How downplaying or exaggerating crime severity in a confession affects perceived guilt

Glenys A. Holt

School of Psychology, University of Chester

Matthew A. Palmer

Department of Psychology, University of Tasmania

Corresponding author: Dr Glenys Holt, School of Psychology, University of Chester,

Parkgate Road, Chester, CH1 4BJ, United Kingdom, g.holt@chester.ac.uk

Abstract

This study investigated how judgments of guilt are influenced by factual errors in confessions that either amplified or downplayed the severity of the crime. Participants read a confession statement and a police report. Information in the confession statement either was consistent with the facts of the crime in the police report, the suspect admitted to a worse crime than described in the police report, or the suspect admitted to a lesser crime than described in the police report. Mediation analyses showed that, compared to consistent confessions, both types of directional errors reduced judgments of guilt. Inconsistencies that made the suspect look better—but not those that made the suspect look worse—also increased judgments of guilt via a direct effect. Confessions that contain errors that appear to exaggerate the severity of the crime prompt no higher judgments of suspect guilt than confessions that are consistent with the facts of the crime. However, errors in confessions that are perceived to downplay the severity of the crime can prompt an increased perception of suspect guilt, when compared to a consistent confession.

Keywords: attribution theory, false confession, juror decision-making, inconsistencies, wrongful conviction

Word count: 5414

How downplaying crime severity in a confession affects judgments of guilt

False confessions have resulted in a surprising number of wrongful convictions, with the percentage of false confessions implicated in DNA exoneration cases estimated anywhere between 16% (Garrett, 2011) and 25% ("Innocence Project," 2017). These numbers stand in contrast to the commonly held belief that innocent people will not confess unless mentally ill or tortured (Leo & Ofshe, 1998). Contributing to wrongful conviction rates is the difficulty police investigators and jurors have in discovering and discounting false confessions, with a general inclination to simply trust that a confession is true. Drizin and Leo (2004) found that 30 out of 37 proven false confessors (81%) who decided to go to trial, were found guilty by the jury, even after pleading not guilty. This highlights not only the overall acceptance of confessions, but the unparalleled power that confessions hold in courts (Kassin & Neumann, 1997), even when retracted (Kassin & Sukel, 1997; Ofshe & Leo, 1997).

Jurors are not always best equipped with the knowledge or skills needed to effectively apply court instructions, evidence, and legal arguments to their decisions, and might instead use heuristics that are not reliable in the unique circumstances of a court case (Bornstein & Greene, 2011). One reason why people find confessions so convincing is the strength of the heuristic that assumes that, as there is no advantage to confessing, the confession must be true. However, there is conflicting evidence in the literature about whether this heuristic is robust in the face of inconsistent confession evidence. While inconsistencies in eyewitness testimony have been shown to reduce the credibility of the witness testimony and the likelihood of conviction (e.g. Berman, Narby & Cutler, 1995), inconsistencies in confessions are often perceived as having minimal impact on jurors' verdicts (Malloy & Lamb, 2010). However, some laboratory studies have found that inconsistencies can lead participants to reject a confession, resulting in a lower conviction rate (Henderson & Levett, 2016; Jones, Bandy, & Palmer Jr., 2019; Palmer, Button, Barnett, & Brewer, 2016; Woesthoff &

Meissner, 2016). Such conflicting findings in the context of confessions may be due to different presentations of inconsistencies being interpreted differently by jurors, such as one suspect giving multiple contradictory statements, versus a suspect giving a single confession that contradicts the facts of the crime (e.g., Jones et al., 2019). There is also the possibility that jurors might process confession inconsistencies differently according to the direction of the inconsistency - that is, where the inconsistency appears to increase or decrease the severity of the crime.

Directional errors

The direction of confession inconsistencies can be important if we think about how false confessions come about. Innocent people sometimes confess to a crime without adequate consideration of the consequences. This occurs because the person believes confessing is the only way to stop a stressful interrogation and begin the process of clearing their name (Kassin, 2005, 2012; Leo & Drizin, 2010). In order to produce a confession believable enough to satisfy the interrogator, the innocent suspect creates a story using the information they have at hand. This might include details that have been learned during the interrogation process, seen in the media, or those which could be deduced using logic or common sense. However, the innocent suspect is at a disadvantage when creating an accurate confession as they do not have first-hand knowledge of the crime. In their ignorance of the specifics of the crime, the suspect might inadvertently understate or overstate key details, such as admitting to shooting the victim too few or too many times compared to the actual wounds inflicted. While research has been conducted into the effects of confession inconsistencies on perception of suspect guilt, the effect of directional errors that might reasonably be expected of an innocent suspect, have not to our knowledge been investigated.

In particular, understanding the effect of understating the crime (inadvertently or otherwise), is important as some police interrogation techniques may increase the possibility

of a suspect confessing to an understated version of a crime. The commonly used Reid technique (Inbau, Reid, Buckley, & Jayne, 2013) advocates the use of minimization in the police interrogation process, in which the police officer attempts to elicit a confession by downplaying the severity or consequences of the crime and providing the suspect with face-saving excuses for their involvement. By lessening the perception of the severity of the crime, minimization could inadvertently encourage false narratives with inconsistencies that indicate the desire for minimal culpability. The use of such minimization techniques has been shown to increase the likelihood of both true and false confessions alike (Russano, Meissner, Narchet, & Kassin, 2005), providing an important reason for investigating the effect of directional errors on how jurors perceive the suspect's level of guilt when processing confession evidence.

The aim of the present study was to investigate how factual errors in confessions that either amplified or downplayed the severity of the crime influenced judgments of guilt.

Attribution Theory

Kelley's (1973) attribution theory holds that the inferences an observer draws from another's behavior depends on attributions about the motive underlying that behavior. An important implication of this principle is that a particular behavior can lead to very different inferences depending on the attributions made by the observer. An example of this principle, provided by Reeder, Vonk, Ronk, Ham & Lawrence (2004), involves an observer who sees a student helping a professor. Without additional context, the observer may attribute this behavior to the student's helpful nature. However, if the observer is aware of an ulterior motive (e.g., the professor is evaluating the student's scholarship application), the helping behavior may be attributed to the pursuit of this ulterior motive rather than the student having a helpful nature. In this case, the perceived presence of an ulterior motive leads to *discounting*; that is, the helping behavior is discounted when making inferences about the

disposition of the student. In contrast, some contextual information can lead to *augmentation*, whereby an observed behavior has a stronger effect on dispositional attributions. For example, if a student helps a professor without the professor's knowledge, this might enhance an observer's attributions about the student having a helpful nature; if the professor is unaware, it reduces the scope for any ulterior motive to explain the behaviour. Thus, the same behavior (a student helping a professor) can lead to very different inferences about the underlying cause of the behavior, depending on the attributions made by the observer.

Although the principles of discounting and augmentation were developed in the context of making dispositional attributions, we can consider how they might help explain how jurors interpret directional inconsistencies in a confession. We assume that jurors are motivated to explain inconsistencies in confessions and will seek explanations for them (Palmer et al., 2016). Further, we expect that the inferences drawn from inconsistencies in confession evidence will vary systematically depending on the direction of the inconsistency. For example, if the evidence shows that the suspect has confessed to a crime that is less severe than the facts would imply (e.g., admitting to firing one shot when there is evidence three were fired), the juror might reason that the suspect is lying in an attempt to make themselves look less guilty to attract a lesser penalty. If so, then according to the discounting principle, inconsistencies that imply a less severe version of the crime might not translate into reduced perceptions of guilt (relative to a confession without inconsistencies). However, if the evidence shows that the suspect has confessed to a worse version of the crime (e.g., firing 10 shots), the juror might reasonably wonder what the suspect has to gain by lying in such a manner. As there is no perceived benefit in confessing to a crime of greater magnitude, the juror might suspect that the confession is the result of an unknown motive unrelated to guilt (e.g., perhaps the defendant is innocent and was pressured to confess). In turn, inconsistencies that imply a more severe version of the crime will lead to reduced perceptions of guilt

(relative to a confession without inconsistencies). In this way, attribution theory provides a framework for investigating how jurors not only rationalise the presence of errors in a confession, but also how they investigate the types of motivations that might underpin the directionality of those errors.[1]

Although no prior studies have directly compared inconsistencies that imply a more severe versus less severe crime, two published studies have included manipulations of inconsistencies with some directional properties. Palmer et al. (2016) manipulated inconsistencies by including three details in a confession statement that differed from verified facts in a police case file. Two of these concerned non-directional details (e.g., the time of crime occurred) but the third detail implied that the suspect confessed to a crime more severe than actually occurred: the confession stated that the victim had been shot five times whereas the police file stated that the victim had been shot once. In two experiments that included this manipulation, participants were less likely to convict if they read a confession that was inconsistent with the case file than one that was consistent with the case file.

Henderson and Levett (2016) had participants read a confession that contained errors that downplayed the severity of the crime. The suspect confessed to stabbing the victim 1-2 times, and the police case report stated that the victim was stabbed either 1-2 times (consistent condition) or 38 times (inconsistent condition). In two experiments using this manipulation, the manipulation of inconsistency had a marginal or null effect on verdicts. In a third experiment, the direction of the inconsistency was counterbalanced such that some participants read a confession that exaggerated the severity of the crime and others read a confession that downplayed the severity of the crime. However, comparing inconsistencies of

---

[1] We note that these ideas are consistent with the Story Model of juror decision-making (Pennington & Hastie, 1986, 1992). According to this model, jurors construct an internal narrative about a case and then use individual pieces of evidence to confirm or alter this narrative. The notion that jurors make attributions about the reason for inconsistencies in confession evidence—and that these attributions shape the inferences drawn from the confession evidence—fits well in the story model.

different directions was not the focus of that study, and these two conditions were collapsed together when assessing the effects of the inconsistency manipulation. With the two directional conditions combined, inconsistency reduced guilty verdicts.

Together, these results align with the rationale outlined above—when confessions contain inconsistencies that exaggerate that severity of the crime, guilty verdicts are reduced (as in Palmer et al., 2016). When confession contain inconsistencies that downplay the severity of the crime, the effects on verdicts are smaller and sometimes null (as in Henderson & Levett, 2016, Experiments 1 and 2). However, these studies did not conduct comparisons between inconsistencies of different directions; this was the focus of the present research.

Hypotheses

In line with Palmer et al.'s (2016) study, we hypothesised that confessions containing inconsistencies would be rated as less consistent than confessions that were consistent (assuming people noticed the inconsistencies). We further hypothesised that juror perception of lower confession consistency would influence judgments of guilt dependent on how the jurors attributed the reasons for the errors in the confession. That is, if the confession errors increased the severity of the crime and acted to make the suspect look *worse*, then we expected greater inconsistency to be associated with lower judgments of guilt. According to attribution theory, because there is no obvious ulterior motive for the *worse* inconsistencies, jurors might assume that the inconsistencies are an indicator that the person might not have committed the crime, (*"why would someone make themselves look even worse? Maybe he doesn't know what happened because he didn't commit the crime?"*).

However, if the inconsistencies decreased the crime severity and acted to make the suspect seem *better*, we predicted that while the presence of inconsistencies would reduce perceptions of consistency of the confession, this reduction would not necessarily translate to lower judgments of guilt. In the case of a *better* confession, there is a clear ulterior motive for

the inconsistencies, with the suspect self-servingly downplaying the severity of the crime,

(*"he's making errors on purpose to make himself look better. Of course he did it").*

Method

Participants

Ninety-four participants (67 female, 26 male, 1 undisclosed), aged 18 to 63 years (*M*

= 26.49, *SD* = 9.21), volunteered from undergraduate psychology classes at [redacted], as

well as from the greater university community. The majority were enrolled at university full

time (98.9%). Participants were awarded with partial course credit, or AUD$10

remuneration. The majority of participants (83%) spoke English as their main language at

home. Two participants were excluded from the data set for misunderstanding the

instructions resulting in nonsensical answers – such as believing that the police report had

been written by the suspect. Participants who provided partial answers were retained in the

data set, resulting in some analyses with a total N < 94.

Procedure

This study was given ethical approval by, and in accordance with the Human

Research and Ethics Council guidelines of [redacted], which adheres to the National

Statement on Ethical Conduct in Human Research (NHMRC, 2007, Updated May 2015).

Supervised testing took place in a laboratory setting in small groups (*n=2-3*), with

participants giving informed consent. Participants acted as individual jurors and no

deliberation or discussion took place. Random allocation was used to place participants in

one of three confession conditions (*worse, consistent, better*) in a between-groups design.

After reading two pieces of written evidence (a typed confession statement, and police

report), participants answered the pen and paper questions in order, with instructions not to

read ahead or change their answers once given. Completion time for the task was

approximately 20 minutes.

Stimulus materials

Participants were given two pieces of evidence from a fictional case: a confession statement and a police summary report. The confession was presented as a typed statement in which the suspect admits to committing armed robbery at a service station and discharging a firearm at the service station attendant. The typed police summary gave detailed information from the service station CCTV footage, including the number of shots fired and confirmation that the safe was completely emptied in the robbery. The report confirmed that the safe contents had been counted and verified by two staff members prior to the robbery taking place. This was to ensure that participants did not dismiss any incorrect statement about the money stolen as being due to an administrative error or unrelated theft by a staff member. Materials were made to appear as if they were copies of evidence from a real case.

Directional errors manipulation. To test the directional effect of inconsistencies, key information in the confession statements was altered so that the confession matched the facts of the crime in the police statement (*consistent*), or the suspect admitted to a greater crime than that outlined in the police report (made the participant look *worse*), or the suspect admitted to a lesser crime than that outlined by police (made the participant look *better*). The confession errors were related to key facts of the crime (amount of money stolen, and number of shots fired) to ensure that they were salient enough to attract notice and warrant a juror questioning the truthfulness of the confession. For example, the police report stated that three shots were fired and $2,100 was stolen. In the *better* condition, the confession stated that one shot was fired and several hundred dollars stolen. In the *worse* condition, the confession stated that around 10 shots were fired and $10,000 stolen.

Measures. After reading the confession statement and police report, participants were asked to give a dichotomous (not guilty, guilty) verdict and rate the confidence in their verdict on a scale of *1 = not confident at all* to *10 = completely confident*. Previous research

has found that the majority of people see a confession as an honest admission of guilt and will render a guilty verdict accordingly (Drizin & Leo, 2004). Therefore, dichotomous verdicts may not prove a subtle enough measure of the possible variation in juror belief in suspect guilt. To reveal more nuanced perceptions of guilt, we utilised a common practice in juror research (see Appleby & Kassin, 2016; Leippe, Eisenstadt, Rauch, & Seib, 2004; Sauer, Palmer, & Brewer, 2017; Sommers & Kassin, 2001; Tenney, MacCoun, Spellman, & Hastie, 2007), creating a continuous variable of *judgment of guilt* by combining the dichotomous verdict and confidence in verdict measures. The resulting scalar variable measures *judgment of guilt* on a scale from -9.5 (complete confidence in a not guilty verdict) to 9.5 (complete confidence in a guilty verdict). Participants who score close to zero on this scalar variable are those who have little strength of belief that their verdict is correctFollowing the verdict and confidence questions, participants were asked to rate the consistency of facts of the crime between the two pieces of evidence (from *1 = not consistent at all, to 10 = completely consistent*). To prime participants to think more closely about the consistency of the evidence, they were asked to list why they thought differences (if any) between the pieces of evidence might have occurred. Two raters coded a sample of verbatim participant responses into those which indicated a belief that the suspect's error were a ploy to make themselves look better, reduce perception of the severity of the crime, or lessen the consequences of the crime. These were coded as '*downplaying*', with responses that did not meet this criterion coded as '*no downplaying*'. Full agreement between the two raters was achieved.

Results

Judgment of guilt

Table 1 shows the means, standard deviations and confidence intervals for judgments of guilt and perception of consistency ratings for each of the conditions. There was no significant overall difference between the judgment of guilt scores for the consistent, better, and worse condition, with the mean scores for all three groups indicating that participants were reasonably confident that the suspect was guilty.[2] The manipulation check of consistency indicated that participants were able to perceive that the two inconsistent conditions (*better, worse*) contained confession inconsistencies, and that the consistent condition did not, $F(2, 90) = 56.10$, $p < .001$.

[Approximate placement of Table 1]

Mediation analyses were conducted to test whether the effect of the directional manipulations on *judgments of guilt* were mediated by perceived consistency (from the consistency manipulation check). Perceived consistency was considered an important mediator as the actual presence of inconsistencies might be unrelated to whether participants consciously noticed the inconsistencies or not. A similar concept was explored in Palmer et al. (2016), where the degree of perceived consistency (as measured by the number of inconsistencies reported) mediated the effect of their manipulation on juror verdicts. To test the directional inconsistencies in the present study, two separate mediation analyses were conducted: *consistent* versus *better*, and *consistent* versus *worse.*

When broken down into its components, the results of the *consistent* versus *better* conditions ($n = 61$) suggest that there were two different effects occurring (see Fig. 1). First, there was a significant indirect effect, in which the presence of inconsistencies reduced perceived confession consistency, (*a*), $b = -3.71$, $t = -8.98$, $p < .001$, [-4.55, -2.89]. In turn, as

---

[2] In terms of dichotomous verdicts, as expected, the majority (85.1%) of participants gave a guilty verdict. There was no significant difference in guilty verdicts between the consistent (90.6%), better (83.3%), and worse (81.3%) conditions, $n = 94$, $\chi^2(2) = 1.22$, $p = .54$.

perception of confession consistency reduced, judgments of guilt were similarly reduced (*b*),

$b = 1.34$, $t = 3.67$, $p < .001$, [.61, 2.07]. The significant indirect effect of inconsistencies on

judgments of guilt (*ab*) confirms that confession errors overall reduced judgments of guilt by

decreasing perceived consistency, $b = -4.99$, [-8.42, -1.72]). This finding replicates the

indirect effect of inconsistencies found in Palmer et al. (2016) and suggests that the greater

the perceived inconsistency, the larger the reduction in guilt ratings.

[Approximate placement of Figure 1]

Second, and separate to the indirect effect, the direct effect of inconsistencies (*c'*)

operated in the opposite direction ($b = 3.84$, $t = 2.15$, $p = .036$, [.26, 7.42]), with inconsistent

testimony increasing judgments of guilt. Opposing direct and indirect effects might seem

counterintuitive, but these are not uncommon, and are indicative of a mediation model in

which two opposing mediational processes result in a relationship between the predictor and

outcome variables that is effectively zero (MacKinnon, Fairchild, & Fritz, 2007). In the case

of our study, the results suggest that two opposing mechanisms influenced judgments of guilt.

One involves the indirect effect described above, whereby the presence of inconsistencies

reduced judgments of guilt via differences in perceived consistency of the confession

evidence. The second—opposing—mechanism is linked by the direct effect, whereby the

presence of inconsistencies that made the suspect look better were associated with higher

judgments of guilt. We return to this issue after reporting the mediation for the *consistent*

versus *worse* conditions.

The results of the *consistent* versus *worse* conditions ($n = 62$) reflected similar

findings to the *consistent* vs *better* conditions for the indirect effect, but not the direct effect

(see Fig. 2). For the *consistent* versus *worse* conditions, the presence of inconsistencies

significantly reduced perceived consistency (a), $b = -2.10$, $t = -10.56$, $p < .001$, [-2.50, -1.71],

with lower perceived consistency in turn resulting in lowered judgments of guilt (b), $b = 1.06$,

$t = 2.53$, $p = .014$ [.22, 1.90]. The significant indirect effect of inconsistencies on judgments

of guilt (*ab*), $b = -2.23$, [-4.23, -.33], also aligns with the indirect effect of inconsistencies on

verdicts found by Palmer et al. (2016), whereby confession inconsistencies reduced

judgments of guilt. However, unlike in the *consistent* vs *better* condition, there was no

significant direct effect of the manipulation on judgments of guilt in the *consistent* vs *worse*

conditions ($b = 1.38$, $t = 1.27$, $p = .209$ [-.79, 3.56]). Therefore, there was no evidence that

confession inconsistencies that exaggerated the severity of the crime affected judgments of

guilt separate to the mechanism of inconsistencies reducing perception of confession

consistency.

[Approximate placement of Figure 2]

Downplaying of guilt

The comparison of the *consistent* versus *better* conditions yielded evidence for two

opposing mechanisms, one linked to the indirect effect (whereby inconsistencies reduced

judgments of guilt) and one to the direct effect (whereby inconsistencies increased judgments

of guilt). We suggest that the most plausible interpretation of these results is that

inconsistencies that made the suspect look better caused participants to become suspicious

that the suspect was deliberately making mistakes to downplay the severity of the crime.

This, in turn, led to higher guilt ratings, independent of perceived consistency of the

confession.

As a preliminary test of this explanation, we conducted additional, exploratory

analyses. In order to explore what attributions participants made about the directional

confession inconsistencies, participants had been asked to note if they found any

inconsistencies between the two pieces of evidence, and explain why they thought those

inconsistencies may have occurred. Responses were coded as to whether the participant

believed that the confession inconsistencies were due to the suspect deliberately making his

actions seem less severe. Examples of responses coded as downplaying include, *"cash amount differed – this was probably due to the perpetrator deliberately trying to downplay his actions"*, and *"the person who robbed the place is trying to make the crime a lot less serious than it actually was"*.

Chi square analysis revealed a significant difference between the *consistent* and *better* conditions on the number of participants who gave downplaying as a reason for the confession inconsistencies, $\chi^2(1) = 11.23$, $p < .001$. Nine of the 30 participants in the *better* condition responded that confession inconsistencies were due to the suspect deliberately making his actions seem less severe. No participants in the *consistent* or the *worse* conditions believed that the suspect was deliberately trying to downplay the crime. Responses in the *worse* condition were too diverse to be coded into clear categories for the purpose of further analysis. Responses in the *worse* condition generally reflected a broad perception that the suspect was an inexperienced criminal, who was not capable of estimate large sums of money, or remaining calm enough to count the shots he fired. Examples of responses included, *"if this was his first offence, $2,000 may look like $10,000 to him"*, and *"this might be due to the robber being in a hurry to leave and didn't count or was illiterate"*.

A t-test was used to analyse differences in judgments of guilt between those in the *better* condition who specifically mentioned the suspect deliberately downplaying the crime, and those in the same condition who did not mention this possibility. The mean judgment of guilt was significantly higher for those who believed the suspect to be deliberately attempting to downplay the crime ($M = 8.28$), compared to those who did not mention downplaying as a reason for the confession inconsistencies ($M = 4.74$), $t(22.68) = -2.52$, $p = .019$, $d = 0.67$, 95% CI [-6.45, -.626].

These results are consistent with the idea that participants attributed confession inconsistencies that made a suspect appear better (but not worse) to the suspect having the

ulterior motive of trying to reduce the perceived consequences of their crime. The lack of a similar direct effect (opposing the indirect effect) in the *consistent* vs *worse* conditions adds further weight to this evaluation of the results, in that there is no apparent ulterior motive for someone to deliberately make themselves look *worse*, whereas there is a plausible ulterior motive for wanting to downplay involvement in a crime.

Discussion

Results showed that participants were clearly able to see when inconsistencies were present in the confessions. However, the direction of confession inconsistencies affected juror decisions in a way that we had not hypothesised. We predicted that the two types of confession inconsistencies (*better, worse*) would both affect perceptions of confession consistency, but that they would differ in their relationship between perceived consistency and judgments of guilt. We hypothesised that greater perceived inconsistency would translate to lower judgments of guilt for *worse* confessions, but not for *better* confessions. Instead, results showed that both types of inconsistencies were well detected by participants, which translated to lower judgments of guilt for both the *worse* and *better* conditions. The effect of confession inconsistencies is more robust than expected, with participants noticing and acting on inconsistent evidence, regardless of the direction of the inconsistency. This finding is consistent with Woestehoff and Meissner (2016), whose findings suggest that juror beliefs and understandings might be changing in a way that gives promise of a more thoughtful and knowledgeable juror than previously thought.

In all conditions, the effect of the consistency manipulation on judgments of guilt was mediated by perception of consistency. This indicates that the effect of inconsistencies is reliant on the participant actually noticing the inconsistencies and then perceiving them to be inconsistent. The important role that perception of consistency might play in affecting judgments of guilt can be related to theoretical reasoning about how jurors might use

information in their decision making process. Theories of decision making imply that a juror's internal narrative of the crime adapts to accommodate information that has not specifically been provided, such as motivation (Kelley, 1973; Pennington & Hastie, 1986), and this information may act to alter their judgments of guilt. In the present study motivation for the confession can be extended to encompass the suspect's reasons for giving a confession that is partly inconsistent with the facts of the crime.

Crucially, there was a difference in the direct effect of inconsistencies on judgments of guilt between the *better* and *worse* conditions. In the *better* condition (but not the *worse*) there was a direct effect indicating that inconsistencies acted to increase judgments of guilt through a mechanism unrelated to the degree of perceived consistency. This suggests that, regardless of the extent to which the confession was perceived as inconsistent, inconsistencies that were seen as an attempt to deliberately downplay the severity of the crime triggered higher judgments of guilt.

The backlash effect of the *'better'* confession

While mediation analysis showed that inconsistencies overall acted to reduce judgments of guilt, when the suspect confessed to a lesser version of the crime there was a backlash in which judgments of guilt increased. The apparent mechanism behind the backlash is that, if the juror believed that the inconsistencies were an attempt by the suspect to deliberately downplay the severity of the crime, then the suspect would be given a higher judgment of guilt than if the juror did not believe the suspect was lying for their own gain. Verbatim responses about why there were inconsistencies between the suspect's confession and the police report frequently indicated a perception that the suspect was an experienced criminal, who deliberately lied about the crime in order to reduce his sentence. Participants in the *better* condition who specifically mentioned that the inconsistencies between the confession and police report could be explained by the suspect deliberately downplaying the

crime severity, judged the suspect as guiltier than those who did not mention downplaying as a possible explanation for the inconsistencies.

This backlash suggests that, while errors in a confession could act to supress automatic judgment heuristics in jurors (where a confession is analogous to guilt), this suppression is less likely if the inconsistencies make the suspect look like they are admitting to a lesser version of the crime. When a suspect appears to be denying the severity of the crime, thus reducing their admitted culpability, the denial acts to increase the belief that the confession is an admission of guilt, with the denial of responsibility for the full crime rejected on the basis of a strong ulterior motive to deceive. This finding highlights an additional reason why the use of minimization techniques in police interrogations are problematic. That is, not only does minimization increase the risk of an understated false confession by suggesting to the suspect that the crime is not so terrible after all, but the resulting false confession can then inflate jurors' perceptions of that suspect's guilt later on if the confession details a lesser version of the crime.

Limitations

A common criticism of lab-based juror studies is that court cases are, by their nature, infinitely more variable and complex than can be replicated in a laboratory setting. In the present study there is no attempt to claim that the methodology captures an entirely realistic jury experience and that findings can be mapped directly on to any single case involving a false confession. For example, participants in Kassin and Kiechel's (1996) experiment were coerced into falsely confessing to crashing a computer program in a laboratory. While those participants are clearly not in the same position as a person being wrongfully accused of murder in a police interrogation, the findings of that experiment allowed greater understanding of the situational pressures that can increase the likelihood of an innocent person confessing. Similarly, our work aims to illustrate the situations in which people are

able to consider the inconsistencies in confessions in a way that inspires new thinking about how jurors might be processing evidence that is of varying quality. The present study also helps explain how people might process directional errors when they come across them in confession evidence, and how that might then affect their perception of a suspect's guilt. Ongoing study into the source of variance in juror decision making allows researchers to make advances in understanding how and why jurors sometimes make wrong decisions based on poor-quality evidence. The hope is that such research can eventually be applied to the ongoing improvement and evolution of legal processes.

Summary

This study applied principles of attribution theory to an important applied issue, that of the ways in which jurors process confession evidence. The way in which a juror might make attributions about the errors in a suspect's confession is important when considering why some false confessions are accepted as truthful by juries, and others are dismissed. The present study shows that a person confessing to the crime will be overwhelmingly judged as guilty, but if they admit to a lesser version of the crime, some jurors will be even more certain that they are guilty. As a jury would not know if the confession was false, the accidental error that reduces the severity of the crime could act to make jurors even more certain that the person has committed the crime, although the error itself should make them question the confession's veracity.

CRIME SEVERITY INCONSISTENCIES IN CONFESSIONS

References

Appleby, S. C., & Kassin, S. M. (2016). When self-report trumps science: Effects of

confessions, DNA and prosecutorial theories on perceptions of guilt. *Psychology,*

Public Policy, and Law, 22(2), 127-140. doi:10.1037/law0000080

Berman, G. L., Narby, D .J. & Cutler, B. L. (1995). Effects of inconsistent eyewitness

statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant

culpability and verdicts. *Law and Human Behavior, 19*, 79–88.

doi:10.1007/BF01499074

Bornstein, B. H., & Greene, E. (2011). Jury decision making: Implications for and from

psychology. *Current Directions in Psychological Science, 20*(1), 63-67.

Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in the post-DNA world.

The North Carolinea Law Review(82), 891-1008.

Garrett, B. L. (2011). False confessions. *Litigation, 37*(4), 54.

Henderson, K. S., & Levett, L. M. (2016). Can expert testimony sensitize jurors to variations

in confession evidence? *Law and Human Behavior, 40*, 638-649.

doi:10.1037/lhb0000204

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). The Reid Nine Steps of

Interrogation *Criminal Interrogation and Confessions* (5th ed., pp. 185-328).

Burlington, U.S.A.: Jones & Bartlett Learning.

Innocence Project. (2017). Retrieved from http://www.innocenceproject.org/causes/false-

confessions-admissions/

Jones, E. E., Bandy, A. D., & Palmer Jr., P. G. (2019) "I did it, but not like that": Effects of

factually incorrect confessions on juror judgments. *Psychiatry, Psychology and Law*,

26, 553-570. doi:10.1080/13218719.2018.1519467

Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*(3), 215-228. doi:10.1037/0003-066X.60.3.215

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist, 67*(6), 431-445. doi:10.1037/a0028212

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7*(3), 125-128. doi:10.2307/40062928

Kassin, S. M., & Neumann, K. (1997). On the power of confession evidence: An experimental test of the fundamental difference hypothesis. *Law and Human Behavior, 21*(5), 469-484. doi:10.1023/A:1024871622490

Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the "harmless error" rule. *Law and Human Behavior, 21*(1), 27-46. doi:10.1023/A:1024814009769

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 107-128.

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Seib, H. M. (2004). Timing of eyewitness expert testimony, jurors' Need for Cognition, and case strength as determinants of trial verdicts. *Journal of Applied Psychology, 89*(3), 524-541. doi:10.1037/0021-9010.89.3.524

Leo, R. A., & Drizin, S. A. (2010). The three errors: Pathways to false confession and wrongful conviction. In G. D. Lassiter & C. Meissner (Eds.), *Police Interrogations and False Confessions: Current Research, Practice, and Policy Recommendations* (pp. 9-32). Washington D.C.: American Psychological Association.

Leo, R. A., & Ofshe, R. J. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *The Journal of Criminal Law and Criminology, 88*(2), 429-496. doi:10.2307/1144288

CRIME SEVERITY INCONSISTENCIES IN CONFESSIONS

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 1-22. doi:10.1146/annurev.psych.58.110405.085542

Malloy, L. C., & Lamb, M. E. (2010). Biases in judging victims and suspects whose statements are inconsistent. *Law and Human Behavior, 34*, 46-48. doi:10.1007/s10979-009-9211-y

NHMRC. (2007, Updated May 2015). *National Statement on Ethical Conduct in Human Research*. Canberra: National Health and Medical Research Council.

Ofshe, R. J., & Leo, R. A. (1997). The social psychology of police interrogation: The theory and classification of true and false confessions. *Studies in Law, Politics and Society, 16*, 189-251.

Palmer, M. A., Button, L., Barnett, E., & Brewer, N. (2016). Inconsistencies undermine the credibility of confession evidence. *Legal and Criminological Psychology, 21*(1), 161-173. doi:10.1111/lcrp.12048

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*(2), 242-258. doi:10.1037/0022-3514.51.2.242

Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. Journal of Personality and Social Psychology, *86*, 530–544. doi:10.1037/0022-3514.86.4.530

Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science, 16*(6), 481-486. doi:10.1111/j.0956-7976.2005.01560.x

Sauer, J. D., Palmer, M. A., & Brewer, N. (2017). Mock-juror evaluations of traditional and ratings-based eyewitness identification evidence. *Law and Human Behavior, 41*,375-384. doi:10.1037/lhb0000235

Sommers, S. R., & Kassin, S. M. (2001). On the many impacts of inadmissable testimony: Selective compliance, Need for Cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin, 27*(10), 1368-1377

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46-50. doi:10.111/j.1467-9280.2007.01874.x

Woestehoff, S. A., & Meissner, C. A. (2016). Juror sensitivity to false confession risk factors: Dispositional vs. situational attributions for a confession. *Law and Human Behavior, Online First Publication*, 1-16. doi:10.1037/lhb0000201

CRIME SEVERITY INCONSISTENCIES IN CONFESSIONS

Table 1

Mean judgment of guilt and perception of confession consistency scores

	n	*M (SD)*	95% CI
Judgment of Guilt			
Consistent	31	6.95 (4.50)	[5.30, 8.60]
Worse	32	5.27 (5.77)	[3.19, 7.35]
Better	30	5.80 (5.47)	[3.76, 7.84]
Perception of Consistency			
Consistent	32	8.98 (.97)	[8.63, 9.34]
Worse	31	4.77 (1.95)	[4.06, 5.49]
Better	30	5.27 (2.08)	[4.50, 6.04]

Note. 95% CI = 95% confidence intervals

CRIME SEVERITY INCONSISTENCIES IN CONFESSIONS

Perceived Consistency
(Mediator)

a ***

b ***

Inconsistencies
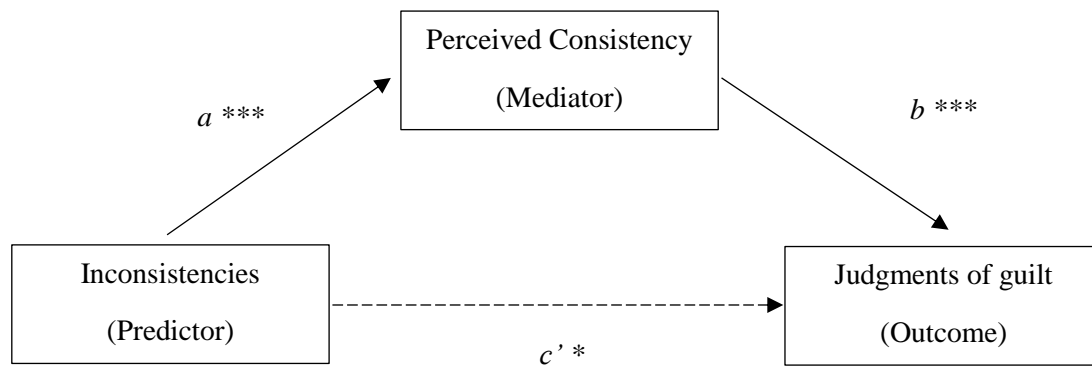(Predictor)

Judgments of guilt
(Outcome)

c' *

Figure 1. Mediation model testing the effect of inconsistencies (*consistent* vs *better*), on judgments of guilt, mediated by perceived consistency of the confession. * ≤.05, **≤.01, ***≤.001
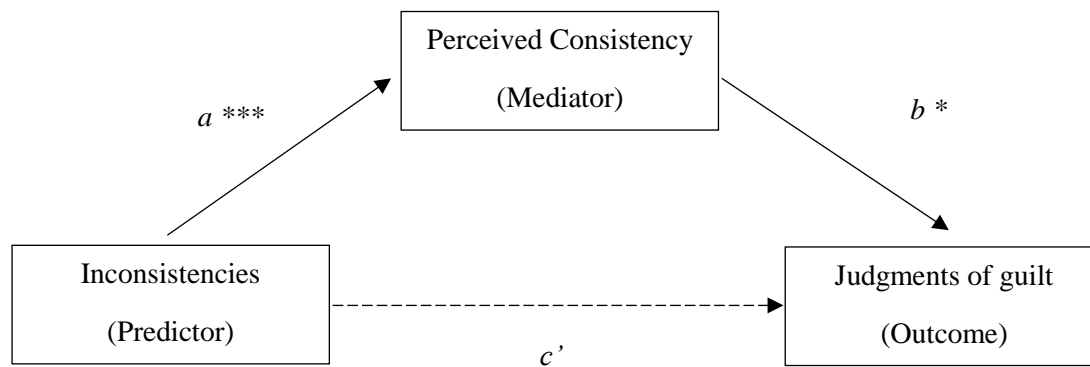
Figure 2. Mediation model testing the effect of inconsistencies (*consistent* vs *worse*) on judgments of guilt, mediated by perceived consistency of the confession. * ≤.05, **≤.01, ***≤.001