Judgment and Decision Making, Vol. 15, No. 2, March 2020, pp. 214–229

Is justice blind or myopic? An examination of the effects of meta-cognitive myopia and truth bias on mock jurors and judges

Myrto Pantazi* Olivier Klein† Mikhail Kissine‡

Abstract

Previous studies have shown that people are truth-biased in that they tend to believe the information they receive, even if it is clearly flagged as false. The truth bias has been recently proposed to be an instance of meta-cognitive myopia, that is, of a generalized human insensitivity towards the quality and correctness of the information available in the environment. In two studies we tested whether meta-cognitive myopia and the ensuing truth bias may operate in a courtroom setting. Based on a well-established paradigm in the truth-bias literature, we asked mock jurors (Study 1) and professional judges (Study 2) to read two crime reports containing aggravating or mitigating information that was explicitly flagged as false. Our findings suggest that jurors and judges are truth-biased, as their decisions and memory about the cases were affected by the false information. We discuss the implications of the potential operation of the truth bias in the courtroom, in the light of the literature on inadmissible and discredible evidence, and make some policy suggestions.

Keywords: truth bias, meta-cognitive myopia, legal decision-making, accountability

1 Introduction

Judges and jurors are the "fact-finders" of the judicial system. They are charged with the task of determining whether the evidence adduced in a trial corresponds to facts, and they then have to reach a verdict and propose a sentence accordingly (Alschuler, 1982; Frankel, 1975, 1982; Freedman, 1975; Robbennolt, MacCoun & Darley, 2010; Wilson, 1963). Distinguishing reality from pretense is, arguably, a very hard task; but can the fact-finders disregard information when they know it to be false?

There is ample empirical evidence that human decision-making suffers from meta-cognitive myopia: a tendency to be extremely sensitive to primary information but "stubbornly" resist the relevant meta-information concerning its history and accuracy (Fiedler, 2012). In many judgment and decision-making contexts people promptly utilize the information they have in hand (Fiedler, 2007; Juslin, Winman & Hansson, 2007), but fail to critically assess its special characteristics, sources and history, even when such "meta-information" is readily available to them (Fiedler, 2000, 2012; Kareev, Arnon & Horwitz-Zeliger, 2002). Meta-cognitive myopia serves as an umbrella mechanism that can account for various well-documented heuristics and biases. Examples include: a) confirmation bias (the tendency to selectively rely on information confirming one's prior beliefs); b) repetition effects (the tendency of repeated information to exert increased impact on people's judgments; and c) the fundamental attribution bias (the tendency to infer an individual's characteristics from non-diagnostic situational information). All these phenomena, which can individually have serious consequences in the courtroom, can be uniformly accounted for by an inability to assess the quality and history of information at hand.

A conspicuous effect of meta-cognitive myopia is the truth bias (Fiedler, 2012), the tendency to believe information regardless of its actual accuracy. The truth bias is detectable when meta-information present in the environment suggests that primary information at hand is unreliable or false. In such cases, people seem to unduly make judgments and inferences about themselves and others based on explicitly discredited information (Anderson, 1983; Anderson, Lepper & Ross, 1980; Guenther & Alicke, 2008; Pantazi, Kissine & Klein, 2018; Ross, Lepper & Hubbard, 1975; Schul & Burnstein, 1985; Thorson, 2015). Analogous effects can be detected in explicit ratings of truth: People are more likely to misremember as true a piece of information they have been told is false, than to misremember as false a piece of

This research was supported by the Mini-ARC "Project" grant, *At the sources of faith*, from the Université libre de Bruxelles. We are grateful to Patrick Mandoux for his great help with recruitment and his valuable advice on aspects of legal decision-making.

*Center for Social and Cultural Psychology & Center of Research in Linguistics LaDisco, Université Libre de Bruxelles. Myrto Pantazi is now at Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, UK. Email: myrto.pantazi@oii.ox.ac.uk

†Center for Social and Cultural Psychology, Université Libre de Bruxelles

‡Center of Research in Linguistics LaDisco, Université Libre de Bruxelles

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

true information (Begg, Anas & Farinacci, 1992; Levine, Park & McCornack, 1999; Pantazi, Kissine & Klein, 2018; Street & Masip, 2015). Therefore, meta-cognitive myopia, among other things, makes people take information as true by default, leading to a detectable truth bias in their judgments and memory (Fiedler, Armbruster, Nickel, Walther & Asbeck, 1996; Pantazi et al., 2018; Street & Masip, 2015).

1.1 Is Justice Blind or Myopic?

The general propensity for people to believe messages they encounter has been argued to be evolutionarily efficient, as in real life truthful information is presumably more prevalent than false or deceitful information (Kissine & Klein, 2013; Levine, 2014; Reber & Unkelbach, 2010). This means that under normal circumstances, where truthfulness in the environment prevails, the truth bias does not lead to erroneous memory and judgments: After all, from a normative perspective people ought to trust and believe true information, and if such information in a given environment prevails, a "bias" towards accepting information as true could be "ecologically rational". The adverse effects of the truth bias on judgment and decision accuracy are detectable in contexts with high occurrence of inaccurate or false information, like courts. Defendants and plaintiffs can be reasonably expected to provide biased, and occasionally inaccurate information to protect themselves or promote their interests. It is not infrequent to see a party in a trial presenting meta-information that discounts a testimony previously provided by the opposite party (e.g., because a witness is untrustworthy, or because what they say contradicts previously presented evidence). From a normative perspective, the fact-finders will have to use the meta-information thus provided to draw an accurate picture of the facts and disregard the discounted testimony in their final decision. The question we ask in this paper is how plausible is this interpretation of "blind justice" from a cognitive point of view.

1.2 The case of legal decision-making.

Especially within legal professional circles, legal reasoning has long been thought to be special and different from lay reasoning (Spellman & Schauer, 2012). From a psychological perspective, legal reasoning may be expected to be less myopic than that of other individuals on social, motivational, as well as cognitive grounds (Spellman, 2007). Experimental studies suggest that participants who expect to be tricked may become more vigilant and resistant to misinformation (Chambers & Zaragoza, 2001). Moreover, fact-finders have been claimed to be driven by "accuracy motivation" (Fleming, Wegener & Petty, 1999), an inherent motivation to render a just verdict (Sommers & Kassin, 2001). The high expectation of being misled and the high accuracy motivation within a courtroom setting may increase vigilance in jurors and judges and pressure them to adopt more systematic information processing (see Evans, 2008) that relies on meta-informational cues.

Another special characteristic of legal fact-finders, especially judges, is that they are generally thought to be accountable for their decisions (Braman, 2010). Accountability is "the implicit or explicit expectation that one may be called on to justify one's beliefs, feelings, and actions to others" (Lerner & Tetlock, 1999, p. 255). The appellate process, which may highlight judicial errors, typically renders judges accountable. The fact that, in some countries, judges may be elected to their position, is also expected to increase their accountability to the public. A large literature suggests that people who feel accountable are more likely to overcome a number of cognitive biases (Lerner & Tetlock, 1999; Simonson & Nye, 1992), arguably due to more vigilant and complex information processing (Tetlock, 1983a).

Finally, a case could be made for judges' potentially special capacity to routinely take meta-information into account and rely only selectively on presented evidence. Whether in bench trials or appellate trials, judges have to sort out the evidence they encounter relative to a case, evaluate it, and accordingly attribute to it specific weight for an appropriate decision to be made. Thus, unlike jurors or laypeople, who rarely need to rely on meta-information, judges can be expected to have considerable expertise in weighing information in the context of their profession. Experts' judgments are often thought of as superior, as experts seem able to resort to more efficient task-specific strategies compared to other people (Klein, Shneiderman, Hoffman & Ford, 2017). By extension, judges may be especially able to make judgments and decisions while remaining unaffected by erroneous evidence. The assumption that judges have special cognitive abilities in handling meta-information and weighing primary information is, in fact, manifest in certain legal systems, as judges are allowed to access inadmissible evidence that jurors are not, on the assumption that they can resist it (see Spellman, 2007). All the evidence reviewed so far in this section would suggest that fact-finders, especially judges, are less myopically prone to the truth bias than other populations.

Nevertheless, there are also good reasons to assume that judges and especially jurors will not be any different from other people. Jurors are generally members of the general public with no special training or expertise in weighing information and making just decisions. As for the judges, since they do not receive feedback on their decisions unless they are reviewed by appellate courts, it is unclear how their training and professional expertise might make them any less bias-prone (see Spellman, 2007). Several experimental studies indeed suggest that jurors' and judges' decision-making is driven by biases and heuristics just like lay decision-making (Daftary-Kapur, Dumas & Penrod, 2010; Kassin & Sommers, 1997; Lieberman & Arndt, 2000; Pickel, Karam & Warner, 2009).

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

Studies focusing on jurors' and judges' capacity to disregard inadmissible evidence are especially relevant here. Inadmissible evidence is evidence that either hinders accurate fact-finding, or interferes with a policy interest, and should thus be excluded from a trial (Wistrich, Guthrie & Rachlinski, 2005). Judges know that inadmissible evidence should be ignored in a trial and clearly instruct jurors to do so. Yet, experimental studies consistently show that both jurors and judges are affected by such inadmissible evidence (Daftary-Kapur et al., 2010; Kassin & Sommers, 1997; Landsman & Rakos, 1994; Lieberman & Arndt, 2000; Pickel et al., 2009; Wistrich et al., 2005).

Since inadmissible evidence has to be disregarded due to rules of evidence, independently of whether it is accurate or not, it is unclear whether inadmissible evidence effects may be an instance of meta-cognitive myopia, that is, an inability to take meta-information into account, or rather an indication of fact-finders' inherent motivation to render a just verdict (see Fleming et al., 1999; Sommers & Kassin, 2001). The distinction between these two possible explanations of the effects of inadmissible evidence amounts to a distinction that has already been drawn in the literature, namely between fact-finders' inability vs. unwillingness to ignore inadmissible evidence (see Sommers & Kassin, 2001; Spellman, 2007). According to the inability/myopia explanation, fact-finders have certain (inadmissible) information about a case at their disposal and are unable to use meta-information pertaining to the rules of evidence, to disregard it in reaching a verdict. According to the unwillingness/just-verdict-motivation explanation, the effects of inadmissible evidence might reflect the fact-finders' unwillingness to ignore or disregard a piece of evidence that is relevant and true, but not admissible under the rules of evidence. For instance, a judge and probably even more so a juror, might think that a defendant who has admitted being guilty to his attorney should be convicted, despite the fact that this piece of information is protected by the attorney-client privilege and should thus be rejected as inadmissible evidence — the defendant is indeed guilty, after all.

1.3 The present research

Our goal in the present article was to test the effects that the truth bias may exert in a courtroom. To achieve this, we asked whether mock jurors and professional judges are truth-biased when given information that is blatantly flagged as false in the context of a given case. Questioning the truth of the evidence we used eliminates the possibility that fact-finders are willing or motivated to rely on it. Thus, any truth bias effect detectable in our studies would clearly reflect a myopic reliance on the presented evidence and could not be explained on motivational grounds.

1.3.1 General method

Across the two studies reported below, we relied on a paradigm that has been previously used in the truth-bias literature (Gilbert, Tafarodi & Malone, 1993; Pantazi et al., 2018). We used the four crime reports from Pantazi et al. (2018). The four reports were constructed based on two criminal cases, both involving an armed robbery, by each time incorporating false information that was either aggravating or mitigating the crime described (see Appendix for the full reports). The true evidence in the reports (27 statements), which also constituted the reports' main body, described the main events around the armed robberies.[1] In each report, interspersed among the true evidence were 7 false statements that constituted either aggravating circumstances or mitigating ones.

Each participant listened to two reports, each concerning a different case. The one report contained aggravating false evidence and the other report contained mitigating false evidence. For example, while the false evidence in one report would suggest that the defendant was a regular offender and depicted the robbery as a heinous crime, the false evidence in the other report would suggest that the defendant did not have a criminal record (admissible evidence in Belgium where the study was conducted) and showed signs of remorse. The true-evidence and false-evidence statements in each report were distinguishable as the two types of statements were uttered by speakers of different genders. Participants were explicitly told that one speaker provided truthful evidence while the evidence provided by the other speaker was false and taken from other reports that were unrelated to the present case.[2] Participants were not told whether the false statements were taken verbatim from other reports or created by mixing-up chunks from several other reports. Since, for reasons of cohesion the false statements of each report mentioned the names of the defendant and victim, in retrospect it may have been difficult for participants to view the false statements as totally "unrelated" to the case.

In line with past studies, to assess whether participants were truth-biased, we first asked them to propose a prison term for the two defendants and judge them on several dimensions. If participants managed to efficiently rely on meta-information about the speakers and their reliability and disregard the false evidence, they should have judged the two defendants equivalently. More severe judgments for the falsely aggravated than the falsely mitigated crime, on the other hand, would signal that participants were myopic regarding the meta-information about the accuracy of presented evidence. We also asked participants to recall whether pieces of evidence contained in the reports were true or false.

[1] According to a pre-test, the true evidence was rated as equally serious across the two reports (see Pantazi et al. 2018 for the pre-test information).

[2] We adopted this strategy from Gilbert et al. (1993) in order to ensure that participants would not infer that negating the false statements would create true statements.

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

Misremembering more false evidence as true than true evidence as false would be another indication of truth bias, given that the number of true and false statements in our memory test was equal.

Building on this general paradigm, in Study 1 we tested a sample of student mock jurors. To increase our study's external validity, and to test whether accountability might be a moderating factor of meta-cognitive myopia in court, we included a manipulation aimed at participants' perceived accountability, borrowed from Tetlock (1983b). In Tetlock (1983b), accountability was found to moderate belief-perseverance effects. Specifically, participants tended to be more severe towards a defendant if they first received incriminating evidence than if they first received exculpating evidence. However, this effect disappeared when participants were made to feel accountable before receiving the evidence (pre-accountable group), as opposed to participants who were not rendered accountable (control group) or were rendered accountable only after having received the evidence (post-accountable group). Given the potential of accountability to play a role in the courtroom setting, in Study 1 we included a similar accountability manipulation. Study 2 did not manipulate accountability, but it tested professional judges using the aforementioned truth-bias paradigm.

2 Study 1

2.1 Method

Each participant listened to the aggravating version of one report and the mitigating version of the other report, each containing similar true evidence, but false evidence that was either aggravating or mitigating. For half of the participants, it was the female speaker who provided true information in the reports and the male speaker who provided the false information. For the remaining participants, the speaker/truth-value combination was reversed. Since in Pantazi et al. (2018) the speaker/truth-value combination did not affect whether people were truth-biased or not and this question was out of the scope of the present research, we treated this as a randomization rather than an experimental factor of our design and disregarded it in our analyses.

Participants were randomly assigned to one of three groups. "Pre-accountable" participants were told, at the beginning of the experiment, that they would be video-recorded while they orally justify their judgments about the defendants to the experimenter; "post-accountable" participants received the same information after listening to the reports but before making their judgments; a third control group simply performed the task without being asked to justify their judgments. Since we were not really interested in participants' justifications of their responses, but only in the effect of accountability on them, we did not actually video-record participants' judgments. Instead, after they had responded to the judgment and memory questions, participants in the two accountable groups were told that, finally, they had been assigned to a control group and would not be video-recorded while justifying their responses, but rather asked to write down their justifications. The accountability manipulation aimed at ensuring high levels of accountability at least in some of our mock-juror participants, thereby increasing the external validity of the study, while also testing a potential moderating role of accountability. The distinction between the pre- and post-accountable groups tested whether the potential moderating role of accountability may be due to the elicitation of different information processing, or rather by urging jurors to adjust their responses after they had processed case-related information.

As mentioned in the general methods section, we employed two measures of truth bias, judgment and memory, for all subjects. The design for the judgment analysis was a mixed 2 (false evidence: aggravating vs. mitigating; within-subjects) × 3 (group: pre-accountable vs. post-accountable vs. control; between subjects). The design for the memory analysis was a mixed 2 (statement type: true vs. false; within-subjects) × 3 (group: pre-accountable vs. post-accountable vs. control; between subjects).

2.2 Measures

2.2.1 Judgments

Mock jurors were asked to make five judgments for each defendant. Specifically, they provided (a) a prison term (0–5 years) and (b) an index of punishment severity (extremely light 0–10 extremely severe), and reported (c) their feelings towards him (total indifference 0–10 total aversion), (d) how dangerousness he was (not at all dangerous 0–10 extremely dangerous), and (e) how likely he was to recidivate (not at all probable 0–10 extremely probable).

2.2.2 Memory

Similar to Pantazi et al. (2018), all participants were presented with two 24-statement lists, one for each report (see Supplement). Each of the two lists consisted of 4 true, 4 false, and 16 new items that had not been presented but were plausible in the context of the report. Participants were asked to decide for each statement whether it appeared in the report or not and, if so, whether it was true or false. We were interested in the comparison of the percentage of true statements misremembered as false and false evidence misremembered as true.

Based on past research, we employed judgments and memory as complementary measures of the truth-bias effects and expected a significant correlation between the percentage of false statements that participants misremembered as true and the difference between their judgments of (falsely) ag-

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

gravated and mitigated defendants (see Gilbert, Tafarodi & Malone, 1993; Peter & Koch, 2015, Pantazi et al., 2018).

2.3 Participants and Procedure

According to G*power (3.1; Faul, Erdfelder, Albert-George & Buchner, 2007), to detect an accountability effect similar to Tetlock's (f = 0.49; reflected in false statement × group interaction for judgments and statement type × group interaction for memory) with .95 power at the .05 alpha level, we would need 27 and 24 participants for the two analyses, respectively. To detect the memory (f = 0.22) and judgment (f = 0.34) truth bias within-subject effects reported by Pantazi et al. (2018) with .90 power at the .05 alpha level, we would need 87 participants (see the Supplement for a full description of the Power analysis). We recruited 73 first-year psychology students to serve as our mock jurors in exchange for course credits. Participants were randomly assigned to one of the three groups (pre-accountable vs. post-accountable vs. control). One participant was excluded because of being dyslexic and the responses of one subject failed to be recorded. The final sample comprised 71 participants (63 female, 9 male; M_{Age} = 19.66, SD_{age} = 5.08; 23–24 participants per group).

Participants were tested in individual booths of an experiment room, in groups of maximum eight. After filling in an informed consent form, all participants were given the instructions describing their task (and implementing the accountability manipulation for the pre-accountable group). All participants were instructed to adopt the role of a trial judge and listen carefully (and only once) to the two crime reports. Participants were informed that they would be asked to judge the defendants and to remember some of the reports' details. Participants were explicitly informed that they should listen to the reports very carefully because the evidence provided by the one speaker (e.g., the male) was accurate and truthful, while that provided by the other speaker (e.g., the female) was inaccurate and drawn from the reports of other cases. The post-accountable group manipulation was implemented in the questionnaire, while the pre-accountable group was "reminded" of its accountability in the questionnaire. All participants listened to the two reports in the same order (audios lasting between 98–110 seconds). Yet, since each participant listened to the aggravating version of the one report and the mitigating version of the other report (which was counterbalanced across participants), half of the participants listened to the aggravated report first, and the other half listened to the mitigated report first. After listening to both reports, participants answered a computer-based questionnaire, in which they had to judge the defendants and complete the memory test. Because the reports were relatively short — shorter than the information that actual jurors receive in the context of a trial — we thought that asking participants to respond after listening to both reports would provide a relatively stricter test than asking them to respond to each report right after listening to it.

2.4 Results

Since our memory variable was categorical, only our judgment data were scanned for outliers per false evidence condition, using Leys, Ley, Klein, Bernard and Licata's (2013) method of Median Absolute Deviation with a constant of 3.[3] Thirty out of the total 710 judgments (4.2%) were excluded as outliers. All pairwise comparisons presented are Bonferroni-corrected, and Cohen's d effect sizes and 95%CI are reported for all pairwise comparisons. We analyzed participants' judgments and memory using mixed-effect models, starting with Judd, Westfall and Kenny's (2017) recommendations for model choice, and eventually deleting any random term that would appear to be redundant based on their estimated variance components.

2.4.1 Judgments

In line with Pantazi et al. (2018), we treated the five judgments as different items measuring the mock jurors' opinions about the defendants (the prison term value was multiplied by two for compatibility with the other judgments). We ran a generalized linear mixed model (using the lme4 [Bates, Machler, Bolker & Walker, 2015] and lmerTest [Kuznetsova, Brockhoff & Christensen, 2017] packages in R) with false evidence, group, and their interaction as fixed factors. For random effects, we initially included: intercepts for participants and items, slopes for participants and items per false evidence condition, intercepts for participants-by-items, slopes for items per group condition, and slopes for items per group-by-false evidence conditions. Finally, we also included covariances between all the estimated random slopes and intercepts (see Judd et al., 2017). As fitting this model resulted in a singular variance/covariance matrix, we adapted the model, excluding from the random portion the per item slopes and interactions involving the group and false evidence condition (for which variances were extremely low). This change did not alter the fixed effects. Since the true evidence described two very similar crimes of pretested similar seriousness in the two reports, a main effect of false evidence (more severe judgments for the aggravated than the mitigated defendant) would signal the effect of meta-cognitive myopia and truth bias. Moderating effects of accountability would be indicated by a significant false evidence × group interaction.

[3]Following Leys et al.'s (2013) suggestions, we excluded data points according to the following rule: $M - 3 \cdot MAD < x_i < M + 3 \cdot MAD$, where x_i is the ith data point, M is the median of the data distribution, and MAD is the median absolute deviation of the observations from the median multiplied by the constant b = 1.4826 linked to the assumption of normality of the data. The threshold of 3 we selected is proposed by Leys et al. (2013) as a conservative criterion.

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

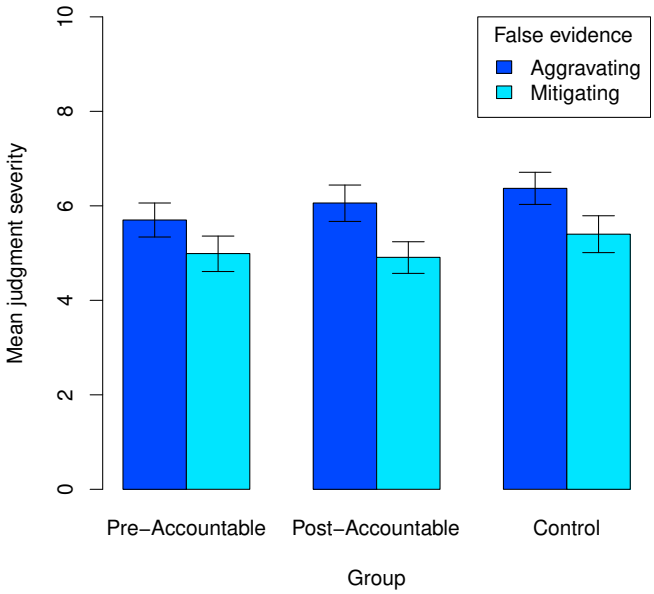

FIGURE 1: Study 1: Mean judgment severity as a function of whether the false information contained in the reports was aggravating or mitigating. Judgments are displayed separately for each group. Error bars represent 95% CIs.

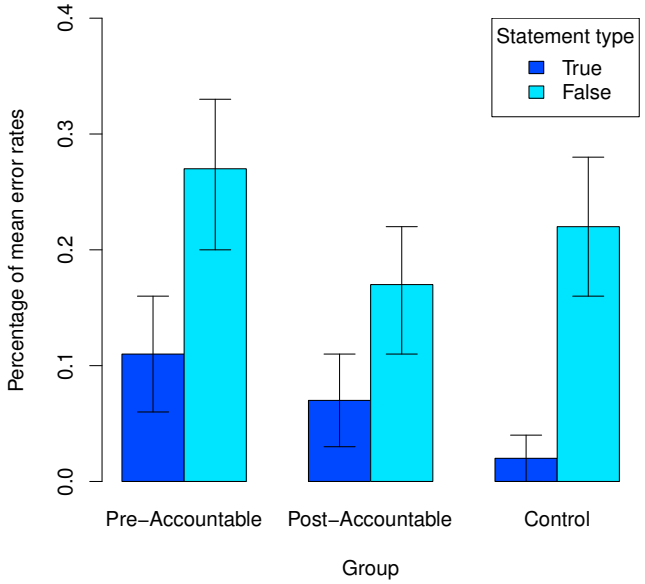

FIGURE 2: Study 1: Mean error rates for the true and false information in the memory test, separately for each group. Black bars represent true statements misremembered as false and grey bars represent false statements misremembered as true. Error bars represent 95% CIs.

Figure 1 plots the severity of participants' judgments for the aggravated and mitigated perpetrator by group (numerical mean judgments and their SDs appear in Table S1 of the Supplement. Participants were more severe towards the defendant for whom they received incriminating false information than towards the defendant for whom they received mitigating false information ($F(1, 9.29) = 10.49$, $p = .009$, mean difference = 0.94, 95% CI [0.23, 1.48], $d = 0.47$), regardless of their group ($F(1, 68.11) = 0.22$, $p = .637$ for the main effect of group; $F(1, 67.50) = 0.72$, $p = .396$ for the interaction).

2.4.2 Memory

We used a Generalized Linear Mixed Model for binomial data to analyze raw memory responses for true and false statements remembered as having been presented (GENLINMIXED procedure in SPSS; Quené & van den Bergh, 2008).[4] Statement type, group, and their interaction were included as fixed factors. As random factors, we included intercepts of subjects and statements, slopes for subjects per statement type condition and slopes for statements per group condition, as well as slope-intercepts covariances (see Judd et al., 2017 for the model specification). The truth bias would be demonstrated in a higher error rate for false than for true statements. A reduced truth bias under accountability would be illustrated by a significant statement type × group interaction, signaling that the difference in the error rates of the true and false statements is smaller in (either of) the accountable groups. As revealed in Figure 2 (see Table S2 for numerical percentages and SDs), false evidence was significantly more likely to be misremembered as true than true evidence as false ($F(1, 1138) = 14.43$, $p < .001$, mean difference = .15, 95%CI [.05, .25], $d = 0.44$). There was also a main effect of group ($F(2, 1138) = 3.16$, $p = .043$) qualified by a statement type × group interaction ($F(2, 1138) = 3.06$, $p = .047$). The interaction was not directly related to the mock jurors' truth bias and was rather due primarily to the fact that the pre-accountable group misremembered more true evidence as false than the control group ($t(1, 1138) = 2.28$, $p = .026$, $d = 0.37$). In line with our prediction, memory and judgment results were compatible, as the amount of false evidence that participants misremembered as true significantly correlated with the impact that false evidence had on mock jurors' judgments ($r = .328$, $p = .005$).

To further examine the operation of a truth bias in our participants' memory, we also analyzed memory responses based on Signal Detection Theory. Because the assumptions of normality and equality of standard deviations for the signal and noise distributions are impossible to test in our categorical responses (see Stanislaw & Todorov, 1999), we preferred to use the non-parametric measures of sensitivity and bias, A′ and B″ respectively, using the following formulas to calculate HIT and FA rates:

1. HIT = True statements remembered as True/(True statements remembered as True + True Statements remem-

[4] We report memory patterns for new statements in the Supplement, to exclude the possible operation of a general guessing bias in the memory test.

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

TABLE 1: Mean sensitivity and bias measures and 95% CIs per group in Study 1.

	Pre-Accountable	Post-Accountable	Control
A′	79 [.69,.90]	.88 [.84,.93]	.91 [.88,.95]
B″	−.20 [−.37,−02]	−.14 [−.32,.02]	−.33 [−.48,−.18]

bered as False)

2. FA = False statements remembered as True/(False statements remembered as True + False statements remembered as False)

A′ values normally lie between .5 and 1, indicating indistinguishability and perfect distinguishability of noise and signal, respectively. B″ ranges from -1 to 1, indicating extreme bias towards "true" responses and extreme bias towards "false" responses, respectively (see Stanislaw & Todorov, 1999). As is customary in SDT analyses, the HIT and FA values equal to 0 or 1 were adjusted by .01 (see Kane, Conway, Miura & Colflesh, 2007). Mean sensitivity and bias measures per group and 95% CIs are presented in Table 1. Overall, the sensitivity measure suggested that participants in all groups displayed fairly good discrimination of true and false statements, between 0.79 and 0.91. The bias measure on the other hand, which was negative in all three groups, suggested that participants tended to respond "true" in the memory test. Two one-way ANOVAs comparing sensitivity and bias across groups suggested that, while the three groups did not differ in bias ($F(2, 68) = 1.42$, $p = .247$), they did differ in their ability to discriminate the true from the false statements ($F(2, 68) = 3.46$, $p = .037$). More specifically, pairwise comparisons suggested that the control group displayed significantly better discrimination than the pre-accountable group ($p = .041$). The post-accountable group lay in-between and did not differ significantly from either the pre-accountable group ($p = .192$) or the control group ($p > .99$). Finally, the difference in the judgments for the aggravated and mitigated perpetrators was negatively correlated to participants' discrimination ($r(71) = -.346$, $p = .003$) but not related to participants' bias ($r(71) = -.131$, $p = .278$).

2.5 Discussion

Study 1 suggested that mock jurors may judge defendants based on evidence that they encounter, even if they clearly know this evidence to be false. What is more, mock jurors tend to misremember false evidence as being true more than they remembered true evidence as being false. The two types of measures, judgment and memory, were correlated: The more participants misremembered false statements as true and the less they discriminated between the true and false statements in the memory test, the more their judgments were affected by the false statements. These findings indicate that the truth bias may play a significant role in the decisions of students in mock juries. In other words, mock jurors seem to be meta-cognitively myopic and therefore unable to adjust their beliefs and decisions about cases based on meta-information they receive about case-related evidence. Concerning the hypothesis that the increased accountability jurors may feel in a judicial setting may increase their capacity to disregard explicitly false evidence when reaching verdicts, it could not be verified. Nevertheless, at the memory level it was obvious that, if anything, accountability worsened mock jurors' memory about misinformation since participants who were made to feel accountable when listening to the information had worse memory for the cases' true information. Based on the SDT analyses, we can infer that all three groups were biased towards answering "true", while the poor memory performance of the pre-accountable group was due to a lower ability of this group to discriminate the true from the false statements in the reports.

This unexpected finding is in line with previous studies suggesting that, under specific conditions, accountability may actually backfire, especially if the feeling of accountability depletes the cognitive or emotional resources of decision-makers (see Lerner & Tetlock, 1999). This ironic effect of accountability could also be explained under the lens of an Ironic Mental Processes account (see Wegner, 1994). According to this line of reasoning, pre-accountable participants might have tried to increase cognitive efforts to pay more attention to the case information, but their increased cognitive efforts finally had an ironic effect on their memory, possibly by increasing the accessibility of false evidence. The SDT analyses corroborate such an "ironic" explanation, as the pre-accountable group did not exhibit a significantly different threshold point (bias). This eliminates the possibility that pre-accountable participants were simply more "vigilant", and overall tended to reject more of the encountered information as false.

Regardless of the mechanism behind the effects of the accountability manipulation in the pre-accountable group, the fact that this group had a different memory performance from the other two groups suggests that the manipulation did have an effect, albeit an ironic one. In any event, the important finding of Study 1 is that mock jurors seem to be affected by evidence they know to be false. In Study 2, we went on to test whether professional judges would display analogous effects.

3 Study2

In Study 2 we asked professional judges to perform the same task as the control group in Study 1. We were interested in whether the judges would display a truth-bias effect in the

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

first place. Additionally, we were interested in comparing the judges' performance to that of the mock jurors in Study 1.

3.1 Method

The same criminal reports as in Study 1 were used. To render the task relevant for professional judges, we consulted an experienced judge, who suggested slight adaptations to our materials. First, we reduced the length of the instructions, while keeping their core message. Second, we included only the prison-term measure (now ranging from 0-10 years) and the general index of punishment severity. Our rationale was that only these among our five judgments measures constitute legitimate judgment dimensions for real judges in a court and, to ensure ecological validity, we did not want to lead participating judges to think in a non-professional manner in the context of our task. The memory test was the same as in Study 1.

For each measure, judgments and memory, we conducted two separate analyses. The first analysis tested the operation of the truth bias in the judges' sample alone. The second analyses compared the judges' judgments and memory responses to those of the mock jurors in Study 1. We included only the control and post-accountable groups from Study 1, given that the accountability manipulation did seem to have a small effect on pre-accountable participants' memory. This choice allowed us to have comparable sub-samples across studies, while also relying on a sufficiently large sample to potentially detect differences between the two studies.

3.1.1 Participants and Procedure

According to G*power (3.1; Faul et al., 2007), we would need 45 participants to reach .95 power in detecting the hypothesized within-subject or between-samples effects at the .05 alpha label (see the Supplement for a more detailed power analysis). The judges were recruited through an invitation sent to the e-mail list of the judges at the Law Court of a European capital and participated voluntarily. 142 e-mails were sent in total. We managed to recruit 42 professional judges (31 female and 11 male; M_{age} = 48.68, $range_{age}$ = 35 – 62), meaning that we had a response rate of 30%. Fourteen of the recruited judges were specialized as civil judges, 19 as penal, 2 as juvenile, 1 as commercial, 1 as judge of inquiry,[5] and 4 had more than one of these specializations. The judges in our sample had an average experience of 9.61 years (SD = 6.34). They were tested individually on a laptop in an isolated room of the Law Courts building using E-prime (2.0) to record their responses. The rest of the procedure was the same as for the control group of Study 1.

[5]A judge of inquiry is responsible for conducting the investigative hearing prior to a criminal trial.

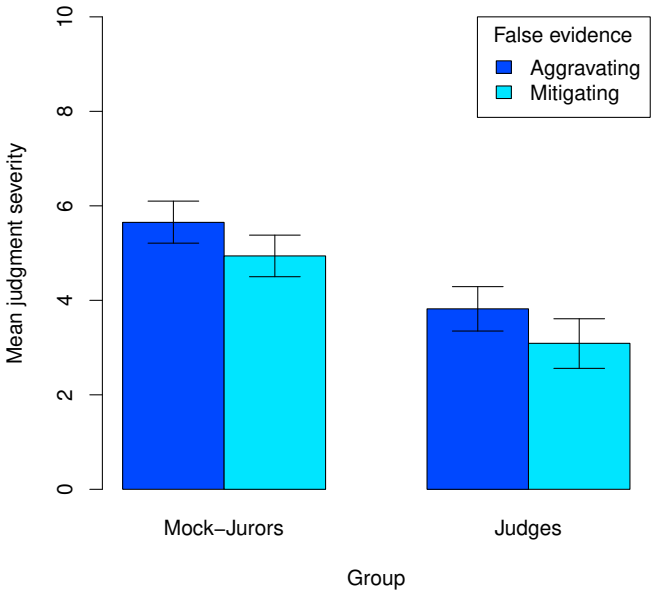

FIGURE 3: Study 2: Mean judgment severity as a function of whether the false information contained in the reports was aggravating or mitigating. The judgments of the judges are pitted against those of the mock jurors in Study 1. Error bars represent 95% CIs.

3.2 Results

3.2.1 Judgments

The judgments of one judge failed to be recorded. Since judges' proposed prison terms turned out to have very different distributions from those of index of punishment severity (despite reflecting analogous numerical scales going from 0–10), we excluded outliers per false evidence condition, separately for the two measures. 13 out of the 164 responses (7.7%) were excluded as outliers, all of which constituted prison terms proposed for the mitigated defendant. Thus, there was considerable variation in the prison term that judges proposed for the same (mitigated) case.

Figure 3 plots the judges' judgments pitted against those of the mock jurors in Study 1. Because in Study 2 we included only 2 judgments per participant, we included judgments (prison term vs. punishment severity) as a fixed factor rather than as a random one. The first analysis included false evidence, judgments, and their interaction as fixed factors. We also added experience in years (centered) as a covariate, as well as the interaction of experience by false evidence. In terms of random effects, we included only intercepts and slopes for subjects per false evidence condition.

The main effect of false evidence was significant (F(1, 36.47) = 5.30, p = .02, mean difference =.73, 95%CI [-0.23, 1.69], d = 0.34), with the aggravated perpetrator judged more severely (M = 3.82, SD = 2.13) than the mitigated one (M = 3.09, SD = 2.15). There was also a main effect of judgment

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

(F(1, 64.63) = 268.80, p < .001, mean difference = 2.83, 95%CI[2.29, 3.36], d = 1.67), with values for punishment severity much higher (M = 4.78, SD = 1.89) than those for prison terms (M = 1. 95, SD = 1.35). Experience did not have a significant effect (F(1, 36.41) = 2.37, p = .131), nor did it interact with condition (F(1, 37.03) = 0.27, p = .606). For the difference in the proposed sentence alone (M = 2.3, SD = 1.34 for the aggravated and M = 1.18, SD = 0.48 for the mitigated perpetrators, respectively), this amounted to more than a year (F (1, 64) = 23.74, p <.001, mean difference = 1.22, 95% CI [0.68, 1.55], d = 1.21), which means that the judges would propose longer sentences after receiving false incriminating evidence in 83% of the times (see Lakens, 2013). Again, experience (F(1, 64) = 0.58, p = .446) did not have a significant effect or interact with condition (F(1, 64) = 0.82, p = .366).

Next, to compare the judges' judgments with those of mock jurors, we ran a mixed model, with false evidence, group, and judgment as fixed factors. We also included participant intercepts and slopes per false evidence condition, as well as intercepts-by-slopes covariances as random factors. The main effect of false evidence was significant (F(1, 78.64) = 17.00, p < .001, mean difference = 0.65, 95% CI [-0.07, 1.37], d = 0.28) and so was the main effect of study (F(1, 81.00) = 50.16, p < .001, mean difference = 1.82, 95%CI [1.09, 2.54], d = 0.78), signaling both a general truth-bias effect across studies, and that the judges were more lenient, overall, towards both defendants. However, the false evidence × group interaction was not significant (F(1, 78.59) = .001, p = .968), signaling that the false evidence appeared to have a comparable impact on the judgments of the judges and the mock jurors. Finally, the judgment factor turned out to be significant once more (F(1, 147.15) = 103.91, p < .001) signaling that the scores for punishment severity were higher (M = 5.31, SD = 1.79) than for proposed prison terms (M = 3.11, SD = 2.49, mean difference = 2.20, 95%CI [1.65, 2.74], d = 0.84).

3.2.2 Memory

Figure 4 presents the judges' memory pattern, pitted against that of the mock jurors. We first ran a Generalized Linear Mixed Model on the judges' memory errors with statement type (true vs. false), experience (centered) and their interaction as fixed factors. Participant and statement intercepts, as well as participant slopes per statement type, were included as random factors. The judges misremembered more false statements as true (M = .25, SD = .43) than true statements as false (M = .05, SD = .21; F(1, 668) = 50.96, p < .001, mean difference = .20, 95%CI [.05, .34, d = .59). As with judgments, years of experience (F(1, 668) = 1.75, p = .191) and its interaction (F(1, 668) = 0.74, p = .390) with statement type were not significant. As in the case of mock jurors, the correlation between the judges' judgments and their explicit memory errors (false items misremembered as true) was significant (r(42) = .368, p = .018).

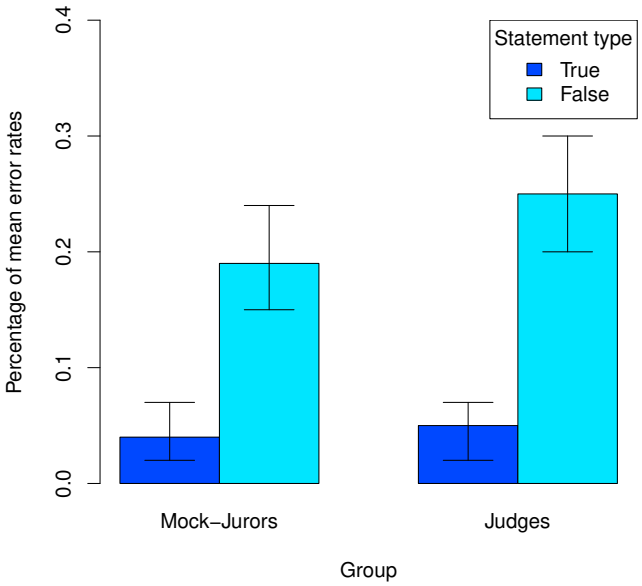

FIGURE 4: Study 2: Mean error rates for the true and false information in the memory test. The memory pattern of the judges is pitted against the memory pattern of the student participants of Study 1. Error bars represent 95% CIs.

As announced, we planned to directly compare the judges' memory pattern to that of the mock jurors (from two of the three conditions) in order to test whether the former may be less truth-biased than the latter. However, the effect sizes in the two populations alone suggest that, if anything, the judges were more truth-biased at the level of memory (d = 0.59) than the mock jurors (d = 0.44). Indeed, in a secondary analysis, we included statement type and group (judges vs. mock jurors) as fixed factors, and intercepts of subjects and statements as random factors. This analysis revealed again a main effect of statement type (F(1, 1428) = 38.95, p < .001). Neither the main effect of group (F(1, 1428) = 0.37, p = .542) nor the statement type × group interaction (F(1, 1428) = 0.75, p = .384) was significant.

In terms of SDT measures, judges exhibited an overall good discriminability for the true and false statements (mean A′ = .86, 95% CI [81, 91]) and, like the mock jurors, they were biased towards responding "true" in the memory test (mean B″ = −.29, 95% CI [−.41, −.17]). According to two one-way ANOVAs, these measures did not differ from those of the control and post-accountability groups of Study 1 (F(1, 87) = 1.89, p = .172 for sensitivity; F(1, 87) = 0.40, p = .529 for bias). Neither of the SDT measures in the judges' sample correlated significantly with the judgment measure (r(41) = −.230, p = .148 for discrimination; (r(41) = −.145, p = .367 for bias).

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

3.3 Discussion

Study 2 suggests that judges were affected by the false information on their verdicts to a similar degree as the two groups of mock jurors were. Interestingly, there was much variability in the prison terms that the judges proposed when they listened to false mitigating information, which in itself suggests some degree of subjectivity in judges' decisions. In any event, the effect size obtained in our study suggests that 83% of the time, the judges would propose longer sentences after receiving false evidence that aggravated the crime than after receiving mitigating false evidence.

At the memory level, judges displayed a tendency to misremember false aggravating and mitigating information as true, just like the mock jurors in Study 1. The SDT analyses suggested that they exhibited levels of discrimination and bias similar to those of mock jurors. Interestingly, the judges' capacity to filter our false information, both at the judgment and the memory level, did not seem to be contingent on their years of experience, which suggests that the truth bias in the courtroom may not easily be amenable to training or experience, at least not as these occur naturally in the course of a judge's career.

4 General Discussion

The two studies presented in this article suggest that judges and jurors may be affected by information they know to be false about a case. They are, in other words, truth-biased. This finding has important theoretical as well as practical implications, which we discuss in the remainder of this article.

4.1 Theoretical Implications

As outlined in the introduction, the account of meta-cognitive myopia has been put forward to theoretically unify otherwise disparate biases and heuristics, by tracing their common origins to a general meta-cognitive failure to assess and consider available meta-information about information present in the environment. By attesting to the potential operation of the truth bias in courtrooms, the present article, therefore, does not only contribute to the literature on the effects of cognitive biases in judicial contexts (e.g., Daftary-Kapur et al., 2010; Englich, Mussweiler & Strack, 2006). It additionally extends the implications of the meta-cognitive myopia and the truth-bias literature (Fiedler, 2012; Pantazi et al., 2018; Street & Masip, 2015) to the field of legal decisions. We, thus, hope to put forth meta-cognitive myopia as a common basis for the systematic and comprehensive assessment of the operation of cognitive biases and heuristics in judicial decision-making. The finding that, in our paradigm, participants took the false information into account, despite this being clearly designated as false in the context of the cases in hand, indicates participants' limited ability to take meta-information into account. Similar meta-cognitive limitations are likely to lead to the operation of various other biases in a courtroom. For example, the fundamental attribution error, the confirmation bias, and the repetition effect have already been proposed to have significant effects in fact-finders' decisions (Foster et al., 2012; Neuschatz, Lawson, Swanner, Meissner & Neuschatz, 2008; Peer & Gamliel, 2011). We suggest that the lack of a systematic assessment of available information lies at the core of such biases, thus offering an alternative account of their operation, not in terms of limited cognitive capacity or lack of motivation but rather on the basis of a meta-cognitive failure (see Fiedler, 2012). Naturally, a better understanding of the reasons why biases may operate in a judicial context is a prerequisite for designing successful interventions combating them (Arkes, 1991).

A question that may arise with respect to the interpretation of the SDT results is the extent to which the bias index actually reflects a "truth bias" rather than an "ideal acceptance threshold", given that the priors of true vs. false statements in the reports were approximately .75 to .25. As already noted, we posit that the truth bias results from a largely legitimate and ecological information-processing mode, adapted to many real-life situations in which people are more likely to encounter truthful than untruthful information. Given that in real life, truthful information is more common than untruthful information, a tendency towards believing can be considered not as a "bias", but as an ecologically adapted threshold to the signal. However, the frequencies of true and false statements in our memory test were equal, which means that, in this context, the tendency towards responding "true" does reflect a bias in strict SDT terms. The threshold that participants used in the memory test could of course indicate that participants had prior assumptions favoring the "true" response, based on their perceptions of the truthfulness/falsehood ratio in the reports material or in real life. In other studies, we have replicated the same findings using report materials that had an equal mix of true and false statements (Pantazi et al., 2018). This means that our results more likely reflect participants' general thresholds of acceptance rather than thresholds adapted to the reports' materials. It is exactly this inability to adapt acceptance thresholds to any given environment that, in our view, generates the "truth bias". In any event, even with the insights that SDT measures can offer, it is unclear whether our results overall reflect a low acceptance threshold at the moment of initial information processing, an inability to tag or retain and recall false tags, or even a source-monitoring limitation alongside a bias to respond "true" in the memory test (see Johnson, Hashtroudi & Lindsay, 1993). Ideally, more comprehensive tests of the truth bias, along the lines of multinomial processing trees, are necessary to further clarify such aspects of the phenomenon.

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

The present findings also bear some relevance to interpretations of inadmissible evidence effects (e.g., Fleming et al., 1999; Landsman & Rakos, 1994; Schul & Manzury, 1990). Inadmissible evidence effects have generally been described either as instances of psychological reactance (Brehm, 1966), whereby jurors feel that instructions to disregard inadmissible but relevant information limit their freedom, and react to this limitation by indeed taking such information into account (Lieberman & Arndt, 2000); or as the result of ironic processes of mental control (see Lieberman & Arndt, 2000; Wegner, 1994), assuming that jurors' hypothetical efforts to follow instructions to ignore specific information have the ironic effect of increasing the accessibility of this information. The fact that our to-be-disregarded material was flagged as false suggests that our results reflect our participants' inability, rather than unwillingness, to disregard it, thus pointing to a cognitive rather than a motivational account. It would be good for future studies on inadmissible evidence to try disentangling the two possibilities.

Our studies did not provide any support for the hypothesis that accountability or any other factor specific to the courtroom may lead jurors and judges to adopt a special processing style that would render them more resistant to false evidence. Accountability did not have an effect on the threshold participants in the memory task used to classify statements as true, although it did seem to affect their discrimination. Therefore, the present findings are clearly incompatible with previous claims that jurors and judges may adopt a special, vigilant, processing strategy because they expect testimonies and evidence presented in courts to be generally inaccurate (see Schul & Manzury, 1990). Our findings show that such an expectation, if active at all, does not suffice to counter the effects of meta-cognitive myopia in mock jurors and in judges.

4.2 Policy Implications

Our results and the meta-cognitive myopia account that we put forward suggest that the mere presence of specific evidence or testimony in a judicial context can have a strong impact on the fact-finders' decisions. Crucially, the impact of such information may be exerted regardless of whether any of the implicated parties in a trial challenge its accuracy. Our studies reveal that meta-information about presented evidence, such as a party's objection to that evidence or a judge's decision to sustain or overrule an objection, is not bound to prevent presented evidence from affecting the triers of fact. Our findings thus challenge the efficacy of objections as a means to ensure that false testimonies or evidence presented in a trial do not ultimately affect the fact-finders' decisions.

Fortunately, the judiciary system is to some extent shielded by intrusions of illegitimate evidence, since objections are most often raised before a witness's answer or piece of evidence is presented in court. Therefore, most of the time, inadmissible or false evidence is prevented from entering the fact-finders' mental representations of a case in the first place. Nevertheless, objections can also be raised after a witnesses' response has been given. Such objections may not actually protect the fact-finders from the information that has already been presented. An important question that remains open from a policy perspective is therefore how we are to safeguard the rules of evidence, given the fact-finders' inability to take such meta-information into account.

Previous accounts of bias correction in jurors have pointed out the importance of carefully designing juror instructions, e.g., in terms of clarity and persuasiveness (see Lieberman & Sales, 1997; Wegener, Fleming, Kerr & Petty, 2000) and have specifically insisted on the necessity of providing sufficient justification for excluding specific evidence or testimonies. In our material, the justification for disregarding the information was straightforward: it was false. Yet such meta-information was not enough to eliminate the student mock jurors' and judges' reliance on it. This finding raises questions as for the utility of any meta-information provided as justification for disregarding evidence in a courtroom. We, therefore, believe that the only certain way of protecting procedures from false evidence is to deter its presentation in the first place. This could be accomplished if the content of evidence and testimonies that the opposing parties aim to present in a trial could be somehow pre-screened and approved as for their admissibility and accuracy.

Deterrence of illegitimate evidence presentation could also be accomplished by rendering parties and witnesses who may recklessly present such evidence in court more accountable. For example, witnesses might need to be better informed a priori about the kind of admissible and inadmissible evidence they can provide, and be explicitly asked to avoid presenting inadmissible evidence before agreeing to testify in court. While better knowledge might be a restorative strategy to prevent witnesses from presenting inadmissible evidence, this may not be the case for legal practitioners. There are reasons to believe that prosecutors and lawyers may intuit the irreversible impact that incorrect evidence may have in a trial and purposefully try to present it even if they realize that it will be objected to. We suggest that legal practitioners should somehow be kept liable in such cases, for example by keeping a record of a lawyer's tendency to resort to such techniques, which will result in an official complaint if this number becomes excessive. Although Study 1 did not provide evidence for an effect of accountability on the side of the receptor, other scholars have found that it may deter the emission of falsehood (Nyhan & Reifler, 2014).

In view of the above suggestions, a special mention should be made for the cases of perjury or recantation of testimony. "Witnesses have violated their judicially administered oaths to tell the whole truth since the beginning of American jurisprudence. . . " (Salzman, 1977, p. 273). A witness who

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press

recants their testimony is the major real-life counterpart of our experimental design: Jurors and judges listen to a piece of evidence which is (then) explicitly signaled to be false. Even if a defendant is lucky enough and the recantation of a false adversarial testimony takes place during the trial and not after years of the defendant's incarceration (e.g., People v. Dotson, 1981) our results suggest that the original, later-recanted testimony may still play a role in the fact-finders' decision. Probably an extreme example of the irreversible effect of a recanted testimony can be found in People v. Rivera (2011), where the defendant's false confession during investigation, which he later recanted in court, acted as the major basis for Rivera's conviction, since there was no other solid factual evidence linking the defendant to the crime.

Unfortunately, as Anne Bowen Poulin points out, "the law does not provide adequate protection from convictions based on lies" (Poulin, 2011, p. 331). Even if testimony or evidence used as basis for a conviction is proved to be false, it is quite challenging for defendants to be granted a new trial, once a conviction has been made. This is because courts often (a) require that the prosecutor personally knows that the evidence or testimony was false, (b) restrict the definition of false testimony to cases of perjury, and (c) shift the burden of proof to defendants, requiring them to prove the materiality of the false evidence or testimony for a new trial to begin. Even more sadly, although perjury is supposed to constitute a serious criminal offense, in many real cases false accusations are not actually responded to with any serious repercussions (as in the infamous case of Gary Dotson who served 10 years in prison for rape and aggravated kidnapping before his complainant recanted her testimony; see Cobb, 1986). We believe that by highlighting the subtlety of the possibly involuntary impact that information presented in the course of a trial may have, our results provide additional support for the materiality of false evidence in the outcome of trials. Therefore, we side with Poulin's suggestions to extend the prosecutor's responsibility for the presentation of false evidence to cases where the prosecutor was not aware of the evidence falsity, to extend the characterization false testimony to cases other than perjury, and to apply more lenient standards of materiality for a defendant to be granted a new trial.

In any event, we hope that our findings will raise awareness of the impact of primary evidence presented in a courtroom, an impact that can hardly be moderated or countered by instructions aiming at fact-finders' meta-cognition. This awareness could be especially relevant within legal circles, which tend to have an idealized image for the functioning of justice (see Spellman & Schauer, 2012). Granted that lawyers seem to already know and exert the power of inadmissible and even questionable evidence in courts, motivated by their duty to protect their clients (Gershman, 1995; Stuntz, 1993), we believe that the legal system might need to adjust its capacity to punish and deter the presentation of such evidence in court.

References

Alschuler, A. (1982). The Search for truth continued, the privilege retained: A response to judge Frankel. *University of Colorado Law Review, 54*, 67–81.

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19*(2), 93–108. http://doi.org/10.1016/0022-1031(83)90031-8.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*(6), 1037–1049. http://doi.org/10.1037/h0077720.

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

Bates, D. M., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*(4), 446–458. http://doi.org/10.1037//0096-3445.121.4.446.

Brehm, J. W. (1966). *A theory of psychological reactance.* Oxford, England: Academic Press.

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*(8), 1120–1129. http://doi.org/10.3758/BF03206381.

Cobb, S. (1986). Gary Dotson as victim: The legal response to recanting testimony. *Emory Law Journal, 35*(4), 969–1009. http://doi.org/10.3366/ajicl.2011.0005.

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

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction. *Journal of Applied Research in Memory and Cognition, 6*(2), 185–192.http://doi.org/10.1016/j.jarmac.2017.01.014.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality*

and Social Psychology Bulletin, 32(2), 188–200. http://doi.org/10.1177/0146167205282152.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. http://doi.org/10.1146/annurev.psych.59.103006.093629.

Faul, F., Erdfelder, E., Albert-George, & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659–676. http://doi.org/10.1037/0033-295X.107.4.659.

Fiedler, K. (2007). Construal level theory as an integrative framework for behavioral decision-making research and consumer psychology. *Journal of Consumer Psychology, 17*(2), 101–106. http://doi.org/10.1016/S1057-7408(07)70015-3.

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. *Psychology of Learning and Motivation, 57*, 1–55. http://doi.org/10.1016/B978-0-12-394293-7.00001-7.

Fiedler, K., Armbruster, T., Nickel, S., Walther, E., & Asbeck, J. (1996). Constructive biases in social judgment: Experiments on the self-verification of question contents. *Journal of Personality and Social Psychology, 71*(5), 861–873. http://doi.org/10.1037/0022-3514.71.5.861.

Fleming, M. A., Wegener, D. T., & Petty, R. E. (1999). Procedural and legal motivations to correct for perceived judicial biases. *Journal of Experimental Social Psychology, 35*(2), 186–203. http://doi.org/10.1006/jesp.1998.1375.

Foster, J. L., Garry, M., & Loftus, E. F.(2012). Repeated information in the courtroom. *Court Review, 48*, 44–47.

Frankel, M. (1975). The search for truth: an umpireal view. *University of Pensylvania Law Review, 123*(5), 1031–1059. http://doi.org/10.3868/s050-004-015-0003-8.

Frankel, M. (1982). The search for truth continued: More disclosure, less privilege. *University of Colorado Law Review, 54*, 51–66. http://doi.org/10.3868/s050-004-015-0003-8.

Freedman, M. H. (1975). Judge Frankel's search for truth. *University of Pensylvania Law Review, 123*, 1060–1066. http://doi.org/10.3868/s050-004-015-0003-8.

Gershman, B. L. (1995). Prosecutorial misconduct in presenting evidence: "Backdooring" hearsay. *Criminal Law Bulletin, 31*(2), 99–112.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221–233.

Guenther, C., & Alicke, M. (2008). Self-enhancement and belief perseverance. *Journal of Experimental Social psychology, 44*(3), 706–712.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. http://doi.org/10.1037/0033-2909.114.1.3.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1), 601–625. http://doi.org/10.1146/annurev-psych-122414-033702.

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114*(3), 678–703. http://doi.org/10.1037/0033-295X.114.3.678.

Kane, M. J., Conway, A. R. a, Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(3), 615–622. http://doi.org/10.1037/0278-7393.33.3.615.

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 131*(2), 287–297. http://doi.org/10.1037/0096-3445.131.2.287.

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.http://doi.org/10.1177/01461672972310005.

Kissine, M., & Klein, O. (2013). Models of communication, epistemic trust and epistemic vigilance. In J. Laszlo, J. Forgas, & O. Vincze (Eds.), *Social Cognition and Communication.*, pp. 139–154. New York: Psychology Press.

Klein, G., Shneiderman, B., Hoffman, R. R., & Ford, K. M. (2017). Why expertise matters: A response to the challenges. *Intelligent Systems, 32*(6), 67–73. http://doi.org/10.1109/MIS.2017.4531230.

Kuznetsova, A., Brockhoff, B. P., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(3), 1–26.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1–12. http://doi.org/10.3389/fpsyg.2013.00863.

Landsman, S., & Rakos, R. F. (1994). A preliminary inquiry into the effect of potentially biasing information on judges and jurors in civil litigation. *Behavioral Sciences & the Law, 12*, 113–126. http://doi.org/10.1002/bsl.2370120203.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*(2), 255–275. http://doi.org/10.1037/0033-2909.125.2.255.

Levine, T. R. (2014). Truth-default theory (TDT): a theory of human deception and deception detection. *Journal of Language and Social Psychology, 33*(4), 378–392. http://doi.org/10.1177/0261927X14535916.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: documenting the "veracity effect." *Communication Monographs, 66*(2), 125–144. http://doi.org/10.1080/03637759909376468.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviations around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. http://doi.org/http://dx.doi.org/10.1016/j.jesp.2013.03.013.

Lieberman, J. D.,& Arndt, J. (2000). Understanding the limits of limiting instructions: Social psychological explanations for the failures of instructions to disregard pretrial publicity and other inadmissible evidence. *Psychology, Public Policy, and Law, 6*(3), 677–711. http://doi.org/10.1037/1076-8971.6.3.677.

Lieberman, J. D., & Sales, B. D. (1997). What social science teaches us about the jury instruction process. *Psychology, Public Policy, and Law, 3*(4), 589–644. http://doi.org/10.1037/1076-8971.3.4.589.

Neuschatz, J. S., Lawson, D. S., Swanner, J. K., Meissner, C. A., & Neuschatz, J. S. (2008). The effects of accomplice witnesses and jailhouse informants on jury decision making. *Law and Human Behavior, 32*(2), 137–149. http://doi.org/10.1007/s10979-007-9100-1.

Nyhan, B., & Reifler, J. (2014). The effect of fact-checking on elites: A field experiment on U.S. state legislators. *American Journal of Political Science, 59*(3), 628–640. http://doi.org/10.1111/ajps.12162.

Pantazi, M., Kissine, M., & Klein, O. (2018). The power of the truth bias: false information affects memory and judgment even in the absence of distraction. *Social Cognition, 36*(2), 167–198. http://doi.org/10.1521/soco.2018.36.2.167.

Peer, E., & Gamliel, E. (2011). Heuristics and b2iases in judicial decisions. *Court Review, 49*, 114–118.

Peter, C., & Koch, T. (2015). When debunking scientific myths fails (and when it does not): the backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. *Science Communication, 38*(1), 1–23. http://doi.org/10.1177/1075547015613523.

Pickel, K. L., Karam, T. J., & Warner, T. C. (2009). Jurors' responses to unusual inadmissible evidence. *Criminal Justice and Behavior, 36*(5), 466–480. http://doi.org/10.1177/0093854809332364.

Poulin, A. B. (2011). Convictions based on lies: Defining due process protection. *Penn State Law Review, 116*, 331–401.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413–25. http://doi.org/10.1016/j.jml.2008.02.002.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology, 1*(4), 563–581. http://doi.org/10.1007/s13164-010-0039-7.

Robbennolt, J. K., MacCoun, R. J., & Darley, J. M. (2010). Multiple contraint satisfaction in judging. In *The Psychology of Judicial Decision-Making*, pp. 27–39. Oxford University Press: New York, NY.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Pesonality and Social Psychology, 32*(5), 880–892.

Salzman, A. (1977). Recantation of perjured testimony. *Journal of Ciminal Law and Criminology, 67*(3), 273–286.

Schul, Y., & Burnstein, E. (1985). When discounting fails: conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology, 49*(4), 894–903. http://doi.org/10.1037//0022-3514.49.4.894.

Schul, Y., & Manzury, F. (1990). The effects of type of encoding and strength of discounting appeal on the success of ignoring an invalid testimony. *European Journal of Social Psychology, 20*, 337–349.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology, 39*(06), 127–161. http://doi.org/10.1016/S0065-2601(06)39003-X.

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes, 51*, 416–446.

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. http://doi.org/10.1177/01461672012710012.

Spellman, B. (2007). On the supposed expertise of judges in evaluating evidence. *University of Pensylvania Law Review, 157*(1), 1–9.

Spellman, B., & Schauer, F. (2012). Legal reasoning. In K. J. Holyoak, & R. G. Morrison, *The Oxford Handbook of Thinking and Reasoning (2nd ed.)*, pp. 719–735). Oxford University Press: New York, NY.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. http://doi.org/10.3758/BF03207704.

Street, C. N. H., & Masip, J. (2015). The source of the truth bias: heuristic processing? *Scandinavian Journal of Psychology, 56*(3), 254–263. http://doi.org/10.1111/sjop.12204.

Stuntz, W. J. (1993). Lawyers, deception, and evidence gathering. *Virginia Law Review, 79*(8), 1903–1956. http://doi.org/10.1145/379437.379836.

Tetlock, P. E. (1983a). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*(1), 74–83.

Tetlock, P. E. (1983b). Accountability and the perseverance of first impressions. *Social Psychology Quarterly, 46*(4), 285– 292.

Thorson, E. (2015). Belief echoes: the persistent effects of corrected misinformation. *Political Communication, 33*(3), 460–480. http://doi.org/10.1080/10584609.2015.1102187.

Wegener, D. T., Fleming, M. A., Kerr, N. L., & Petty, R. E. (2000). Flexible corrections of juror judgments: Implications for jury instructions. Psychology, Public Policy, and Law, 6(3), 629–654. http://doi.org/10.1037/1076-8971.6.3.629

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*(1), 34–52. http://doi.org/10.1037/0033-295X.101.1.34.

Wilson, W. A. (1963). A note on fact and law. *The Modern Law Review, 26*(6), 609–624. http://doi.org/10.1111/j.1468-2230.1963.tb02231.x.

Wistrich, 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\T1\textendash1345.http://doi.org/10.2307/4150614.

Appendix

The four reports used in Studies 1 and 2. Statements in normal font are true, whereas statements in bold font are false. Each participant either read an aggravating report for Etienne and a mitigating report for Dimitri, or an mitigating report for Etienne and an aggravating report for Dimitri

Dimitri-Aggravating

The night of 28th November, 2011 Dimitri abruptly left his home in south Brussels after his wife had threatened to call the police. Dimitri and his wife had a fight, **which erupted over Dimitri's extramarital affairs,** and Dimitri screamed at his wife many times **and threatened to punch her.** Neighbors later testified that Dimitri and his wife had frequent loud disputes. While leaving his house, Dimitri had a brief fight with his brother in law who arrived in the meantime after having been called by Dimitri's wife. When his brother in law tried to stop him from deserting the house Dimitri pushed him and rushed in in his blue Renault. Consequently, with his car he rushed out of the parking and went to a friend's place who lived nearby. **At his friend's place Dimitri consumed a significant amount of cocaine.** Some hours later, Dimitri started heading home. He stopped his car in front of a night shop to buy cigarettes. When he arrived at the counter he took out a 9 millimeter. **He placed it against the clerk's head** and asked for money. The clerk silently rendered him 510 euros in small bills. **In the meantime, Dimitri sexually harassed one of the clients in the night shop.** While coming out, Dimitri overturned a large magazine rack **by violently kicking it out of his way.** After Dimitri left the night shop the clerk looked out of the window to see the license plate **and had the impression that Dimitri was laughing in his car.**

Dimitri-Mitigating

The night of 28th November, 2011 Dimitri abruptly left his home in south Brussels after his wife had threatened to call the police. Dimitri and his wife had a fight, **which erupted over Dimitri's wife's extramarital affairs,** and Dimitri screamed at his wife many times **and left his house so that his children would not witness such fights.** Neighbors later testified that Dimitri and his wife had frequent loud disputes. While leaving his house, Dimitri had a brief fight with his brother in law who arrived in the meantime after having been called by Dimitri's wife. When his brother in law tried to stop him from deserting the house Dimitri pushed him and rushed in in his blue Renault. Consequently, with his car he rushed out of the parking and went to a friend's place who lived nearby. **At his friend's place Dimitri confessed to his friend that he had long-lasting marital problems.** Some hours later, Dimitri started heading home. He stopped his car in front of a night shop to buy cigarettes. When he arrived at the counter he took out a 9 millimeter. **He told the clerk to stay calm** and asked for money. The clerk silently rendered him 510 euros in small bills. **In the meantime, Dimitri explained that he needed the money for a serious operation that his daughter should have.** While coming out, Dimitri overturned a large magazine rack **while he excused himself for what he had just done.** After Dimitri left the night shop the clerk looked out of the window to see the license plate **and had the impression that Dimitri was crying in his car.**

Etienne-Aggravating

The night of 16_{th} January, 2010, Etienne left his neighbor's apartment in Waterloo, after having spent the evening there. **During that time, he had been drinking beer and scotch.** He walked north toward highway N5, through a busy central street. From time to time, he turned to face traffic and extended his thumb in order to hitch a ride. Before reaching the highway, he was picked up by Victor, **an old man who was handicapped.** Victor was also heading north. Etienne indicated that he wanted to get to Brussels. Travelling along the highway, Victor noticed that Etienne was nervous, preoccupied and strangely silent. In order to remedy the tension, Victor started recounting a humorous incident that he had recently witnessed. Moments into the story, Etienne

pulled out a knife. **He held the knife tight against Victor's throat** and demanded Victor's wallet, watch and rings. Victor obeyed without saying a word. Having the valuables in his possession, **Etienne began muttering that he thought crippled people were really disgusting.** He then ordered Victor to take the first exit, directed him through a residential area, and forced him to stop at a deserted corner. As Etienne turned to get off the car, Victor began pleading with him to at least return him his wedding ring. **Before running off, Etienne threatened to slit Victor's throat if he tried to follow.**After having committed his crime Etienne returned to Waterloo **and broke into a house.** Etienne's children testified in his trial **and declared that Etienne committed crimes often.**

Etienne-Mitigating

The night of 16_{th} January, 2010, Etienne left his neighbor's apartment in Waterloo, after having spent the evening there. **That night Etienne had found out that his wife cheated on him.** He walked north toward highway N5, through a busy central street. From time to time, he turned to face traffic and extended his thumb in order to hitch a ride. Before reaching the highway, he was picked up by Victor, **a man who had robbed Etienne in the past.** Victor was also heading north. Etienne indicated that he wanted to get to Brussels. Travelling along the highway, Victor noticed that Etienne was nervous, preoccupied and strangely silent. In order to remedy the tension, Victor started recounting a humorous incident that he had recently witnessed. Moments into the story, Etienne pulled out a knife. **He told Victor he was ashamed of what he had to do and** and demanded Victor's wallet, watch and rings. Victor obeyed without saying a word. Having the valuables in his possession, **Etienne apologized by saying that his family was very poor.** He then ordered Victor to take the first exit, directed him through a residential area, and forced him to stop at a deserted corner. As Etienne turned to get off the car, Victor began pleading with him to at least return him his wedding ring. **Before running off, Etienne returned all the valuables to Victor.** After having committed his crime Etienne returned to Waterloo **and went to the police station to surrender.** Etienne's children testified in his trial **and declared that until the day of the crime, Etienne had been a good family guy.**

https://doi.org/10.1017/S1930297500007361 Published online by Cambridge University Press