prior evidence, which happens when the judge does not investigate the statement or when investigation verifies the statement and provides no additional evidence

11. The distinctiveness of verifiable statement S can be inversely captured by the odds ratio $\frac{1-r_S}{r_S}$ of verification being unreliable. This ratio ranges from 0 for $r_S = 1$ (maximal distinctiveness), to 1 for $r_S = \frac{1}{2}$.

against them, they receive a payoff of 0 (so the imposed sentence leads to a payoff reduction of 1). If they get convicted after their statement S was investigated and falsified, they receive an additional obstruction penalty π_S, for which we assume that $0 \leq \pi_V \leq \pi_P$.[12] We also assume that this obstruction penalty is only applied when a suspect is eventually convicted.[13]

The preferences of the judge are modeled in the following way. The judge gets 1 for reaching a correct verdict, that is to acquit an innocent suspect and to convict a guilty suspect. In case the judge makes a mistake, she receives a lower payoff that depends on the type of mistake made. We normalize the payoff of acquitting a guilty suspect to 0 and set the payoff of convicting an innocent suspect to $-\alpha$. The assumption that $\alpha \geq 0$ captures the notion that the judge (weakly) prefers to let go of a guilty suspect over sending an innocent suspect to jail. Higher values of α result in a tighter threshold on the judge's belief for her to prefer conviction over acquittal; it thereby essentially quantifies exactly what is meant by "beyond any reasonable doubt" and sets the standard of proof. In particular, with these payoffs the (updated) belief that the suspect is innocent should exceed the tipping point of $\frac{1}{2+\alpha}$ for the judge to acquit.[14]

Besides obtaining the above payoffs the judge has to pay a positive cost $c > 0$ when she investigates statement S. These costs not only reflect that investigating the truthfulness of statements is costly in terms of the resources needed (time and detectives), but may also capture other, more indirect types of costs. For instance, in criminal cases of high importance that receive widespread public attention, society often really disapproves of cases that last

12. This penalty can be interpreted in various ways. If the lying was under oath, then the defendant may be charged with perjury (US Sentencing Commission, 2018, §2J1.3.). If the lying significantly impeded official investigation, then the defendant may be charged with obstruction of justice (US Sentencing Commission, 2018, §3C1.1.). The sentencing guidelines also recommend a reduction of penalty if a defendant provided substantial assistance in the investigation, for example by giving truthful, complete and reliable testimony (see §5K1.1.). In this case, the penalty can be interpreted as the difference between the full and the reduced sentence.

13. A prosecutor will very often drop a criminal charge if it is determined that the evidence against the accused is not strong enough, see Cohen (1992).

14. Given beliefs b, the judge's expected payoffs from acquit equal $1 \cdot b + 0 \cdot (1-b) = b$ and thus increase with b, while the expected payoffs from convict equal $1 \cdot (1-b) - \alpha \cdot b = 1 - (1+\alpha)b$ and decrease with b. At $b = \frac{1}{2+\alpha}$ these expected payoffs coincide.

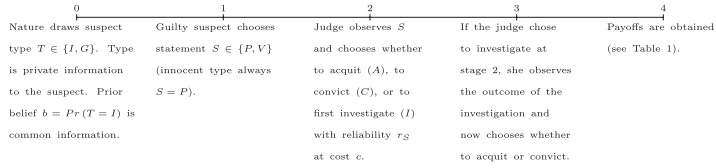


Figure 1. Timeline of the Game

for years, so our cost parameter could also be seen as pressure to reach a verdict faster. Note that our assumptions regarding the judge's payoffs arguably make these largely aligned with what society would seem to require. Her expected payoffs could thus potentially serve as a first approximation to a more encompassing welfare analysis.

To simplify the exposition we finally assume that the innocent type always provides a precise statement. Telling the truth—and in full detail if asked to do so—comes as a default to innocent people who have no incentive to lie (Verschuere and Shalvi, 2014), and innocent people even waive their right to remain silent due to their belief that their truth will shine (Kassin and Norwick, 2004). In an earlier version of this article, we did not make this simplifying assumption and analyzed the model assuming that the innocent type is also a strategic agent who endogenously chooses between a vague and a precise statement as well. The single notable difference is that in that case additional equilibria may exist alongside the other equilibria in which both suspect types always provide a vague statement. These pooling equilibria generally do not survive standard equilibrium refinements based on payoff dominance or on restricting out-of-equilibrium beliefs, like for example, the divinity concept of Banks and Sobel (1987). Our simplifying assumption essentially solves the multiplicity of equilibria issue in a simpler way without losing much nuance.

Figure 1 provides a succinct summary of the order of moves in the strategic interaction between the suspect and the judge and Table 1 summarizes the payoffs of all agents.

4. Equilibrium Analysis

4.1. Perfect Bayesian Equilibria

Besides her prior belief, the judge in principle has two information sources available: investigation (at cost c) of the actual statement made

Table 1. Payoffs of Suspect and Judge for All Type-Action Combinations

	Convict		Acquit	
	Suspect	Judge	Suspect	Judge
Innocent type:				
Precise w/out verification	0	$-\alpha$	1	1
Precise with verification	$-\pi_P$	$-\alpha - c$	1	$1 - c$
Guilty type:				
Vague w/out verification	0	1	1	0
Vague with verification	$-\pi_V$	$1 - c$	1	$-c$
Precise w/out verification	$-\lambda_P$	1	$1 - \lambda_P$	0
Precise with verification	$-\lambda_P - \pi_P$	$1 - c$	$1 - \lambda_P$	$-c$

Notes: By assumption $\alpha \geq 0, c > 0, 0 \leq \pi_V \leq \pi_P$ and $\lambda_P \geq 0$.

by the subject and the potentially different strategies the two types of suspects employ in making statements. In this section we explore the extent to which these different information sources are actually drawn upon in equilibrium and how they interact, by providing an encompassing (perfect Bayesian) equilibrium analysis.

For the judge, the main goal of the entire process is to get a better idea of whether the suspect is guilty or not. Given the assumptions made, a vague statement can only be coming from a guilty suspect and, consequently, leads to immediate conviction without further costly investigation. Starting from a prior belief b that the suspect is innocent, after seeing a precise statement the judge updates her initial belief based on the strategic behavior of the suspect. Let p denote the probability that the guilty type gives a precise statement. Using Bayes' rule, a rational judge then updates her belief that the suspect is innocent to:

$$b^P \equiv Pr(T = I|S = P) = \frac{b}{b + (1 - b)p}. \tag{1}$$

Note that $b \leq b^P \leq 1$. The more the guilty suspect lies, that is, the higher p, the closer the posterior belief is to the prior. Likewise, the less the guilty suspect lies, the closer the posterior gets to 1.

Having seen a precise statement, the judge convicts, investigates, and acquits with respective probabilities q_C, q_I, and q_A. (As noted, after a vague statement the judge convicts for sure given the assumptions made.) In case

the judge investigates, she obtains additional information that allows her to update her beliefs another time, based on the outcome of the investigation. From the given reliability of the investigation process and again Bayes' rule, we immediately obtain that these beliefs equal:

$$b^{P+} \equiv Pr(T = I|S = P \text{ and verified}) = \frac{b^P r_P}{b^P r_P + (1 - b^P)(1 - r_P)} \quad (2)$$

$$b^{P-} \equiv Pr(T = I|S = P \text{ and falsified}) = \frac{b^P (1 - r_P)}{b^P (1 - r_P) + (1 - b^P) r_P}. \quad (3)$$

From these expressions, together with $r_P > \frac{1}{2}$, it follows that $b^{P-} \leq b^P \leq b^{P+}$. Falsification of the statement made by the suspect thus lowers the judge's belief that he is innocent, while a verified statement increases this belief.

Because investigating a precise statement is costly to her, the judge is willing to do so only if it yields her valuable information. That is, the information received should be *influential*; the judge's optimal decision whether to acquit or convict should (strictly) vary with the outcome of the investigation process.[15] Otherwise the judge could better immediately opt for the decision she would in the end take anyway and avoid costly investigation altogether. Recall from the previous section that the tipping point (in terms of beliefs) for the judge to prefer acquit over convict equals $\frac{1}{2+\alpha}$. Influential information thus requires that updated beliefs satisfy $b^{P-} < \frac{1}{2+\alpha} < b^{P+}$, such that the judge acquits when the suspect's precise statement is verified and convicts when the precise statement is falsified.[16] Lemma 1 details this requirement in terms of the posterior belief b^P.

15. Note that the notion of the investigation being *influential* is stronger than that it being *informative*. The latter holds as long as long as the outcome of the investigation is more likely to be aligned with the truth, which is guaranteed by our assumption that $\frac{1}{2} < r_V < r_P$. Sobel (2020) provides an insightful discussion of the differences between the definitions of informative and influential.

16. When either $b^{P-} = \frac{1}{2+\alpha} < b^{P+}$ or $b^{P-} < \frac{1}{2+\alpha} = b^{P+}$, the judge would be indifferent between acquit and convict after either falsification or verification, respectively. In both cases (which cannot happen simultaneously), the judge essentially always weakly prefers either acquit or convict, irrespective of the outcome of the investigation; she thus would not be willing to invest $c > 0$ in it. That is why we require the optimal outcome to *strictly* vary with the outcome of the investigation.

LEMMA 1 Investigating a precise statement is influential iff: $\frac{1-r_P}{\alpha r_P+1} < b^P < \frac{r_P}{\alpha(1-r_P)+1}$. In that case the judge would acquit if a precise statement were to be verified and convict if a precise statement were to be falsified.

Proof. Investigating a precise statement is influential as long as $b^{P-} < \frac{1}{2+\alpha} < b^{P+}$. Using expressions (2) and (3) for b^{P+} and b^{P-} above and rewriting immediately gives the result. □

Intuitively, investigation can be influential only if, after having just heard a precise statement and correctly inferring the suspect's strategic behavior (in particular, probability p with which a guilty suspect makes such a statement), the judge is still insufficiently confident about the suspect's type. That is, she is neither sufficiently convinced that the suspect is guilty (b^P is not very low), nor sufficiently convinced that the suspect is innocent (b^P is neither very high).

Obtaining influential information is a necessary requirement for the judge to investigate, yet it is not a sufficient. The expected benefits from the influential information received should also outweigh the costs of investigation c. Lemma 2 precisely characterizes this requirement and pins down the judge's optimal choice for any posterior belief $b^P \in [0, 1]$ she might have.

LEMMA 2 Define $\underline{b}(r_P, c; \alpha) \equiv \min\left\{\frac{(1-r_P)+c}{\alpha r_P+1}, \frac{1}{2+\alpha}\right\}$ and $\overline{b}(r_P, c; \alpha) \equiv \max\left\{\frac{r_P-c}{\alpha(1-r_P)+1}, \frac{1}{2+\alpha}\right\}$. Moreover, let $\hat{c}(r_P; \alpha) \equiv \frac{1+\alpha}{2+\alpha} \cdot (2r_P - 1)$. After a precise statement and based on updated belief b^P, the judge's optimal choice of action equals:

(1) convict if $b^P < \underline{b}(r_P, c; \alpha)$;
(2) investigate if $b^P \in (\underline{b}(r_P, c; \alpha), \overline{b}(r_P, c; \alpha))$;
(3) acquit if $b^P > \overline{b}(r_P, c; \alpha)$.

The interval $\left(\underline{b}(r_P, c; \alpha), \overline{b}(r_P, c; \alpha)\right)$ is nonempty and equals $\left(\frac{(1-r_P)+c}{\alpha r_P+1}, \frac{r_P-c}{\alpha(1-r_P)+1}\right)$ iff $c < \hat{c}(r_P; \alpha)$. In that case, the judge is indifferent between convict and investigate if $b^P = \underline{b}(r_P, c; \alpha)$, and indifferent between investigate and acquit if $b^P = \overline{b}(r_P, c; \alpha)$. If $c > \hat{c}(r_P; \alpha)$ and thus $\underline{b}(r_P, c; \alpha) = \overline{b}(r_P, c; \alpha) = \frac{1}{2+\alpha}$, the judge is indifferent between convict and acquit when $b^P = \frac{1}{2+\alpha}$.

Proof. For updated beliefs b^P that the suspect is innocent, immediate acquittal after a precise statement yields the judge b^P in expected payoffs while immediate conviction yields her $1-(1+\alpha)b^P$ in expectation. Acquittal thus dominates conviction iff $b^P > \frac{1}{2+\alpha}$. Given that an investigation is costly ($c > 0$), the judge is only willing to do so if it is influential (cf. Lemma 1); it then leads to an expected payoff of $r_P - b^P(1-r_P)\alpha - c$. This exceeds the payoff of convicting if $b^P > \frac{(1-r_P)+c}{\alpha r_P+1}$ and the one of acquitting if $b^P < \frac{r_P-c}{\alpha(1-r_P)+1}$. For these thresholds, it holds that $\frac{(1-r_P)+c}{\alpha r_P+1} \leq \frac{1}{2+\alpha}$ and $\frac{r_P-c}{\alpha(1-r_P)+1} \geq \frac{1}{2+\alpha}$ iff $c \leq \hat{c}\,(r_P;\alpha)$. Hence, if $c < \hat{c}\,(r_P;\alpha)$, the interval $(\underline{b}(r_P,c;\alpha), \overline{b}(r_P,c;\alpha))$ is nonempty and in this range the judge prefers investigation. □

The belief interval where costly investigation pays off collapses when the verification is completely inaccurate. Put differently, the break-even cost threshold equals $\hat{c}\,(r_P;\alpha) = 0$ for $r_P = \frac{1}{2}$. Recall from the Introduction that nonverbal deception detection methods are almost indistinguishable from chance as their accuracy is close to 50%. Thus, relying on such methods, while costly, does not facilitate information revelation. Verbal deception detection methods can achieve higher accuracy which benefits the judge. Intuitively, the range of beliefs $b^P \in \left(\underline{b}(r_P,c;\alpha), \overline{b}(r_P,c;\alpha)\right)$ for which investigation pays off widens if the verification process becomes more reliable, that is, when r_P increases, and when investigation becomes cheaper (lower c). If nonempty, the interval always contains the tipping point $\frac{1}{2+\alpha}$ between convicting and acquitting. The further away beliefs b^P are from this point of indifference, the more confident the judge is to solely act on the basis of the existing evidence—that is, the prior belief and the statements per se—and to skip costly investigation altogether.

Turning to the guilty type of suspect, in equilibrium he chooses a best response to the judge's anticipated behavior. Our next lemma characterizes his optimal choice of statement when he anticipates that the judge responds with (q_A, q_I, q_C) to a precise statement.

LEMMA 3 Define $\hat{\lambda}\,(r_P,\pi_P) \equiv 1 - r_P\,(1+\pi_P)$. If the judge chooses (q_A, q_I, q_C) in response to a precise statement, the guilty's type optimal choice of statement equals:

(1) a precise one $S = P$ if $\lambda_P < q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P)$;
(2) a vague one $S = V$ if $\lambda_P > q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P)$.

The guilty type is indifferent between a precise and vague statement iff $\lambda_P = q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P)$.

Proof. With the judge's response (q_A, q_I, q_C), the expected payoffs from choosing a precise statement equal $q_A \cdot 1 + q_I \cdot ((1 - r_P) \cdot 1 - r_P \cdot \pi_P) - \lambda_P = q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P) - \lambda_P$. Choosing a vague statement leads to immediate conviction and thus payoffs equal to 0. Comparing these payoffs gives the result. □

Providing a vague statement leads to immediate conviction and a payoff of 0. The guilty type is then only willing to make a precise statement if the cognitive costs of doing so are not prohibitively large compared to the expected benefits of a potentially more favorable decision (than conviction) such a statement might bring. The relevant threshold for λ_P thus depends on the judge's response to a precise statement. If the judge would always acquit ($q_A = 1$), a precise statement would yield the guilty type a payoff of 1. The expected benefits relative to the benchmark of conviction (yielding 0) then equal 1. If the judge would always investigate ($q_I = 1$) after a precise statement, it would yield the guilty type an expected payoff equal to $(1 - r_P) - r_P \pi_P$. This expression follows because with probability $(1 - r_P)$ the guilty type gets away with his lie and is acquitted, while with the remaining probability r_P he is caught lying and, besides conviction, is imposed obstruction penalty π_P. The overall expected benefits from a precise statement then equal $\hat{\lambda}(r_P, \pi_P)$. Note that $\hat{\lambda}(r_P, \pi_P)$ falls short of $\frac{1}{2}$ given $r_P > \frac{1}{2}$ and decreases with both r_P and π_P (and becomes negative for π_P large). As Lemma 3 illustrates, for a general anticipated response from the judge the cost–benefit analysis for the guilty type compares the direct lying costs λ_P with the appropriately weighted average of the two relevant thresholds 1 and $\hat{\lambda}(r_P, \pi_P)$.

Based on the best responses in Lemmas 2 and 3, *mutual* best responses—and thereby equilibrium outcomes—can now be intuitively understood. First observe from Lemma 3 that if $\lambda_P > 1$, the guilty type will choose a vague statement for sure. Put differently, if the cognitive costs of fabricating a false precise statement are prohibitively high, the guilty type necessarily

chooses to willingly expose himself by making a vague statement. A precise statement then provides conclusive evidence that the suspect is innocent, inducing the judge to acquit for sure after such a statement. We thus immediately obtain a unique separating equilibrium in this case. In this separating equilibrium, the strategic behavior of the two suspect types is fully revealing and the judge always reaches the correct verdict, without the need to ever verify the statements made.

Arguably, in criminal cases the conditions for full separation are often not met. A guilty suspect may either be cognitively able to produce a detailed (but false) statement, or can afford the legal expertise to help him produce one. In those instances where $\lambda_P < 1$ and the guilty type in principle would be willing to provide a precise statement, completely revealing equilibria do not exist. In that case, the evidence of the case as captured by the prior belief b determines the extent to which he actually will do so in equilibrium. Given that a precise statement always induces the judge to update her belief upwards (i.e., $b^P \geq b$ by equation (1)), making such a statement can always ensure acquittal if the prior belief would already do so. From Lemma 2, it follows that this happens when $b > \overline{b}(r_P, c; \alpha)$. In that case, the guilty suspect can safely lie and completely get away with it. This yields a pooling equilibrium in which no additional information at all is obtained and the judge reaches a verdict purely based on her prior belief.

For completeness, we formally describe these two—arguably unrealistic—"corner" equilibria in the following proposition. Here, we omit the choice of the innocent type as we have assumed he always provides a precise statement. We also omit the choice of the judge after a vague statement as we established earlier that a vague statement leads to immediate conviction. Therefore, the equilibria are described with the probability that the guilty suspect provides a precise statement (p), the posterior belief of the judge after she observes such a precise statement (b^P), and the subsequent decision of the judge right after updating her beliefs (q_A, q_I, q_C).

Proposition 1 Consider the case with either $\lambda_P > 1$ or $b > \overline{b}(r_P, c; \alpha)$. Then there exists a unique Perfect Bayesian equilibrium which is either separating (Sep) or pooling (Pool) and characterized as follows.[17]

17. Here and in the sequel, we focus on "generic" cases. In nongeneric cases, multiple equilibria may exist side by side. For instance, in the knife-edge case where $\lambda_P = 1$ (and $b > \overline{b}(r_P, c; \alpha)$) equilibria Sep and Pool co-exist.

Sep Suppose $\lambda_P > 1$. Then the guilty type always gives a vague statement and the judge always acquits after a precise statement. Formally: $p = 0$, $b^P = 1$, $q_A = 1$.

Pool Suppose $\lambda_P < 1$ and $b > \overline{b}\,(r_P, c; \alpha)$. Then the guilty type always gives a precise statement and the judge always acquits after a precise statement. Formally: $p = 1$, $b^P = b$, $q_A = 1$.

Proof. If $\lambda_P > 1$, then $p = 0$ from Lemma 3. In turn, $b^P = 1$ from equation (1) and thus $q_A = 1$ from Lemma 2. This gives the separating equilibrium Sep. Next, let $\lambda_P < 1$. If $b > \overline{b}\,(r_P, c; \alpha)$, then $q_A = 1$ necessarily from Lemma 2 and $b^P \geq b$. From $\lambda_P < 1$ and Lemma 3, the guilty type's best response then equals $p = 1$. This yields equilibrium Pool. □

If neither the cognitive costs nor the prior beliefs are that high and thus the conditions of Proposition 1 are not met, necessarily some but not all information is revealed in equilibrium.[18] The amount of information revelation, as well as the information source effectively drawn upon in equilibrium, then depends on how the prior belief b and the characteristics of the investigation technology as reflected by parameters $\left(\lambda_P, c, r_p, \pi_p\right)$ compare to the relevant thresholds $\underline{b}(r_P, c; \alpha)$ and $\hat{c}\,(r_P; \alpha)$ from Lemma 2, and $\hat{\lambda}\,(r_P, \pi_P)$ from Lemma 3. For each distinct class of parameter combinations, Proposition 2 characterizes the unique informative equilibrium that exists. The numbering of these equilibria reflects their desirability from the perspective of the judge (to which we return in the next subsection).

PROPOSITION 2 Consider the case with $\lambda_P < 1$ and $b < \overline{b}\,(r_P, c; \alpha)$. Then, in the generically unique perfect Bayesian equilibrium the judge necessarily obtains some information beyond her prior beliefs b. This Informative equilibrium corresponds to one from the list below.

18. This follows because no information revelation would require that the guilty type always makes a precise statement (i.e., $p = 1$) such that the statement per se reveals no information and thus $b^P = b$. For $b < \overline{b}(r_P, c; \alpha)$ it then follows from Lemma 2 that the judge either convicts or investigates after a precise statement. The latter is incompatible with the judge not getting additional information beyond her prior. But if the judge would always convict after a precise statement, the guilty type would not be willing to bear the cognitive costs λ_P.

Inf.1 Suppose $c < \hat{c}(r_P;\alpha)$ and $\lambda_P > \hat{\lambda}(r_P,\pi_P)$. Then the guilty type mixes between a vague and a precise statement and a partially pooling equilibrium results. The judge mixes between acquit and investigate after a precise statement. Formally: $p = \frac{b}{1-b} \cdot \frac{(1-r_P)(1+\alpha)+c}{r_P-c}$, $b^P = \overline{b}(r_P,c;\alpha)$, $q_A = \frac{\lambda_P - \hat{\lambda}(r_P,\pi_P)}{1-\hat{\lambda}(r_P,\pi_P)}$, $q_I = 1 - q_A$.

Inf.2 Suppose $c < \hat{c}(r_P;\alpha)$, $\lambda_P < \hat{\lambda}(r_P,\pi_P)$ and $b > \underline{b}(r_P,c;\alpha)$. Then the guilty type always gives a precise statement and a pooling equilibrium results. The judge always investigates after a precise statement. Formally: $p = 1$, $b^P = b$, $q_I = 1$.

Inf.3 Suppose $c < \hat{c}(r_P;\alpha)$, $\lambda_P < \hat{\lambda}(r_P,\pi_P)$ and $b < \underline{b}(r_P,c;\alpha)$. Then the guilty type mixes between a vague and a precise statement and a partially pooling equilibrium results. The judge mixes between investigate and convict after a precise statement. Formally: $p = \frac{b}{1-b} \cdot \frac{r_P(1+\alpha)-c}{1-r_P+c}$, $b^P = \underline{b}(r_P,c;\alpha)$, $q_I = \frac{\lambda_P}{\hat{\lambda}(r_P,\pi_P)}$, $q_C = 1 - q_I$.

Inf.4 Suppose $c > \hat{c}(r_P;\alpha)$. Then the guilty type mixes between a vague and a precise statement and a partially pooling equilibrium results. The judge mixes between acquit and convict after a precise statement. Formally: $p = \frac{b}{1-b} \cdot (1+\alpha)$, $b^P = \frac{1}{2+\alpha}$, $q_A = \lambda_P$, $q_C = 1 - q_A$.

Proof. See Appendix A. □

For prohibitively high investigation costs $c > \hat{c}(r_P;\alpha)$, the interval of posterior beliefs for which the judge prefers to investigate a precise statement is empty; that is, $\underline{b}(r_P,c;\alpha) = \overline{b}(r_P,c;\alpha) = \frac{1}{2+\alpha}$ and $q_I = 0$ (cf. Lemma 2). In that case, when the prior belief favors conviction ($b < \frac{1}{2+\alpha}$), the guilty type necessarily uses a mixed strategy in equilibrium. This follows because always making a precise statement would induce a posterior belief equal to the prior, and thus a payoff $0 - \lambda_P \leq 0$. If instead the guilty type would always make a vague statement, the judge would acquit for sure after a precise statement given that then $b^P = 1$, providing the guilty type a strong incentive to deviate (given $\lambda_P < 1$ in the case considered here). The mixed strategy the guilty type employs in equilibrium makes the judge indifferent between convict and acquit after receiving a precise statement. Vice versa, the judge's equilibrium probability of acquittal equal to λ_P makes the guilty

type indifferent between the two statements. This yields equilibrium Inf.4 in which some information is revealed only via the statements per se.

Only when the investigation costs are sufficiently low (i.e., $c < \hat{c}\,(r_P; \alpha)$), the judge may potentially want to investigate after a precise statement. If she would always do so, then by Lemma 3 the guilty type would be deterred from making a precise statement iff

$$\lambda_P > \hat{\lambda}\,(r_P, \pi_P) \quad (\equiv 1 - r_P(1 + \pi_P))\,. \tag{4}$$

If condition (4) holds, the guilty type is effectively deterred away from *always* lying. In equilibrium he then only does so occasionally (i.e., $0 < p < 1$) as to ensure the judge will mix between acquitting and investigating after a precise statement. This yields Inf.1 in which both information sources are drawn upon.

When condition (4) is not met, the guilty type prefers to make a precise statement even if such a statement would always be verified. If the guilty type would indeed always lie, the judge's posterior belief b^P equals prior b and always verifying a precise statement is a best response only in case the prior is intermediate, that is, if $\underline{b}(r_P, c; \alpha) < b < \overline{b}(r_P, c; \alpha)$. This gives pooling equilibrium Inf.2 in which the judge only obtains additional information via investigation. If instead the prior is low and favors conviction ($b < \underline{b}(r_P, c; \alpha)$), the guilty type necessarily employs a mixed strategy. Loosely put, the best he can do is then to convince the judge to not always convict but occasionally investigate instead. This yields Inf.3.

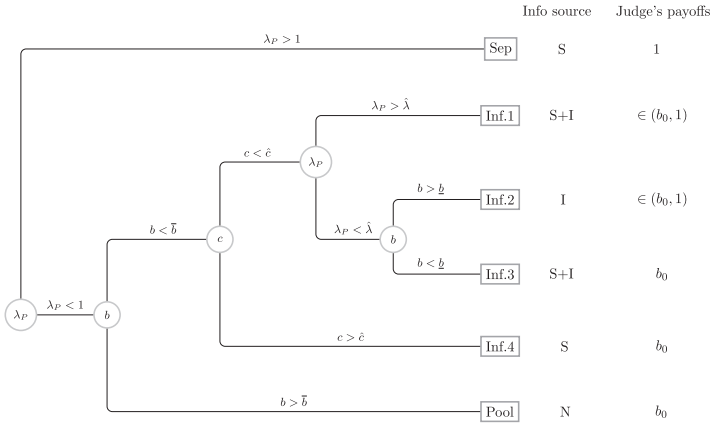
In the partially pooling equilibria Inf.1 and Inf.3, the judge's two information sources complement each other. In both equilibria, strategic information revelation by the suspect allows the judge to update her belief that he is innocent upwards ($b^P > b$) after having received a precise statement. This induces her to now and then verify such a statement and, if she indeed does so, to acquit if verified and convict if falsified. The two equilibria differ in what happens if the judge does not investigate though: acquittal in case of Inf.1 and conviction in case of Inf.3. Therefore, while in Inf.3 a *verified* precise statement is necessary to get acquitted, in Inf.1 an *unchecked* precise statement already suffices. The comparative statics of how the equilibrium behavior of the guilty type (i.e., p) and the judge (i.e., q_I) varies with the characteristics of the verification technology as reflected by λ_P, r_P, π_P, and c, is also opposite in the two equilibria (cf. the next subsection).

Condition (4) intuitively captures the deterrence effect of the potential verification of precise statements and the verification technology more broadly. Investigation becomes a stronger threat the more reliable it is (higher r_p) and the higher the obstruction penalty π_P becomes. This deterrence effect comes on top of the cognitive costs λ_P of formulating a precise (but false) statement, effectively creating an interdependence between the two. The higher these cognitive costs, the lower r_P and π_P can be for condition (4) still to be met. Note, for instance, that even in the absence of an obstruction penalty ($\pi_P = 0$), the condition still holds as long as the direct lying costs are high enough: $\lambda_P > 1 - r_P$ ($= \hat{\lambda}(r_P, 0)$). The cognitive costs thus play a supporting role for information revelation even when full separation cannot be achieved (i.e., when $\lambda_P < 1$). Also note from condition (4) that a high reliability r_P is by itself not a sufficient deterrent for the guilty type to refrain from always making a precise statement. It should be complemented with either sufficient cognitive costs of lying ($\lambda_P > 0$) or a sufficiently high obstruction penalty ($\pi_P > 0$) for the guilty type to be discouraged to always mimic the innocent type.

An illustrative summary of the conditions under which each equilibrium arises is provided in tree form in Figure 2. The tree splits into separate branches with respect to how the values of the lying cost λ_P, the prior belief b and the investigation cost c compare to the relevant thresholds. For generic parameter values, there is a unique equilibrium outcome; the labels in the boxes just refer to the equilibria listed in Propositions 1 and 2. The tree is augmented with two additional columns: the information source drawn upon along the equilibrium path (statements per se or investigation) and the expected equilibrium payoffs of the judge. The latter are explained in the next subsection, where we explore in detail how the amount of (valuable) information revelation and the effective reliance on different information sources varies with the characteristics of the verification technology, as captured by parameters λ_P, r_P, π_P, and c.

4.2. Improvements in the Investigation Technology and Valuable Information Revelation

4.2.1. *The effect of information on the judge's equilibrium payoffs.* The judge does not want to obtain just any information per se, but rather influential information that is instrumental to her decision. The effective



Info source: S=statement, I=investigation, S+I=statement+investigation, N=no information
Abbreviations: $\overline{b} = \overline{b}(r_P, c; \alpha)$, $\underline{b} = \underline{b}(r_P, c; \alpha)$, $\hat{c} = \hat{c}\,(r_P; \alpha)$, $\hat{\lambda} = \hat{\lambda}\,(r_P, \pi_P)$

Figure 2. All Equilibria with Conditions for Existence and Payoffs to the Judge
Info source: S = statement, I = investigation, S + I = statement+investigation, N = no information
Abbreviations: $\overline{b} = \overline{b}(r_P, c; \alpha)$, $\underline{b} = \underline{b}(r_P, c; \alpha)$, $\hat{c} = \hat{c}\,(r_P; \alpha)$, $\hat{\lambda} = \hat{\lambda}\,(r_P, \pi_P)$

value of such information can be inferred from how her payoffs are affected. In the absence of any additional information beyond her prior belief, the best the judge can achieve in terms of expected payoffs is

$$b_0 \equiv \max\{1 - b(1 + \alpha), b\},$$

that is, the best from either convicting or acquitting for sure.[19] Relative to this, getting additional information will always make her weakly better off. If the judge would always take the right decision (without bearing further investigation costs), she would get her maximum payoffs equal to 1. Proposition 3 ranks (for a given level of b) the payoffs of the judge in the various equilibria by comparing these with the lower bound b_0 and the upper bound of 1.

PROPOSITION 3 For the judge's equilibrium payoffs it holds that:

19. The analysis presented in this subsection applies for any $\alpha \geq 0$, thus also for $\alpha = 0$. This effectively implies that the exact same conclusions are obtained if we just focus on the probability of taking the correct decision instead, rather than on the judge's payoff function (which weighs taking the correct decision differently in different eventualities).

(a) In Sep the judge's expected payoffs equal the upper bound of 1;
(b) In Inf.1 and Inf.2 the judge's expected payoffs are strictly in between b_0 and 1. Holding prior belief b constant, the judge earns strictly more in Inf.1 than in Inf.2;
(c) In Inf.3, Inf.4 and Pool the judge's expected payoffs equal the lower bound b_0.[20]

Proof. The equilibrium payoffs in Sep and thus part (a) follow immediately. In Pool the judge always acquits and obtains b in expected payoffs. This equals b_0 for the range $b > \overline{b}(r_P, c; \alpha) \geq \frac{1}{2+\alpha}$ where Pool exists. In equilibria Inf.3 and Inf.4 the judge chooses $q_C > 0$. Conviction is thus always a best response (for the given equilibrium behavior p of the guilty type) and equilibrium payoffs for the judge coincide with those of always choosing conviction for sure, that is, $1 - b(1 + \alpha)$. This corresponds to b_0 under the conditions of existence for these equilibria, which require $b < \frac{1}{2+\alpha}$. This yields part (c).

The equilibrium payoffs in Inf.2 equal $r_P - b(1 - r_P)\alpha - c$. From Lemma 2 it immediately follows that these strictly exceed b_0 on the range $\underline{b}(r_P, c; \alpha) < b < \overline{b}(r_P, c; \alpha)$ where this equilibrium exists. With $\overline{b}$ as a shorthand for $\overline{b}(r_P, c; \alpha)$, these payoffs can be rewritten as $r_P - b(1 - r_P)\alpha - c = r_P - c - b[(1 - r_P)\alpha + 1] + b = \left(\frac{\overline{b}-b}{\overline{b}}\right) \cdot [r_P - c] + b$. Given $q_A > 0$ in Inf.1 and thus acquit being a best response (taking the equilibrium p as given), the judge's equilibrium payoffs there coincide with always acquitting after a precise statement. In that case, the judge only arrives at the wrong verdict if the suspect is indeed guilty and makes a precise statement, which happens with probability $(1 - b)p$. Hence, the judge's expected payoffs in Inf.1 equal $1 - (1 - b)p$, with $p = \frac{b}{1-b} \cdot \frac{(1-r_P)(1+\alpha)+c}{r_P-c} = \left(\frac{b}{1-b}\right) \cdot \left(\frac{1-\overline{b}}{\overline{b}}\right)$ from Proposition 2. Rewriting gives expected payoffs of $\left(\frac{\overline{b}-b}{\overline{b}}\right) + b$ in Inf.1. From $r_P - c < 1$, it follows that these strictly exceed the payoffs in Inf.2 derived above. This gives part (b). □

20. Note that, although they all reach lower bound b_0, equilibrium Pool on the one hand and Inf.3 and Inf.4 on the other hand cannot all occur for a given level of b; Pool requires $b > \frac{1}{2+\alpha}$ and thus $b_0 = b$, while Inf.3 and Inf.4 require $b < \frac{1}{2+\alpha}$ and thus $b_0 = 1 - b(1 + \alpha)$.

As Proposition 3 reveals the judge's payoffs are equal to lower bound b_0 in Inf.3, Inf.4, and Pool. This is immediate in the pooling equilibrium Pool in which she does not get any additional information from either the statements per se or the verification thereof. Yet, perhaps somewhat surprisingly, strategic information revelation and potential verification do not guarantee higher payoffs to the judge, as Inf.3 and Inf.4 exemplify. In equilibrium Inf.4 the guilty suspect mixes between a vague and a precise statement. This information is influential because it affects the choice the judge makes, but effectively immaterial as her expected payoffs do not improve. A similar observation holds with respect to Inf.3. Here, the guilty suspect again mixes between a vague and a precise statement and the judge occasionally investigates the latter. Despite both the statements per se and the investigation revealing influential (i.e., decision relevant) information, the judge again gains nothing in expected payoffs terms (as the benefits of a better verdict cancel out against the investigation costs c borne).

The judge does strictly improve upon deciding on her prior belief in the remaining equilibria. In separating equilibrium Sep she does so to the fullest extent possible and obtains her maximum payoff equal to 1. The incremental value $1 - b_0$ of the information received can be solely attributed to the statements per se. In Inf.1 and Inf.2, the judge also strictly benefits from the additional information obtained, albeit to a smaller extent. Holding prior belief b constant, the judge's expected payoffs are higher in Inf.1 than in Inf.2 (and, similarly so, higher in Inf.1 than in Inf.3 and Inf.4 for a given b). The intuition here is that in both equilibria it is a best response for the judge to investigate a precise statement, making that the judge is equally well off if such a statement is indeed received. Yet only in Inf.1, the guilty type now and then sends a vague statement ($0 < p < 1$ vs. $p = 1$ in Inf.2) and the judge does strictly better in those instances. Overall, in Inf.1 the judge thus obtains valuable information via both the statements per se as well as from (occasional) investigation, while in Inf.2 the judge only obtains valuable information via investigation.

4.2.2. *Comparative statics in the verification technology* For generic parameter values, there is a unique equilibrium outcome. Taking the prior level of evidence (as captured by b) and thus the extent of the investigation problem as given, the judge may benefit from shifts in the parameters that characterize the verification technology. These may either induce a

shift towards a "better" equilibrium as ranked in Proposition 3, or improve the judge's expected payoffs within a given equilibrium. Proposition 4 formally characterizes both these extensive and intensive margin (comparative statics) effects.

PROPOSITION 4 Shifts in the parameters $(\lambda_P, r_P, \pi_P, c)$ of the verification technology may have both intensive margin (within equilibrium) and extensive margin (shift to a different equilibrium) effects on the judge's equilibrium payoffs.

(a) Shifts in λ_P and π_P have extensive margin effects only. An increase in λ_P makes a beneficial shift towards either Sep or Inf.1 more likely, while an increase in π_P makes a beneficial shift towards Inf.1 more likely;

(b) Shifts in r_P and c have both extensive and intensive margin effects:
(Ext) both an increase in r_P and a decrease in c make a beneficial shift towards either Inf.1 or Inf.2 more likely;
(Int) the judge's payoffs within Inf.1 and Inf.2 are increasing in r_P and decreasing in c.

Proof. From Proposition 3 the judge's equilibrium payoffs in Int.3, Int.4, and Pool equal b_0 and those in Sep equal 1. These payoffs are independent of $(\lambda_P, r_P, \pi_P, c)$. Hence intensive margin effects only concern Inf.1 and Inf.2. The judge expected equilibrium payoffs in Inf.2 equal $r_P - b\,(1 - r_P)\,\alpha - c$ and thus increase with r_P, decrease with c and are independent of λ_P and π_P. From the proof of Proposition 3, the equilibrium payoffs in Inf.1 equal $1 - (1 - b)p = 1 - b \cdot \frac{(1-r_P)(1+\alpha)+c}{r_P - c}$. Also these increase with r_P and decrease with c and are independent of λ_P and π_P. This yields the claims about intensive margin effects in part (a) and (b.Int).

The extensive margin effects in both part (a) and (b.Ext) follow from the payoff ranking of equilibria in Proposition 3, together with the conditions for existence in Propositions 1 and 2, and the comparative statics of the relevant thresholds in $(\lambda_P, r_P, \pi_P, c)$. In particular, $\underline{b}(r_P, c; \alpha)$ decreases with r_P and increases with c, $\overline{b}(r_P, c; \alpha)$ increases with r_P and decreases with c, $\hat{c}\,(r_P; \alpha)$ increases with r_P and $\hat{\lambda}\,(r_P, \pi_P)$ decreases with both r_P and π_P. □

Clearly, the judge would prefer a verification technology that allows for the highest ranked equilibrium as identified in Proposition 3. She thus

would prefer the cognitive lying costs λ_P to be prohibitively high for the guilty type, such that Sep materializes. Otherwise it would be best for her to have low verification costs c and "sorting" condition (4) to be satisfied, as to enable Inf.1. In meeting condition (4) a high λ_P is again conducive, but also a sufficiently harsh obstruction penalty π_P helps (because threshold $\hat{\lambda}(r_P, \pi_P)$ decreases with π_P). Beyond their extensive margin effects of enabling Inf.1, however, an increase in either λ_P or π_P provides no additional benefits. The main intuition here is that the guilty type's statement strategy within a given equilibrium (as reflected by p) does not vary with these parameters and hence neither do the judge's equilibrium payoffs.

In contrast, variations in r_P and c do have both extensive, as well as intensive margin—that is, within equilibrium—effects. The extensive margin effects follow from how compliance with the relevant thresholds $\underline{b}(r_P, c; \alpha)$, $\overline{b}(r_P, c; \alpha)$, $\hat{c}(r_P; \alpha)$, and $\hat{\lambda}(r_P, \pi_P)$ is affected. Both an increase in r_P and a decrease in c facilitate a beneficial shift from a lower ranked equilibrium towards either Inf.2 or Inf.1 (including for r_P a potential shift from Inf.2 to Inf.1). The intensive margin effects derive from two different causes. In equilibrium Inf.2, they follow from how changes in r_P and c affect the cost-effectiveness of the actual verification process itself. To illustrate, the judge's expected payoffs in Inf.2 can be decomposed as:[21]

$$r_P - b(1 - r_P)\alpha - c = \underbrace{b_0}_{\text{prior}} + \underbrace{0}_{\text{statements per se}} + \underbrace{[r_P - b(1 - r_P)\alpha - b_0]}_{\text{verification}} - c.$$

Since both suspect types always make a precise statement in Inf.2, observing such a statement per se does not provide any information and, thus, has zero incremental value. Relative to deciding without first verifying, which would yield b_0, verification of the precise statement received has two opposing effects. On the one hand, it improves decision making if it corrects a would-be wrong verdict based on the prior belief alone.

21. To intuitively understand the expected payoffs on the l.h.s., note that in Int.2 the judge always verifies and thus always bears cost c. She arrives at a correct verdict and thus a payoff of 1 with probability r_P. With the remaining probability $(1 - r_P)$ she takes the wrong decision, with negative payoffs $-\alpha$ (only) if she wrongly convicts an innocent suspect (whose frequency of occurrence in the population is b).

On the other hand, it worsens decision making in those instances where it wrongly overturns a would-be correct verdict based on b alone. The overall net informational effect—reflected within square brackets—is positive and outweighs the costs of verification c. This informational value of verification increases with r_P and is independent of (λ_P, π_P, and) c.

In contrast, in equilibrium Inf.1, the benefits from improvements in the investigation technology via r_P and c effectively follow entirely from their spill-over effects on the strategic behavior of the guilty type, because such improvements induce him to mimic the innocent type less often. To illustrate, the effective reliance on the different information sources can again be inferred from the decomposition of the judge's equilibrium payoffs:[22]

$$1-(1-b)\,p = \underbrace{b_0}_{\text{prior}} + \underbrace{(1-\sigma_P)+b-b_0}_{\text{statements per se}} + \underbrace{\sigma_P q_I\left[r_P - b^P(1-r_P)\alpha - b^P\right]}_{\text{verification}} - \sigma_P q_I c,$$

where $\sigma_P \equiv b + (1-b)\,p$ denotes the overall probability that a precise statement is made in equilibrium. The final two terms cancel out, reflecting that in Inf.1 the judge is indifferent between verifying a precise statement and immediately acquitting for sure. Based on just the statements per se, the judge would convict after a vague statement and acquit after a precise one, yielding $(1-\sigma_P)+b$ in expected payoffs. The incremental value $(1-\sigma_P)+b-b_0$ increases with r_P and decreases with c (and is independent of λ_P and π_P). This reflects the indirect, "deterrence" effect of potential verification. If verification becomes either more reliable or less costly, it deters the guilty type from mimicking the innocent type often, that is, it lowers p and thus σ_P. This in turn makes a precise statement per se more informative. The direct informational value of now and then checking on a precise statement made equals $\sigma_P q_I\left[r_P - b^P(1-r_P)\alpha - b^P\right]$. This informational value *decreases* with r_P and *increases* with c.[23]

22. As explained in the proof of Proposition 3, given the judge's indifference in Inf.1 between convict and acquit after a precise statement, her equilibrium payoffs coincide with those of always acquitting after such a statement (keeping the equilibrium p fixed). In that case the judge only arrives at a wrong verdict if the suspect is guilty and makes a precise statement, which happens with probability $(1-b)\,p$. In all other instances she takes the right decision, yielding 1. This explains her expected payoffs $1-(1-b)p$ in Inf.1.

23. Increases in cognitive lying costs λ_P or obstruction penalty π_P have no intensive margin effects in Inf.1 as they bring no overall net benefits to the judge (cf.

Intuitively, if the verification technology becomes more reliable (higher r_P), overall more valuable information is obtained in Inf.1. But perhaps somewhat counter-intuitively, as the above decomposition reveals this beneficial impact is completely driven by the incremental benefits from the statements per se. *Less* valuable information is actually obtained from verification the higher r_P is. The driving force here is that actual verification occurs less often if r_P increases.[24] This reflects the general intuition that a more effective stick works as a stronger deterrent and thus in the end needs to be used less often. Similarly so, a decrease in c causes that also less valuable information is obtained from verification, both because precise statements are made less often by the guilty type and—as a result—their actual verification then yields less additional information.

In summary, higher cognitive costs of lying and a higher obstruction penalty are beneficial to the judge to the extent that these enable more informative equilibria in which also the suspect's statements per se provide valuable information. The latter requires that the guilty type is effectively deterred away from always lying. Once the relevant threshold for this is met, however, increasing the lying costs or the obstruction penalty further does not increase the provision of valuable information. Improvements in the verification technology that make it more reliable or less costly do have an impact beyond meeting the relevant threshold, however. Even if the guilty type is willing to always lie, such improvements make actual verification more cost effective (cf. Inf.2). And as soon as the threshold that deters the guilty type from always lying is met (cf. condition (4)), such improvements enlarge the deterrence effect. This creates a positive spill-over effect because the guilty type then reveals more information via the statement per se and the

---

Proposition 4). Such marginal changes do not impact the amount of (valuable) information the statements per se reveal in Inf.1, that is, do not strengthen the deterrence effect. This follows because in Inf.1 an increase in either λ_P or π_P reduces the frequency of actual verification q_I. The latter also implies that actually *less* (valuable) information is obtained from occasional verification. This is counterbalanced by incurring the costs of verification equally less often.

24. The informational value of actually verifying a precise statement received equals the term within square brackets $[r_P - b^P(1-r_P)\alpha - b^P]$. The judge's indifference in Inf.1 between whether or not to verify a precise statement implies that this term equals c and thus is independent of r_P. Comparative statics of the direct informational value of occasional verification w.r.t. r_P thus solely follow from how $\sigma_P q_I$ is affected; this term is strictly decreasing in r_P.

actual verification process itself actually yields less valuable information (cf. Inf.1).

5. Model Extensions

In this section, we discuss two extensions that add additional realism to the model: (i) incorporating the possibility of plea bargaining and (ii) accounting for a right to silence. The overall conclusion that follows from the discussion is that these extensions leave the main insights obtained from our basic setup largely unaffected.

5.1. Plea Bargaining

In practice, a very high percentage of cases—up to 95%, see US Bureau of Justice Statistics (2003)—never reach the courtroom and is settled through some sort of plea bargaining. In this case, the prosecutor offers a penalty reduction in exchange for the suspect pleading guilty. In the literature, plea bargaining has been studied as having (among other things) an informational role in the screening of suspect types.

To incorporate this realistic element in our setup, we allow a third option to the suspect: besides making either a vague or a precise statement and the case going to court, he can also choose to confess and immediately receive a payoff of m, with $0 < m < 1$. Confession then yields strictly more than providing a vague statement (inducing immediate conviction) does.[25] A direct implication of this added choice option is that in equilibrium the guilty type no longer provides a vague statement and effectively chooses between confessing and providing a precise statement only.[26]

25. The imposed sentence after confession thus leads to a lower payoff reduction than the imposed sentence after conviction without confession does: $1-m < 1$. Although in practice the prosecutor may have some discretion in the size of the penalty reduction offered, this discretion may be considerably restricted by binding guidelines, see for example, the 2017 "Reduction in sentence for a guilty plea: Definitive guideline" from the sentencing council in the UK (UK Sentencing Council, 2017). Existing game theoretical models of plea bargaining typically allow the prosecutor to endogenously choose the penalty reduction; qualitatively this leads to the same conclusions with respect to amount of information revelation in equilibrium, see the discussion below.

26. We maintain the assumption that innocent suspects always provide precise statements. Dropping this assumption leads to the existence of additional pooling equilibria where both types choose to confess. Analogously to the main model, such pooling

The single substantive difference with the equilibrium analysis in Section 4 is that the relevant benchmark for λ_P in Lemma 3 has to be adapted from $q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P)$ originally to $q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P) - m$ now. This accounts for the fact that the relative benefits of a potentially more favorable decision after a precise statement are now—compared to the new and better alternative of confession—an amount m lower than before. Put differently, the original benchmark $q_A + q_I \cdot \hat{\lambda}(r_P, \pi_P)$ now applies to $\lambda_P + m$ rather than just λ_P before. Acknowledging this, we immediately obtain the following corollary from our main analysis.

COROLLARY 1 Suppose that, besides making a statement $S \in \{P, V\}$, the guilty type can also Confess and immediately receive payoff m, with $0 < m < 1$. The guilty type then never chooses a vague statement. Propositions 1 through 4 immediately apply when we replace λ_P by $\lambda_P + m$ and let $1 - p$ now reflect the probability with which the guilty type confesses.

In the presence of plea bargaining, full separation (Sep) is achieved if $\lambda_P > 1 - m$. Similarly, the condition for potential verification of precise statements to have a sufficiently strong deterrent effect now becomes:

$$\lambda_P > \hat{\lambda}(r_P, \pi_P) - m \tag{5}$$

Compared to condition (4), the opportunity costs m of lying are subtracted from the r.h.s., to account for the fact that the benefits of a plea bargain (yielding m) are foregone if the guilty type decides to give a precise statement instead. With plea bargaining, the judge has an additional tool in trying to induce guilty suspects to come forward. The penalty reduction m complements cognitive lying costs λ_P and the obstruction penalty π_P in facilitating more informative equilibria. For the guilty type to refrain from always mimicking the innocent type, one can either make mimicking less attractive (i.e., a higher λ_P or π_P), or otherwise make the alternative of not mimicking more attractive (which is essentially what the plea bargain does). Beyond meeting the threshold for enabling information revelation via the statements per se, an increase in m has no beneficial impact though.

---

equilibria do not survive standard equilibrium refinements (Banks and Sobel, 1987). The maintained assumption essentially solves the multiplicity of equilibria issue in a simpler way.

In an early game theoretic analysis of plea bargaining, Grossman and Katz (1983) showed that—if a prosecutor could *commit* to proceed to court if the plea offer is rejected—the plea offer can be used as a screening device to fully separate the guilty types from the innocent ones. A similar observation was made by Reinganum (1988) when extending the framework of Grossman and Katz (1983) by assuming that the prosecutor has private information regarding the strength of the case. Baker and Mezzetti (2001) have challenged this equilibrium separation possibility, as the underlying commitment on which it is based "...is inherently noncredible because any defendant that the prosecutor knows for sure is innocent will never stand trial" Baker and Mezzetti (2001, p. 151). Models that drop this possibility to fully commit to go to trial all find that plea bargaining is (at most) essentially semi-separating, with the plea offer accepted by the guilty type with some probability but rejected by the innocent type for sure.[27] In this equilibrium, the prosecutor still proceeds to trial with probability one if the plea offer is rejected (although this is now based on an equilibrium best response rather than on an ex ante commitment as in the earlier papers). The partially pooling equilibria in our setup are qualitatively similar in terms of the guilty type using a mixed strategy, but differ in the judge/prosecutor doing so as well. This causes that if either λ_P, π_P, or m increases, only the judge adapts her behavior, while leaving the behavior of the suspect unaffected. Such changes thus do not have positive intensive margin (within equilibrium) effects (cf. Section 4.2).[28]

27. See Baker and Mezzetti (2001), Bjerk (2007), Kim (2010), and Tsur (2017). A remaining criticism of some of these models is that the behavior of the judge/jury is assumed to be purely exogenous and does not react to (the information revealed by) the behavior of the prosecutor and the suspect. This arguably provides another unrealistic commitment possibility, viz. to a mechanical conviction rule. Bjerk (2007) and Tsur (2017) endogenize the behavior of the judge/jury and obtain the same type of semi-separating equilibrium (though a multiplicity of these may exist). Note that our simplified setup with a unitary judiciary actor essentially corresponds to the case where different representatives of the judiciary share the same information and beliefs, and endogenously act on these; the probability of conviction is thus entirely the result of equilibrium strategies.

28. Although the obstruction penalty and the penalty reduction play a similar deterrence role in incentivizing the guilty type to sometimes either implicitly (via a vague statement) or explicitly confess, their payoff implications for the suspect are quite different. He is clearly better off with higher penalty reductions than with higher obstruction penalties. From a broader social welfare perspective, however, society might dislike

5.2. Right to Silence

In our baseline model, the judge can use both the strategic behavior of the suspect as well as the outcome of the potential investigation to update her belief about the suspect's innocence and act accordingly without any restrictions. In particular, the suspect's choice of making a vague statement can be fully held against him and lead to immediate conviction. Traditional common law systems, however, typically give the suspect the "right to remain silent"; if a suspect refuses to answer any questions, the verdict must solely be based on other evidence and the suspect's silence cannot be considered evidence of his guilt (*Miranda v.Arizona*, 1966). Effectively, this right thus works as a commitment to ignore some of the suspect's strategic information revelation.

To analyze the implications of a right to silence for our analysis, we again introduce a third option to the suspect: besides making a precise ($S = P$) or a vague ($S = V$) statement as in the baseline model, he can now also remain silent ($S = \phi$). Since no viable leads are obtained at all, a silent statement is even more difficult to investigate than a vague one, so we assume $\frac{1}{2} \leq r_\phi < r_V$.[29] Moreover, by remaining silent the suspect is not obstructing justice in any way (except perhaps in highly unusual circumstances), implying that $\pi_\phi = 0 \leq \pi_V$. We continue to assume that the innocent suspect always makes a precise statement. Therefore, observing $S = \phi$ is a clear indication of being guilty and the introduction of this additional option to the suspect per se has no impact on equilibrium outcomes if no further assumptions are made. Within our setup, prior belief b can be interpreted as the evidence collected by the judge before any statement is received. If silence cannot be held against the subject, this then constitutes all the evidence there is. We thus incorporate a right to silence in the following way.

---

penalty reductions as they allow offenders to largely "get away with it" and rather prefer penalties for obstruction (Fagan, 1981; Cohen and Doob, 1989; Herzog, 2003; Johnson, 2019). A reduced form way to incorporate such broader considerations in our model would be to let the judge's expected payoffs depend on m and π_P as well.

29. Clearly, with a silent statement there is nothing to verify. Investigation of a silent statement thus should be interpreted as additional independent investigation by the judge not inspired by the empty statement made.

ASSUMPTION RTS Under a right to silence the judge's choice of action after a silent statement $S = \phi$ should be guided by a restricted posterior belief $b^{\phi} = b$, rather than by a Bayesian posterior belief $b^{\phi} = 0$ that applies in the absence of such a right.

Similar to analysis in the previous subsection, the single substantive difference with the baseline model is that the benchmark payoffs to which the relative benefits of making a precise statement have to be compared are now (potentially) different. Note that with a RTS, making a vague statement is (weakly) dominated by remaining silent for the guilty type. The relevant benchmark payoffs are thus given by the judge's choice of action after $S = \phi$. For this we can immediately apply Lemma 2 when we replace r_P with r_ϕ and b^P with b. Hence, if $b < \underline{b}(r_\phi, c; \alpha)$, the judge convicts for sure after remaining silent and the equilibrium analysis coincides with the one of the baseline model. A RTS is then inconsequential. In case $b \in (\underline{b}(r_\phi, c; \alpha), \overline{b}(r_\phi, c; \alpha))$, a RTS effectively forces the judge to investigate in case of silence. By the equivalent of Lemma 1, the guilty type is then acquitted with probability $1 - r_\phi$ when keeping silent. This gives the guilty type an expected payoff of $1 - r_\phi$ after $S = \phi$, rather than 0. These opportunity costs $1 - r_\phi$ from making a precise statement now come on top of the direct lying costs λ_P, but apart from that the analysis is as before. Finally, if $b > \overline{b}(r_\phi, c; \alpha)$, a RTS ensures that the suspect is always acquitted after silence. As this also happens after a precise statement (made by the innocent type), the equilibrium outcome in terms of the judge's verdict is then the same as in equilibrium Pool in Proposition 1. From these observations we immediately obtain the following corollary.[30]

30. For listing the different equilibria that exist in the various subcases of part (b), we have used that $\left(\underline{b}\left(r_\phi, c; \alpha\right), \overline{b}\left(r_\phi, c; \alpha\right)\right) \subset \left(\underline{b}\left(r_P, c; \alpha\right), \overline{b}\left(r_P, c; \alpha\right)\right)$ given that $\underline{b}$ decreases with r, $\overline{b}$ increases with r and $r_\phi < r_P$. This also implies that $\left(\underline{b}\left(r_\phi, c; \alpha\right), \overline{b}\left(r_\phi, c; \alpha\right)\right)$ is empty if $\left(\underline{b}\left(r_P, c; \alpha\right), \overline{b}\left(r_P, c; \alpha\right)\right)$ is, that is, for $c > \hat{c}\left(r_P; \alpha\right)$. Moreover, for case (b.2) we have used that sorting condition (6) discussed below always holds.

COROLLARY 2 Suppose that, besides making a statement $S \in \{P, V\}$, the guilty type can also remain silent, that is, $S = \phi$, with $\frac{1}{2} \leq r_\phi < r_V < r_P$ and $\pi_\phi = 0 \leq \pi_V \leq \pi_P$.

(a) Without a RTS, Propositions 1 through 4 immediately apply when we let $1 - p$ now reflect the probability with which the guilty type either chooses $S = V$ or $S = \phi$ (both leading to immediate conviction);

(b) With a RTS, the guilty type never chooses a vague statement. Letting $1 - p$ now reflect the probability with which the guilty type chooses $S = \phi$, it then holds that:

(b.1) if $b < \underline{b}\left(r_\phi, c; \alpha\right)$, then Propositions 1 through 4 continue to apply (for b in this range) and we either have Sep, Inf.1, Inf.2, Inf.3, or Inf.4;

(b.2) if $\underline{b}\left(r_\phi, c; \alpha\right) < b < \overline{b}\left(r_\phi, c; \alpha\right)$, then Propositions 1 through 4 continue to apply (for b in this range) when we replace λ_P with $\lambda_P + \left(1 - r_\phi\right)$ and we either have Sep or Inf.1. The judge now always investigates in case of silence;

(b.3) if $b > \overline{b}\left(r_\phi, c; \alpha\right)$, then the guilty suspect always remains silent and is always acquitted (outcome equivalent to equilibrium Pool).

In case (b.2) of Corollary 2, potential verification of precise statements is a sufficiently powerful deterrent if:

$$\lambda_P > \hat{\lambda}\left(r_P, \pi_P\right) - \left(1 - r_\phi\right) \quad \left(= r_\phi - r_P(1 + \pi_P)\right) \tag{6}$$

From $r_\phi < r_P$ the r.h.s. is negative. The condition is thus always satisfied, irrespective of λ_P and π_P. Therefore, only equilibria that are equivalents of Sep and Inf.1 remain to exist (and Inf.2 and Inf.3 disappear), which correspond to these after replacing λ_P with $\lambda_P + (1 - r_\phi)$. The probability p with which the guilty type makes a precise statement stays exactly the same as without a RTS, but the judge now always investigates after a silent statement and immediately acquits after a precise statement with increased probability $q_A = \frac{\lambda_P + (1 - r_\phi) - \hat{\lambda}(r_P, \pi_P)}{1 - \hat{\lambda}(r_P, \pi_P)}$ in Inf.1.

The above shifts in equilibria are in line with the effects of a right to silence identified by the game theoretic analyses of Seidmann (2005) and

Leshem (2010) (see also Seidmann and Stein, 2000; Mialon, 2005). In particular, the innocent type benefits from such a right in two ways. A first, direct benefit is that it provides "innocent suspects, who are otherwise compelled to speak, with the alternative of silence" (Leshem, 2010, page 400). In our simplified setup this effect is reflected by the nonexistence of the informative equilibria in case $b > \overline{b}(r_\phi, c; \alpha)$; the judge is then compelled to acquit in the absence of further information and only an outcome equivalent to Pool remains. In general, exercising the right to silence provides the innocent type a safe alternative to making a precise statement, as with the latter he runs the potential risk of his statement being wrongly falsified. A second, indirect benefit is that innocent types who choose to make a precise statement are less likely to be wrongfully convicted. This effect is exemplified by the increased probability of immediate acquittal q_A in Inf.1 above.

More generally, Corollary 2 reveals that if $b < \underline{b}\left(r_\phi, c; \alpha\right)$ a RTS is immaterial for the equilibrium payoffs of both subject types and the judge. For a sufficiently high prior belief $b > \overline{b}\left(r_\phi, c; \alpha\right)$ a RTS either increases the equilibrium payoffs of the innocent and the guilty type or leaves these unaffected. The opposite holds for the judge; she then either earns the same or loses (cf. Proposition 3). In the intermediate range where $\underline{b}\left(r_\phi, c; \alpha\right) < b < \overline{b}\left(r_\phi, c; \alpha\right)$ introducing a RTS again always (weakly) benefits the innocent and the guilty type. But, interestingly, the effect for the judge then can go either way. The typical case remains that the judge loses.[31] Yet the opposite may happen (only) when introducing a RTS induces

31. In case (b.2) of Corollary 2, we either have Sep, Inf.1 or Inf.2 in the absence of a RTS (note that $c < \hat{c}\,(r_P; \alpha)$ for the belief range to be nonempty). Now if Inf.2 applies, then introducing a RTS necessarily leads to a shift to Inf.1. This follows because $\lambda_P < \hat{\lambda}\,(r_P, \pi_P) \implies \lambda_P + 1 - r_\phi < 1$, given that $\hat{\lambda}\,(r_P, \pi_P) < \frac{1}{2}$ and $r_\phi \geq \frac{1}{2}$. This shift benefits both the innocent (as q_A increases) and the guilty type (now convicted with smaller probability $p \cdot r_P + (1-p) \cdot r_\phi < r_P$ and bearing λ_P less often), but harms the judge. She gets $(1-b) \cdot (1-p) \cdot (r_P - r_\phi)$ less in Inf.1 with a RTS as compared to Inf.2 without. (As $q_I > 0$ in Inf.1 and thus investigation a best response, the judge's equilibrium payoff can be calculated as if $q_I = 1$. In that case only the outcome for a guilty type is different from Inf.2 in the instances that he now remains silent in Inf.1.) In case Sep applies without a RTS, it continues to apply with a RTS. This leaves the innocent type unaffected, benefits the guilty type (given now investigation after silence) but harms the judge (as the guilty type is now sometimes acquitted). The same—guilty wins, judge loses, innocent unaffected—holds if Inf.1 applies both without and with a

a shift from Inf.1 to Sep. The induced change in the guilty type's behavior then makes that the judge can convict him with probability r_ϕ under a RTS at investigation costs c to her, compared to probability $1-p$ before (where p is given in Proposition 2 for Inf.1). Depending on parameter values, we either have $r_\phi - c < 1-p$ or $r_\phi - c > 1-p$.[32] In the latter case the judge is strictly better off in Sep under a RTS than in Inf.1 without a RTS.[33] Unlike Seidmann (2005) and Leshem (2010), therefore, we do obtain instances in which the judge explicitly benefits from an ex ante commitment to block adverse (but correct!) inferences from silence.

Most important for our purposes, however, is that the qualitative features of the informative equilibria are robust to introducing a right to silence. Although such a right diminishes the role for strategic information revelation when the judge is initially inclined to acquit, this role essentially remains the same when this is not the case. Strategic information revelation via the statements per se thus continues to play an important role in affecting the judge choice of action and remains complementary to the judge now and then checking on messages. Moreover, a right to silence reinforces the attractiveness of improvements in reliability, as neither the lying costs nor the obstruction penalty have a supportive role when the judge is a priori insufficiently confident about what the appropriate verdict would be. The verifiability approach to lie detection thus continues to have a strong bite even in the presence of a right to silence.

6. Conclusion

In this article, we analyze the strategic interaction between a speaker who wants to convince an investigator of his innocence and an investigator who wants to know the truth, that is, whether the speaker is guilty or

RTS. This leaves the case where Inf.1 applies without a RTS and Sep with a RTS, which is discussed in the main text.

32. To illustrate, consider the following numerical example. Let $b = \frac{1}{3}$, $\alpha = 1$, $\lambda_P = \frac{3}{5}, c = \frac{1}{20}, \pi_P = 0$ and $r_\phi = \frac{11}{20}$. Then $\underline{b}\left(r_\phi, c; \alpha\right) \approx 0,323$ and $\overline{b}\left(r_\phi, c; \alpha\right) \approx 0,345$ and thus case (b.2) indeed applies. For these parameters, $r_\phi - c = \frac{1}{2}$. Now for $r_P > \frac{7}{10}$ it holds that $1-p > \frac{1}{2}$ and for $r_P < \frac{7}{10}$ that $1-p < \frac{1}{2}$.

33. Because $q_A > 0$ in Inf.1, the judge payoffs can be calculated as if $q_A = 1$ and only the outcome for the guilty type is different in Inf.1 and Sep.

innocent. In our model, the investigator can check the specific details in the statement of the speaker at some cost. This yields informative, but imperfect evidence. The more detailed the speaker's statement is, the more reliable the examination of this statement becomes. This encourages innocent speakers to be forthcoming in providing many verifiable details in their statement, while guilty types would prefer to remain vague, also because fabricating a precise but false statement is cognitively costly to them. If, on the basis of an investigation, the investigator concludes that the speaker is lying, an additional obstruction penalty is imposed on the speaker.

We show that complete separation is possible only if lying is prohibitively costly to a guilty speaker. Full information revelation then takes place via the statements per se. With lower cognitive costs of fabricating a precise but false statement, the speaker's statement is partially revealing at best and provides valuable information only if its potential verification is a sufficiently strong deterrent. The latter requires that the joint effect of the lying costs, the reliability of verification, and the obstruction penalty is strong enough to potentially tip the balance in the trade-off for a guilty speaker. If this is indeed the case, a partially pooling equilibrium exists in which the guilty type mixes between making a vague and making a precise statement (with the innocent type making a precise statement for sure). Precise statements are now and then investigated by the investigator to verify their veracity. In this equilibrium verification and strategic information revelation by the speaker thus go hand in hand.

Our analysis allows us to understand the behavioral patterns observed for lie detection methods. It explains the shortcomings of the early approaches that were based on a speaker's microexpression of emotional cues that do not convey sufficient reliable information. In particular, our model explains why no beneficial information will be revealed in equilibrium when the observer's investigation is not sufficiently reliable and mimicking an innocent type is cognitively not prohibitively costly. For recent advances with the verifiability approach, the picture is more promising. By judging the frequency of precise verifiable details in a speaker's statement, more reliable information is acquired. In such settings our analysis suggests that a partial pooling equilibrium is most plausible. This equilibrium agrees with empirical observations, in which innocent types furnish their statements with precise, verifiable details, whereas guilty types face a difficult trade-off that

they solve by sometimes imitating the innocent types and by remaining vague at other times.

Our analysis also offers some insights that go beyond what has been observed in the recent psychological literature on lie detection. The overall amount of information provision in the partially pooling equilibrium is especially facilitated by an improved reliability of the verification technology. This renders verification more informative per se and (thus) makes the investigator more willing to investigate. Realizing this, the guilty type reduces the likelihood with which he makes a precise statement, in turn providing the investigator actually less incentives to investigate. The overall net effect is that, when reliability improves, more can be learned from the strategic behavior of the speaker and actually less is learned via actual verification. In contrast, not much is accomplished by enhancing the obstruction penalty further. Once the deterrence-by-verification condition for the existence of the partial pooling equilibrium is met, such an increase has no further impact on the usefulness of a lie detection method. An increase in the obstruction penalty then leads the investigator to investigate less, but leaves the amount of strategic information revelation unaffected. The investigator—and thus also "truth"—is better served by an improved reliability of the verification technology. A similar remark applies to the cognitive lying costs. These only have extensive margin effects in enabling more informative equilibria, but leave the amount of *valuable* information transmission within a given equilibrium unaffected.

In our approach, the quality of the verification technology is exogenous to the model. In practice, the relevant actors can make decisions that affect the quality of the investigation technology. The legal system may benefit from novel scientific insights and investments therein, such as in the area of the development of DNA identification or in the area of verbal detection methods. A judge can also order earlier searches of a suspect's house before any evidence is destroyed. Alternatively, a mother suspecting her son is using drugs can search his phone before asking him. After such actions, any statement made by the son or the suspect can be verified or falsified more accurately. However, such endogenous improvements in accuracy do not come as a free lunch. Searching her son's phone may destroy trust in the relationship between the mother and the son; sweeping a suspect's house before pressing charges may violate their right against unreasonable

searches and may be deemed as inadmissible evidence in court. To mitigate such adverse effects, a mother can only check the whereabouts of her son after hearing his explanations, and a judge can increase the number of witnesses to examine. Allowing the relevant actor to endogenously decide the scope of investigation and taking the adverse effects of the increase in accuracy into account goes beyond the scope of our article, but constitutes a fruitful direction for future research.

Appendix A: Proof of Proposition 2

Proof. Let $\lambda_P < 1$ and $b < \overline{b}(r_P, c; \alpha)$. Observe first that then $p = 0$ cannot happen in equilibrium; this would induce $q_A = 1$ by Lemma 2, in turn providing the guilty type an incentive to deviate to $p = 1$ per Lemma 3. Hence, necessarily either $p > 0$ or $p = 1$.

We next consider the various mutually exclusive parameter ranges in turn. First consider the case $c > \hat{c}(r_P; \alpha)$. From Lemma 2 then $q_I = 0$ necessarily and thus $q_A = 1 - q_C$. Now suppose $p = 1$. Then we would have $b^P = b < \overline{b}(r_P, c; \alpha) = \frac{1}{2+\alpha}$ and thus $q_A = 0$ as well. But for $q_C = 1$ the guilty type would want to choose $p = 0$ by Lemma 3, contradicting $p = 1$.[34] Hence, $0 < p < 1$ necessarily. The required indifference of the guilty type between making a vague and a precise statement then implies for q_A that:

$$0 = q_A - \lambda_P \implies q_A = \lambda_P$$

In turn, $0 < q_A < 1$ requires that the judge is indifferent between acquit and convict after $S = P$. From Lemma 2 and equation (1), we then obtain that:

$$b^P = \frac{1}{2+\alpha} \implies p = \frac{b}{1-b} \cdot (1+\alpha)$$

This yields equilibrium Inf.4.

From now on assume $c < \hat{c}(r_P; \alpha)$. In that case $\underline{b}(r_P, c; \alpha) < \frac{1}{2+\alpha} < \overline{b}(r_P, c; \alpha)$. From Lemma 2, this implies that the judge may potentially mix

34. In the nongeneric case, $\lambda_P = 0$ the guilty type is willing to choose $p > 0$ even when $q_C = 1$. Then multiple equilibria exist, which are all payoff and outcome equivalent to Inf.4 (with $q_A = 0$) as derived here.

in equilibrium between two options (from convict, investigate and acquit) at most, because b^P cannot meet more than one of these different thresholds at the same time.[35]

Consider the case where $\lambda_P > \hat{\lambda}(r_P, \pi_P)$. Suppose $p = 1$. Then we would have $b^P = b < \overline{b}(r_P, c; \alpha)$ and thus $q_A = 0$ from Lemma 2. But then the guilty type would want to choose $p = 0$ per Lemma 3, a contradiction. Hence, $0 < p < 1$ necessarily. The required indifference of the guilty type then implies by the same lemma:

$$0 = q_A + (1 - q_A) \cdot \hat{\lambda}(r_P, \pi_P) - \lambda_P \implies q_A = \frac{\lambda_P - \hat{\lambda}(r_P, \pi_P)}{1 - \hat{\lambda}(r_P, \pi_P)}$$

In turn, $0 < q_A < 1$ requires that the judge is indifferent between acquit and investigate after $S = P$. From Lemma 2 and equation (1), we then obtain that:

$$b^P = \overline{b}(r_P, c; \alpha) \implies p = \frac{b}{1 - b} \cdot \frac{(1 - r_P)(1 + \alpha) + c}{r_P - c}$$

This yields equilibrium Inf.1.

Finally, consider the case where $\lambda_P < \hat{\lambda}(r_P, \pi_P)$ (besides $c < \hat{c}(r_P; \alpha)$). First assume $b > \underline{b}(r_P, c; \alpha)$. From $b^P \geq b$ by equation (1) it then follows that $q_C = 0$ by Lemma 2. In turn, by Lemma 3, we obtain that $p = 1$ necessarily. Hence, $b^P = b$ and $q_I = 1$ by Lemma 2. This yields equilibrium Inf.2.

Next assume $b < \underline{b}(r_P, c; \alpha)$. Suppose $p = 1$. Then, we would have $b^P = b < \underline{b}(r_P, c; \alpha)$ and thus $q_C = 1$ by Lemma 2. But then the guilty type would want to choose $p = 0$ per Lemma 3, a contradiction. Hence, $0 < p < 1$ necessarily. The required indifference of the guilty type then implies:

$$0 = q_I \cdot \hat{\lambda}(r_P, \pi_P) - \lambda_P \implies q_I = \frac{\lambda_P}{\hat{\lambda}(r_P, \pi_P)}$$

35. Mixing between all three options convict, investigate and acquit would require both $b^P = \frac{1}{2+\alpha}$ to make the judge indifferent between convict and acquit, as well as $c = \hat{c}(r_P; \alpha)$ to ensure indifference with investigate. This thus can happen in nongeneric knife-edge cases only.

In turn, $0 < q_I < 1$ requires that the judge is indifferent between investigate and convict after $S = P$. From Lemma 2 and equation (1), we then obtain that:

$$b^P = \underline{b}\,(r_P, c; \alpha) \implies p = \frac{b}{1-b} \cdot \frac{r_P(1+\alpha) - c}{1 - r_P + c}$$

This yields equilibrium Inf.3. □

References

Agranov, M., and A. Schotter, 2012. "Ignorance is bliss: an experimental study of the use of ambiguity and vagueness in the coordination games with asymmetric payoffs," 4 *American Economic Journal: Microeconomics* 77–103.

Baker, S., and C. Mezzetti. 2001. "Prosecutorial resources, plea bargaining, and the decision to go to trial," 17 *Journal of Law, Economics, and Organization* 149–67.

Balbuzanov, I. 2019. "Lies and consequences: the effect of lie detection on communication outcomes," 48 *International Journal of Game Theory* 1203–40.

Banks, J. S., and J. Sobel. 1987. "Equilibrium selection in signaling games," 55 *Econometrica* 647–61.

Bell, B. E., and E. F. Loftus. 1989. "Trivial persuasion in the courtroom: the power of (a few) minor details," 56 *Journal of Personality and Social Psychology* 669–79.

Bjerk, D. 2007. "Guilt shall not escape or innocence suffer? The limits of plea bargaining when defendant guilt is uncertain," 9 *American Law and Economics Review* 305–29.

Bond Jr, C. F., and B. M. DePaulo. 2006. "Accuracy of deception judgments," 10 *Personality and Social Psychology Review*, 214–34.

Cohen, A. 1992. *The Living Law: A Guide to Modern Legal Research*. Rochester, N.Y.: Lawyers Cooperative.

Cohen, S. A. and A. N. Doob. 1989. "Public attitudes to plea bargaining," 32 *Criminal Law Quarterly* 85–109.

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. 509 U.S. 579-601.

Decker, J. F. 2004. "The varying parameters of obstruction of justice in american criminal law," 65 *Louisiana Law Review* 49–130.

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. "Cues to deception," 129 *Psychological Bulletin* 74–118.

Dziuda, W., and C. Salas. 2018. Communication with detectable deceit. Working paper.

Ekman, P., W. V Friesen, and O'sullivan, M. 1988. "Smiles when lying," 54 *Journal of Personality and Social Psychology* 414–20.

Fagan, R. W. 1981. "Public support for the courts: an examination of alternative explanations," 9 *Journal of Criminal Justice* 403–17.

Glazer, J. and A. Rubinstein. 2012. "A model of persuasion with boundedly rational agents," 120 *Journal of Political Economy* 1057–82.

Glazer, J., and A. Rubinstein. 2014. "Complex questionnaires," 82 *Econometrica* 1529–41.

Grossman, G. M., and M. L. Katz. 1983. "Plea bargaining and social welfare," 73 *The American Economic Review* 749–57.

Harvey, A. C., A. Vrij, G. Nahari, and K. Ludwig. 2017. "Applying the verifiability approach to insurance claims settings: exploring the effect of the information protocol," 22 *Legal and Criminological Psychology* 47–59.

Herzog, S. 2003. "The relationship between public perceptions of crime seriousness and support for plea-bargaining practices in Israel: a factorial survey approach," 94 *Journal of Criminal Law and Criminology* 103–32.

Holm, H. 2010. "Truth and lie detection in bluffing," 76 *Journal of Economic Behavior and Organization* 318–24.

Ispano, A. and P. Vida. 2021. Designing interrogations. Working paper.

Jehiel, P. 2021. "Communication with forgetful liars," 16 *Theoretical Economics* 605–38.

Johnson, T. 2019. "Public perceptions of plea bargaining," 46 *American Journal of Criminal Law* 133–56.

Jupe, L. M., S. Leal, A. Vrij, and G. Nahari. 2017. "Applying the verifiability approach in an international airport setting," 23 *Psychology, Crime & Law* 812–825.

Kartik, N. 2009. "Strategic communication with lying costs," 76 *Review of Economic Studies* 1359–95.

Kassin, S. M., and R. J. Norwick. 2004. "Why people waive their miranda rights: the power of innocence," 28 *Law and Human Behavior* 211–21.

Kim, J.-Y. 2010. "Credible plea bargaining," 29 *European Journal of Law and Economics* 279–93.

Kleinberg, B., G. Nahari, and B. Verschuere. 2016. "Using the verifiability of details as a test of deception: a conceptual framework for the automation of the verifiability approach" in Tommaso Fornaciari, Eileen Fitzpatrick, and Joan Bachenko, eds., *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*. San Diego, CA: Association for Computational Linguistics. p. 18–25.

Leshem, S. 2010. "The benefits of a right to silence for the innocent," 41 *RAND Journal of Economics* 398–416.

Mialon, H. M. 2005. " An economic theory of the fifth amendment," 36 *RAND Journal of Economics* 833–48.

Miranda v.Arizona. 1966. 384 U.S. 436-545.

Nahari, G., A. Vrij, and R. P. Fisher. 2014a. "Exploiting liars' verbal strategies by examining the verifiability of details," 19 *Legal and Criminological Psychology* 227–39.

Nahari, G., A. Vrij, and R. P. Fisher. 2014b. "The verifiability approach: countermeasures facilitate its ability to discriminate between truths and lies," 28 *Applied Cognitive Psychology* 122–28.

O'Connor, C. 2014. "The evolution of vagueness," 79 *Erkenntnis* 707–27.

Reinganum, J. F. 1988. " Plea bargaining and prosecutorial discretion," 78 *The American Economic Review* 713–28.

Seidmann, D. J. 2005. "The effects of a right to silence," 72 *Review of Economic Studies* 593–614.

Seidmann, D. J., and A. Stein. 2000. "The right to silence helps the innocent: a game-theoretic analysis of the fifth amendment privilege," 114 *Harvard Law Review*, 430–510.

Serra-Garcia, M., E. Van Damme, and J. Potters. 2011. "Hiding an inconvenient truth: lies and vagueness," 73 *Games and Economic Behavior* 244–61.

Sobel, J. 2020. "Lying and deception in games," 128 *Journal of Political Economy* 907–47.

Sorochinski, M., M. Hartwig, J. Osborne, E. Wilkins, J. Marsh, D. Kazakov, and P. A. Granhag. 2014. "Interviewing to detect deception: when to disclose the evidence," 29 *Journal of Police and Criminal Psychology* 87–94.

Suchotzki, K., B. Verschuere, B. Van Bockstaele, G. Ben-Shakhar, and G. Crombez. 2017. "Lying takes time: a meta-analysis on reaction time measures of deception," 143 *Psychological Bulletin*, 428.

Tsur, Y. 2017. "Bounding reasonable doubt: implications for plea bargaining," 44 *European Journal of Law and Economics* 197–216.

UK Sentencing Council. 2017. *Reduction in Sentence for a Guilty Plea*. United Kingdom Department of Justice. https://www.sentencingcouncil.org.uk/wp-content/uploads/Reduction-in-Sentence-for-Guilty-Plea-definitive-guideline-SC-Web.pdf.

US Bureau of Justice Statistics. 2003. *Sourcebook of Criminal Justice Statistics*. United States Department of Justice. https://www.ncjrs.gov/pdffiles1/Digitization/208756NCJRS.pdf.

US Sentencing Commission. 2018. *Guidelines Manual*. United States Department of Justice. https://www.ussc.gov/sites/default/files/pdf/guidelines-manual/2018/GLMFull.pdf.

Verschuere, B. and S. Shalvi. 2014. "The truth comes naturally! Does it?," 33 *Journal of Language and Social Psychology* 417–23.

Vrij, A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*. Wiley.

Vrij, A. 2018. "Verbal lie detection tools from an applied perspective" in *Detecting Concealed Information and Deception*. London, United Kingdom: Elsevier. p. 297–327.

Vrij, A. 2019. "Deception and truth detection when analyzing nonverbal and verbal cues," 33 *Applied Cognitive Psychology* 160–67.

Vrij, A., R. P. Fisher, and H. Blank. 2017. "A cognitive approach to lie detection: a meta-analysis," 22 *Legal and Criminological Psychology* 1–21.

Vrij, A., S. Mann, S. Kristen, and R. P. Fisher. 2007. "Cues to deception and ability to detect lies as a function of police interview styles," 31 *Law and Human Behavior* 499–518.

Zuckerman, M., B. M. DePaulo, and R. Rosenthal. 1981. "Verbal and nonverbal communication of deception," 14 *Advances in Experimental Social Psychology*, 1–59.