Engel, Christoph; Golder, Jasmin; Rahal, Rima-Maria

Working Paper

Who is afraid of the pink elephant? Character evidence, wiretapping, and debiasing interventions

Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2024/17

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Engel, Christoph; Golder, Jasmin; Rahal, Rima-Maria (2024) : Who is afraid of the pink elephant? Character evidence, wiretapping, and debiasing interventions, Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2024/17, Max Planck Institute for Research on Collective Goods, Bonn, https://hdl.handle.net/21.11116/0000-0010-3C9F-9

This Version is available at:
https://hdl.handle.net/10419/307698

MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

CHRISTOPH ENGEL
JASMIN GOLDER
RIMA-MARIA RAHAL

Discussion Paper
2024/17

WHO IS AFRAID OF THE PINK ELEPHANT? CHARACTER EVIDENCE, WIRETAPPING, AND DEBIASING INTERVEN-TIONS

Who is Afraid of the Pink Elephant?
Character Evidence, Wiretapping, and Debiasing Interventions

Christoph Engel*, Jasmin Golder‡ & Rima-Maria Rahal**

Abstract

Defendants should be judged on the merits of the case, not on prejudice, rumors, or evidence obtained through questionable methods. This is why criminal law of procedure regulates which information can be introduced in a trial. Two types of prohibited evidence are the criminal history of the defendant (the defendant shall not be considered more likely guilty since he had earlier been convicted for another crime), and information harvested from an unauthorized wiretap. In a series of online vignette experiments involving 1432 US participants, we show that character evidence never makes it significantly more likely that the defendant is judged guilty, whereas wiretap evidence has a strong effect. Various interventions aimed at debiasing the adjudicator have an effect, but this effect is insufficient to neutralize the bias.

JEL: C91, D02, D84, D91, K14, K41, K42

Keywords: criminal procedure, character evidence, wiretap, bias, debiasing

* Correspondence should be addressed to: Christoph Engel, Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Straße 10, 53113 Bonn, Germany, engel@coll.mpg.de
‡ University of Heidelberg
** Max Planck Institute for Research on Collective Goods and University of Heidelberg

The presumption of innocence is the cornerstone of the criminal justice system. The jury may only declare the defendant guilty if it has come to the conclusion that guilt is "beyond reasonable doubt". The jury shall reach its verdict based solely on the evidence presented during trial. In this spirit the Federal Rules of Evidence (2023) regulate the *admissibility of evidence*. Evidence is only considered admissible if it is relevant, reliable, and lawfully obtained (Garner & Black, 2014). This evidence forms the exclusive basis for the jury's decision in criminal proceedings. To illustrate, consider a criminal trial in which defendant faces charges of assault. During the proceedings, prosecution introduces a video recording from a surveillance camera capturing the defendant at the location of the assault, engaging in a confrontation with the victim. This video is deemed admissible evidence because it is relevant, reliable, and lawfully obtained.

On the contrary, *inadmissible evidence (IE)* is information that does not meet established legal standards for reliability, relevance or violates procedural rules and therefore may not be considered in reaching a verdict (Garner & Black, 2014). Examples of such IE include the disclosure of evidence based on prejudice and confusion (Rule 403), hearsay (Rule 802) or character evidence and evidence of other crimes, wrongs, or acts (Rule 404). The jury is required to disregard inadmissible evidence when rendering its verdict (Federal Rules of Evidence, 2023). Returning to the assault example, imagine prosecution seeks to introduce witness testimony that relates to the defendants' prior conviction of theft. Criminal procedure requires evidence about the crime with which defendant is charged. It does not want the jury to infer that defendant has committed new crime because he has a criminal career. This is why information about criminal record is classified as character evidence under Rule 404 and therefore deemed inadmissible.

Yet the strict rules about the admissibility of evidence are sometimes violated (eg. Winstrich et al., 2005). IE might, for example, be introduced by inadvertent comments, questions that overstep bounds, or through deliberate interventions, such as pre-trial media coverage (eg. Bornstein & Greene, 2011; Otto et al., 1994). While the jury can be admonished and such evidence can be excluded from formal deliberations (Daftary-Kapur et al., 2010), there is a growing recognition of the possibility that jurors may not be able to neutralize the influence of IE, and instead inadvertently be biased by it in their decision making (eg. London & Nunez, 2000). Even if people are sternly instructed to disregard such an attention attractor, it is doubtful whether this goal can be accomplished. Indeed, the effectiveness of admonitions to the jury to ignore IE is often compared to the effectiveness of the instruction to not think about a pink elephant: close to zero (e.g., *Tammy Sowell vs. John Walker, No. 98-CV-1172*, 2000).

For decades, this assumption has been tested by numerous studies demonstrating that admonitions have a small or no effect when it comes to effectively eliminating IE bias when rendering a verdict (eg. Cush & Delahunty, 2006; Freedman et al., 1998; Greene & Dodge, 1995; Lloyd-Bostock, 2000). Admonitions may even be counterproductive, for example by significantly increasing the jury's desire to be allowed to consider IE in their decision-making, compared to a condition where the jury was not admonished (eg. Cox & Tanford, 1989; Kramer et al., 1990; Wolf & Montgomery, 1977), or lead to overcorrection (Sommers & Kassin, 2001).

A meta-analytic investigation of IE bias (Steblay et al., 2006) supported these findings, based on 48 studies with a combined N = 8,474 participants, demonstrating that "inadmissible evidence has a reliable effect on verdicts" and that "judicial instruction to ignore the inadmissible evidence does not effectively eliminate IE impact" (ibd., p.470).

But how much of an issue is IE in the first place? Is all evidence that is declared inadmissible by the law of criminal procedure equally detrimental for the defendant? The question is of particular relevance for information about the defendant's criminal history. In some jurisdictions, as notoriously in Germany, the defendant's crime register is routinely read out loud by the prosecutor. We investigate the question experimentally, and compare two standard means of evidence that would be illegal in the US: the criminal record, and wiretapping without advance judicial authorisation. Are both of them (equally scary) "pink elephants" that deflect the jury from finding the truth?

Sources of Bias

Explaining IE bias is the subject of much theoretical debate, focusing on both motivational and cognitive factors. Some triers of fact tend to be more motivated to strictly follow the law and disregard IE while others find it more important to render an accurate verdict rather than exactly following the law and therefore use IE that they perceive as relevant to the case (Sommers & Kassin, 2001). Admonition has also been discussed to trigger psychological reactance by being perceived as a threat to decision making authority (Cush & Delahunty, 2006).

C*ognitive processes* also appear to be influential in this context. Hindsight bias, i.e. the tendency for individuals to perceive events as having been more predictable after they have occurred (Fischhoff, 1975), might explain the effect of inadmissible evidence (eg. Smith & Greene, 2005). Other studies claim that admonition makes inadmissible information more salient: trying to suppress information, which is the goal of the juror's admonition, may lead to a rebound and make the suppressed thought even more accessible (eg. Wegner & Erber, 1992). More generally, any evidence presented, including IE, might become integral part of the mental representation of the case and would then be more difficult to disregard (Mallard & Perkins, 2005).

Types of Inadmissible Evidence

Most legal systems agree that not all evidence is admissible. But jurisdictions differ substantially in the definition of inadmissible evidence, for instance when it comes to prior convictions. As mentioned, in the German legal system, details of a defendant's prior convictions are routinely mentioned in criminal cases. The Strafprozessordnung [StPO, German Code of Criminal Procedure] (Strafprozeßordnung (StPO), 2024) outlines that such prior convictions may be used to increase the sentence (§ 243 StPO) and to justify warrants for police investigation and pretrial detention (due to increased flight risk and suspicion of new offenses, § 112 StPO).

This practice contrasts starkly with the United States legal system, where evidence about prior convictions is typically excluded to avoid prejudicing the jury or judge against the defendant under the Federal Rules of Evidence (2023 Rule 404(b); see e.g. *Old Chief v. United States, 519 U.S. 172*, 1997). The underlying assumption in the U.S. is that knowledge of prior convictions might introduce undue bias ("we know the defendant did it once, so they might have done it again") leading to a verdict based on character rather than the evidence related to the specific crime being tried. Evidence about prior offenses may be admissible for other purposes, such as to prove motive or opportunity, but the argument of having committed a prior offense may also not be used to question the credibility of a witness (Rule 609).

Given these contrasting approaches, our research seeks to empirically investigate whether prior convictions indeed exert a significant influence on judicial outcomes. By examining the effects of prior conviction evidence on verdicts and sentencing, we aim to assess whether the German approach introduces bias and whether the U.S. system's cautionary measures are indeed necessary to ensure fair trials.

We contrast the introduction of character evidence with another means of evidence that both jurisdictions agree should not be used: what defendant has said over the phone in a private conversation (unless wiretapping had explicitly been authorized by the competent judge: see Electronic Communications Privacy Act; § 100a StPO).

Interventions to Debias

Despite clear rules outlining which evidence is admissible, inadmissible evidence may sometimes be introduced in court proceedings although it should not have been, be it inadvertently or in an underhand attempt to influence the jury. In such cases, the most conservative approach would be to declare mistrial and start proceedings anew, with a fresh judge and jury unexposed to the inadmissible evidence. Instead, in most cases, the judge merely admonishes the jury to disregard evidence declared as inadmissible (eg. *Carter v. Kentucky, 450 U.S. 288*, 1981; *People v. Goldsberry, 509 P.2d 801, 803*, 1973). Judges explicitly remind jurors to disregard inadmissible evidence and to focus solely on the facts of the case that have been established with legally permissible means of evidence. Yet, psychological research has raised significant doubts whether individuals consistently ignore information that should be ignored (eg. Dietvorst & Simonsohn, 2019), particularly in jury decision-making (Steblay et al., 2006). Instructions to disregard inadmissible evidence, while necessary, may not always be sufficient to eliminate bias, as jurors can struggle to (want to) "unhear" or ignore information that has already been presented.

To ensure a fair trial, therefore, substantial research has been devoted to designing interventions that may prove more effective (compared to mere admonitions) at debiasing judges and jury after exposure to inadmissible evidence. Among these types of instructions, three are most noteworthy.

First, we are interested in instructions that encourage the jury to neutralize the influence of the inadmissible evidence, explicitly acknowledging that this is a difficult undertaking. This intervention rests on the assumption that limited success in ignoring inadmissible evidence is driven by limited elaboration (i.e., failure to exert sufficient effort) to do so (Kennedy, 1992; Wilson & Brekke, 1994). Therefore, activating reflective thinking by imposing time delay (eg. Capraro et al., 2019; Rand, 2016) may counteract bias (but see Isler et al., 2020).

Second, we focus on instructions that highlight the normative importance of ignoring the evidence to avoid an unfair trial. This intervention is built on the assumption that effective jury instructions must enhance jurors' understanding of the reasons behind the inadmissibility of certain evidence. This intervention consists of providing reasons (e.g., unreliability (Kassin & Sommers, 1997; Oakes et al., 2021) or explaining the normative background (Diamond & Casper, 1992; Dietvorst & Simonsohn, 2019; but see Pickel, 1995). Both may be instrumental in persuading jurors to ignore inadmissible evidence.

Our third intervention would not be feasible in the court room, but serves as a particularly strong test. It exploits that we had data available from mock jurors who were not exposed to inadmissible evidence in the first place. We pledged a monetary bonus for making the same decision as the majority in the condition without exposure to inadmissible evidence. Prior research notes that introducing incentives is not necessarily successful at debiasing decisions (eg. Arkes, 1991; Fischhoff et al., 1977). But our intervention combines the monetary incentive with the instruction to consider the alternative state of the world where no influence of inadmissible evidence can exist. This is in the spirit of a “consider the opposite” debiasing strategy (Lord et al., 1984; Steblay et al., 2006).

Present Research

We had started this research with the expectation that character evidence would have a strong and clear biasing effect. Our initial focus was on alternative strategies for mitigating the bias. To our growing surprise, despite a whole consecutive battery of experimental interventions, we never established a significant bias. After multiple replications (which all have been preregistered), we are now convinced that merely learning about the fact that defendant has previously been convicted for a similar crime does not make it more likely that laypersons come to the conclusion, in a deliberately ambiguous case, that defendant is guilty. Only after having firmly established this finding have we started to compare character evidence with another standard case of inadmissible evidence, wiretapping. With this second intervention we immediately and easily established the bias. This learning experience of us experimenters explains the apparent richness of the design, and the imbalance between character evidence and wiretapping. We consider it important to report the complete evidence as it shows how robust the surprising finding actually is.

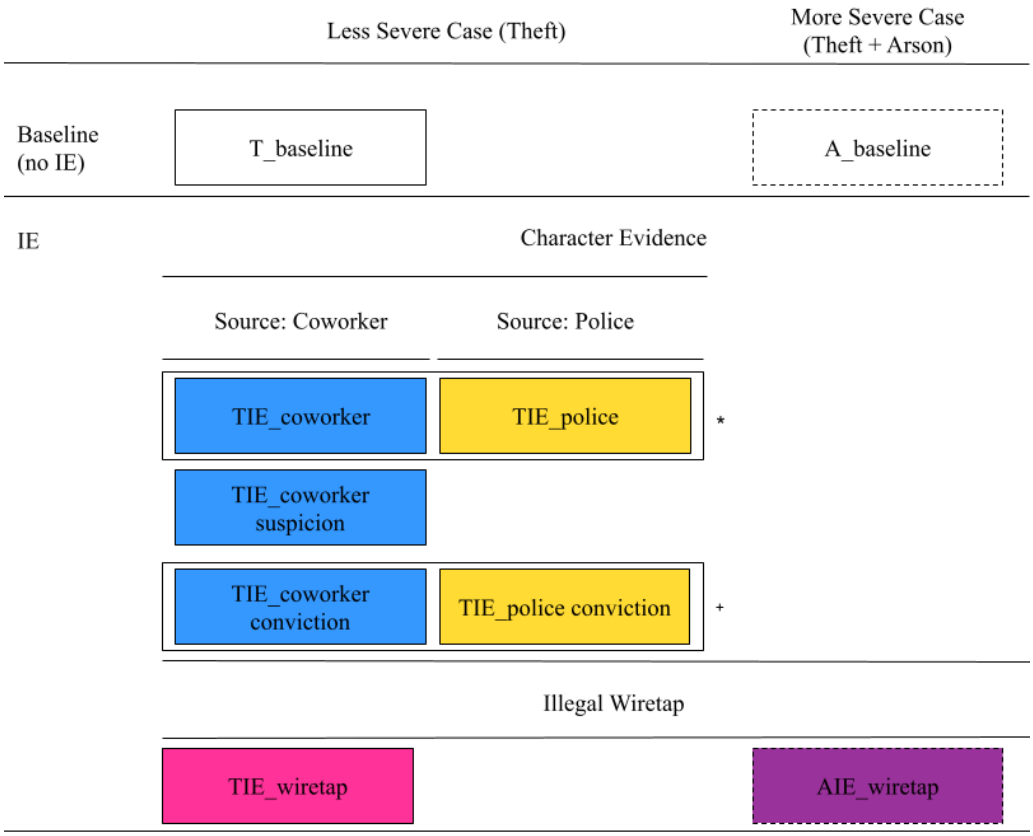
In the light of this experience, our study has two objectives. We first investigate whether inadmissible evidence, either in the form of character evidence or of wiretapping, biases laypersons

to the detriment of defendant. Second we test how effective one of four alternative interventions is in mitigating the detriment.

Figure 1 summarizes the design of the experiment. We test participants on two versions of a criminal case. In the first version, the defendant is accused of theft, while in the second the charge comprises arson as well. In the *baseline,* we test either version without introducing inadmissible evidence. In further conditions, we add either *character evidence* or an unauthorized *wiretap*. For character evidence, we additional manipulate whether it is brought by a witness (*coworker*) or by a *police* officer, how close the prior conviction is to the present charge (unrelated to the victim, or in the same premises), and whether the defendant has only been suspected of the earlier crime, or convicted.

Figure 1

Overview of baseline and IE conditions



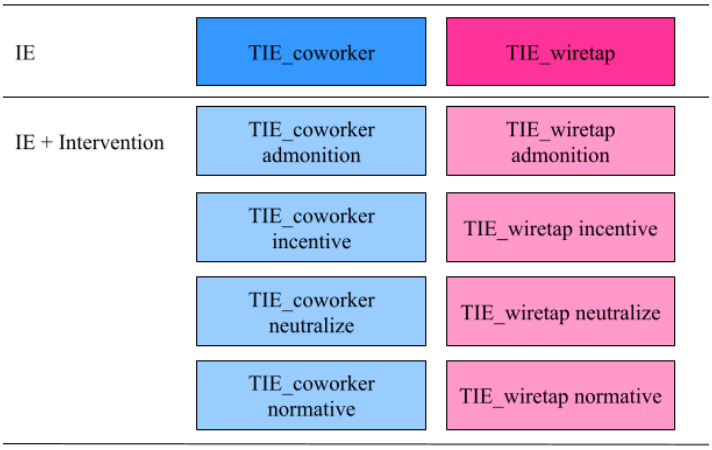
Note. * indicates conditions that share prior convictions outside of Construction Ltd; + indicates conditions that share prior conviction for offenses at Construction Ltd.

Moreover, we compare the effectiveness of different interventions to reduce bias after inadmissible evidence has been introduced (preregistered, Figure 2). We expected lower rates of guilty charges when any of the four instructions (*ignore, neutralize, normative, incentive*) is presented, compared to the condition in which inadmissible evidence is presented without instructions to ignore. Specifically, we expected that the instructions *neutralize, normative* and *incentive* are more effective than the mere instruction to ignore. Note that we had no specific

expectations about the relation of the *neutralize* vs. *normative* vs. *incentive* instructions. On the other hand, we expected the effect of all four instruction types to be limited, such that the rate of guilty verdicts would be lower when no inadmissible evidence was presented than when inadmissible evidence was presented, but an attempt was made to mitigate its effect.

Figure 2

Overview of intervention conditions with relevant comparison group (top line)



Method

We collected data in five waves, each of which was preregistered. Materials, data and code are available on the OSF.

Participants and Design

Participants (N_{total}= 1432, M_{age} = 39.80, SD_{age} = 13.51, 645 female, 70 diverse) were recruited via Prolific and were eligible to participate if they were US residents, were fluent in English and had a desktop computer capable of playing audio files available. They received a flat fee of £ 3.00 for participating in online surveys that took about 20 min to complete. Participants were randomly assigned to one of 16 between-subjects conditions in a fractional factorial design. We manipulated whether IE was introduced (baseline vs. IE), what the charge was (theft vs. theft and arson), and which instructions to disregard IE were presented (none vs. admonition vs. incentive vs. neutralize vs. normative).

In wave 1, 474 participants were recruited and randomly allocated to one of six conditions (*T_baseline, TIE_coworker, TIE_coworker admonition, TIE_coworker incentive, TIE_coworker neutralize, TIE_coworker normative*). Wave 2 added data from 80 participants in one additional treatment (*TIE_police*). In wave 3, data from 240 further participants was collected, who were randomly allocated to one of three conditions (*TIE_coworker suspicion, TIE_coworker conviction, TIE_police conviction*). In wave 4, we collected data from 264 participants randomly allo-

cated to three additional conditions (*A_baseline, TIE_wiretap, AIE_wiretap*), where 132 participants were recruited from pools that had previously indicated democratic or republican political attitudes, respectively. In wave 5, we collected data from 352 participants randomly allocated to four additional conditions (*TIE_wiretap admonition, TIE_wiretap incentive, TIE_wiretap neutralize, TIE_wiretap normative*), where 176 participants were recruited from pools that had previously indicated democratic or political attitudes, respectively.

Materials and Procedure

Participants reported their Prolific ID, age, gender and indicated whether they liked watching movies or series about courtroom drama, lawyers or judges (yes/no) and if they had ever sat on a jury for a criminal case in court (yes/no). In wave 4, they additionally indicated their highest level of education from a 10-option list ordered between no schooling completed and having obtained a doctoral degree, and political ideology (republican, democrat, independent, or other with an opportunity to give details). Finally, they underwent an audio check.

Participants were then introduced to their role and the case. They were asked to imagine that they were on jury duty in a criminal case and asked to decide whether the defendant, Jason Wells, was guilty as charged. They were instructed about the presumption of innocence and the concept of reasonable doubt, as well as explicitly instructed to only pronounce the defendant as guilty if they were convinced that he had stolen the money, and informed that they would otherwise have to pronounce Jason Wells as not guilty.

They were then introduced to the charge. They were told that evidence on the case would be read out to them, involving both evidence from prosecution and defense. In addition, participants were told that all witnesses had been sworn in and informed that they would commit perjury if they did not tell the truth. Participants were also informed that the information presented would be summarized in keywords visible while the evidence was presented.

Participants then moved on to the presentation of evidence, where they listened to an audio recording (transcripts and original files available on the OSF) detailing the evidence for about 5 min. The basic descriptions contained six pieces of evidence each of inculpating and exculpating evidence.

Our stimulus material consisted of a case that had originally been designed by Simon (Simon 2004). It has repeatedly been used to test legal decision making by laypersons. The case is deliberately ambiguous, so that different participants can come to different conclusions, and that there is room for manipulations, such as standard of proof (Glöckner Engel 2013), professional roles (Engel Glöckner 2013) or the order in which the parties plead their case (Engel Glöckner Timme 2020). Building on the same basic case, we presented two versions varying in the severity of the crime, associated with higher potential sentences.

Less Severe Crime

For the less severe crime, the defendant was charged with theft of $5200 from his employers' safe. Regarding this version of the case, data was collected in 14 between-subjects conditions which manipulated whether IE was introduced (no IE in *baseline*), what type of IE was introduced and by whom, whether there was an instruction to disregard the evidence, and if so, what type of intervention was chosen. We refer to conditions relating to this less severe theft case starting with the letter T (e.g., *T_baseline*).

More Severe Crime

In the more severe crime condition, the defendant was additionally charged with arson because a torch placed near the safe was identified as the cause of a fire, which had almost destroyed the CCTV evidence and endangered three members of the cleaning staff who were in the building. Regarding the case involving both theft and arson, data were collected in 2 between-subjects conditions which manipulated whether IE was introduced, but not whether jurors could be debiased. We refer to conditions relating to this more severe case involving both theft and arson starting with the letter A (e.g., *A_baseline*).

Baseline Conditions without Inadmissible Evidence

In the *T_baseline* and *A_baseline* conditions, no IE was presented (see Figure 1). **Introducing the Pink Elephant: Conditions with Inadmissible Evidence and No further Instructions**

In seven conditions, IE was introduced to the case without further instructions (i.e., no instruction to disregard the evidence, see Figure 1).

Character Evidence. As we had been surprised by the fact that, in our first attempt, we did not find an effect of character evidence on verdict, we replicated the (non) effect multiple times. To that end, we collected data about the influence of character evidence on jurors' judgments of the defendants' guilt in five conditions. Three conditions used a coworker as the source of the IE. Two further conditions used a police officer as the source of IE, on the hypothesis that the additional authority of a state official might strengthen the bias.

Coworker as Source. In the *TIE_coworker* condition, a coworker mentions that the defendant was convicted 2 years ago for having tried to break into an apartment. Two further conditions placed potential prior crimes closer to the victim in the current case. In the *TIE_coworker suspicion* condition, a coworker indicates that since the defendant had started working at the company, they had experienced a surge in alleged crimes, including disappeared valuables and stolen office goods, without being able to identify the culprit. In the *TIE_coworker conviction* condition, the coworker reports the same evidence as in *TIE_coworker suspicion*, but adds that in one case of a customer's stolen bag, the defendant had been charged with theft and found guilty.

Police Officer as Source. In three further conditions, the source of IE was a police officer. In the *TIE_police* condition, a police officer mentions that the defendant has a criminal record and had previously been convicted for three charges of various crimes against others' property. The *TIE_police conviction* condition placed potential prior crimes closer to the victim in the current case. In addition to the information of *TIE_police*, it included the information that the defendant had been charged with theft and found guilty in a case of a customer's stolen bag, at the same premise.

Wiretap Evidence. In two further conditions, IE was presented via a wiretap (*TIE_wiretap, AIE_wiretap*) of the defendants' phone that contained an admission of having stolen the money.

Fighting the Pink Elephant: Instructions to Disregard Inadmissible Evidence

For two of the conditions in which we expected inadmissible evidence to bias the assessment of guilt, we tested alternative interventions meant to neutralize the bias. For character evidence, we did so if the prior conviction had been mentioned by a coworker heard as witness (*TIE_coworker*). We also added these interventions if inadmissible wiretap evidence had been presented (*TIE_wiretap*; see Figure 2). In each of these additional treatments, participants were instructed by the judge to disregard the evidence. They were told that the defense attorney protested and the judge agreed that the questionable evidence should not have been introduced. We tested four interventions:

In the *admonition* conditions, participants were informed that information about prior convictions was inadmissible and were instructed to disregard this information.

In the *neutralize* conditions, participants were told that disregarding information that one has heard is difficult. In the interest of fair trial, they were asked to make an effort to prevent this information from biasing their judgements.

In the *coworker normative* condition, participants were told that the presumption of innocence also holds for defendants with a criminal record. They were informed that a prior conviction should therefore not interfere with the question whether it has been proven that the defendant has committed the crime for which he was charged.

In the *wiretap normative* condition, participants were told that the wiretap had been obtained without a warrant and during a conversation the defendant could reasonably expect to be private. They were informed that the interception of this communication constituted a violation of the defendant's constitutional rights under the Fourth Amendment.

In the *incentive* conditions, participants were told that they could obtain a bonus payment of $ 0.50 if they make the same decision as the majority of participants in the otherwise identical study who had not been exposed to IE. Note that no other condition involved monetary incentives.

Judgements

After engaging with the case, participants indicated whether they found the defendant guilty (yes/no). In addition, they reported how certain they were to have made the right judgment on a 10-point scale (ranging from totally uncertain to totally certain), their estimation of the probability that the defendant had taken the money from the safe (percentage between 0 and 100) and indicated how high the likelihood for the defendant to have taken the money would have to be for them to judge him as guilty (percentage between 0 and 100).

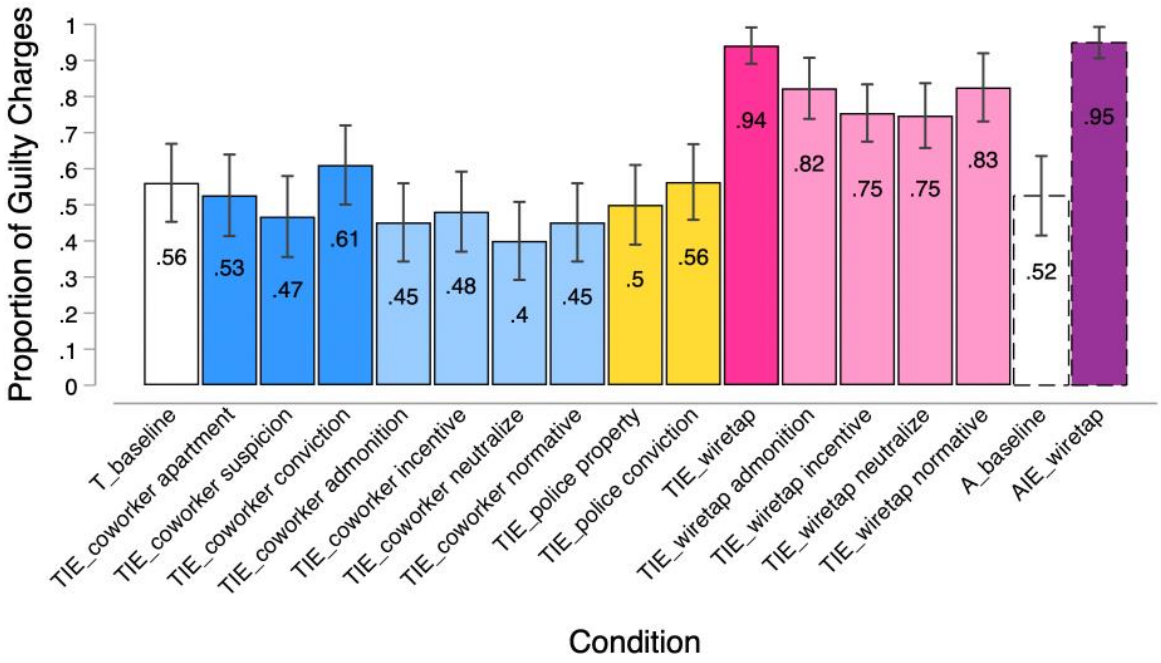
Results

Introduction of Inadmissible Evidence: Increased Guilty Verdicts Only For Wiretap

We had expected that including inadmissible evidence on the defendants' prior convictions would increase the rate of guilty verdicts, compared to the *baseline* case where no such evidence was present. However, we only partly support this hypothesis (Figure 3). Despite the fact that we have run multiple (conceptual) replications, we have found no evidence of increased conviction rates compared to the *baseline* when *character evidence* was introduced (Table 1, Model 1). Only when wiretap evidence was presented did the odds ratio of guilty verdicts significantly increase compared to the relevant baseline, both for the less severe (*TIE_wiretap:* OR = 12.52, z = 4.94, p < 0.001) and the more severe case (*AIE_wiretap:* OR = 17.01, z = 5.55, p < 0.001, Table 1, Model 2).

Figure 3

Main Result: Proportion of Guilty Charges, per Conditi



Note.
Error bars are 95% Confidence Intervals.

Table 1

Logistic Regression predicting Guilty Judgments in the Absence of Debiasing Attempts

(1 = guilty, 0 = not guilty) in Conditions with IE compared to Baselines (no IE)

	(1)		(2)	
	OR	*z*	*OR*	*z*
T_baseline				
TIE_coworker	0.87	-0.44		
TIE_coworker suspicion	0.69	-1.18		
TIE_coworker conviction	1.23	0.63		
TIE_police	0.78	-0.78		
TIE_police conviction	1.01	0.03		
TIE_wiretap	12.52***	4.94		
A_baseline				
AIE_wiretap			17.01***	5.55
Constant	1.28	1.10	1.11	0.45
Observations	564		179	

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Instructions to Ignore: Decreased Guilty Verdicts Only for Neutralizing Wiretap

We had expected that instructing the jury to disregard the inadmissible evidence would lead to a reduction of guilty verdicts. However, in the conditions comparing guilty verdicts when character evidence was introduced (*coworker* condition), we found no evidence that any of the four variations of the instructions to disregard this evidence were successful at lowering the rate of convictions (Table 2, Model 1). In a way, this is a comforting finding: if a piece of evidence has actually not distorted judgment, there is no need for debiasing. Had we found an effect, the interventions would even have been counterproductive (albeit the effect would have been less worrisome, given the presumption of innocence).

Yet, instructions to disregard the inadmissible *wiretap* evidence successfully reduced guilty ratings (Table 2, Model 2). All four instructions significantly lowered the odds of finding the defendant guilty (*TIE_wiretap admonition*: OR = 0.29, z = −2.26, p= 0.024; *TIE_wiretap incentive*: OR = 0.19, z = -3.24, p = 0.001; *TIE_wiretap neutralize*: OR = 0.18, z = -3.25, p = 0.001, *TIE_wiretap normative*: OR = 0.30, z = -2.15, p = 0.032).

We had expected that instructing the jury to ignore inadmissible evidence would have an effect, but that the effect would be limited, in that the rate of guilty verdicts would still be higher when IE had been presented and the jury was asked to disregard it, rather than the jury not having heard the evidence in the first place. To test this hypothesis, we compared the odds of guilty ratings following admonitions to ignore the IE with the *baseline* case in which no IE was presented. When the coworker introduced the inadmissible character evidence into the proceedings, the only effect we observed suggested that, compared to the control condition, participants who were instructed to neutralize inadmissible evidence showed decreased odds ratios of finding the defendant guilty (OR = 0.52, z = 2.04, p = 0.05), while there was no evidence that the other manipulations affected guilty verdicts (Table 2, Model 3). Note that this effect is no longer significant when adjusting the alpha error for multiple comparisons.

When IE was introduced in the form of an illegal *wiretap*, however, the odds of finding the defendant guilty were higher than in the scenario without IE for all four types of instructions to disregard the IE (*TIE_wiretap admonition*: OR = 3.63, z = 3.49, p < 0.001; *TIE_wiretap incentive*: OR = 2.40, z = 2.82, p = 0.005; *TIE_wiretap neutralize*: OR = 2.31, z = 2.56, p = 0.011, *TIE_wiretap normative*: OR = 3.70, z = 3.27, p = 0.001). Hence the debiasing interventions had an effect. But for no intervention, the effect was strong enough to completely neutralize the bias.

Table 2

Logistic Regressions Predicting the Effect of Debiasing Interventions

(1 = guilty, 0 = not guilty) in Conditions with Instructions to Ignore Character Evidence (Model 1) and Wiretap (Model 2) compared to Case with IE and no further instructions, and compared to baseline (No IE, Model 3)

	(1)		(2)			(3)	
	OR	*z*	*OR*	*z*		*OR*	*z*
TIE_coworker					*T_baseline*		
TIE_coworker admonition	0.74	-0.94			TIE_coworker admonition	0.64	-1.40
TIE_coworker incentive	0.83	-0.56			TIE_coworker incentive	0.73	-1.01
TIE_coworker neutralize	0.60	-1.58			TIE_coworker neutralize	0.52*	-2.04
TIE_coworker normative	0.74	-0.94			TIE_coworker normative	0.64	-1.40
TIE_wiretap							
TIE_wiretap admonition			0.29*	-2.26	TIE_wiretap admonition	3.63***	3.49
TIE_wiretap incentive			0.19**	-3.24	TIE_wiretap incentive	2.40**	2.82
TIE_wiretap neutralize			0.18**	-3.25	TIE_wiretap neutralize	2.31*	2.56
TIE_wiretap normative			0.30*	-2.15	TIE_wiretap normative	3.70**	3.27
Constant	1.11	0.46	16.00***	6.01		1.28	1.10
Observations	399		432			752	

Note. + $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

We had originally been interested in testing the effectiveness of alternative methods to debias jury members who have been exposed to inadmissible evidence. To our growing surprise, in a series of increasingly worrisome pieces of clearly inadmissible character evidence, we could never establish the bias, and hence the baseline for the originally planned experiment. Only if we switched to another quintessential piece of inadmissible evidence, wiretapping, did we find the bias. Turning back to our original research question, we found that instructions to disregard wiretap confessions were effective in reducing the odds of finding the defendant guilty compared to a scenario without such instructions. However, the effect was limited, such that the odds of finding the defendant guilty were still significantly and substantially higher following instructions to disregard inadmissible evidence than if inadmissible evidence had never been introduced.

Procedural fairness is important in its own right. The way how individuals feel treated in their typically exceedingly rare direct interactions with judicial authority has a spillover effect on law abiding behavior in totally different domains, as famously shown by Tom Tyler (Tyler 2006). If he hears his criminal record read out loud, the defendant may (possibly wrongly) deem his case hopeless, and refrain from defending himself effectively against an unwarranted accusation. The legislator, or the judiciary, may also have deontological reasons to care about procedural fairness. For all these reasons, the law may well want to ban character evidence in criminal procedure altogether. All our experiment contributes to this debate in one data point: in an experiment that tests members of the general population, learning that defendant has a criminal record does not stack the odds to his detriment. If the law is chiefly interested in the accuracy of a guilty verdict, banning character evidence is not a precondition.

References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498. https://doi.org/10.1037/0033-2909.110.3.486

Bornstein, B. H., & Greene, E. (2011). Jury Decision Making: Implications For and From Psychology. *Current Directions in Psychological Science*, *20*(1), 63–67. https://doi.org/10.1177/0963721410397282

Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics*, *79*, 93–99. https://doi.org/10.1016/j.socec.2019.01.007

Carter v. Kentucky, 450 U.S. 288 (U.S. Supreme Court 9. März 1981). https://supreme.justia.com/cases/federal/us/450/288/

Cox, M., & Tanford, S. (1989). Effects of evidence and instructions in civil trials: An experimental investigation of rules of admissibility. *Social Behaviour*, *4*(1), 31–55.

Cush, R. K., & Delahunty, J. G. (2006). The Influence of Limiting Instructions on Processing and Judgments of Emotionally Evocative Evidence. *Psychiatry, Psychology and Law*, *13*(1), 110–123. https://doi.org/10.1375/pplt.13.1.110

Daftary-Kapur, T., Dumas, R., & Penrod, S. D. (2010). Jury decision-making biases and methods to counter them. *Legal and Criminological Psychology*, *15*(1), 133–154. https://doi.org/10.1348/135532509X465624

Diamond, S. S., & Casper, J. D. (1992). Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury. *Law & Society Review*, *26*(3), 513–563. https://doi.org/10.2307/3053737

Dietvorst, B. J., & Simonsohn, U. (2019). Intentionally "biased": People purposely use to-be-ignored information, but can be persuaded not to. *Journal of Experimental Psychology: General*, *148*(7), 1228–1238. http://dx.doi.org/10.1037/xge0000541

Federal Rules of Evidence. (2023). LII / Legal Information Institute. https://www.law.cornell.edu/rules/fre

Fiedler, K., Hütter, M., Schott, M., & Kutzner, F. (2019). Metacognitive myopia and the overutilization of misleading advice. *Journal of Behavioral Decision Making*, *32*(3), 317–333. https://doi.org/10.1002/bdm.2109

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Freedman, J. L., Martin, C. K., & Mota, V. L. (1998). Pretrial publicity: Effects of admonition and expressing pretrial opinions. *Legal and Criminological Psychology*, *3*(2), 255–270. https://doi.org/10.1111/j.2044-8333.1998.tb00365.x

Garner, B. A., & Black, H. C. (Hrsg.). (2014). *Black's law dictionary* (10th ed). Thomson Reuters.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of „blind" auditions on female musicians. *American Economic Review*, *90*(4), 715–741. https://doi.org/10.1257/aer.90.4.715

Greene, E., & Dodge, M. (1995). The influence of prior record evidence on juror decision making. *Law and Human Behavior*, *19*(1), 67–78. https://doi.org/10.1007/BF01499073

Isler, O., Yilmaz, O., & Dogruyol, B. (2020). Activating reflective thinking with decision justification and debiasing training. *Judgment and Decision Making*, *15*(6), 926–938. https://doi.org/10.1017/S1930297500008147

Kassin, S. M., & Sommers, S. R. (1997). Inadmissible Testimony, Instructions to Disregard, and the Jury: Substantive Versus Procedural Considerations. *Personality and Social Psychology Bulletin*, *23*(10), 1046–1054. https://doi.org/10.1177/01461672972310005

Kennedy, S. J. (1992). *Debiasing audit judgment with accountability: A framework and experimental results* [Ph.D., Duke University]. https://www.proquest.com/docview/303974973/abstract/ 18AAC70FFEC541C9PQ/1

Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior*, *14*(5), 409–438. https://doi.org/10.1007/BF01044220

Lloyd-Bostock, S. (2000). *The effects on juries of hearing about the defendant's previous criminal record: A simulation study*. *85*(6), 932–939.

London, K., & Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence. *Journal of Applied Psychology*, *85*(6), 932–939. https://doi.org/10.1037/0021-9010.85.6.932

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Mallard, D., & Perkins, D. P. (2005). Disentangling the Evidence: Mock Jurors, Inadmissible Testimony and Integrative Encoding. *Psychiatry, Psychology and Law*, *12*(2), 289–297. https://doi.org/10.1375/pplt.12.2.289

Oakes, M. A., Crosby, C. A., McCallops, K., McDonald, B. R., & Schwarz, A. C. (2021). Judge, jurors, and gendered instructions to disregard evidence: Stereotype-congruent judicial instructions increase compliance. *Psychology, Crime & Law*, *27*(10), 933–955. https://doi.org/10.1080/1068316X.2020.1867132

Old Chief v. United States, No. 519 U.S. 172 (7. Januar 1997). https://www.oyez.org/cases/1996/95-6556

Otto, A. L., Penrod, S. D., & Dexter, H. R. (1994). The biasing impact of pretrial publicity on juror judgments. *Law and Human Behavior*, *18*(4), 453–469. https://doi.org/10.1007/BF01499050

People v. Goldsberry, No. 509 P.2d 801, 803 (14. Mai 1973). https://law.justia.com/cases/colorado/supreme-court/1973/25397.html

Pickel, K. L. (1995). Inducing jurors to disregard inadmissible evidence: A legal explanation does not help. *Law and Human Behavior*, *19*(4), 407–424. https://doi.org/10.1007/BF01499140

Rand, D. G. (2016). Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. *Psychological Science*, *27*(9), 1192–1206. https://doi.org/10.1177/0956797616654455

Simon, Dan (2004). A Third View of the Black Box. Cognitive Coherence in Legal Decision Making. *University of Chicago Law Review* **71**: 511-586.

Smith, A. C., & Greene, E. (2005). Conduct and its consequences: Attempts at debiasing jury judgments. *Law and Human Behavior*, *29*(5), 505–526. https://doi.org/10.1007/s10979-005-5692-5

Sommers, S. R., & Kassin, S. M. (2001). On the Many Impacts of Inadmissible Testimony: Selective Compliance, Need for Cognition, and the Overcorrection Bias. *Personality and Social Psychology Bulletin*, *27*(10), 1368–1377. https://doi.org/10.1177/01461672012710012

Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis. *Law and Human Behavior*, *30*(4), 469–492. https://doi.org/10.1007/s10979-006-9039-7

Strafprozeßordnung (StPO) (2024). https://www.gesetze-im-internet.de/stpo/

Tammy Sowell vs. John Walker, No. 98-CV-1172 (Appeal from the Superior Court of the District of Columbia 22. Juni 2000). https://cases.justia.com/district-of-columbia/court-of-appeals/98-cv-1172-6.pdf?ts=1396116000

Tyler, Tom R. (2006). Why People Obey the Law. New Haven, Yale University Press.

Wegner, D. M., & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology*, *63*(6), 903–912. https://doi.org/10.1037/0022-3514.63.6.903

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142. https://doi.org/10.1037/0033-2909.116.1.117

Winstrich, A. J., Guthrie, C., & Rachlinski, J. J. (2005). Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding. *University of Pennsylvania Law Review*, *153*(4), 1251. https://doi.org/10.2307/4150614

Wolf, S., & Montgomery, D. A. (1977). Effects of Inadmissible Evidence and Level of Judicial Admonishment to Disregard on the Judgments of Mock Jurors [1]. *Journal of Applied Social Psychology*, *7*(3), 205–219. https://doi.org/10.1111/j.1559-1816.1977.tb00746.x