How Guilt Presumptions and Evidence-Strength influence Interviewers Judgements of Guilt, Judgements of Believability, and their Intention to Further Investigate

Jonah Sauer

S2298007

First Supervisor: Dr. Steven Watson

Second Supervisor: Dr. Iris Sintemaartensdijk

Bachelor Thesis: Conflict, Risk and Safety Psychology

Faculty of Behavioural, Management and Social Sciences, University of Twente

04.07.2022

UNIVERSITY OF TWENTE.

Abstract

Innocent suspects and investigators themselves need and benefit from good investigative decisions. The PEACE framework, implemented in 1992 in England, focused on an unbiased and ethical approach towards investigative interviews. Important factors found guiding perceptions of interviewers about the suspect are Evidence-Strength and Guilt Presumptions. By guiding perceptions, they influence investigative decisions. This study aimed to show the influence of Evidence-Strength and Guilt Presumptions on guilt judgements, judgements about believability of innocent suspects narratives and the intention to keep up investigations against them. For investigation, a scripted interview was carried out with 70 participants, mostly students of Dutch or German nationality, accompanied by pre- and post-questionnaires. The experimental design was a 2x2 factorial design with Evidence-Strength (*weak* vs. *strong*) and Guilt Presumptions (*guilty* vs. *not guilty*) as independent variables. Guilt Judgements were shifted after exposure to explanations from the innocent suspect. Although Guilt Judgements were significantly reduced, only portions changed from guilty to innocent. Evidence-Strength had a significant effect on Guilt Judgements, which was that strong evidence resulted in higher guilt judgements. The manipulation of Evidence-Strength largely functioned as a manipulation of Guilt Presumptions in that it inflicted confirmation bias too. Since information retrieval practices were held consistent through scripted interviews, they reduced confirmation bias effects. Apart from uncovering a need for further research into confirmation bias and confidence in investigative judgements, the present study illustrated possible reduction of poor investigative decisions by keeping information retrieval practices consistent. Additionally, it was shown that strong evidence increases investigators perceptions of guilt.

Table of Contents

Introduction

To secure fair, objective criminal investigations, it may logically follow that police officers in investigative interviews should stay objective toward the suspect. This is highlighted by the findings of Hill et al. (2008) regarding negative influences of subjectivity. However, remaining objective represents a major challenge due to various biases. Bias might hinder police officers in following the required objective approach in the investigative interview and result in poor investigative decisions (Hill et al., 2008), for example as shown with confirmation bias influencing expert judgements (Kassin et al., 2013). Within investigative interviews, assumptions of guilt often elicit bias. This leads to investigators being mainly concerned with confirming their own beliefs, thereby inhibiting the collection of accurate information (Shawyer et al., 2013) and possibly misinterpreting evidence. Further, this leads to detrimental effects for innocent suspects or even wrongful convictions.

To reduce the number of wrongful convictions and improve interrogation procedures, police officers in England and Wales questioned current methods and began to use a new framework for investigative interviews with suspects to counteract bias, the PEACE model, in 1992 (Clarke et al., 2011). A fundamental principle within this framework is the avoidance of guilt assumptions. In other words, it aims at preventing confirmation bias.

PEACE is an acronym, representing five stages of the interview process, namely planning and preparation, engagement with interviewee and explaining of the process, gaining an account, closure, and evaluation (Clarke et al., 2011). It emphasises ethics by using open questions rather than closed and (mis-)leading questions, shifting the focus from being confession-seeking to being information-acquiring (Clarke et al., 2011). This can also be observed in the shift in terming from "interrogation" to "investigative interview", turning away from the confrontational, manipulative, and aggressive persuasion to a less intrusive and more cooperative approach focused on enquiry (Shawyer et al., 2013).

However, evaluation of the actual implementation of the PEACE framework shed a light on deficiencies that occurred within the police. Clarke & Milne (2001) uncovered that PEACE was often not incorporated into the training of investigators. Additionally, PEACE principles were often missing in interviews in practice (Clarke & Milne, 2001). More recently, researchers have shown its principles to be missing in various countries (Snook et al., 2010; Adam & van Golde, 2019). Consequently, interviews are poorly conducted, bias is still frequently prevalent, and evidence misinterpreted. This can be observed in the case of Andrew Mallard, an Australian man falsely convicted for murder in 1994, or in the case of Gene Gibson who wrongfully served five years in prison until 2017. Both men fell victim to pressuring interview techniques and misinterpreted information gathered in presumptive interviews (Adam & van Golde, 2019).

Not adhering to the PEACE framework might show detrimental effects for the investigative process and the application of justice. For the police, the process might be complicated in that coercive and manipulative interrogation tactics lead to less effective information collection and less cooperation from suspects (Goodman-Delahunty & Martschuk, 2018). It might even lead to collection of incorrect information, which then might again lead to poor investigative decisions (Gudjonsson, 2003). Further, coercive and manipulative interrogation behaviours often result from presumptions of guilt manifesting themselves in the behaviour of interviewers (Meissner & Kassin, 2004). This, in turn, indicates a misuse of available human resources (Scherr et al., 2018), in the sense that police officers would do more efficient and valuable work when adhering to non-presumptive tactics like in the PEACE approach.

For the suspects, especially when innocent, there are also many detrimental effects when investigative interviews are not carried out according to PEACE. For example, they might experience stigma when their social environment takes notice of them being under

suspicion of having committed a crime, or they might suffer psychologically from long interview procedures and might even be wrongfully convicted due to investigators being misled by their presumptions, which again represents a misuse of available human resources (Scherr et al., 2018). Coercive and guilt presumptive interviewer behaviour might even lead to false confessions and in the worst case to wrongful convictions (Villalobos & Davis, 2016).

Investigative Decision Making

Investigative decision making is based on information gathered from the investigative interview and previously gathered information about the case. Several judgements influence the ultimate decision that decides on how the police are going to proceed their investigation, including a possible second interview or releasing the suspect. The behaviours focused on here are judgements being made consciously or unconsciously by the investigator in a real situation.

To decide on how to proceed with the suspect, the investigating officers` own feeling of the guiltiness of the suspect is crucial. Research investigated judgements of guilt of hypothetical legal decision-makers and found them to be very prone to bias, since they were influenced by minor factors such as the vividness of phrasing (Bensi et al., 2003) or emotions in statements (Dahl et al., 2007). Therefore, guilt judgements constitute a major part in investigative decision making that is prone to being easily influenced, making further investigation of it important to enable more precise prediction-making.

Resembling guilt judgements is the confidence the deciding person has in his own judgement. Research showed that confidence is not necessarily a valuable predictor for accuracy (Saraiva, 2020). However, Wixted & Wells (2017) identified confidence to be an important factor in legal decision-making. The confidence-accuracy relationship shown by Saraiva (2020) might also indicate that investigators who are highly confident in their incorrect judgement are less willing to pursue approaches not in line with their previous

beliefs. The indication of confidence into a judgement may show a more profound understanding of how strong the individuals' actual judgement of guilt is, delivering a more valuable aspect of a judgement rather than a binary statement. However, a binary statement might be useful for having a direct measurement accompanying, possibly better usable for quantifications.

Another factor inherent to every investigative decision based on utterances from investigative interviews is whether the story or narrative the interviewee tells is deemed plausible or believable by the investigating person. Research found that the believability of a story in legal settings is dependent on how detailed a narrative is given and how much emotion is prevalent in it (Justice & Smith, 2018). Thus, further investigation of other factors influencing the judgements of plausibility of the suspect's narrative might fill a niche in scientific knowledge.

Lastly, the theory of planned behaviour names the intention of a person to act out a certain behaviour the direct predecessor of that very behaviour (Ajzen, 2011). The investigators perception of the innocent suspect will ultimately lead him to a personal decision on how to proceed. This decision on how to proceed has two main pathways. Either the investigator believes the suspect and judges him innocent and releases him, or he judges that he is not convinced of his innocence. The latter probably entails further investigation on him to clarify. Investigating this intention by directly or indirectly asking for it will give an impression of how the investigator is inclined to decide.

Guilt Presumptions

Previous research concluded that investigators sometimes base their working hypotheses regarding criminal investigation not only on solid facts, but expectations and preconceptions (Ask & Granhag, 2005). Further, it is stated that presumptions of guilt can inflict serious harm on innocent suspects (Ask & Granhag, 2005), and that guilt presumptions

may cause poor investigative decisions and behaviour, such as using coercion, opposing the PEACE approach (Kassin et al., 2003). Adams-Quackenbush et al. (2020) confirmed that there is an influence of guilt presumption on interviewer behaviour. Research has shown that accusatory and presumptive behaviour might decrease suspect cooperation and can have a detrimental effect on the interviewer's ability to gain relevant information (Adams-Quackenbush et al., 2020). The influence of such presumptions might even expand beyond effects in the interview, leading up to serious detrimental effects for suspect and police in the following process. For example, the police could obtain false information (Goodman-Delahunty & Martschuk, 2018). The suspect, on the other hand, might fall victim to a wrong conviction (Villalobos & Davis, 2016). The remaining question is whether there are different stages within the investigative interview that could help uncovering the onset of effects and what this effect is.

Previous research identified three distinct processes that take place within the investigating individual to ultimately arrive at decisions about how to proceed: *Information retrieval, inference drawing,* and *decision-making* (Ask & Fahsing, 2019). *Information retrieval* pertains to the retrieval of pieces of information. *Inference drawing* describes the cognitive process of combining information and making up one's mind about what the acquired information adds to the uncovering of the case. *Decision making* pertains to the act of deliberately deciding about how to proceed, based on the new information combined with what was known before.

Previous research into guilt presumptions mostly let participants in the information retrieval stage already change their behaviour, thereby allowing for this phase to become biased already. For example, a previous study let participants in the role of the investigator formulate the interview questions themselves (Adams-Quackenbush et al., 2020). Further, Kassin et al. (2013) showed that guilt presuming interviewers often choose presumptive

questions. However, it remains subject to investigation if the effect of bias holds if participants have no free choice in the *Information retrieval* stage (enforcing standardized, high quality information retrieval) and if bias is only allowed to affect the later stages. Kassin et al. (2013) showed that confirmation bias, as reflected in guilt presumptions, can have great effects on expert judgements, which is also confirmed by Ask & Fahsing (2019). These expert judgements are most likely to set in the *Inference drawing* stage and are ultimately made in the *decision-making* stage. Therefore, effects of confirmation bias might even be reduced when keeping the *Information retrieval* stage consistent.

Generally, confirmation bias seems to be the cognitive effect that is central to the detrimental effect of guilt presumptions. Confirmation bias is defined as "an agent`s tendency to seek, interpret and use evidence in a manner biased towards confirming her existing beliefs or hypotheses" (Charness & Dave, 2017, p. 1). As Kassin et al. (2003) state, investigators presuming guilt were more often using presumptive questions and more likely to judge the innocent suspect as guilty, compared with no guilt presuming investigators. Therefore, guilt presuming interviewers will likely be more inclined to judge the innocent suspect as guilty and show more poor investigative decisions than interviewers who are not presuming guilt, for example by deeming suspects as less trustworthy as they are (Shawyer & Milne, 2015).

Evidence-Strength

Another important factor that needs to be examined is the strength of evidence. Moston & Engelberg (2011) indicate it to be the most influential factor for the outcome of an investigative interview. Evidence about a certain case can be categorised into two distinct dimensions. The first one being the type of evidence, which largely divides itself into "direct evidence" (e.g., the suspect possessing a particular good that's relevant for the case) or "statements from different sources" (e.g., witnesses, police officers, or victims) (Moston &

Engelberg, 2011). The second distinction pertains to the strength of evidence, which can be divided into *weak* and *strong* evidence (Moston & Engelberg, 2011).

The Evidence Framing Matrix (Granhag et al., 2012) helps with the categorization of strong and weak evidence. It explicitly names two dimensions important in evaluating the strength of evidence. The first dimension pertains to the *strength of the source of evidence,* while the second dimension is termed *degree of precision of the evidence,* or *specificity* (Granhag et al., 2012). Other research also conceptualised evidence on two dimensions: *reliability* and *proximity* (Brimbal & Luke, 2021). These terms closely resemble the dimensions of Granhag et al. (2012), as *reliability* is equal to the *strength of the source of evidence*, while *proximity* is equal to *degree of precision of evidence/specificity.* The former can be described as the degree to which the source of the evidence is one that is trustworthy, while the latter pertains to the degree to which the piece of evidence explicitly links the suspect to the crime. For example, a source with low reliability would be a young child, as opposed to an adult (Pozzulo & Dempsey, 2009). An example for high proximity evidence would be surveillance camera footage, as opposed to an eyewitness statement. Evidence with high reliability and specificity is considered strong, opposed to sources with low reliability and specificity being considered weak.

Moston et al. (1992) found that there is an influence of evidence-strength on the behaviour of the interviewer. As they argued, the interviewer's behaviours are largely determined by their attitudes and beliefs. Those in turn, are determined by pre-existing contextual factors, like the evidence and thus the corresponding strength of it (Moston et al., 1992). This does take place in the *Information retrieval* stage, as implied by the influence on interviewer behaviour. The question remains if it can influence investigative decision making when the onset of the influence is set in the *Inference drawing* stage by confirmation bias exerting influence when evaluating information collected in the interview and not in the

Information retrieval stage when formulating questions. Keeping the *Information retrieval* stage and corresponding interviewer behaviours consistent in a systematic way might reduce effects of confirmation bias, which would stand in line with the spirit of the PEACE framework. Further, presenting interviewers with an innocent suspect that provides good explanations might further isolate the effect of confirmation bias.

It might be the case that an interviewer who is presented with an innocent suspect that has good explanations for the evidence is more likely to judge him guilty when the evidence is strong than when it is weak, irrespective of good explanations. This is based on the proposed confirmation bias taking place, since the interviewer prior to the interview only has the evidence as information to make up his mind. However, the intention to further investigate might be higher for interviewers that are presented with weak evidence, since Smith et al. (2011) found that weak evidence results in greater use of heuristics. One fitting heuristic here might be a representative heuristic, being a mental shortcut for individuals in which they project stereotypical backstory and characteristics onto a person sitting in front of them and fit them into a certain class (Tversky & Kahneman, 1974). It might well be this, in this case, leads the investigator to think of the suspect as a typical criminal, being prepared to answer and explain away allegations.

It might additionally be that a realistic explanation of weak evidence is deemed less plausible than a realistic explanation of strong evidence. An availability heuristic (Tversky & Kahneman, 1974) might tell investigators that there were/are many cases in which suspects have perfectly sound appearing explanations against weak evidence and are ultimately found guilty anyway. Therefore, they might transfer this experience onto the case at hand and make a poor investigative judgement.

Since presumptions of guilt are built on a basis of preliminary evidentiary information, the question arises whether stronger evidence hinders investigators more to overcome

presumptive bias than weak evidence and if it might influence their investigative decision making. Accordingly, interaction effects of Evidence-Strength and Guilt Presumptions might occur. The guilt presuming interviewer might generally be more inclined to think that the innocent suspect is guilty based on weak evidence than the interviewer who is not presuming guilt. Additionally, the guilt-presuming interviewer presented with weak evidence will likely be more interested in pursuing the investigation on the individual, thereby trying to confirm his own beliefs rather than staying objective.

Purpose of this study

The present study tries to set up a critical examination of the effects of *Guilt Presumptions* and *Evidence-Strength* on investigative decision-making and perceptions of the interviewer within the context of an investigative interview. The key dependent variables are *Guilt Judgements, Plausibility of the Suspect's Story,* and *Intention to Further Investigate*. Additionally, it will be tested if there is an interaction effect of *Guilt Presumptions* and *Evidence-Strength* on *Guilt Judgements* and *Intention to further Investigate*. This will add to the scientific knowledge base and potentially highlight whether there is a potential problem with bias in investigative interviews and where this lies. In doing so, it might potentially improve circumstances for police and innocent suspects in future investigative interviews.

The study will employ mock interviews to mimic investigative interviews and manipulate *Guilt Presumptions* and *Evidence-Strength* prior to the interviews with an information sheet/vignette. The interview will be scripted, enabling the attribution of the effects of the independent variables to the stages of investigative decision-making, particularly the *Inference drawing* and *Decision-making* stages. Thus, it contains equal questions and answers for all participants (apart from evidence mentioned therein), depending on their corresponding condition. All suspects will be innocent and provide realistic explanations for the interviewers' questions. This will facilitate the investigation of the

influences of the independent variables on the interviewers' judgements and decisions. The research question in this study is: *What are the effects of guilt presumptions and evidence-strength on important investigative judgements and perceptions of the investigating officer about the suspect?*

Hypotheses

1. *Interviewers presuming guilt will be more likely to judge the innocent suspect as guilty.*

2. *Interviewers presuming guilt will be less likely to judge the innocent suspects story as plausible then interviewers not presuming guilt.*

3. *Interviewers presented with strong evidence are more likely to judge the innocent suspect as being guilty than interviewers presented with weak evidence.*

4. *Interviewers presented with strong evidence will be less likely to have a higher intention to further investigate the suspect than interviewers presented with weak evidence.*

5. *Interviewers presented with strong evidence will deem the narrative of the suspect as more plausible than interviewers presented with weak evidence.*

6. *Interviewers presented with weak evidence will be more likely to judge the suspect as innocent when they are presuming guilt as when they are not presuming guilt.*

7. *Interviewers that are presuming guilt will be more likely to have a higher intention for further investigation when presented with weak evidence than interviewers not presuming guilt.*

Methods

Design

The present study employed a 2x2 factorial design with the independent variables Guilt Assumption (*Presuming Guilt* vs. *Not Presuming Guilt*) and Evidence-Strength (*strong* vs. *weak*). The dependent variables tested were Guilt Judgement/ Confidence Judgement, Intention to Conduct a Second Interview, and Plausibility of the Suspect's Story.

Participants

Participants were partly recruited via SONA ($n = 9$), which is a university-based platform designed to help researchers find participants for their projects. In return, participants are granted credits needed to complete their programmes. Mostly, social media and social environment of the researchers were consulted for recruitment ($n = 61$). Those participants did not receive any incentives or rewards. Inclusion criteria were that individuals were at least 18 years of age and understand and speak English.

Some data was excluded because progress on the questionnaire was only 8% ($n = 3$). The resulting total sample size after this exclusion was 67. Majority of the sample was German ($n = 58$), followed by Dutch ($n = 2$) and Spanish ($n = 2$) and French *($n = 1$)*, Jordanian *($n = 1$)*, Filipino *($n = 1$)* and Belgian *($n = 1$)*. The sample largely consisted of students ($n = 41$), of which 66 % ($n = 27$) study in the Netherlands. Each participant was sequentially allocated to one out of four experimental conditions: (*A*) = No Guilt Presumptions and Weak Evidence *($n = 17$)*; (*B*) = Guilt Presumptions and Weak Evidence *($n = 16$)*; (*C*) = No Guilt Presumptions and Strong Evidence *($n = 20$)* and (*D*) = Guilt Presumptions and Strong Evidence *($n = 14$)*. Data confidentiality and anonymity was granted, explicitly stating that data will be stored anonymised in a secured place. The research was approved by the BMS Ethics Committee (reference number: 220428).

Materials and Procedure

Vignette

Prior to the experiment sessions, participants received a mail. The mail included information on how to join the session, information about their task to role-play a police officer interviewing a suspect, and a Vignette (Appendix A). The Vignette was intended to prepare participants for the session. It included information on the scenario, the alleged crime, the planned interview procedure, and the interview script itself, with the instruction to stick to the presented script. Online questionnaires that accompany the interview session were announced. Potential risks, confidentiality, and their right to withdraw were mentioned.

Evidence Strength. The vignette included five pieces of evidence information, differing in strength dependent on the experimental condition. To correspondingly frame the evidence regarding its strength, the Evidence Framing Matrix (Granhag et al., 2012) was used. It ascribes two dimensions to evidence-strength, namely *strength of the source of evidence,* and *specificity/degree of precision of evidence* (sometimes referred to as *reliability* and *proximity*, respectively (Brimbal & Luke, 2021)). Pozzulo & Dempsey (2009) found children witnesses to be perceived as less reliable as adult witnesses by jurors. The same was found for people with mental health conditions (Reavey et al., 2016). Therefore, in the weak evidence condition, statements of a child and a person with a mental health condition (weak sources which might have mistaken the suspect for someone else) were included, as opposed to surveillance camera footage (strong source that precisely shows suspect). The other three pieces of information each differed in *specificity* between the condition. The differential pieces of evidence can be seen in table 1.

Table 1

Pieces of evidence in the Weak Evidence condition and opposing Strong Evidence condition

Weak Evidence condition	Strong Evidence condition
An old woman (suffering from schizophrenia) saw someone that looked like the suspect together with the Mrs. Brown in the park, 5 minutes before and after Mrs. Brown dealt the drugs	There is surveillance camera footage that the suspect was together with Mrs. Brown in the park, 5 minutes before and after Mrs. Brown dealt drugs
A 12-year-old child (suffering from autism) saw a man that might have been the suspect driving together with Mrs. Brown to the crime scene, shortly before Mrs. Brown dealt the drugs	There is surveillance camera footage that the suspect drove together in a car with Mrs. Brown to the crime scene, shortly before Mrs. Brown dealt the drugs
There was one phone call from Mrs. Brown on the suspect´s phone on the day of the crime, though the content of this call is unknown	There were 2 specific messages from Mrs. Brown on the suspect´s phone, these messages specify the time and place of the drug deal
Mrs. Brown says that she knows the suspect but refused to disclose the nature of their relationship or whether the suspect is directly involved in her drug dealing	The suspect and Mrs. Brown are friends on Facebook and follow each other on Instagram and Mrs. Brown claims the suspect provided her with the drugs she sold
There were traces of marijuana found in the car of the suspect	There were traces of the same drugs (Heroin, morphine, LSD, Marijuana) in the car from exactly the same batch as Mrs Brown dealt

Note. Based on Evidence Framing Matrix (Granhag et al., 2012)

Guilt Presumptions. Additionally, the manipulation of guilt presumption was included, mentioning at the end of the Vignette that the interviewer's supervisor is either "very sure that the suspect is guilty" (*Guilt Presumption* condition) or is "not sure whether the suspect is guilty" (*No Guilt Presumption* condition), again depending on the experimental condition. Moreover, questions in all the conditions were similar, except that the evidence mentioned in them was corresponding to the experimental condition.

Pre-Interview

After participants joined the online video-call, they were greeted and asked to shortly read through the information sent to them beforehand again, offering them opportunity to ask

questions in the case of misunderstandings. As soon as they communicated to the researchers to be ready, they were told to start with the first part of the questionnaire, accessible via the link the researcher provided them.

Pre-Questionnaire

Questionnaires were carried out on Qualtrics. One of the first questions participants saw after accessing the link was the manipulation check. The Guilt Presumption was checked by the item "*How sure are you about the suspects guilt*?", assessed on a 5-point Likert scale (1 = *very sure the suspect is innocent*, 5 = *very sure the suspect is guilty*). Therein, scores above 3 are considered *Guilty,* while scores below 3 are considered *Not Guilty.* Additionally, there was a binary measure of *Guilt Judgement,* asking "*Based on the Information I have about the current case, I believe the suspect is…"* with the answer options *Guilty* or *Not Guilty.* This latter item was included to enable assessment of percentages of participants that presume guilt before the interview. Evidence-Strength manipulation was tested by the item "*I perceive the evidence against the suspect as …*" answerable on a 5-point Likert scale (1 = *very weak*; 5 = *very strong*).

Ultimately, the participant arrived at a page stating that the first part of the questionnaire was finished and that he should leave this page open and head back to the video-call and notify the researcher.

Interview

For the next step, participants were questioned again whether there were any uncertainties about the following interview. When participants communicated that there were none and that they were ready, the researcher turned off his camera and microphone and the other researcher, role-playing the suspect, joined the meeting. It was ensured that the participant did not know the researcher role-playing the suspect who was told to be another

participant, thereby deceiving the participant. This was done to prevent a possible rapport induced distortion of results, as preliminary research suggests an influence of rapport on information uptake (Weiher, 2020). Further, this was done to prevent possible biases (e.g., social desirability bias when answering questionnaire) to occur. If participants would have known the other person, they could have been inclined to answer questionnaires in a socially desirable way out of the fear that researchers might identify their data and judge them personally based on their answers.

As soon as both were "facing" each other, the participant started asking questions. The questions and answers were all scripted and pre-defined. This method was chosen to ensure comparability between the conditions via isolating the effect of the independent variables. Scripted questions and answers therein guarantee that each participant experiences the interview equally and experiences similar effects regarding the suspects utterances, except for the variable-related utterances (that is, everything related to the evidence) who differ across the experimental conditions. For example, in the *weak* evidence condition, a question was "*An old woman saw that you were together with Mrs. Brown in the park, 5 minutes before and after she dealt drugs.* Can you explain why you were with Mrs. Brown if you do not know her?", while the corresponding question in the strong evidence condition was "*There is surveillance camera footage that you were together with Mrs. Brown in the park, 5 minutes before and after she was confirmed to have dealt drugs.* Can you explain why you were with Mrs. Brown if you do not know her?".

In total, the interview part of the study took about 5 minutes per participant. Measures of the entire process displayed a median of 20.18 minutes (14.2 – 29.73) per participant. As soon as the interview was finished, the researcher joined back into the meeting to tell the participant that the interview is over and ask the participant to fill out the final part of the questionnaire. Additionally, the researcher told the participant that he would stay in the

meeting in case of upcoming questions. To keep the deception of the other person being a participant too, the researcher gave both persons (confederate & participant) identical information (i.e., told both to fill out the rest of the questionnaire). The confederate filled out a different questionnaire, measuring trust and rapport, which will not be described or analysed here since it is part of a different paper.

Post-Questionnaire for Participants

Guilt Judgements. Following up to the interview, participants instantly started with the post-questionnaire. The first question, pertaining to the variable *Guilt Judgement,* was asking "*How sure are you about the suspects guilt?"*, with a 5-point scale as answer option ranging from 1 = *Very sure the suspect is innocent* to 5 = *very sure the suspect is guilty*. Here, the cut-off point is the same as in the manipulation check (> 3 = *Guilty;* < 3 = *Not Guilty*). As this was the same item as was used in the manipulation check for *Guilt Presumptions,* it enabled later comparison as pre-interview and post-interview measures.

Evidence-Strength. After answering the items on *Guilt* Judgements, participants had to complete the statement "*Now, I perceive the evidence against the suspect as…"* with answer options on a 5-point scale from 1 = *very weak* to 5 = *very strong*. This concerned checking whether the interview changed the participants perception of *Evidence-Strength.* This item was the same as the manipulation check of *Evidence-Strength* in the pre-questionnaire, again enabling later comparison.

Plausibility of the Suspect's Story. To measure the dependent variable *Plausibility of the suspect's story,* the Narrative Believability Scale (NBS), consisting of 12 items (Yale, 2013) was employed. The scale consists of four subscales (*plausibility, completeness, consistency,* and *coverage)* which are 7-point Likert-scales (1 = *strongly disagree;* 7 = *strongly agree)*. However, the subscale of *completeness* in this study could be more understood as *comprehensiveness* since it cannot be known by an investigator whether a

narrative is *complete* if he/she does not exactly know what happened. However, he/she can assess how much detail and depth is conveyed, which is denoted in *comprehensiveness*.

Reliability of the whole scale can be deemed very good with a Cronbach`s alpha of .82. Corresponding values for the subscales were for *plausibility* .89, for *comprehensiveness* .71, for *consistency* .86, and for *coverage* .77. Therefore, reliability of *plausibility* and *consistency* is very good, reliability for *comprehensiveness* and *coverage* is good (Streiner, 2003). Originally intended to measure believability of a narrative in a courtroom setting, the scale was adopted to investigative interviews since they measure the same judgement, namely whether a told narrative is believable or not. Some items were reverse-coded, such as "*There were lots of "holes" in the suspects narrative*" (1 = *strongly agree*; 7 = *strongly disagree*). A higher score on the full scale indicated that the participant had a stronger belief in the plausibility of the suspect's story.

Intention to further Investigate. The scale for the independent variable *Intention to further Investigate* was created. It consisted of six items, answerable on a 7-point Likert-scale, concentrated on different aspects underlying the construct of intention. The operational definition of intention here being "a mental state that guides and organizes behaviour" and that is "a determination to act in a certain way or to bring about a certain state of affairs" (Shultz, 1980). Aspects covered were, for example, the guidance of behaviour ("*I think this person should be further investigated*") or the determination to achieve particular states of affairs ("*I believe that a second investigative interview would contribute to the progress in this case*"). Some were positive coded, focused on the direct questioning of intentions (e.g., "*I think this person should be further investigated*"), while others were negatively coded, such as "*I believe the suspect should be released without further questioning*". On the whole scale, a high score would ultimately indicate a higher intention for further investigation.

Demographics and Immersion. Lastly, participants had to state their nationality, age, whether they were students and where they were studying. They were placed at the end of the questionnaire to prevent priming and/or stereotype threat to happen and to improve participant retention (Fernandez et al., 2016). Additionally, there were questions regarding the participants own feelings and opinions about the interview and immersion. These were four 5-point Likert-Scale items ("*I took the task seriously*"; "*I could imagine myself as an interviewer*"; "*The interview was somewhat realistic*"; "*The suspect played his/her part well*") answerable from 1 = *strongly disagree* to 5 = *strongly agree*. Further, there was a question "*Do you have anything left to say to the researchers?*", which was answerable with an open text.

Debrief and Conclusion

After all questions were answered, the participants got provided the debriefing, containing information about the actual purpose of the study. There, it was also communicated to them that they have been deceived in being told that the interviewee was another participant. After the debriefing withdrawal of their participation was offered again. However, no participant stated any kind of distress or disturbance, and no-one withdrew their consent. The debriefing was included in the Qualtrics online form (Appendix B). After they read the debrief, participants headed back to the meeting, where the researcher asked one last time whether participants still had some questions. When this got denied, the session was concluded.

Data Analysis

To analyse the data, the program SPSS 26 was employed. The alpha-level for all analyses was 0.05. T-tests were used to test the manipulations of Guilt Presumptions and Evidence-Strength. A two-way ANOVA was run to analyse the differences in the dependent variables (*Guilt Judgement, Plausibility of the Suspect's Story,* and *Intention to Further*

Investigate), caused by the independent variables. Further, the two-way ANOVA showed possible interaction effects of the two independent variables on the dependent variables.

Results

Descriptive Statistics

The scores of *Guilt Judgement (pre-interview)* were rather high ($M = 3.55$, $SD = 0.96$), while those of *Guilt Judgement (post-interview)* were lower, but still on the positive side ($M = 2.90$, $SD = 1.17$). Both were normally distributed. The values of *Plausibility of the Suspect's Story* were normally distributed ($M = 4.42$, $SD = 1.12$). The mean of *Intention to further investigate* was high ($M = 5.36$, $SD = 1.30$) and the scores were approximately normally distributed. A full overview on descriptive statistics and correlations can be seen in table 2.

Guilt Judgment (pre-interview) was positively correlated with *Guilt Judgement (post-interview)*, $r(65) = .30$, $p = .015$. *Guilt Judgment (pre-interview)* was not significantly correlated with *Plausibility of the Suspect`s Story*, $r(64) = -.13$, $p = .301$, but significantly positively correlated with *Intention to further Investigate,* $r(64) = .34$, $p = .006$. *Guilt Judgement (post-interview)* was significantly negatively correlated with *Plausibility of the Suspect's Story*, $r(64) = -.51$, $p < .001$ and significantly positively correlated with *Intention to further Investigate*, $r(64) = .50$, $p < .001$. Further, *Intention to further Investigate* and *Plausibility of the Suspect's Story* were significantly negatively correlated, $r(64) = -.46$, $p < .001$. Collectively, the correlations show that *Guilt Judgements (pre-interview)* were positively associated with *Guilt Judgements (post-interview)* and *Intention to further Investigate.* Additionally, *Plausibility of the Suspect's Story* was negatively associated to *Guilt Judgements (post-interview),* and *Intention to further Investigate.* Lastly, *Intention to further Investigate* was positively associated with *Guilt Judgements (post-interview).*

Table 2

Descriptive Statistics and Correlations of Dependent Variables

	N	M	SD	Guilt Judgement (pre-interview	Guilt Judgement (post-interview)	Plausibility of the Suspect's Story
Guilt Judgement (pre-interview)	67	3.55	0.96			
Guilt Judgement (post-interview)	67	2.90	1.17	.30*		
Plausibility of the Suspect's Story	66	4.42	1.12	-.13	-.51*	
Intention to further Investigate	66	5.36	1.30	.34*	.39*	-.46*

Note. $* = p < .05$

Frequencies of the binary measures of *Guilt Judgements* displayed that before the interview, 68.7% ($n = 46$) of participants judged the suspect *Guilty* and 31.3% ($n = 21$) judged the suspect *Not Guilty.* After the interview, only 43% ($n = 29$) of participants judged the suspect *Guilty* and 56.7% ($n = 38$) judged the suspect *Not Guilty.*

Effectiveness of Manipulations

T-tests were employed for manipulation checks. The *Guilt Presumption* manipulation check indicated no significant differences between groups in their scores of *Guilt Judgements (pre-interview)* between the *No Guilt Presumption* ($M = 3.57$, $SD = 0.93$) and *Guilt Presumption* ($M = 3.53$, $SD = 1.01$) conditions, $t(65) = 0.14$, $p = .886$. The manipulation check of *Evidence-Strength* indicated a significant difference in the perceptions of *Evidence-Strength* between the groups of the *Weak Evidence* ($M = 2.88$, $SD = 1.09$) and *Strong*

Evidence (*M* = 3.60, *SD* = 1.19) conditions, *t* (67) = -2.60, *p* = .011. Thus, the *Guilt*

Presumption manipulation failed while the *Evidence-Strength* manipulation succeeded.

Hypothesis-Testing

To account for the failure of the *Guilt Presumption* manipulation, the analysis of *Guilt*

Judgements was changed from a two-way ANOVA to a mixed measure ANOVAs with the

within-subjects factor *Time* (*pre-interview* vs. *post-interview*) and between-subjects factor

Evidence-Strength (*weak* vs. *strong*). The distinction between pre-interview and post-

interview was chosen since most participants assumed guilt before the interview, so this

allowed to check if perceptions changed due to the interview. The mixed ANOVA tested for

main effects of the interview itself (*Time*) and *Evidence-Strength*, as well as for an interaction

effect between *Time* and *Evidence-Strength*. However, since *Plausibility of the Suspect's*

Story and *Intention to further Investigate* were only measured post-interview, their analysis

was conducted according to the initial plan via a two-way ANOVA, keeping the factor *Guilt*

Presumptions. This tested for independent main effects of the *Guilt Presumptions* and

Evidence-Strength. Additionally, these ANOVAs checked whether there was an interaction

effect between the two independent variables on the corresponding dependent variable.

Guilt Judgements

All assumptions for running the ANOVA were met (see Appendix C). The mixed

ANOVA showed a significant difference in means between the Pre-Interview (*M* = 3.55, *SE* =

.11) and Post-Interview (*M* = 2.89, *SE* = .14) measurements, indicating that Guilt Judgements

did reduce from pre-interview to post-interview (*F* (1,65) = 17.54, *p* < .001, partial η^2 = .212).

While the mean was below three (*Not Guilty)* after the interview, it was still close to the cut-

off point. While this indicates that the interview was able to change participants' guilt

perception, the proximity to the cut-off point hints towards the previous notion to be taken

cautiously.

Further, the results of the analysis of main effects showed the means to be significantly different between the *Weak* ($M = 2.99$, $SE = .15$) and the *Strong* ($M = 3.46$, $SE = .14$) evidence conditions, $F (1,65) = 5.38$, $p = .024$, partial $\eta^2 = .076$.

Additionally, there was no significant interaction effect of *Time*Evidence-Strength*, Wilks` Lambda = .99, $F (1,65) = 0.26$, $p = .612$, partial $\eta^2 = .004$ (see figure 1).

Figure 1

Interaction effect of Evidence-Strength and Time on Guilt Judgements
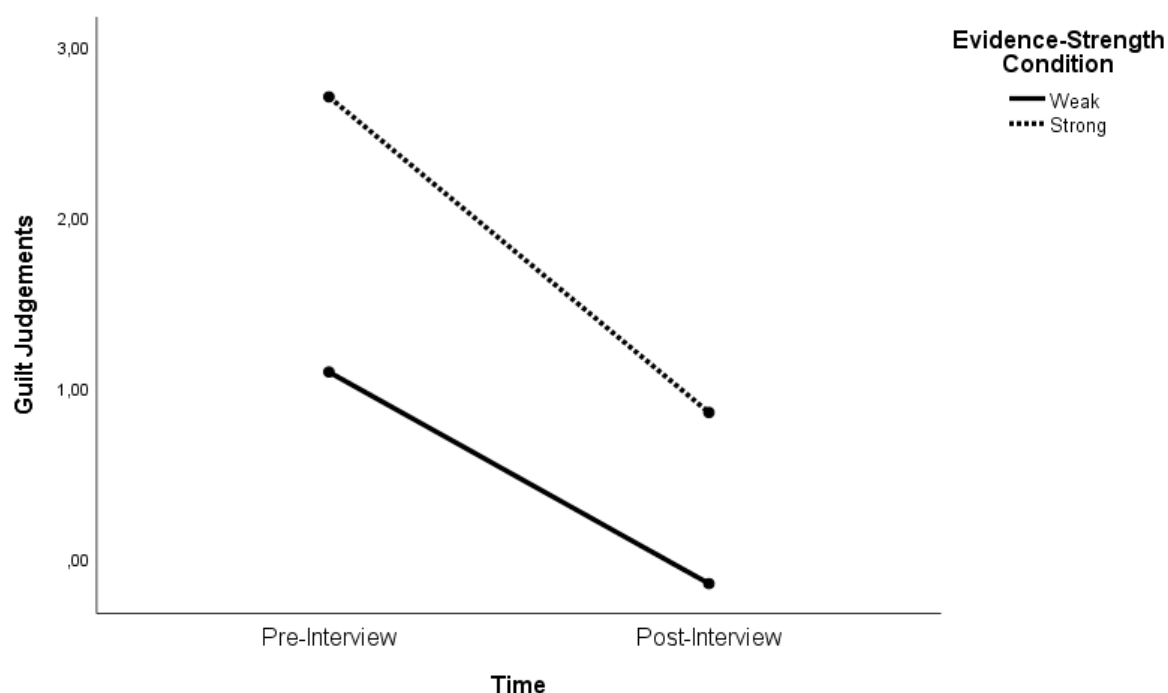
The original hypothesis *H1 ("Interviewers presuming guilt will be more likely to judge the innocent suspect as guilty.")* could not be tested due to manipulation failure. However, since most participants did presume guilt in the first place, a manipulation happened in the sense that most participants, irrespective of conditions, presumed guilt. This enables a checking of whether those presumptions changed over the course of the interview. Therefore,

the new hypothesis shall be *H1a* ("*Interviewers perceptions of guilt will change due to the interview itself.*"). The null hypothesis of *H1a* could be rejected, as perceptions of guilt, reflected in *Guilt Judgements*, did change significantly from pre-interview to post-interview.

The null hypothesis of *H6* (*"Interviewers presented with weak evidence will be more likely to judge the suspect as innocent when they are presuming guilt as when they are not presuming guilt."*) could not be rejected due to failed manipulation. However, it can be said that the means of both groups are still indicating that participants believe the suspect to be guilty after the interview, especially in the *strong* evidence condition.

Further, the null hypothesis of *H3* ("*Interviewers presented with strong evidence are more likely to judge the innocent suspect as being guilty than interviewers presented with weak evidence.*") could be rejected.

Plausibility of the Suspect's Story

All assumptions for the ANOVA were met (see Appendix E). The analysis of the main effects showed that there was no significant difference in means between the *Guilt Presumption* ($M = 4.62$, $SE = 0.21$) and *No Guilt Presumption* ($M = 4.23$, $SE = 0.18$) conditions, $F(1,62) = 2.05$, $p = .158$, partial $\eta^2 = .032$. Further, there was no display of a significant difference in means between *Strong Evidence* ($M = 4.28$, $SE = 0.20$) and *Weak Evidence* ($M = 4.57$, $SE = 0.19$) conditions, $F(1,62) = 1.16$, $p = .287$, partial $\eta^2 = .018$.

Results displayed no significant interaction effect between *Guilt Presumptions* and *Evidence-Strength* on *Plausibility of the Story*, $F(1,62) = 1.82$, $p = .182$, partial $\eta^2 = .029$ (see figure 2).

Figure 2

Interaction effect of Evidence-Strength and Guilt Presumption on Plausibility of the Suspect's Story.
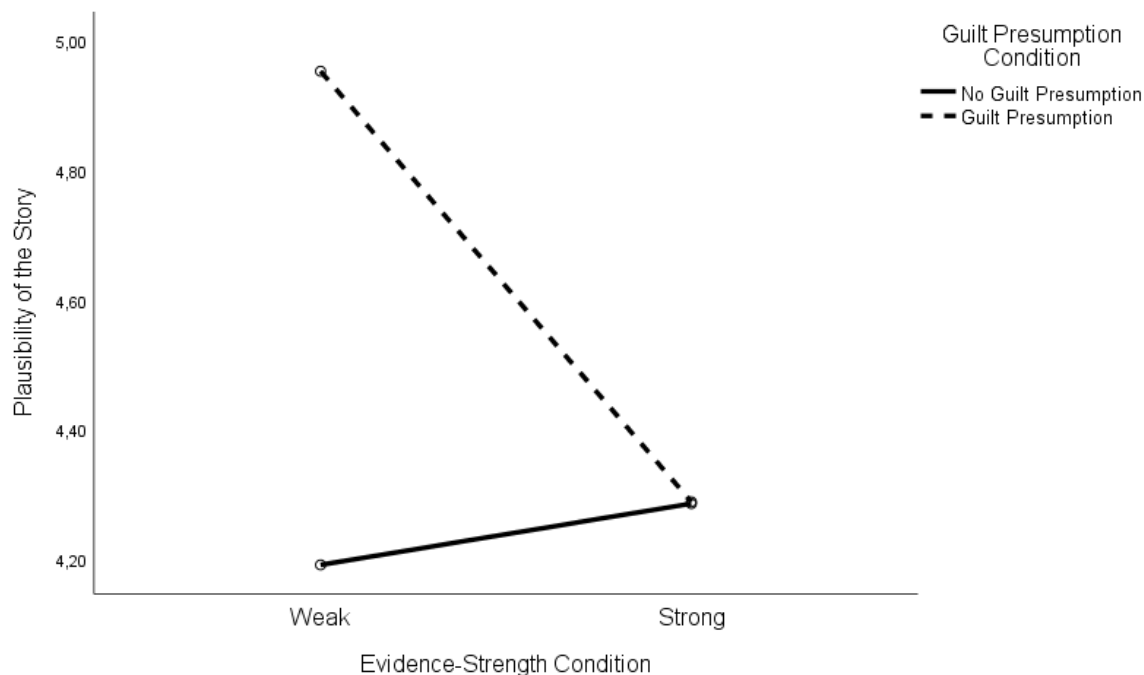
Therefore, the null hypotheses of *H2* ("*Interviewers presuming guilt will be less likely to judge the innocent suspects story as plausible then interviewers not presuming guilt.*") could not be adequately tested. The null hypothesis of *H5* ("*Interviewers presented with weak evidence will deem the narrative of the suspect as less plausible than interviewers presented with strong evidence.*") could not be rejected.

Intention to further Investigate

There were 3 outliers in the No Guilt Presumption condition detected in the boxplots (Appendix F). These outliers were unusual data, therefore kept in the dataset. Therefore, ANOVA could still be conducted.

There were no significant differences between means in the *Guilt Presumption* ($M = 5.15$, $SE = 0.24$) and *No Guilt Presumption* ($M = 5.56$, $SE = 0.21$) conditions, $F(1,62) = 1.68$, $p = .200$, partial $\eta^2 = .026$. Additionally, there was no display of significant differences between the *Strong Evidence* ($M = 5.64$, $SE = 0.23$) and *Weak Evidence* ($M = 5.06$, $SE = 0.22$) conditions, $F(1,62) = 3.34$, $p = .073$, partial $\eta^2 = .051$. However, the effect of *Evidence-*

*Strength* was close to being significant (*p* = .073), possibly indicating a tendency for individuals in the *Strong Evidence* condition to be more likely to want a further investigation.

Therefore, the null hypothesis corresponding to *H4* ("*Interviewers presented with weak evidence will be more likely to have a higher intention to further investigate the suspect than interviewers presented with strong evidence.*") could not be rejected.

Results displayed no significant interaction effect between the variables *Guilt Presumption* and *Evidence-Strength* on *Intention to further Investigate, F* (1,62) = 1.50, *p* = .225, partial η^2 = .024 (see figure 3).

Figure 3

Interaction effect of Evidence-Strength and Guilt Presumption on Intention to further Investigate.
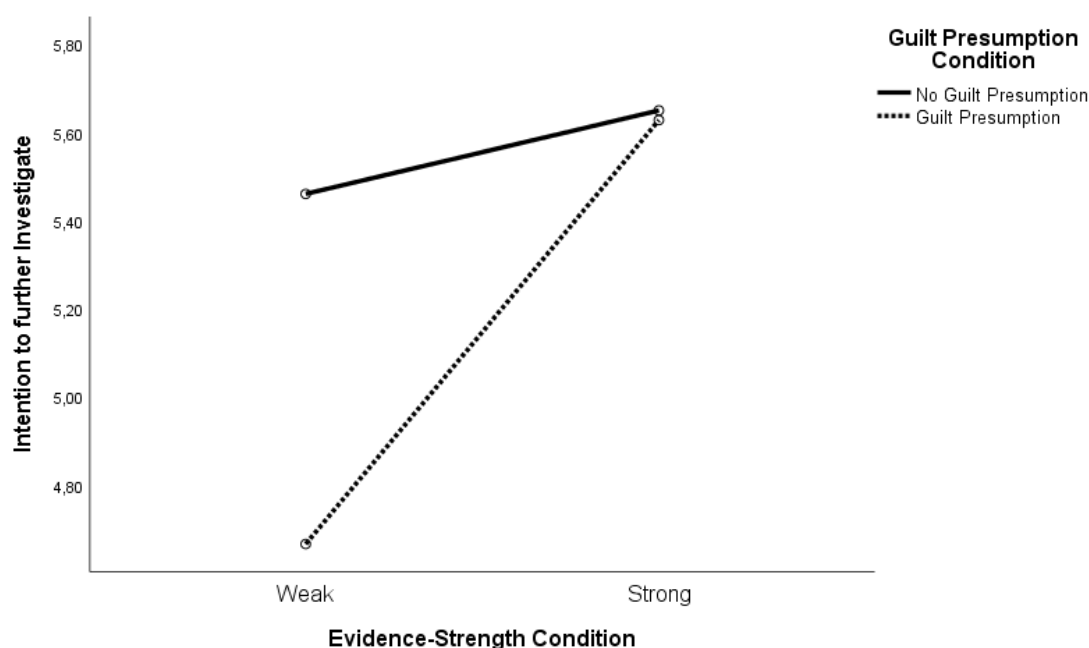
Based on this, the null hypothesis regarding H7 ("*Interviewers that are presuming guilt will be more likely to have a higher intention for further investigation when presented with weak evidence than interviewers not presuming guilt.*") could not be rejected.

Exploratory Analysis

To follow up on the found significant effect of *Evidence-Strength* on *Guilt Judgement,* and to determine where these differences might originate from, it was decided to include an analysis on *Immersion.* It might have been the case that participants perceived the interview different in the two *Evidence-Strength* conditions. This could be attributable to the different evidence pieces (e.g., *surveillance camera* vs. *witness*) included within the interview. To explore whether there was a difference of the perception of the interview, it was checked whether the *Immersion* score differed between groups of *Evidence-Strength*. This was done by employing an independent samples Wald t-test with *Immersion* as the test-variable and *Evidence-Strength* (*Weak* vs. *Strong*) as the grouping variable. Wald t-test was chosen since the test of heterogeneity of variances displayed a significant difference. However, the test revealed no statistically significant differences of *Immersion* scores between the conditions ($M = 4.02$, $SD = 0.79$ vs. $M = 3.99$, $SD = 0.74$, $t (63.71) = .161$, $p = .873$). Therefore, it can be concluded that the differential pieces of evidence included did not cause a difference in perceptions of the interview itself as displayed in *Immersion*.

Additionally, it was checked whether any of the questions about immersion of participants correlated with any of the dependent variables. Results of the correlational analysis are displayed in table 3. A statistically significant correlation was observed between "*The interview was somewhat realistic*" and *Intention to further investigate* ($r (64) = .30$, $p = .015$). Further, "*I could imagine myself as an interviewer*" was statistically significant negatively correlated with *Plausibility of the Suspect's Story* ($r (64) = -.27$, $p = .03$).

Table 3

Descriptive Statistics of Immersion Questions and Correlations with Dependent Variables

	N	M	SD	Intention to further Investigate	Plausibility of the Suspect's Story	Guilt Judgement (pre-interview)	Guilt Judgement (post-interview)
"I took the task seriously"	65	4.66	0.74	.065	-.03	.01	.05
"I could imagine myself as an interviewer"	66	3.77	1.15	.09	-.27*	.22	.22
"The interview was somewhat realistic"	66	4.09	0.92	.30*	-.16	.19	.21
"The suspect played his/her part well"	66	4.15	0.90	.20	.04	-.01	.07

Note. * = p < .05

Since the manipulation check of *Guilt Presumptions* failed, but the scores of *Guilt Judgement (pre-interview)* were still high, indicating most participants deeming the suspect guilty before the interview, a t-test was performed to check whether the scores of *Guilt Judgement (pre-interview)* differ across the two dimensions of *Evidence-Strength.* The Wald t-test, chosen due to heterogeneity of variances, showed a significant difference in means between the two experimental conditions, t (63.64) = 2.44, p = .017.

Discussion

The aim of the current study was to investigate how *Evidence-Strength* and *Guilt Presumptions* influence several investigative judgements in an investigative interview. Scripted interviews with participants role-playing police officers were carried out. The results showed that there were significant main effects from *Evidence-Strength* and the interview itself on *Guilt Judgements*. While the interview itself decreased *Guilt Judgements,* participants in the *weak* evidence condition displayed lower means in *Guilt Judgements* than those in the *strong* evidence condition. The latter also points towards *Evidence-Strength* being an alternative manipulation of Guilt Presumptions, due to the clear effect it had on the pre-interview measurements of Guilt Judgements. Apart from this, there were no statistically significant main or interaction effects found. However, the effect of *Evidence-Strength* on *Intention to further Investigate* was close to being significant.

Additionally, there was significant correlation between the pre-interview and post-interview measures of *Guilt Judgements.* This supports the notion that higher baseline guilt assumptions lead to higher *Guilt Judgements* even when presented with explanations. Further, *Guilt Judgements (post-interview)* were significantly negatively correlated with *Plausibility of the Suspect's Story,* indicating that even though participants were presented with a plausible explanation, they still perceived the story as less plausible, possibly due to confirmation bias setting in and thereby making participants seek to confirm their assumption that the suspect is

guilty by deeming the story unplausible. Both measures of *Guilt Judgement* were significantly positively correlated with *Intention to further Investigate*. Another statistically significant negative correlation was found between *Plausibility of the Suspect's Story* and *Intention to further Investigate*, showing that the more participants deemed the story of the suspect plausible, the less they intended to further investigate on him/her. These correlations are supporting of the proposed influence of *Guilt Presumptions* on investigative judgements, even though the independent variable itself exerted no influence.

Guilt Presumptions

It was possible to extend the findings of Kassin et al. (2003) that assumptions of guilt influence expert judgements. It was shown that constant information retrieval behaviour in the first stage of investigative decision-making (*Information* retrieval) allows people to change their minds to a certain extent, even if guilt is assumed prior to the interview. The extent therein being that it was possible to decrease the judgement of guilt, to such a degree that the mean was below the cut-off value. However, it was very close to the cut-off value, thus the effect of constant information retrieval was sufficient to make all people believe that an innocent suspect is innocent. This might be a hint towards that guilt presumption indeed does prevent people from changing their minds about guiltiness, irrespective of provision of good explanations against evidence, which apparently mostly influences their confidence. Thus, even though the manipulation failed, there might still be confirmation bias about guilt taking place, but induced by the very case itself, possibly by the manipulation of *Evidence-Strength*. The high mean scores on pre-interview and post-interview measures of *Guilt Judgements* and the correlations between the pre-measure and *Intention to further Investigate* underline this.

Considering the initially failed manipulation, it might be inferred that a simple one sentence statement stating the supervisor's opinion might just be not enough to act on

individuals' feelings about the case. Therefore, a more thorough approach should be followed when trying to manipulate presumptions in an experiment in general. Possibly, a normative statement indicating a percentage might work better, as previous studies successfully used this kind of manipulation (Hill et al., 2008). Perhaps this could be combined with a statement about supervisor opinions, thereby getting more weight in participants considerations. An example for this could be "Your supervisor just told you that he thinks the suspect is guilty and reminded you of the fact that 90% of drug related allegations connected with the location were found to be true.".

The significant effect of the interview itself shows that the interaction with the suspect exerted an influence on the investigator. The case here is most likely that participants were doubting their judgement due to the explanations made by the suspect. Further supporting this explanation is that, based on the item used, most likely participants confidence in their judgements had notably decreased. Therefore, the explanations given by the suspect were deemed plausible, as seen in the *Plausibility of the Suspect's Story* score. They were equally plausible in both conditions, *weak* and *strong*, but participants in the *strong* condition had initially higher scores on *Guilt Judgement (pre-interview),* making them still deem the suspect guilty after hearing explanations.

Evidence-Strength

Building up on the notions of Moston & Engelberg (2011) of evidence-strength being one of the most influential factors in investigative interviewing, it was hypothesized that it would significantly influence every single investigative judgement included. The only significant effect found here was on *Guilt Judgements.* It must be noted that it might have especially influenced the *confidence,* which was included as a measure of *Guilt Judgements.* However, this confirms the hypothesis that stronger evidence will lead to higher *Guilt Judgement* scores, even though plausible explanations are provided. This finding is in line

with other findings that suggest that decision-makers mostly base their decisions on situational factors (e.g., evidence) when the evidence is indeed strong (Smith et al., 2011). The effect size of *Evidence-Strength* being rather high additionally supports this.

Apart from this, the logical pathway by which the evidence-strength influences *Guilt Judgements* or any confounding variables may be subject to further investigation. Again, confirmation bias is possibly playing a major role and may be investigated in how it behaves with different information/evidence presented to participants. It might for example be that confirmation bias distorts the interpretation of information, which is shown by recent research in connection with polarizing topics, as which a crime investigation could broadly be understood, as being prone to being influenced by confirmation bias (Vedejová & Čavojová, 2021).

Intention to further Investigate was close to being significantly influenced by *Evidence-Strength.* It might be that participants, although they indicated to perceive the evidence as strong (or weak), do not really have had an idea of just how strong (or how weak) the evidence presented was. The effect might even be genuine, which would indicate that the strength of evidence presented exerts a direct influence on the *Intention to further Investigate,* making strong evidence result in higher *Intention to further Investigate* and weak evidence result in lower *Intention to further Investigate.* Speaking for this is the criteria of evidence elasticity found by Ask et al. (2008), who mention that criteria for assessing the reliability of evidence are highly prone to influences of context and are highly malleable. Applying this to the current case, participants perhaps would have been equipped with a better idea of strength of the evidence when they would have had a reference or contrast case. This could potentially be an unrelated case with the corresponding other strength of evidence prevalent.

Further, the lack of a significant effect of *Evidence-Strength* on *Plausibility of the Suspect's Story* might reflect the researcher's intention to make the suspects` explanations in

the mock interview as plausible as possible. Based on this, it was the case that most people deemed the story of the suspect as plausible, but this was just not enough for them to believe the suspect to be innocent. This is supported by the frequency of very high values (> 4) on *Plausibility of the Suspect's Story,* which was prevalent in 43 out of 67 cases, with 23 in the *weak* evidence condition and 20 in the *strong* evidence condition (see Appendix G).

Limitations

Limitations of the present study pertain to the failure of manipulation of guilt presumptions, the sample size, methods, and external validity. First, the manipulation of the main independent variable, *Guilt Presumption,* failed. This constitutes the major limitation of this study, since it was hypothesized to be very influential, and its addressing was intended to fill a niche within scientific knowledge. However, the influence of *Guilt Presumptions* was still included in the analysis to test the core ideas that were hypothesized before. Further, the fact that, in the exploratory analysis, it was found that means on *Guilt Judgement (pre-interview)* differed significantly across the two dimensions of *Evidence-Strength* points toward *Evidence-Strength* having been a manipulation of *Guilt Presumptions* and was thereby influencing *Guilt Judgements.*

Another limitation might pertain to the method, namely the interview and how it was conducted. Appearances of researchers in their roles as suspects might have influenced participants. Although no information was deliberately given to participants about the suspects persona beforehand, age or interaction with the other researcher might have given hints about them being students too. Since most of the participants were also students, sympathy based on in-group affiliation might have happened (Dolcini & Adler, 1994) and caused distortion. Additionally, participants who were recruited via SONA might have read the names of researchers before and recognized them in the ZOOM call.

Further, a bigger sample size would perhaps have brought about more confidence about the here near significant influences of *Evidence-Strength* (e.g., on *Intention to further Investigate*) being truly not genuine and would have decreased the probability of making Type II errors. This goes hand in hand with the third limitation to this study, namely the external validity of the results. Additionally, there might be differences in the sample and the target population (investigating police officers). Professional investigators most likely have learned different ways of interpretation of verbal or non-verbal impressions, and additionally might have different approaches in reappraising the value of the given evidence when presented with suspect statements. These differences between target population and sample might be decreased with recruiting actual interviewers as participants, rather than students.

Implications for Future Research

Future research should generally try to further address the topic of presumptions of guilt and evidence strength and their influence on investigative decisions within an investigative interview, therein paying special attention to consistent information retrieval behaviours and confidence of investigators. Given the results of the present study, especially the relationship between confidence and information retrieval practices and judgements of guilt should be focused on. This study showed there is a causal relationship between them. For better understanding of this relationship and the causal pathways, future research should try to find mediating factors between those variables.

A possible candidate for investigation of those causal pathways might be the confidence that the suspect displays in telling his story. Previous research showed that confidence of one's opposite might increase confirmation bias in one's own judgement to take place (Rollwage et al., 2020). This might be interesting in the context of an investigative interview, considering suspects might base their counter-interrogation strategies on the assumption that appearing confident about telling a narrative might make it more believable.

A research design focusing on evidence-strength (*weak* vs. *strong*) and confidence of suspect (low vs. high) influencing perceptions of the interviewer might further extend the present finding. Confidence on the investigators side should also be investigated further, since it seems crucial and is, based on the present results, influenced by information retrieval behaviour.

Further, future research should pay attention to creating a believable and immersive environment in which participants feel more immersed. Not having to rely on online interviews might be beneficial. If possible, future research should try to investigate the issues at hand in a sample that more closely resembles the target population.

Conclusion

The present study tried to investigate effects of guilt presumptions and Evidence-Strength on investigative judgements and interviewer perceptions. The main finding is that the interview itself decreased guilt judgements, as measured by the confidence in those. This can be well attributed to standardized information retrieval procedures, which in turn reinforces the notion that systematic approaches towards investigative interviewing are reducing the effect of confirmation bias, thereby acting in the spirit of the PEACE framework, and allowing for a change of mind. Further, it was shown that the strength of evidence presented has an influence on judgements about guilt. *Evidence-Strength* seemed to act as an alternative manipulation of *Guilt Presumptions*, as it likely exerted influence via confirmation bias, just as expected from *Guilt Presumptions*. Future research should build on these findings and investigate more on confidence in investigative judgements and the causal pathways that allow standardized/consistent information retrieval procedures to reduce effects of confirmation bias. Closer examination of factors influencing interviewers *Intention to further Investigate* might help establish whether the here observed nearly significant effect is real.

References

Adam, L., & van Golde, C. (2019). Police practice and false confessions: A search for the

implementation of investigative interviewing in Australia. *Alternative Law Journal,*

45(1), 52-59. https://doi.org/10.1177/1037969x19874415

Adams-Quackenbush, N., Vrij, A., Horselenberg, R., Satchell, L., & van Koppen, P.

(2020). Articulating guilt? The influence of guilt presumption on interviewer and

interviewee behaviour. *Current Psychology, 41*(4), 2139-2151.

https://doi.org/10.1007/s12144-020-00726-z

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology*

& Health, 26(9), 1113-1127. https://doi.org/10.1080/08870446.2011.613995

Ask, K., & Fahsing, I. (2019). Investigative decision making. In *The Routledge International*

Handbook of Legal and Investigative Psychology (pp. 84-101). New York:

Routledge.

Ask, K., & Granhag, P. (2005). Motivational sources of confirmation bias in criminal

investigations: the need for cognitive closure. *Journal Of Investigative*

Psychology And Offender Profiling, 2(1), 43-63. https://doi.org/10.1002/jip.19

Ask, K., Rebelius, A., & Granhag, P. (2008). The 'elasticity' of criminal evidence: a

moderator of investigator bias. *Applied Cognitive Psychology, 22*(9), 1245-1259.

https://doi.org/10.1002/acp.1432

Bensi, L., Nori, R., Strazzari, E., & Giusberti, F. (2003). Vividness in Judgements of Guilt.

Perceptual And Motor Skills, 97(3_suppl), 1133-1136.

https://doi.org/10.2466/pms.2003.97.3f.1133

Brimbal, L., & Luke, T. (2021). Deconstructing the evidence: The effects of reliability and

proximity of evidence on suspect responses and counter-interrogation tactics. *Journal*

Of Applied Research In Memory And Cognition.

https://doi.org/10.1016/j.jarmac.2021.10.001

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games And Economic Behavior*, 104, 1-23. https://doi.org/10.1016/j.geb.2017.02.015

Clarke, C., & Milne, R. (2001). *A national evaluation of the PEACE Investigative Interviewing Course* (p. 187). London: Home office.

Clarke, C., Milne, R., & Bull, R. (2011). Interviewing Suspects of Crime: The Impact of PEACE Training, Supervision and the Presence of a Legal Advisor. *Journal Of Investigative Psychology And Offender Profiling, 8*(2), 149-162. https://doi.org/10.1002/jip.144

Dahl, J., Enemo, I., Drevland, G., Wessel, E., Eilertsen, D., & Magnussen, S. (2007). Displayed emotions and witness credibility: a comparison of judgements by individuals and mock juries. *Applied Cognitive Psychology, 21*(9), 1145-1155. https://doi.org/10.1002/acp.1320

Dolcini, M. M., & Adler, N. E. (1994). Perceived competencies, peer group affiliation, and risk behavior among early adolescents. *Health Psychology, 13*(6), 496. https://doi.org/10.1037/0278-6133.13.6.496

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., & Kirn, A. et al. (2016) More Comprehensive and Inclusive Approaches to Demographic Data Collection. *ASEE Annual Conference & Exposition Proceedings*. https://doi.org/10.18260/p.25751

Goodman-Delahunty, J., & Martschuk, N. (2018). Securing reliable information in investigative interviews: coercive and noncoercive strategies preceding turning points.

Police Practice And Research, 21(2), 152-171.

https://doi.org/10.1080/15614263.2018.1531752

Granhag, P., Strömwall, L., Willén, R., & Hartwig, M. (2012). Eliciting cues to deception by

tactical disclosure of evidence: The first test of the Evidence Framing Matrix. *Legal*

And Criminological Psychology, 18(2), 341-355. https://doi.org/10.1111/j.2044-

333.2012.02047.x

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A*

handbook. John Wiley & Sons.

Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect

interviews: A systematic evaluation. *Legal And Criminological Psychology,*

13(2), 357-371. https://doi.org/10.1348/135532507x238682

Justice, L., & Smith, H. (2018). Memory judgements: the contribution of detail and

emotion to assessments of believability and reliability. *Memory, 26*(10), 1402-1415.

https://doi.org/10.1080/09658211.2018.1484142

Kassin, S., Dror, I., & Kukucka, J. (2013). The forensic confirmation bias: Problems,

perspectives, and proposed solutions. *Journal Of Applied Research In Memory And*

Cognition, 2(1), 42-52. https://doi.org/10.1016/j.jarmac.2013.01.001

Meissner, C., & Kassin, S. (2004). "You're Guilty, So Just Confess!". *Interrogations,*

Confessions, And Entrapment, 85-106. https://doi.org/10.1007/978-0-387-38598-3_4

Moston, S., & Engelberg, T. (2011). The effects of evidence on the outcome of

interviews with criminal suspects. *Police Practice And Research, 12*(6), 518-526.

https://doi.org/10.1080/15614263.2011.563963

Moston, S., Stephenson, G., & Williamson, T. (1992). The Effects of Case Characteristics on Suspect Behaviour during Police Questioning. *The British Journal Of Criminology, 32*(1), 23-40. https://doi.org/10.1093/oxfordjournals.bjc.a048178

Pozzulo, J., & Dempsey, J. (2009). Witness Factors and Their Influence on Jurors' Perceptions and Verdicts. *Criminal Justice And Behavior, 36*(9), 923-934. https://doi.org/10.1177/0093854809338450

Reavey, P., Wilcock, R., Brown, S., Batty, R., & Fuller, S. (2016). Legal professionals and witness statements from people with a suspected mental health diagnosis. *International Journal Of Law And Psychiatry, 46*, 94-102. https://doi.org/10.1016/j.ijlp.2016.02.040

Rollwage, M., Loosen, A., Hauser, T., Moran, R., Dolan, R., & Fleming, S. (2020). Confidence drives a neural confirmation bias. *Nature Communications, 11*(1). https://doi.org/10.1038/s41467-020-16278-6

Saraiva, R. B. (2020). Distinguishing guilty and innocent suspects identifications: The role of confidence, decision time, decision processes and individual differences. In R. Horselenberg, P. Van Koppen, & Keijser, J (Eds), *Bakens in de Rechtspsychologie*. Liber amicorum voor Peter van Koppen (*Beacons in Legal Psychology*. Liber amicorum for Peter van Koppen). Den Haag: Boom Criminologie

Scherr, K., Normile, C., & Putney, H. (2018). Perpetually stigmatized: False confessions prompt underlying mechanisms that motivate negative perceptions of exonerees. *Psychology, Public Policy, And Law, 24*(3), 341-352. https://doi.org/10.1037/law0000163

Shawyer, A., & Milne, R. (2015). PEACE in fraud interviews: Assumptions and

detection of guilt and the impact on interviewer behaviour. *Crime Prevention And*

Community Safety, 17(1), 30-46. https://doi.org/10.1057/cpcs.2014.16

Shawyer, A., Milne, B., & Bull, R. (2013). Investigative interviewing in the UK. In

International developments in investigative interviewing (pp. 50-64). Willan.

Shultz, T. R. (1980). Development of the concept of intention. In *The Minnesota Symposium*

on child psychology (Vol. 13, pp. 131-164).

Smith, L., Bull, R., & Holliday, R. (2011). Understanding Juror Perceptions of

Forensic Evidence: Investigating the Impact of Case Context on Perceptions of

Forensic Evidence Strength. *Journal Of Forensic Sciences, 56*(2), 409-414.

https://doi.org/10.1111/j.1556-4029.2010.01671.x

Snook, B., Eastwood, J., Stinson, M., Tedeschini, J., & House, J. (2010). Reforming

Investigative Interviewing in Canada. *Canadian Journal Of Criminology And*

Criminal Justice, 52(2), 215-229. https://doi.org/10.3138/cjccj.52.2.215

Streiner, D. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and

Internal Consistency. *Journal Of Personality Assessment, 80*(1), 99-103.

https://doi.org/10.1207/s15327752jpa8001_18

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases.

Science, 185(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vedejová, D., & Čavojová, V. (2021). Confirmation bias in information search, interpretation,

and memory recall: evidence from reasoning about four controversial topics. *Thinking*

& Reasoning, 28(1), 1-28. https://doi.org/10.1080/13546783.2021.1891967

Villalobos, J., & Davis, D. (2016). Interrogation and the Minority Suspect: Pathways to True and False Confession. *Advances In Psychology And Law*, 1-41. https://doi.org/10.1007/978-3-319-29406-3_1

Weiher, L. (2020). *The impact of interview style on the development, maintenance, and transfer of rapport*. Lancaster University (United Kingdom).

Wixted, J., & Wells, G. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science In The Public Interest, 18*(1), 10-65. https://doi.org/10.1177/1529100616686966

Yale, R. (2013). Measuring Narrative Believability: Development and Validation of the Narrative Believability Scale (NBS-12). *Journal Of Communication, 63*(3), 578-599. https://doi.org/10.1111/jcom.12035

Appendix A

Example Vignette (Group A):

INSTRUCTIONS

Background information about the case and list of allegations

Imagine yourself to be a police officer at the police station in your city. Your supervisor asks you to interview the suspect involved in a new case. In the following text, your supervisor gives you some more information about the crime the suspect is accused of. This includes the evidence gathered against the suspect.

Alleged offence:

On 14/02/2022 the police arrested a woman named Mrs. Brown for dealing drugs. The woman was caught selling different types of drugs in the park of your town. The women dealt Opiates (Heroin, morphine), Hallucinogens (LSD), and Marijuana. She was arrested at 4.30pm by two police officers who were on street patrol in the park. The suspect is alleged to be her accomplice and therefore is also suspected of dealing drugs.

The evidence gathered against the suspect that may indicate they were implicated in the drug dealing offence of the 14/02/2022 is listed here:

- An old woman (suffering from schizophrenia) saw someone that looked like the suspect together with the Mrs. Brown in the park, 5 minutes before and after Mrs. Brown dealt the drugs
- A 12-year-old child (suffering from autism) saw a man that might have been the suspect driving together with Mrs. Brown to the crime scene, shortly before Mrs. Brown dealt the drugs
- There was one phone call from Mrs. Brown on the suspect´s phone on the day of the crime, though the content of this call is unknown
- Mrs. Brown says that she knows the suspect but refused to disclose the nature of their relationship or whether the suspect is directly involved in her drug dealing
- There were traces of marijuana found in the car of the suspect

Based on this evidence, your **supervisor is not sure whether the suspect is guilty** of drug dealing.

Your task is to question the suspect, who will be played by another participant. To help you, a script has been provided which gives you the questions you should put to the suspect.

You can assume that the introduction part of the interview, where you introduce yourself to the suspect and explain the legal rights to the suspect, is already done. It has also been explained that he is being questioned because of his links to a woman who was arrested for dealing drugs. Now you are only collecting the suspect´s version of events. This means you can directly ask the questions we have provided without having to introduce yourself.

Please **read these questions in order**, think carefully about the suspects responses, and afterwards we will ask you questions about your experience in the interview.

When you will be asked what letter you are, please indicate you are letter "A".

QUESTIONS

Interviewer: Can you please tell me your version of events?

Interviewee: …

Interviewer: Can you tell me what you did on the afternoon of the 14/02/2022?

Interviewee: …

Interviewer: Do you remember any of the clients you had that day?

Interviewee: ...

Interviewer: Do you remember anything else?

Interviewee: …

Interviewer: *An old woman saw that you were together with Mrs. Brown in the park, 5 minutes before and after she dealt drugs.* Can you explain why you were with Mrs. Brown if you do not know her?

Interviewee: …

Interviewer: *A 12-year-old child saw you together with Mrs. Brown in a car, driving to the crime scene, shortly before Mrs. Brown dealt the drugs.* Can you also explain this?

Interviewee: …

Interviewer: *We know that there was one phone call from Mrs. Brown on your phone.* Do you know why this is?

Interviewee: …

Interviewer: Earlier you said that you don't know Mrs. Brown. *However, she said she knows you, but refused to disclose the nature of your relationship or whether you are directly involved in her drug dealing.* Can you explain this?

Interviewee: …

Interviewer: *Why do you think that there were traces of Marijuana found in your car?*

Interviewee: ...

Interviewer: Okay, that is the end of the interview.

Appendix B

Debrief within Qualtrics

Q88

Thank you for your participation! After this study you might ask yourself what will happen with the data. Our study will help us to understand how **presumptions of guilt** affect how interviewers interpret explanations provided by suspects. In our study all the suspects that are interviewed are innocent and provide identical accounts to the interviewer. However, we manipulated **prior guilt assumptions** and the **apparent strength of the evidence** against the suspect to determine whether these made the accounts provided by suspects less likely to be believed and whether this might affect investigative decision making such as whether to continue investigating the suspect. We did this by changing the information you received prior to the interview taking place.

We also told you that the suspect was another participant. In truth, they were part of the research team. We apologise for this deception, but it was necessary in order to have participants focus on the narrative provided by the suspect.

In case you would like to know more about the study, the theoretical background or the study findings, feel free to contact any of the researchers. If you have questions or concerns about the study, please do not hesitate to reach out to the researchers.

If you have enjoyed taking part please feel free to share our contact details with your friends, however to maintain the integrity of the study, please do not share with anyone the information in this form about the specific ideas we are testing or how we test them.

Luca Marie Hülscher; l.m.hulscher@student.utwente.nl
Jonah Sauer; j.a.sauer@student.utwente.nl
Project Supervisor: Dr. Steven Watson; s.j.watson@student.utwente.nl

Appendix C

Diagnostic Output for Guilt Judgements
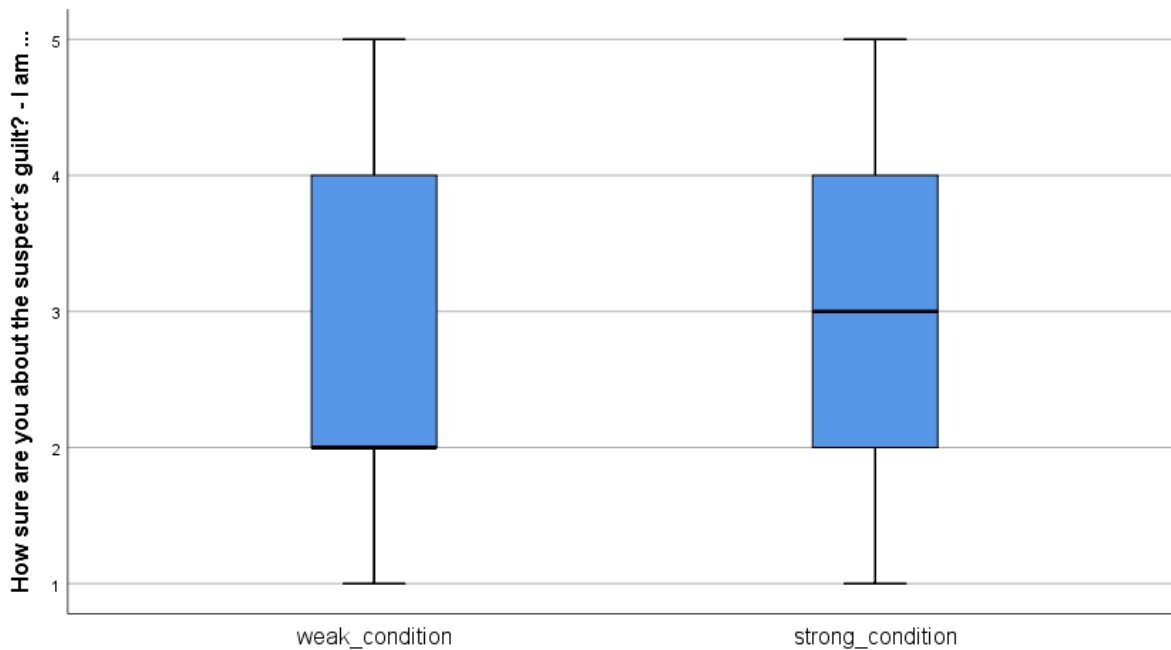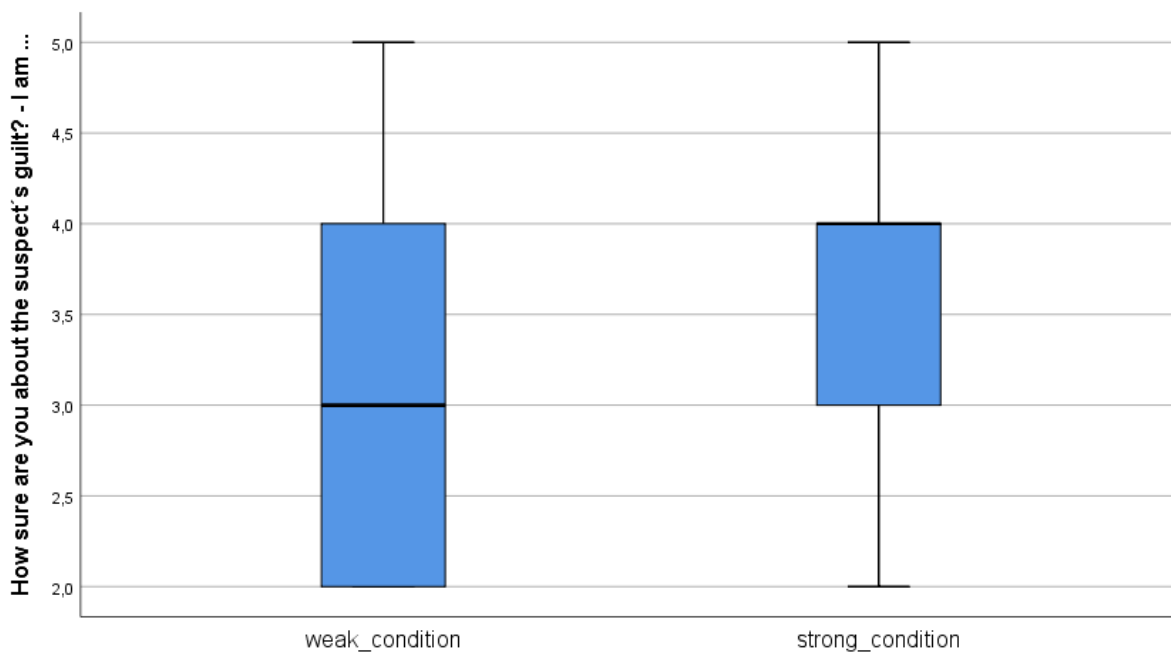
Appendix C1. Boxplots for Guilt Judgements (pre- and post interview)

Levene's Test of Equality of Error Variances

	Levene Statistic	df1	df2	
Based on Mean	2,100	1	65	,152
Based on Median	2,289	1	65	,135

How sure are you about the suspect´s guilt? - I am ...	Based on Median and with adjusted df	2,289	1	64,124	,135
	Based on trimmed mean	2,285	1	65	,135

How sure are you about the suspect's guilt? - I am ...	Based on Mean	,043	1	65	,836
	Based on Median	,025	1	65	,876
	Based on Median and with adjusted df	,025	1	59,076	,876
	Based on trimmed mean	,053	1	65	,818

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept + Evidence_Condition

Within Subjects Design: Time

Appendix C2. Levene`s Test for Equality of Error Variances for Guilt Judgements

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Evidence_Condition	Statistic	df	Sig.	Statistic	df	S
How sure are you about the suspect´s guilt? - I am ...	weak_condition	,226	33	,000	,864	33	
	strong_condition	,286	34	,000	,854	34	
How sure are you about the suspect´s guilt? - I am ...	weak_condition	,246	33	,000	,897	33	
	strong_condition	,202	34	,001	,901	34	

a. Lilliefors Significance Correction

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Evidence_Condition	Statistic	df	Sig.	Statistic	df	S
How sure are you about the suspect´s guilt? - I am ...	weak_condition	,226	33	,000	,864	33	
	strong_condition	,286	34	,000	,854	34	
How sure are you about the suspect´s guilt? - I am ...	weak_condition	,246	33	,000	,897	33	
	strong_condition	,202	34	,001	,901	34	

a. Lilliefors Significance Correction

Appendix C3. Shapiro-Wilk test of normality for Guilt Judgements
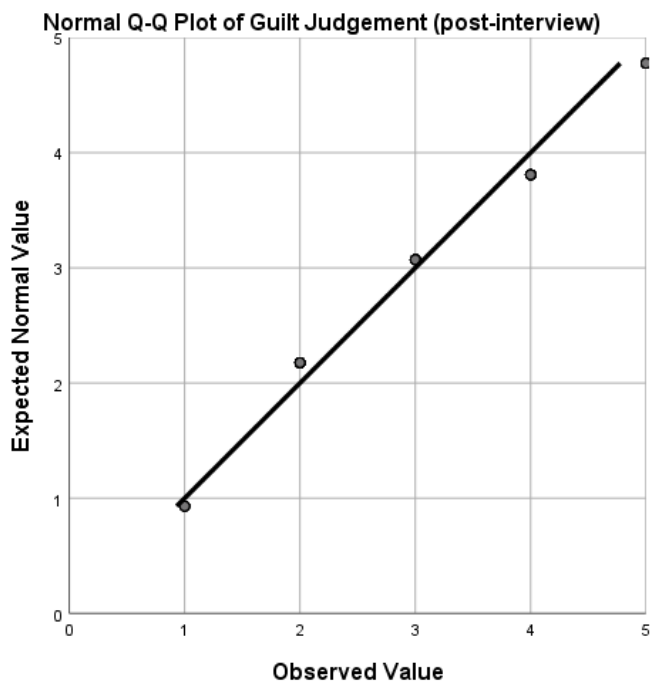
Appendix C4. Q-Q Plot for Guilt Judgement (post-interview)
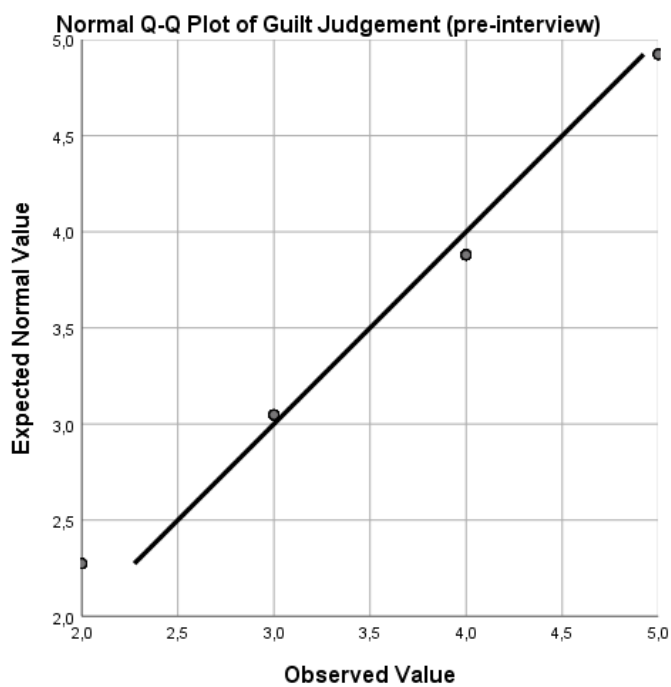
Appendix C5. Q-Q Plot for Guilt Judgement (pre-interview)

Box's Test of Equality of
Covariance Matrices^a

Box's M	7,894
F	2,544
df1	3
df2	783759,150
Sig.	,054

Tests the null hypothesis
that the observed
covariance matrices of the
dependent variables are
equal across groups.
a. Design: Intercept +
Evidence_Condition
 Within Subjects Design:
Time

Appendix C6. Box`s Test of Equality of Covariance Matrices for mixed ANOVA

Appendix D

Complete Original Text of Diagnostic Tests (written before changing Guilt Judgement

variable from binary + scale to only scale)

Guilt Judgments

Boxplots showed five outliers, assessed via looking for datapoints that were more than

1.5 box-lengths away from the box (see Appendix C). However, these outliers were unusual

data but no measurement or data entry errors, therefore still included in the analysis. The

residuals of Guilt Judgement (pre-interview) and Guilt Judgement (post-interview) were

normally distributed as seen in the Q-Q-Plots (Appendix C). Levene`s test of homogeneity of

variances indicated that there was heterogeneity of variances in the Guilt Judgment (pre-

interview) ($p = .016$) and homogeneity of variances in Guilt Judgment (post-interview) ($p =$

.157) (see Appendix C). However, since sample sizes are nearly the same in the *Weak* ($N =$

33) and the *Strong* ($N = 34$) conditions, making it a balanced sample (Tillé, 2011), it can be

assumed that ANOVA is robust to this violation. Box`s test of equality of covariance matrices

indicated homogeneity of covariances ($p = .538$).

Plausibility of the Suspect's Story

The Boxplots (Appendix E) show no outliers and the data was normally distributed, as

shown by the Shapiro-Wilk`s test ($p > .05$). Further, there was homogeneity of variance ($p =$

.581).

Intention to further Investigate

There were 3 outliers in the No Guilt Presumption condition detected in the boxplots

(Appendix F). These outliers were unusual data, therefore kept in the dataset. The data

residuals showed normal distribution, as can be seen in the Q-Q-Plots (Appendix F). Further,

Levene`s test displayed homogeneity of variance ($p = .69$).

Appendix E

Diagnostic Test Output for Plausibility of the Suspect's Story
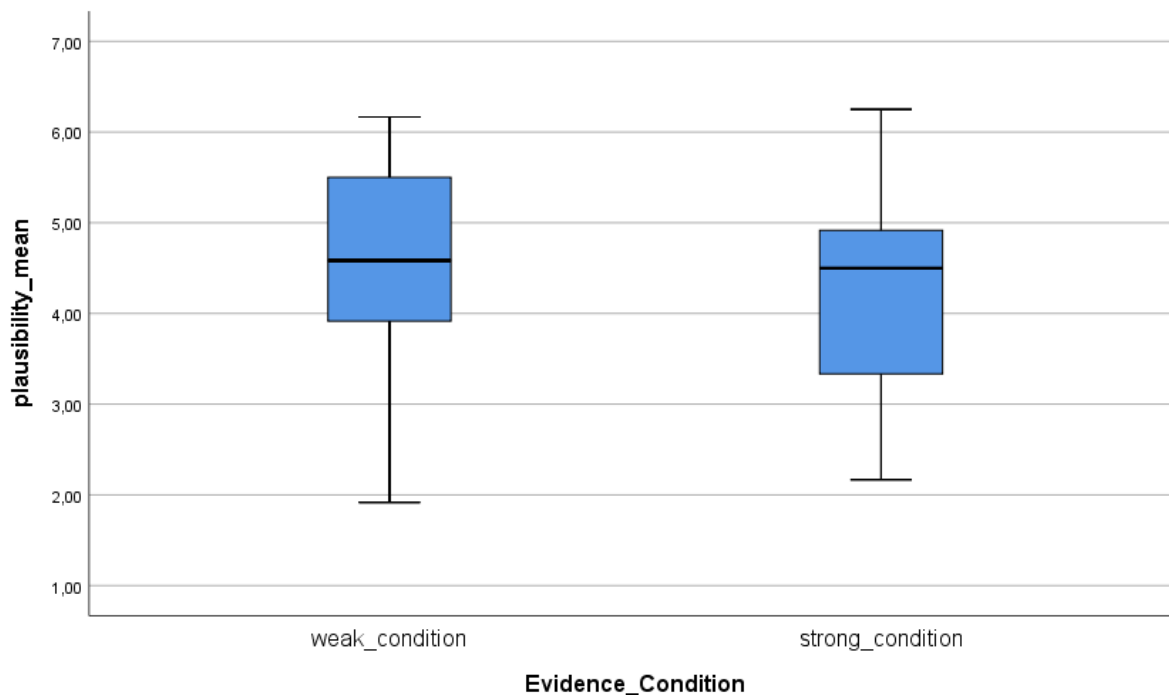
Appendix F1. Boxplots for Plausibility of the Suspect's Story and Evidence-Strength
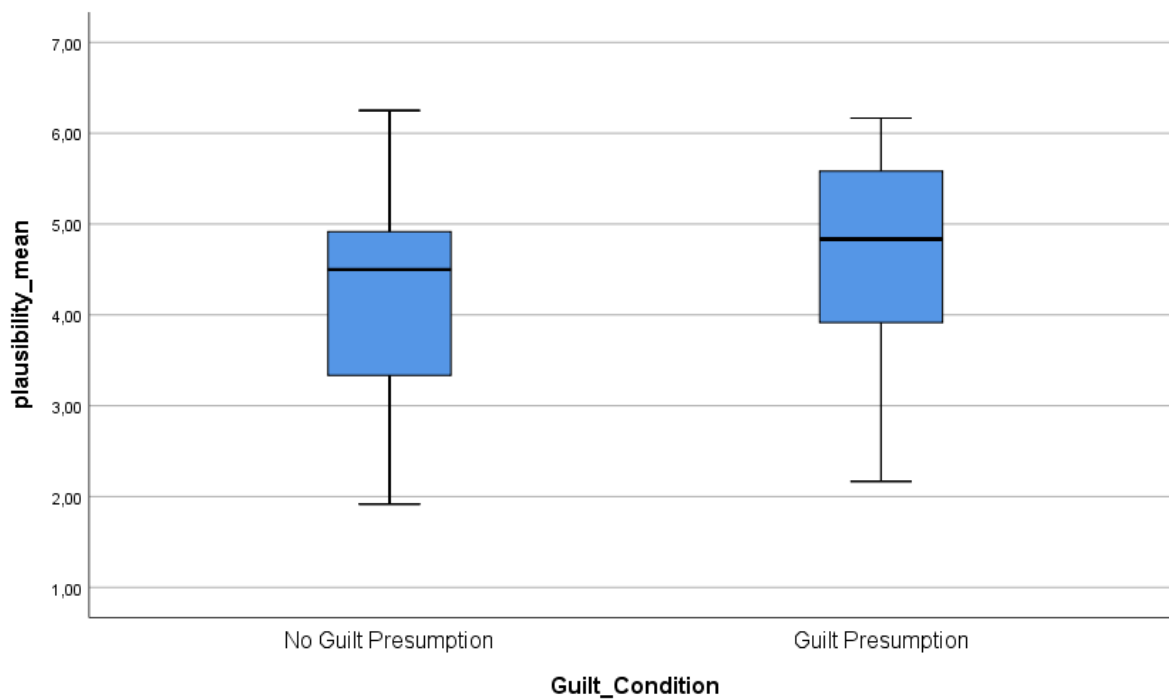
Appendix F2. Boxplots for Plausibility of the Suspect's Story and Guilt Presumption

Tests of Normality

	Evidence_Condition	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
plausibility_mean	weak_condition	,127	33	,190	,940	33	,067
	strong_condition	,127	33	,190	,962	33	,299

a. Lilliefors Significance Correction

Appendix F3. Shapiro-Wilk Test of Normality for Plausibility of the Suspect's Story

Test of Homogeneity of Variance

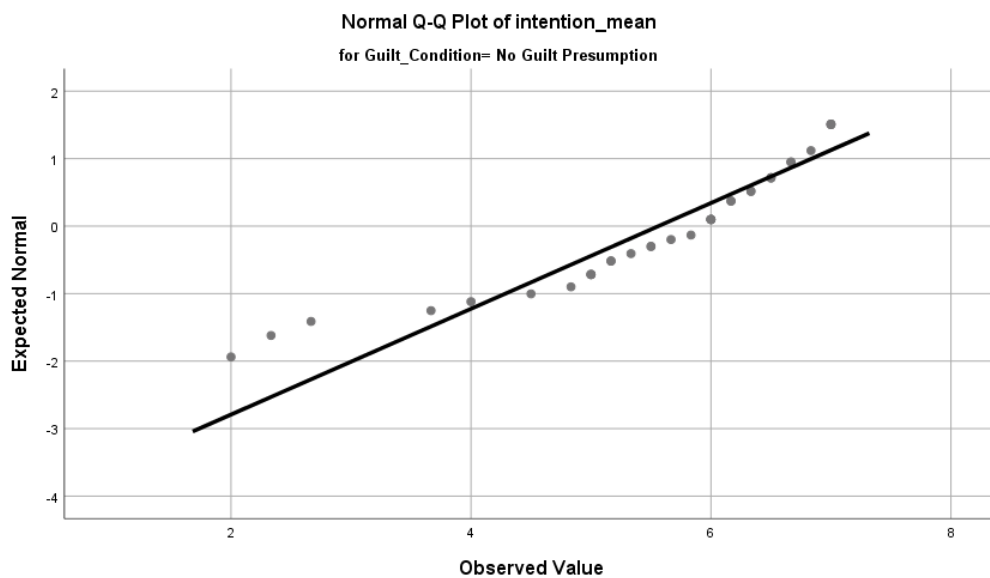|		Levene Statistic	df1	df2	Sig.
plausibility_mean	Based on Mean	,307	1	64	,581
	Based on Median	,106	1	64	,746
	Based on Median and with adjusted df	,106	1	62,847	,746
	Based on trimmed mean	,267	1	64	,607

Appendix F4. Test of Homogeneity of Variance for Plausibility of the Suspect's Story

Appendix F

Diagnostic output for Intention to further Investigate
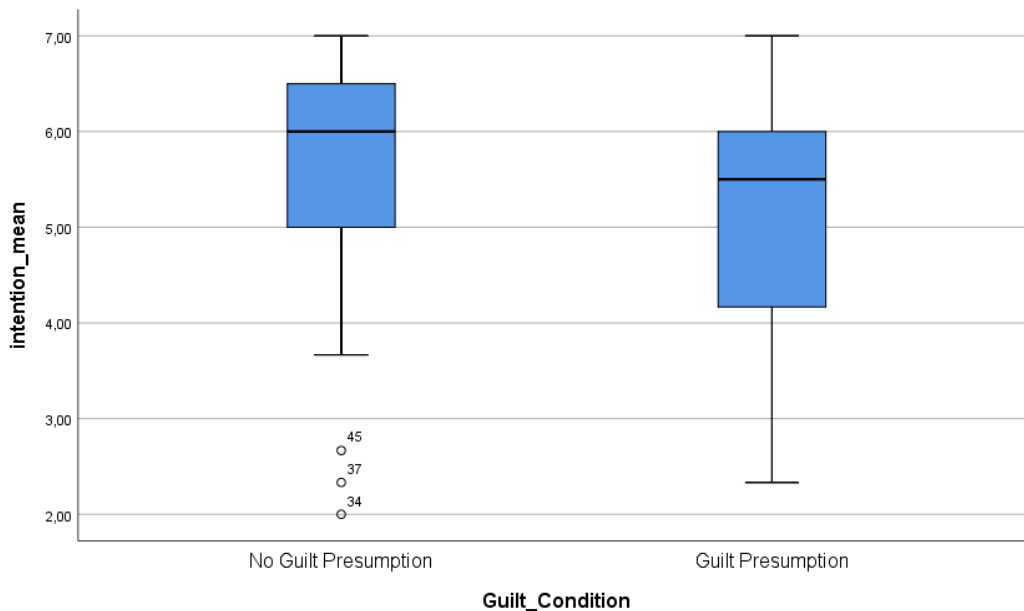
Appendix E1. Boxplots Intention to further Investigate

Appendix E2. Q-Q Plot for Intention to further Investigate

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
intention_mean	Based on Mean	,495	3	62	,687
	Based on Median	,472	3	62	,703
	Based on Median and with adjusted df	,472	3	58,519	,703
	Based on trimmed mean	,498	3	62	,685

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: intention_mean

b. Design: Intercept + Evidence_Condition + Guilt_Condition + Evidence_Condition * Guilt_Condition

Appendix E3. Levene`s Test for Equality of Error Variances for Intention to further Investigate

Appendix G

Frequency of very high values on *Plausibility of the Suspect's Story* according to *Evidence-Strength* condition

*Hohe_plausibility * Evidence_Condition Crosstabulation*

			Evidence_Condition		Total
			weak_condition	strong_condition	
Hohe_plausibility	1,00	Count	23	20	43
		% of Total	53,5%	46,5%	100,0%
Total		Count	23	20	43
		% of Total	53,5%	46,5%	100,0%