Impact of Interviewer Assumptions on the Outcomes of Investigative Interviews

Britt Hudepohl (s2617315)

Department of Psychology
Conflict, Risk, and Safety
June 30, 2023
S.J. Watson
P.W. De Vries

Abstract

Police services rely heavily on information collected through interviewing suspects. Interviewer perceptions are found to not always be as objective as they should be. Factors which were found to influence those perceptions are guilt presumptions, which influence investigative decisions that are made. This could cause miscarriages of justice or could lead to false convictions. The purpose of this study is to show the influence of guilt presumptions on guilt judgements, the intention to further investigate the suspect, the perceived plausibility of the suspect's story, the purpose of an investigative interview and the confidence in one's own ability to determine guilt. Furthermore, the differences in these variables between experts (police officers) and non-experts (lay people) will also be investigated. The experimental design was a between-subjects design with the independent variables guilt presumptions (guilt presumption versus innocence presumption) and experts (experts versus non-experts). The design required participants to conduct a face-to-face interview with a suspect, together with filling in a pre- and post-questionnaire. The sample consisted of 35 participants (non-experts) and 16 police officers (experts). The manipulation of the guilt presumption was not as strong as hoped, especially in manipulating innocence beliefs. However, it is of importance to highlight the finding that two third of the participants in the non-expert group assumed guilt, while almost no participant in the expert group assumed guilt when the interview was finished. Another interesting finding was that the variable of the perceived plausibility of the suspect's story had a significant negative correlation with the pre- and post-interview measures on guilt judgement. This indicates that participants who believed the suspect was guilty perceived the story the suspect told as less plausible, even though they were given a plausible explanation by the suspect. Perceived plausibility of the suspect's story was also correlated with the variable of confidence and the intention to further investigate. Future research should focus on enlarging the sample size of the experts as well as non-experts, so the results can be considered as more reliable. Also, involving police detectives in setting up the experiment, and with that in writing and formatting the interview script, would increase the ecological validity.

Introduction

Miscarriages of justice are present not only today, but, as can be seen in the case of the murdered gas station attendant in the year 1985, in all times. The cheerful gas station attendant Micelle Mooij was found dead next to her car in a Dutch village called Warnsveld. The murder seemed easy to resolve for the detectives, since it took place in a busy street with many potential witnesses and the time was known. However, the murderer was not found, and the case initially remained unsolved (De Jong, 2023). Due to tips from informants, the police reopened the investigation in 2001. After 17 years, four men got arrested for the murder on Micelle Mooij due to a secret tip that was given and they were sentenced to prison for 6 to 8 years. The police detectives started from the idea: "these persons did it" and tried to come up with everything that had something to do with it, thereby ignoring the evidence which debunked it and only searched for confirmation (Cuijpers, 2023). The question remained if these four men were indeed the true murderers. Therefore, documentary makers shed a light on this remarkable case by watching several tapes and using the available material. When investigating this case, the makers of the documentary noticed that during the interrogation of the four men, dubious methods had been used. Examples of these methods are that police detectives invented incorrect information which was presented towards the suspects and that they presented them an attractable deal: if they confessed, a short prison sentence waited for them, whereas not confessing could lead to much more punishment. In the end, the police got a confession (De Jong, 2023). The chance that the four men actually did it is zero percent, as the recorded tapes contain enough exculpatory information (Cuijpers, 2023). Estimates of false confessions vary a lot, from almost 0 to 150 for every 1000 people confessing, but they certainly occur, and every false confession is one too many (Horselenberg & Smeets, 2005). In the worst case, it could lead to a wrongful conviction. A phenomenon which often plays a role at such miscarriages of justice like the one in Warsnveld, are confirmation bias and tunnel vision, which have been suggested to contribute to the wrongful conviction issue (Elaad, 2022). Therefore, the influence of guilt presumptions on the outcomes of an investigative interview with an innocent suspect will be examined in this study.

Tunnel Vision

Tunnel vision can be described as 'the tendency of individuals in the criminal justice system to use short-cuts and heuristics to filter evidence in a selective way to create a case for the conviction of a suspect, while ignoring evidence which indicates the innocence of the suspect' (Elaad, 2022, p. 1). When placed in the criminal justice system, tunnel vision has been suggested to contribute to the false conviction issue. An example where tunnel vision

occurred is in the case of the murder on the pump attendant in Warnsveld. The impression from the conversation with the suspects was that the detectives already made their decision on the guilt of the suspect, by pursuing an accusatorial plan instead of a fact-finding one. They were biased and showed no interest in the denial of the suspect, worked deliberately towards eliciting a confession and only searched for evidence that confirmed the guilt of the suspects (Elaad, 2022).

In conducting a successful investigative interview, interviewers should not allow their personal beliefs affecting the manner in which they deal with suspects, and they should look for useful information from anyone involved, not only the suspect (Gudjonsson, 1994). Tunnel vision is associated with different cognitive biases, such as hindsight bias, outcome bias and confirmation bias (Elaad, 2022). Confirmation bias is the tendency of people to process information and to interpret and use this information in a way that is consistent with their own existing beliefs (Casad & Luebering, 2023). According to a study from O'Brien (2009), students that were asked to investigate a fictional crime and were asked to name their suspect early in the investigation were more likely to seek evidence, compared to those that did not got asked to name their suspect. More specifically, students who named their suspect early in the investigation remembered the given facts as more conforming with their suspect, defended more evidence that was focused on him, asked for more reports focused on him, and moved their opinions slowly about subjects that were relevant to determining guilt in a manner that was supportive of their initial suspicions (O'Brien, 2009).

Considering police detectives, Adams-Quackenbush et al. (2019) found that situational factors, experience, and pre-existing beliefs are often the basis of police officers' decision making, leading to confirmatory thinking. These pre-existing beliefs can interfere with the objectivity of officers, as well as guiding the creation of a scenario when the belief becomes the most logical explanation for the crime. Another source of confirmatory thinking may be the procedural aspect of criminal investigations. The manner in which police officers collect information and evidence is often given as a reason for a higher susceptibility to confirmatory behaviour. Information about criminal activities is overall collected in a sequential manner. This sequential way of collecting information increases confirmation bias, as there is a stronger preference for information which supports earlier theories. Police officers then evaluate the information and make a decision of who should be questioned and in what place they need to look for more information and evidence. When information is discovered that is seen as supportive for a favoured scenario, a confirmatory cycle may be started, resulting in a high commitment to decisions made based on that information. Researchers have also

discovered that officers who shape theories in an early stage of an investigation are more likely to count theory-disconfirming information as less reliable (Adams-Quackenbush et al., 2019).

As described, prior information about a case can influence how individuals evaluate the evidence in it. According to a study conducted by Adams-Quackenbush et al. (2019), interviewers which assume guilt ask low quality and biased questions. The way in which those first pieces of information are judged, be they biased or objectively, may be of great importance in how further information is treated (Adams-Quackenbush et al., 2019). Adams-Quackenbush et al. (2020) showed that guilt assumptions have an influence on the behaviour of the interviewer. The most common finding within the literature mentions that interviewers are more likely to use guilt-presumptive language when they have a belief about guilt of the suspect. So, only naming a suspect and developing ideas for suspicion already worsened the bias (O'Brien, 2009). The study of Adams-Quackenbush et al. (2020) showed that one of the consequences of this bias could be that it can lead to a decreased cooperation of the suspect, as well as a harmful effect on the ability of the interviewer to collect useful information (Adams-Quackenbush et al., 2020). This can be seen in the case of the murder on the gas station attendant, where only confirmatory evidence was gathered based on tips from informants, eventually leading to harmful consequences in the form of 6 to 8 years of prison for the four men. According to Holmberg (2004), the manner in which a suspect evaluates the interviewer affects their response, which could in turn become favourable or unfavourable. This may affect the willingness of a suspect in the process of building a rapport (Holmberg, 2004). Looking at police interviews, Yuille et al. (1999) emphasized that these interviews should be done in a way that minimises any negative impact on the suspect, emotionally as well as personally. This helps in building rapport, which eventually opens or closes doors for further investigation by the police (Yuille et al., 1999).

Police Officers' Assumptions and Criminal Investigations

Police officers have the stressful task of collecting information from several sources, for example from witnesses, victims, and suspects. After gathering this information, it needs to be put into context of potential evidence, followed by verifying and gathering more information. When making decisions in a fast pace in high stress situations, individuals quickly come to their conclusions without using all the evidence which is available to them. In the case of a police officer, a lot of pressure to solve a case can have an influence on their decision-making. This increased pressure can also have an effect on the officer's attempt to put the evidence together in order to create a scenario of how a crime was committed, the

reasons for it and who may have been involved, which could bring about a tunnel vision effect (Adams-Quackenbush et al., 2019).

Questioning Techniques

According to Oxburgh et al. (2010), good questioning is essential in investigative interviewing, thus preventing the detrimental effects of guilt assumptions (Oxburgh et al., 2010). The way in which police officers use good questioning techniques is often different in reality, as they may still hold guilt assumptions and use poor questioning. Much of that attention was due to highly publicised cases in which there was police misconduct or dubious interview practices, such as the murder on the gas station attendant in Warnsveld. Officers use several techniques, such as the PEACE model, that has influenced training around the world and Europe in particular, to avoid guilt presumptive interviewing (Adams-Quackenbush et al, 2019). The PEACE model has several aims, such as obtaining accurate and reliable accounts from all witnesses, suspects and victims that are relevant to an investigation, fairly acting of investigators when questioning interviewees and having an investigative mindset during the investigative interview (Clarke & Becky, 2001). More internationally, the Méndez principles (2021) are also relevant, which has several investigative interviewing principles that are almost equal to the ones mentioned within the PEACE model. One way of enacting these Méndez principles is by ensuring fair and non-guilt presumptive questioning (Méndez, 2021).

As already mentioned, types of questions asked also has as influence on the interview process. Therefore, the PEACE model emphasizes the use of open questions rather than closed, as it elicits more free recall. Open questions produce more accurate and elaborate answers. It also ensures that the chance of imposing the interviewer's own view on the person is minimised (Davison, 2009). Additionally, an investigative interview is designed to construct a rapport in order to prompt a suspect to give information that is valuable. Enforcing a confession is not necessary, obtaining reliable and accurate information to search for the truth is of greater importance. The purpose of investigative interviewing is thus to understand the true narrative of an event (Wakefield & Fleming, 2009). The current evidence base supports non-accusatory interviews, while the Dutch police does not do this at the moment.

Dutch Police Interrogation Techniques

Looking at the Dutch police force, every detective in The Netherlands is expected to be able to apply the standard interrogation strategy. This strategy is based on theories regarding behaviour and influencing behaviour (Hiemstra, 2012). The assumptions of this standard interrogation strategy are that there must be a good working relationship between the interrogator and the suspect, a confrontation with contradictive statements needs to be made

towards the suspect, a sufficient person-oriented attitude of the interrogator must be present and the behaviour of the suspect needs to be rewarded. These are the principles of today's standard interrogation strategy in The Netherlands. The purpose of this strategy is to minimise resistance of the suspect in telling the truth (Hiemstra, 2012). Besides this standard interrogation strategy, several other tactics, such as suggestive interrogation and propping, are used to induce a suspect to speak. According to Hiemstra (2012), these tactics could possibly lead to a coerced false confession, as suspects could confess because they fall for the psychological tricks of the police. During a suggestive interrogation, the interrogator suggests the answer. These interrogations are often characterised by using closed questions. In the case of propping, the police try to give the impression that they know exactly how the suspect has committed the crime. By creating this impression, the police hopes that the suspect will confess (Hiemstra, 2012). These disturbing findings reinforce the need for good questioning techniques, as with these strategies a coerced false confession is lurking.

Since most prior literature has shown that accusatory interviewing and guilt presumptions are bad, at least in part because of the tendency to use leading closed questions, there is still a need to test the effect of guilt presumptions when high quality questions are asked, and a suspect provides a coherent and plausible alternative for the evidence. Question types that use fair and non-guilt presumptive questioning when people assume guilt were not denounced yet in prior research. Therefore, using scripted and high-quality questions in investigative interviewing are of importance since it is essential to know if these are sufficient to remove the known issues of confirmation bias.

Purpose of this study

The goal of the present study is to determine the effect of guilt assumptions on investigative decision-making within an investigative interview. This was done through using scripted interviews where participants role-played police interviewers. Additionally, differences between police officers and lay people on guilt assumptions in the context of an investigative interview will also be tested, in order to determine if lay people are a sufficient stand in for police officers. However, there are good reasons why police officers should be able to make more expert judgements than lay people, since they are more experienced in assessing whether a suspect's testimony changes how compelling the case against them is. Research has already been done on the topic of guilt assumptions and investigative interviewing, but the uniqueness of this study is that the outcomes of police detectives and lay people will be compared with each other, to see if there exist any differences between these two groups when interviewing an innocent suspect. Additionally, the use of a suspect that was

innocent in each interview conducted also adds new insights into the already existing research. The research question in this study is: *"What is the influence of guilt presumptions on the outcomes of investigative interviews with an innocent suspect?"*.

Hypotheses

1. *"Non-experts who presume guilt will continue to presume guilt even after an interview with an innocent person"*
2. *"Experts will be less susceptible to presume guilt than non-experts"*
3. *"Experts will be more likely to change their guilt judgements after the interview"*
4. *"Guilt presumptions before the interview will affect guilt judgements post-interview"*

Methods

Design

The present study consists of a between-subjects design with the independent variables guilt assumption (guilt presumption versus innocence presumption) and experts (experts versus non-experts). The dependent variables tested were guilt judgement, purpose, which measures the thoughts of participants on the purpose of a suspect interview, confidence cues, which measures the amount of confidence participants have in their abilities to detect whether a suspect is guilty or not, the intention to further investigate and plausibility of the suspect's story. The design requires participants to interview confederates, in the guise of fellow participants. All questionnaires used in this study can be found in Appendix C, also the questionnaires of deception detection and guilt and investigative decisions which were not used.

Participants

Participants (non-experts) were recruited through SONA ($n = 6$), which is a platform where researchers can find participants from the university, and through the social environment of the researchers ($n = 29$). Those gathered through SONA were given one credit for their participation, which is needed to complete their programmes. Participants recruited through the social environment of the researchers did not receive any rewards. Through contacting the educational staff at the Police Academy in Apeldoorn, the researcher gathered 16 police officers (experts). The police officers were all detectives and recruited during their training, but before their training on the subject of tunnel vision. Afterwards they were given a flyer with information about the subject of this study, a short description of the murder at the pump attendant at Warsnveld, and other additional information. This flyer and an additional presentation about the results of this study was given at the Police Academy to provide the detectives with information about the study they participated in and as an incentive for taking

part. Inclusion criteria for the non-experts for taking part in this study were that individuals were capable of speaking and understanding English and at least 18 years old. For the police officers, all materials were in Dutch and the inclusion criteria were that all officers were capable of speaking and understanding Dutch and that they were at least 18 years old.

The sample size at the university was 35 participants. The majority of the sample consisted of Dutch ($n = 18$), followed by German ($n = 8$), then followed by participants from other European countries ($n = 6$) and non-European countries ($n = 3$). The sample consisted of 19 females and 16 males, with 20 participants in the age of 18 to 25, 7 participants ranging from 26 to 39, 6 participants in the age of 40 to 59 and 2 participants were above the age of 60. Each participant was allocated to one of the two conditions. Participants in condition A, the guilt presumption condition, ($n = 19$) had a sentence in the case description which said, "your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually *committed* the crime". In condition B ($n = 16$), the innocence presumption condition, participants received the sentence "your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually *did not commit* the crime".

The sample size at the Police Academy consisted of 16 police officers. The majority of the sample was of Dutch origin ($n = 14$), followed by one person originating from the Dutch Antilles, and another person with a Turkish origin. The sample consisted of 8 females and 8 males, with an average age of 36.25 years ($SD = 5.67$), ranging from 28 to 49 years. The average of experience of the participants within the police was 15.8 years (years of education at the Police Academy included), ranging from 9 to 26 years of experience. Each participant was allocated to one of the two conditions: guilt presumption ($n = 8$) and innocence presumption ($n = 8$).

Materials and Procedure

Information Sheet

First, participants received an information sheet where their right to withdraw, potential risks, anonymity and confidentiality were stated, as well as that the research had been approved by the BMS Ethics Committee, which could be found by the reference number 230575. The participants were told that they would play the role of a police-interviewer and needed to interview a suspect, played by another participant. But in reality, they interviewed a researcher from the study. This in order to prevent potential biases to happen, for example social desirability biases, and to have them focused on the suspects behaviour, since prior relationships could have an influence (Weiher, 2020). After reading all information,

participants filled in an informed consent form. Questionnaires that accompany the interview session were also announced.

Case Description

The case description (Appendix B) prepared participants for the session, including information on the scenario, the interview procedure, the alleged crime, and the interview script itself with the instruction to stick to the interview script. The case was described as follows. On 30/03/2023 the police arrested a woman named Mrs Brown for dealing drugs. The woman was caught selling different types of drugs in the park. The suspect who will be interviewed by the participant is alleged to be Mrs Brown's accomplice and is therefore also suspected of dealing drugs.

The case description included five pieces of weak evidence information. The evidence is weak since research found that children witnesses are perceived as less reliable by jurors compared to adult witnesses (Pozzulo & Dempsey, 2009). Reavey et al. (2016) found the same for people suffering from mental health conditions. Based on this, statements of a child and a person having a mental health condition (weak sources that might have mistaken the suspect for someone else) were included. This weak evidence is beneficial for this study, as a good investigator would have thought more quickly that this evidence is not enough for a conviction.

Table 1

Pieces of evidence

Evidence

An old woman with schizophrenia saw someone that looked like the suspect together with Mrs Brown in the park, 5 minutes before and after Mrs Brown dealt the drugs.

A 12-year-old child 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.

Guilt Presumptions

The manipulation of guilt presumptions was also included in the case description. At the end of the case description, a sentence stated, "your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually *committed* the crime" (guilt presumption condition) or "your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually *did not commit* the crime" (innocence presumption condition) (Kassin et al., 2003). Adding a percentage to the statement helps to reinforce the supervisor's statement, since the study of Kassin et al. (2003) successfully used this in their manipulation. Additionally, the questions in all conditions were identical, only the guilt presumptions differed.

Pre-Questionnaire for Participants

Qualtrics was used to collect the measures. After the participants read the case description, they received a question about their guilt judgement. The question stated, "*Based on the information I have about the current case, I believe the suspect is…*", with a scale with answers ranging from 1 = I am very sure the suspect is innocent to 6 = I am very sure the suspect is guilty. This scale has no midpoint, as research conducted by Chyung (2020) showed that participants may use a midpoint as a dumping ground, especially in the case where they do not know enough about the questions asked in the questionnaire. This forces participants to make a choice between guilty or not guilty (Chyung, 2020). This question on guilt judgement served as a manipulation check, to see if the guilt manipulation successfully made people presume guilt or innocence. After they filled in this question, the interview started.

Interview

After reading all information and answering the first question about guilt judgement, the interview between the participant and the suspect started. Before the interview, participants were asked again if there were any uncertainties about the coming interview and got the chance to ask questions too. When the participant communicated that everything was clear to them, the other researcher, who role-played the suspect, entered the room. As mentioned before, it was ensured that the participant did not know the researcher role-playing the suspect, thinking the suspect was another participant, thereby deceiving the participant.

It was told to the participants that the introduction part of the interview, where the interviewer and suspect introduce themselves and where the legal rights are explained to the suspect, has already been done. It has also been explained that the suspect was being questioned because of the suspect's link to a woman who was arrested for dealing drugs. Now

the interviewer was only collecting the suspects version of the events. This means that the participant directly started to ask the questions, without introduction. When the suspect and the participant faced each other, the participant began to start asking questions. All questions and answers were scripted and pre-defined, as this ensured comparability between the two conditions (guilt presumption versus innocence presumption) through isolating the effect of the independent variables (Appendix C). In this way, the control has been kept over the quality of the questions and over the responses of the suspect. It also ensured that every participant experienced the interview and with that the suspects utterances equally. The questions asked to the suspect included the evidence that was stated in Table 1, for example: "*An old woman saw you were together with Mrs Brown in the park, 5 minutes before and after she dealt drugs. So can you explain why you were with Mrs Brown if you do not know her?*". The evidence in the interview script was stated in *italics* to help the interviewer see where the evidence was introduced. Answers of the suspect offered a credible counter narrative for the evidence, for example: "*Mhh... so ... sometimes during my break, which is usually at around 4 pm, I go in the park and walk around a bit. You can ask my colleagues, I have done that since I started working at the taxi company, even when it is a busy day. I am sorry, I forgot to mention that earlier. Sometimes, I feel like talking with other people in the park, so I just ask them how their day was. Most of them are quite friendly, so I have a short chat with them while walking around in the park. Maybe this Mrs Brown could have been one of the people I talked to*".

At the moment the interview finished, the researcher asked the participant to fill in the post-questionnaire and told the participant that the researcher would stay in the room in case of any unclarities or questions. To keep the deception intact of the suspect being a participant too, the researcher gave the suspect as well as the participant both a post-questionnaire. The confederate filled out a different questionnaire than the participant, measuring rapport and trust.

Post-Questionnaire for Participants

Guilt Judgements. After the interview was conducted, participants directly began filling in the post-questionnaire. The first question in the post-questionnaire was the exact same question as the one asked in the pre-interview questionnaire and also stated, "*Based on the information I have about the current case, I believe the suspect is...*", with a scale with answers ranging from 1 = I am very sure the suspect is innocent to 6 = I am very sure the suspect is guilty. Asking the same question twice was done to discover any differences in guilt judgement due to the interview.

Confidence and Purpose Cues. The scale of *Confidence Cues* consisted of three items, with answers given on a 5-point Likert Scale, ranging from 1 = strongly disagree to 5 = strongly agree. An example item is "*How confident are you in your ability to detect whether a suspect is guilty or innocent after an interview*". The purpose of the scale is to measure the amount of confidence the participants have in general in their own abilities to detect whether a suspect is guilty or not.

The scale of *Purpose Cues* consisted of five statements and could be answered on a 6-point Likert Scale, ranging from 1 = never the purpose to 6 = always the purpose. The aim of the scale is to measure the extent to which the participant felt the statements reflected the purpose of a suspect interview in their experience. Thus, whether the purpose is about getting a confession or about gathering information. Example statements are "*To gather new evidence*" and "*To get a confession*". Accusatory questions, such as the latter example statement, have been reversed. A high score reflects good alignment with investigative interviewing, whereas a low score means a tendency to endorse accusatory approaches.

Intention to Further Investigate. The scale of *Intention to Further Investigate* consisted of six items, with answers given on a 7-point Likert scale, ranging from 1 = strongly disagree to 7 = strongly agree. An example of a statement is "*If I was the investigating police officer, I would suggest my supervisor to continue the investigation on this suspect*". The overall purpose of the scale is to measure to what extent the interviewer would indicate that the suspect should be placed under further scrutiny. Overall, a high score on the whole scale indicates a higher intention for further investigation of the suspect.

Plausibility of the Suspect's Story. To measure the extent to which the suspect's story was plausible, the Narrative Believability Scale has been used (Yale, 2013). This scale consisted of 11 items, with four subscales (completeness, consistency, coverage, plausibility) and 7-point Likert scales with answers ranging from 1 = strongly disagree to 7 = strongly agree (Yale, 2013). An example item of the subscale "completeness" is "*It was easy to follow the story from beginning to end*", with "consistency" an example statement is "*All of the facts in the suspect´s narrative agreed with each other*". The item "*There were "no holes" in the suspect´s narrative*" measured the coverage and "*I believe the suspect´s narrative could be true*" is an example statement of the plausibility of the suspect's story. The four subscales will be analysed together.

Demographics and Immersion. At the end, participants answered questions about their age, gender, nationality, profession, highest completed level of education and if they have ever conducted or been a suspect in an investigative interview. These questions were

placed at the end of the questionnaire to prevent stereotype threat and/or priming to occur and to improve retention of the participants (Fernandez et al., 2016). After the demographic questions, they were asked about their own opinions and feelings about the interview and immersion. There were four statements, for example "*The interview felt real to me*", which could be answered using a 5-point Likert scale, ranging from 1 = strongly disagree to 5 = strongly agree. These questions were used as a check in general, to see whether they took the task seriously. After these items there was an open question where participants could say anything they wanted to tell the researchers, using an open text box. These remarks were used in the results section, to see if future research can be further improved.

Debrief

After the participants answered all questions, the debriefing was shown to them, with information about the real purpose of the study in which they participated. Inside the debrief sheet it was also stated that they had been deceived in being told at the start of the experiment that the suspect was another participant, since it actually was a person of the researchers' team. After the debrief, withdrawal of participation was offered again to the participant. No participant has withdrawn after the debrief was given. The debriefing sheet was included within the Qualtrics online form (Appendix D). When the debrief was finished, the experiment was ended and the police officers received the flyer. In total, the entire process had a mean of 37.95 minutes for participants in the non-expert group and 28 minutes for those in the expert group.

Data Analysis

In order to analyse the data, the program RStudio version 1.3.1073 was used. Since the outcome variable is ordinal and not normally distributed, a Mann-Whitney U test was conducted to compare the manipulation of the Guilt Presumption conditions, to test the hypotheses, and to discover any differences between experts and non-experts. A Wilcoxon's signed-rank test was used to compare the pre- and post-interview scores. Correlation tests and t-tests were used to analyse the differences in the other dependent variables (*Plausibility of the Suspect's Story, Intention to Further Investigate, Guilt Judgement and Purpose and Confidence Cues*).

Results

Descriptive Statistics

In table 2, a full description of the correlations and statistics for *Non-Experts* has been given. A quick paired t-test was conducted and showed no significant correlation between *Guilt Judgement* (*pre-interview*) ($M = 3.71$, $SD = 1.11$) and *Guilt Judgement* (*post-interview*)

($M = 3.60$, $SD = 0.99$), t (34) $= .75$, $p = .456$. The pre- and post-interview outcomes are ordinal, so a normality test was not conducted. The correlation between *Guilt Judgement* (*pre-interview*) and *Guilt Judgement* (*post-interview*) was assessed using Kendall's Tau coefficient, which yielded a value of т $= .59$, $p < 0.001$.

The variable *Intention to Further Investigate* ($M = 5.14$, $SD = 1.39$) had a high mean and was also significantly positively correlated with *Guilt Judgement* (*pre-interview*), $r = .41$, $p = .015$, as well as with *Guilt Judgement* (*post-interview*), $r = .59$, $p < 0.001$. *Plausibility of the Suspect's Story* has normally distributed values ($M = 3.87$, $SD = 1.26$) and a significant negative correlation with the *Guilt Judgement* (*post-interview*) condition, $r = -.61$, $p = < 0.001$ and *Guilt Judgement* (*pre-interview*), $r = -.35$, $p = .038$. The variable of *Purpose* ($M = 4.02$, $SD = 1.36$) is significantly positively correlated with *Guilt Judgement* (*pre-interview*), $r = .34$, $p = .046$, and significantly negatively correlated with *Guilt Judgement* (*post-interview*), $r = -.62$, $p < 0.001$.

Table 2

Descriptive Statistics of Non-Experts

	M	*SD*	Guilt Judgement (post-interview)	Confidence Cues	Purpose Cues	Intention to Further Investigate	Plausibility of the Story
Guilt Judgement[1] (pre-interview)	3.71	1.11	.59*	-.21	.34*	.41*	-.35*
Guilt Judgement[1] (post-interview)	3.60	0.99	-	.14	-.62*	.59*	-.61*
Confidence Cues[2]	2.79	1.05	-	-	-.17	.09	-.05
Purpose Cues[3]	4.02	1.36	-	-	-	.32	.13
Intention to Further Investigate[4]	5.36	1.24	-	-	-	-	-.33
Plausibility of the Suspect's Story[4]	3.87	1.26	-	-	-	-	-

*Note: * = p < .05, [1] = scale range 1 (innocent) – 6 (guilty), [2] = scale range 1 (no confidence) – 5 (very much confidence), [3] = scale range 1 (never the purpose) – 6 (always the purpose), ^^^^ = scale range 1 (strongly disagree) – 7 (strongly agree)*

In table 3, a full description of the correlations and statistics for *Experts* has been given. Also here, a quick paired t-test has been conducted, which showed a significant correlation between *Guilt Judgement* (*pre-interview*) ($M = 3.56$, $SD = 0.51$) and *Guilt Judgement* (*post-interview*) ($M = 3.06$, $SD = 0.68$), $t(15) = 2.45$, $p = .027$. The pre- and post-interview outcomes are ordinal, so a normality test was not conducted. The variable of *Plausibility of the Suspect's Story* was statistically significantly negatively correlated with *Confidence Cues*, $r = -.71$, $p = 0.002$, as well as with *Intention to Further Investigate*, $r = -.63$, $p = .009$. Kendall's Tau coefficient was used to assess the correlation between *Guilt Judgement* (*pre-interview*) and *Guilt Judgement* (*post-interview*), which yielded a value of τ = .02, $p = .945$.

Table 3

Descriptive Statistics of Experts

	M	*SD*	Guilt Judgement (post-interview)	Confidence Cues	Purpose Cues	Intention to Further Investigate	Plausibility of the Story
Guilt Judgement[1] (pre-interview)	3.56	0.51	.08	.23	.38	.17	-.35
Guilt Judgement[1] (post-interview)	3.06	0.68	-	.28	-.17	.27	-.24
Confidence Cues[2]	3.35	0.46	-	-	-.08	.49	-.71*
Purpose Cues[3]	3.65	0.65	-	-	-	.04	.14
Intention to Further Investigate[4]	4.95	0.97	-	-	-	-	-.63*
Plausibility of the Suspect's Story[4]	4.59	0.67	-	-	-	-	-

*Note: * = p < .05, [1] = scale range 1 (innocent) – 6 (guilty), [2] = scale range 1 (no confidence) – 5 (very much confidence), [3] = scale range 1 (never the purpose) – 6 (always the purpose), [4] = scale range 1 (strongly disagree) – 7 (strongly agree)*

In table 4, a description of *Guilt Judgements* of *Non-Experts* in both conditions has been given. The table shows that 23 participants judged the suspect as guilty before the interview took place. The majority (*n* = 15) were in the *Guilt Presumption* condition. Participants who declared the suspect innocent before the interview existed of two-thirds of *Innocence Presumption* participants. Looking at the *Guilt Judgements* of *Experts* in table 5, nine participants judged the suspect guilty, whereof 5 in the *Guilt Presumption* condition. Considering the post-interview scores, the *Guilt Judgements* were lower than pre-interview in the *Non-Experts* condition, which went from 23 to 19 participants declaring the suspect as guilty. Additionally, a McNemar test was also conducted to compare the judgments of guilt between the pre- and post-interview among both conditions within the *Non-Experts.* This showed no significant change in outcomes between pre- and post-interview *Guilt Judgements*, X^2 (1, N = 35) = 3.13, p = .077. This was also done for the *Experts* group, where the post-

interview scores on *Guilt Judgement* were considerably lower than they were before the interview. Whereas 9 participants declared the suspect as guilty before the interview, only 2 of them indicated this after the interview. This means that almost all participants (*n* = 14) declared the suspect as innocent after the interview. The McNemar test which was conducted showed no significant changes in guilt judgements between pre- and post-interview *Guilt Judgements*, X^2 (14, *N* = 16) = -1.02, *p* = .327. Since the number of people in each cell was low, not much stock can be put into this outcome.

Table 4

Frequency Table of Guilt Judgements in Guilt Presumption Conditions (Non-Experts)

	Pre-interview (guilty)	Pre-interview (innocent)	Post-interview (guilty)	Post-interview (innocent)
Guilt Presumption	15	4	13	6
Innocence Presumption	8	8	6	10
Total	23	12	19	16

Table 5

Frequency Table of Guilt Judgements in Guilt Presumption Conditions (Experts)

	Pre-interview (guilty)	Pre-interview (innocent)	Post-interview (guilty)	Post-interview (innocent)
Guilt Presumption	5	3	1	7
Innocence Presumption	4	4	1	7
Total	9	7	2	14

Inferential Statistics

Effectiveness of Manipulations

As mentioned earlier, the *Guilt Judgement* (*pre-interview*) variable was not normally distributed, so a Mann-Whitney U test was conducted to check the effect of the manipulations in the *Non-Expert* group. This test indicated almost statistically significant differences in *Guilt*

Judgement (*pre-interview*) between the *Guilt Presumption* (*M* = 4.00, *SD* = 1.05) and the *Innocence Presumption* (*M* = 3.38, *SD* = 1.15) conditions, *W* = 207.5, *p* = .051. There is some indication of a difference, but in both conditions it was clear that the evidence was failing to convince the participants that the suspect was innocent.

Looking at the *Experts* group, a Mann-Whitney U test was also done, and the outcomes made clear that there were no significant differences in *Guilt Judgement* (*pre-interview*) between the *Guilt Presumption* (*M* = 3.63, *SD* = 0.52) and *Innocence Presumption* (*M* = 3.50, *SD* = 0.53) condition, *W* = 36, *p* = .669.

Guilt Judgements (post-interview)

Since the outcomes are ordinal, a Mann-Whitney U test was used to determine whether *Guilt Judgements* (*post-interview*) differed between the *Guilt Presumption* and *Innocence Presumption* condition among *Non-Experts*. This test of the post-interview *Guilt Presumption* condition pointed out that there were no significant differences in *Guilt Judgement* between the *Guilt Presumption* (*M* = 3.79, *SD* = 1.18) and the *Innocence Presumption* condition (*M* = 3.38, *SD* = 0.72), *W* = 198, *p* = .115. This was also done for the *Experts*, and the findings pointed out that there were also no significant differences between the *Guilt Presumption* (*M* = 3.00, *SD* = 0.93) and *Innocence Presumption* (*M* = 3.13, *SD* = 0.35) condition, *W* = 25.5, *p* = .407. In table 6, a description of the mean variables between *Experts* and *Non-Experts* has been given. All variables will be compared between *Experts* and *Non-Experts* in the next section by using t-tests and Mann-Whitney U tests.

Table 6

Mean of variables between Experts and Non-Experts

	Experts	**Non-Experts**
Guilt Judgement (pre-interview)[1]	3.56	3.71
Guilt Judgement (post-interview)[1]	3.06	3.60
Confidence Cues[2]	3.35	2.79
Purpose Cues[3]	3.65	4.02
Intention to Further Investigate[4]	4.95	5.36
Plausibility of the Suspect's Story[4]	4.59	3.87

Note: [1] = scale range 1 (innocent) – 6 (guilty), [2] = scale range 1 (no confidence) – 5 (very much confidence), [3] = scale range 1 (never the purpose) – 6 (always the purpose), [4] = scale range 1 (strongly disagree) – 7 (strongly agree)

Guilt Judgements (pre- and post-interview) between Experimental Conditions

Since the data was not normally distributed, a Wilcoxon's signed-rank test was conducted to check whether the interview changed any guilt judgements among *Non-Experts*. For the *Guilt Presumption* condition, the output indicated that the post-interview *Guilt Judgement* ($M = 3.79$, $SD = 1.18$) scores were not statistically significantly lower than the pre-interview scores ($M = 4.00$, $SD = 1.05$), $W = 44$, $p = 0.308$. When looking at the *Innocence Presumption* condition, the findings indicate that the post-interview scores ($M = 3.38$, $SD = 0.72$) were also not statistically significantly lower in *Guilt Judgement* than the pre-interview scores ($M = 3.38$, $SD = 1.15$), $W = 18$, $p = 1.00$.

A Wilcoxon's signed-rank test was also used to compare the pre-interview *Guilt Judgements* ($M = 3.63$, $SD = 0.52$) with the post-interview *Guilt Judgements* ($M = 3.00$, $SD = 0.93$) for the group of *Experts* in the *Guilt Presumption* condition. The test revealed no statistically significant difference between the two conditions, $W = 18$, $p = .120$. In the *Innocence Presumption* condition, the Wilcoxon's signed-rank test also revealed no

statistically significant differences between the pre-interview *Guilt Judgements* (*M* = 3.50, *SD* = 0.53) and post-interview *Guilt Judgements* (*M* = 3.13, *SD* = 0.35), *W* = 12, *p* = .233.

The hypothesis "*Non-experts who presume guilt will continue to presume guilt even after an interview with an innocent person*" could partly be accepted. There is weak evidence that there is a plausible difference between both conditions in the pre-interview stage (*W* = 207.5, *p* = .051), but this difference was no longer present post-interview. There was a reduction in *Guilt Presumption* from pre- to post-interview, but this was very small and not statistically significant.

The hypothesis *"Experts will be more likely to change their guilt judgements after the interview"* could partly be accepted, since the McNemar test and Wilcoxon's signed-rank test were non-significant, but on the other hand a clear shift in the frequencies of police moving from "most likely guilty" to "most likely innocent" was present.

The hypothesis "*Experts will be less susceptible to presume guilt than non-experts*" could be partly accepted, as the pre-interview correlations on *Guilt Judgement* of *Experts* were not significant but were about the same size as for *Non-Experts*, while the post-interview correlations were clearly smaller. This might indicate that the police officers were equally susceptible to presume guilt as compared to lay people in the pre-interview phase, but less susceptible to presume guilt after the interview took place when compared with lay people.

The hypothesis *"Guilt presumptions before the interview will affect guilt judgements post-interview"* could partly be accepted, since in the *Non-Expert* group no statistically significant differences in pre- and post-interview *Guilt Judgements* were found. However, although not statistically significant, the *Experts* displayed clear reductions in *Guilt Judgement* from 9 police officers judging guilt pre-interview to 2 post-interview.

Plausibility of the Suspect's Story

Even though *Plausibility of the Suspect's Story* (*M* = 3.87, *SD* = 1.26) had a significant negative correlation with *Guilt Judgement* in the pre- as well as post-interview, no significant differences between the *Guilt Presumption* (*M* = 3.85, *SD* = 0.79) and *Innocence Presumption* (*M* = 3.89, *SD* = 0.95) condition were found, *t* (29) = -0.17, *p* = .869. After conducting a t-test, also no statistically significant differences between the *Guilt Presumption* (*M* = 4.42, *SD* = 0.72) and *Innocence Presumption* (*M* = 4.76, *SD* = 0.62) condition within the *Expert* group were found, *t* (14) = -1.02, *p* = .327.

Confidence and Purpose Cues

There was a significant negative correlation between the variables *Plausibility of the Suspect's Story* and *Confidence* among *Experts* and a t-test showed that this finding was

statistically significant, t (27) = -6.10, p < 0.001. Looking at the *Non-Experts*, no statistically significant correlations were found.

In table 7, a full description of the mean scores on the questions of *Purpose* between *Experts* and *Non-Experts* has been given.

Table 7

Mean Scores of Questions on Purpose between Experts and Non-Experts

	Experts	**Non-Experts**
"To gather new evidence"	3.81	4.23
"To test existing evidence"	4.44	4.91
"To get a confession"	3.00	3.57
"To confirm guilt"	3.38	3.40
"To confirm innocence or rule someone out of an investigation"	3.63	4.00

Note: scale range = 1 (never the purpose) to 6 (always the purpose)

Comparing the outcomes on *Purpose* between the *Experts* (M = 3.65, SD = 0.65) and *Non-Experts* (M = 4.02, SD = 1.36), a Mann-Whitney U test showed that there were no statistically significant differences between *Experts* and *Non-Experts, W* = 340.5, p = .221. However, the standard deviation of *Non-Experts* (SD = 1.36) is extremely high, whereas the *Experts* have a considerably lower standard deviation (SD = 0.65). Looking at the question *"To get a confession"*, *Experts* (M = 3.00) scored considerably lower than *Non-Experts* (M = 3.57).

Intention to Further Investigate

A Mann-Whitney U test was done to compare *Non-Experts* (M = 5.36, SD = 1.24) and *Experts* (M = 4.95, SD = 0.97) on the variable of *Intention to Further Investigate* and displayed no statistically significant differences, W = 241.5, p = .439. When looking at the standard deviation, the *Non-Expert* group displayed an extremely high standard deviation (SD = 1.24), and the *Expert* group exhibits a considerably lower standard deviation (SD = 0.97).

Furthermore, in the *Non-Expert* group, a Mann-Whitney U test has also been used and pointed out that there were no statistically significant differences in the *Intention to Further*

Investigate between the *Guilt Presumption* (*M* = 5.01, *SD* = 1.28) and *Innocence Presumption* (*M* = 5.29, *SD* = 0.90) condition, *W* = 148.5, *p* = .921. In the *Expert* group, also no difference was found between the *Guilt Presumption* (*M* = 5.09, *SD* = 1.15) and *Innocence Presumption* (*M* = 4.81, *SD* = 0.81) condition, *W* = 40.5, *p* = .400.

Experiences of Participants

One of the participants mentioned in the open text box, "*An actor is needed to play the suspect and should give the answers himself instead of reading it from a piece of paper*". Participants said several times after the experiment that they found that it would be more believable if the suspect did not read their answers from a paper, but just answered in a more natural way without any prompts or script.

At the police academy, almost all police detectives asked before the interview if they were also allowed to ask their own questions, since they wanted to get more information before they judged the suspect as guilty or not. They also said, *"Based on this information, I find it difficult to make a good decision on innocence or guilt. To make a justified decision, I would further investigate him and ask more of my own questions"*.

Discussion

The purpose of this study was to find out if guilt presumptions had an influence on guilt judgements of individuals in an investigative interview and whether experts were less susceptible in presuming guilt compared to non-experts. The results showed that there were significant correlations between guilt judgements, intention to further investigate, plausibility of the suspect's story, purpose and the variable of confidence, in the expert as well as in the non-expert group. There is weak evidence that there is a plausible difference within the non-expert group between the guilt and innocence presumption condition in the pre-interview stage, but this difference was no longer present post-interview. There was a reduction in guilt presumption from pre- to post-interview, but this was very small and not statistically significant. Furthermore, a significant correlation was found between the pre- and post-interview measure on guilt judgement among non-experts. This underlines the notion that guilt assumptions lead to guilt judgement, even though a plausible explanation was given. These results pointed out that the main ideas were not as expected. On the other hand, in the experts' group there was a clear reduction in guilt judgement from pre- to post-interview, where the frequencies of police moved from "most likely guilty" to "most likely innocent". It is also interesting that even though the pre-interview guilt correlations among experts were not significant, they were about the same size as for the non-expert group, while the post-

interview guilt correlations were clearly smaller. This might indicate that the interview had a larger effect on the experts than on the non-experts.

Guilt Presumptions

The results showed that two third of the participants in the non-expert group tended to assume guilt. The innocence presumption reduced this, but not sufficiently enough to conduct a clear test of the effects of the guilt presumption. This resulted in extremely tentative findings in suggesting the interview may have removed a plausible pre-interview effect. What is shown in this study is that most people in the non-expert group assumed guilt, even after being told that most suspects convicted of this crime are innocent, only having weak evidence against the suspect and the suspect giving an entirely plausible counter narrative for that weak evidence.

The high mean scores on the pre- and post-interview measures of guilt judgement among non-experts could be due to the participants being biased and showing no interest in the plausible narrative of the suspect, thus working deliberately towards eliciting a confession and only searched for evidence that confirmed their guilt assumption of the suspect (Elaad, 2022). This is confirmed by the finding of the variable plausibility of the suspect's story being significantly negatively correlated with both pre- and post-interview measures on guilt judgement. This indicates that participants who believed the suspect was guilty perceived the story the suspect told as less plausible, even though they were given a plausible explanation by the suspect. A possible explanation for this could be that participants want to find a way to confirm their guilt assumptions, through supposing that the story is unlikely. This could be due to confirmation bias happening, where participants seek for a confirmation of their assumption that the suspect is guilty, by regarding the story as unplausible (Casad & Luebering, 2023). This finding shows that confirmation bias has not gone away post-interview. Further, within the experts' group, the variable of plausibility of the suspect's story had a significantly negative correlation with intention to further investigate. This finding shows that the more participants judged the suspect's story as plausible, the less they intended to further investigate the suspect. Another statistically significant negative correlation among non-experts was found between plausibility of the suspect's story and confidence. This finding shows that as guilt judgement increased, the amount of confidence they had in their ability to judge the suspect's innocence or guilt decreased.

A very important finding is that in the expert group, as well as in the guilt presumption as in the innocence presumption condition, there was a shift from "I think the suspect is more likely guilty, but I am not sure" pre-interview to "I think the suspect is more likely innocent,

but I am not sure" post-interview. Despite this, almost all participants, in the expert as well as non-expert group, suggested they would keep up the investigation on this innocent person, causing even more disruption in their life. Additionally, the pre- as well as post-interview measures of guilt judgement had a significant positive correlation with intention to further investigate within the non-expert group. This means that as guilt judgement increased, the intention to further investigate increased too. In the experts' group, the majority of the detectives mentioned after the interview that they wanted to further investigate the story of the suspect, because they wanted to be very sure of their thoughts before declaring the suspect as innocent or guilty. The experts mentioned they wanted to be very certain in order to make the right decision. The detectives also said that if they could think of their own questions, they would have asked other ones than they got provided with in the interview script. This could also clarify the high need to further investigate, since they got scripted questions where no personal involvement was allowed. However, these scripted questions were necessary since all participants needed to be provided with the same information coming from the suspect. The low amount of confidence the experts displayed in their ability to detect whether a suspect is guilty or not could also have influenced the high need for further investigation.

The guilt manipulation did not work out as in the case of Kassin et al. (2003), where it did work. Compared to this study ($n = 35$, $n = 16$), the sample size in the study of Kassin et al. (2003) was much bigger where 52 suspects, being guilty or innocent of a crime, were questioned by 52 interrogators. The participants could choose 6 out of 13 questions of the Interrogation Questions Checklist on their own. In the study of Kassin et al. (2003), no prior information about the crime was given by the interviewer to the suspect in the innocence presumption condition, whilst this was done in the guilt presumption condition. The crime used in the study of Kassin et al. (2003) went about a key that was hidden behind a VCR on the fireplace, which was stolen by someone. This key was then used to open a locked cabinet, which contained $100. The money was stolen, and the question remained who did this, who stole the money and who put the key back in the cabinet. As with this study, Kassin et al. (2003) implemented the guilt manipulation after the case description and the evidence. After this, the interview between the suspect and interviewer took place. The results of the study of Kassin et al. (2003) showed that the guilt manipulation worked out, as interrogators equipped with guilty expectations opposed to those with innocent expectations selected more guilt presumptive questions, exerted more pressure on the suspect to obtain a confession, and judged the suspect more guilty. The effects of the study of Kassin et al. (2003) and this present study may have been so different because of the prior information given in the Kassin

et al. (2003) study to participants in the guilt presumption condition, whereas not in the innocence presumption condition, which was not done in this study (Kassin et al., 2003).

The variable of guilt judgement, pre- as well as post-interview, displayed a significant correlation with purpose among non-experts. The post-interview measure on purpose had a significant negative correlation with guilt judgement, which indicates that while guilt judgement increased, the variable of purpose decreased, showing that participants tended more towards accusatory approaches. However, the high mean of purpose in the non-experts group needs careful interpretation, since the standard deviation was extremely high, which means that one needs to be careful with the interpretation of these scores. Within the experts' group, the standard deviation also displays that there exist different views on what the purpose of an investigate interview is, but no extreme differences were found. Both groups displayed scores that can be interpreted as good alignment with the investigative interviewing principles. An interesting finding was the scores between the experts and non-experts on the question "*To get a confession*". Experts had a considerably lower mean score than non-experts, which could imply that non-experts have more tendency towards eliciting a confession, thereby having the tendency to be more accusing. According to Wakefield & Fleming (2009), the purpose of investigative interviewing is to understand the true narrative of an event. So, enforcing a confession is not necessary, obtaining reliable and accurate information to search for the truth is of greater importance (Wakefield & Fleming, 2009). The findings on the question "*To get a confession*" within the experts' group are in line with the purpose of an investigative interview, which was mentioned by Wakefield & Fleming (2009). This finding could be due to the police detectives in The Netherlands being trained to apply the standard interrogation strategy. Here, eliciting a confession in not of importance, a sufficient person-oriented attitude of the interrogator and the suspect being confronted with contradictive statements are of greater concern.

Limitations

The first limitation relates to the believability of the interview. A remark given by one of the participants clarified the notion of having a person who plays the role of the suspect without any script or prompts. By using an actor, the story could be made more trustworthy and thus making the results more reliable (Dumas et al., 2020). Not only limitations were present within this study, but there were also strengths such as the usage of an experimental approach. By using this approach, the researchers had firm control over their variables to get results. Also, the fact that the researchers could identify the cause and effect of the hypotheses and the possibility of analysing this relationship in further detail can be considered as a great

benefit. The data collected during the experimental approach could also be used to build new research ideas for further studies and it is straightforward, which allows duplication when the same variables are controlled for by others (Gaille, 2019). The experimental approach which has been used in this study also had some weaknesses, such as the results being at risk for human errors since the experimental research used required some specific control of variables. Besides this, the experimental approach used in this study was, compared to other types of research, time consuming since it needed to isolate each variable and conduct testing on it. Third variables could also not be accounted for, since Mother Nature is unpredictable. So, generalising the results to a greater population can be challenging when using an experimental approach (Gaille, 2019).

Second, the sample size was small. If it would be bigger, it would have brought more confidence about the influence of guilt presumptions and the results of the study would become more reliable. A too small sample size may, for example, hinder the outcomes from being extrapolated. The results will be not statistically significant, and no reliable conclusions can be drawn from a sample which is too small (Faber & Fonseca, 2014). As with this study, the standard error will also be high due to the small sample size and ensures that the results will move further and further away from the mean of the entire population. Future research should ensure that the size is big enough, so reliable conclusions can be drawn from it and the results can be generalised to a greater extent.

Another limitation was that the police detectives mentioned their struggles with the questions in the interview script. They said that if they could choose the questions for the interview, they would have chosen different questions. To improve this, future research should involve police detectives in the whole process of this study, and with that the process of writing an interview script. This ensures that they could help with setting up the experiment and with that the improvement of the interview script. This would increase the quality of the experiment and could make the experiment even more believable.

Conclusion

This study investigated the effects of guilt presumptions on guilt judgements and the differences between experts and non-experts in guilt presumptions. The main finding is that participants in the non-expert group had very clear guilt assumptions, even in the innocent condition. These persisted even after interviewing an innocent suspect. There was only weak evidence that the interview did anything at all to change guilt perceptions, even though the evidence presented was specifically chosen to be weak. If any effect did exist, it was very small. Unlike the non-experts, there was a clear shift in guilt judgement among the experts

from pre- to post-interview from thinking the suspect was more likely guilty to having the thought the suspect was more likely innocent. Even though not significant, the pre-interview guilt correlations among experts were about the same size as for the non-experts, while the post-interview guilt correlations were clearly smaller. This finding may indicate a larger effect of the interview on experts than on non-experts. It was also clear that how much someone believed the suspect was guilty, was correlated with how plausible the suspect's story was thought to be, even though the story was exactly the same for all participants. Additionally, for experts as well as non-experts, there was a high need to further investigate the suspect, despite being innocent. The findings of high guilt judgements and the effect of it on the perceived plausibility of the suspect's story in the non-expert group could be attributed to confirmation bias and tunnel vision. Future research should try to gather more participants within the experts as well as non-experts' group, so the findings can be generalised to a greater extent. To conclude, good efforts already have been made, but raising even more awareness for guilt presumptions would contribute to the process of preventing further miscarriages of justice.

Reference list

Adams-Quackenbush, Nicole & Horselenberg, Robert & Tomas, Frédéric & Koppen, Peter. (2019, August 1). Detecting Guilt Presumption in a Police-Suspect Interview: An Evaluation of the Questions in a Dutch Murder Case. 10. 36.

Adams-Quackenbush, N.M., Vrij, A., Horselenberg, R. et al. Articulating guilt? The influence of guilt presumption on interviewer and interviewee behaviour. Curr Psychol 41, 2139–2151 (2020, April 18). https://doi.org/10.1007/s12144-020-00726-z

Baumeister, R. F., & Finkel, E. J. (2010). *Advanced Social Psychology: The State of the Science*. Oxford University Press.

Casad, B. J., & Luebering, J. (2023, February 3). *Confirmation bias | Definition, Examples, Psychology, & Facts*. Encyclopedia

Britannica. https://www.britannica.com/science/confirmation-bias

Chyung, S. Y. (2020). Evidence-Based Survey Design: Exclude or Include a Midpoint in the Likert Scale? *www.linkedin.com*. https://www.linkedin.com/pulse/evidence-based-survey-design-exclude-include-midpoint-chyung

Clarke, C., & Becky, M. (2001, April 1). *National Evaluation of the PEACE Investigative Interviewing Course*.

ResearchGate. https://www.researchgate.net/publication/263127370_National_Evaluation_of_the_PEACE_Investigative_Interviewing_Course

Conversation Management – Forensic Solutions – Professor Eric Shepherd. (2020, October 19). Forensic Solutions – Professor Eric

Shepherd. https://www.forensicsolutions.co.uk/conversation-management/

Cuijpers, L. (2023, February 2). *Nieuwe documentaireserie De Pompmoord: zijn veroordeelden onschuldig?* Omroep

Gelderland. https://www.gld.nl/nieuws/7851068/nieuwe-documentaireserie-de-pompmoord-zijn-veroordeelden-onschuldig

Davison, J. (2009). The science of interviewing – PEACE a different approach to investigative interviewing. In *Forensic Interview Solutions*. Retrieved February 26,

2023, from https://www.fis-international.com/assets/Uploads/resources/PEACE-A-Different-Approach.pdf

De Jong, T. (2023, February 4). *De pompmoord, hoe zat dat ook alweer? Vijf vragen en antwoorden*. Omroep Gelderland. https://www.gld.nl/nieuws/7854870/de-pompmoord-hoe-zat-dat-ook-alweer-vijf-vragen-en-antwoorden

Dolcini, M. M., & Adler, N. E. (1994). Perceived competencies, peer group affiliation, and risk behavior among early adolescents. *Health Psychology*, *13*(6), 496–506. https://doi.org/10.1037/0278-6133.13.6.496

Dumas, D., Doherty, M., & Organisciak, P. (2020). The psychology of professional and student actors: Creativity, personality, and motivation. *PLOS ONE*, *15*(10), e0240728. https://doi.org/10.1371/journal.pone.0240728

Elaad, E. (2022). Tunnel Vision and Confirmation Bias Among Police Investigators and Laypeople in Hypothetical Criminal Contexts. *SAGE Open*, *12*(2). https://doi.org/10.1177/21582440221095022

Elaad, E. (2022). Tunnel Vision and Confirmation Bias Among Police Investigators and Laypeople in Hypothetical Criminal Contexts. *SAGE Open*, *12*(2), 1. https://doi.org/10.1177/21582440221095022

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics*, *19*(4), 27–29. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo

Fernandez, T. M., Godwin, A., Doyle, J. L., Verdín, D., Boone, H., Kirn, A., Benson, L., & Potvin, G. (2016). More Comprehensive and Inclusive Approaches to Demographic Data Collection. In *123rd ASEE Annual Conference and Exposition*. https://doi.org/10.18260/p.25751

Gaille, L. (2019, December 16). *16 Advantages and Disadvantages of Experimental Research*. https://vittana.org/16-advantages-and-disadvantages-of-experimental-research

Gudjonsson, G. H. (1994). Investigative interviewing: Recent developments and some fundamental issues. *International Review of Psychiatry*, *6*(2–3), 237–245. https://doi.org/10.3109/09540269409023262

Hiemstra, T. (2012, September). Beïnvloeding door de politie in het verdachtenverhoor. *Politie Academie*. Retrieved March 10, 2023, from https://www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/PDF/88791.PDF

Holmberg, U. (2004). Police Interviews With Victims And Suspects of Violent and Sexual Crimes; Interviewees' Experiences and Interview Outcomes [Dataset]. In *PsycEXTRA Dataset*. https://doi.org/10.1037/e530022013-001

Horselenberg, R., & Smeets, T. (2005). Valse Bekentenissen. *Advocatenblad 14*, *14*. https://assets.budh.nl/advocatenblad/article_pdf/20084444/basis_pdf_oid.pdf

Kassin, S. M., Goldstein, C. L., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law And Human Behavior*, *27*(2), 187–203. https://doi.org/10.1023/a:1022599230598

M. Adams, N. (2019). *The Influence of Guilt Presumptive Language on Investigative Interview Outcomes* [PhD Proefschrift]. Universiteit van Maastricht.

Méndez, J. (2021). The need for the Principles on Effective Interviewing for Investigations and Information Gathering. *Interviewing Principles*, *31*(3), 117–120. https://doi.org/10.7146/torture.v32i3.128719

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy and Law*, *15*(4), 315–334. https://doi.org/10.1037/a0017881

Oxburgh, G., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews: A review of the literature from a psychological and linguistic perspective. In *The international journal of speech language and the law* (Vol. 17, Issue 1). Equinox Publishing. https://doi.org/10.1558/ijsll.v17i1.45

Pozzulo, J. D., & 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. 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

Wakefield, A., & Fleming, J. (2009). The SAGE Dictionary of Policing. *SAGE Publications
Ltd EBooks*. https://doi.org/10.4135/9781446269053

Watson, S. J. (n.d.). *Interrogation, False Confession and Investigative Interviewing* [Slide
show; Powerpoint].

Weiher, L. (2020). The impact of interview style on the development, maintenance, and
transfer of rapport [PhD]. Lancaster University.

Yale, R. N. (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

Yuille, J. C., Marxsen, D., & Cooper, B. (1999). Training investigative inter- viewing:
Adherence to the spirit, as well ass the letter. *International Journal of Law and
Psychiatry*, *22*(3–4), 323–336.

Appendix A

Guilt Judgement

Based on the information I have about the current case, I believe the suspect is...

1 I am very sure the suspect is innocent.

2 I am quite sure the suspect is innocent.

3 I think the suspect is more likely innocent, but I am not sure.

4 I think the suspect is more likely guilty, but I am not sure.

5 I am quite sure the suspect is guilty.

6 I am very sure the suspect is guilty.

Questionnaire Guilt and Investigative Decisions

For the interviewee (researcher)
Please answer the following questions on the 5-point Likert scale with the response options:
1=strongly agree
2=agree
3=neither agree nor disagree
4=disagree
5=strongly disagree

1. The interviewer really listened to what I had to say
2. The interviewer paid attention to my opinion
3. The interviewer was attentive to me
4. The interviewer was interested in my point of view
5. The interviewer was honest with me
6. The interviewer respects my knowledge
7. The interviewer can generally be trusted to keep their word
8. The interviewer did the job with skill
9. The interviewer performed expertly
10. The interviewer made effort to do a good job
11. The interviewer acted like a professional
12. We have our culture in common
13. The interviewer and I share one ethnicity
14. The Interviewer shares my culture
15. We worked well as a team
16. The communication went smoothly
17. The Interviewer and I got along well

For the interviewer (participant)
Please answer the following questions on the 5-point Likert scale with the response options:
1=strongly agree
2=agree

3=neither agree nor disagree
4=disagree
5=strongly disagree

1. The interviewee really listened to what I had to say
2. The interviewee paid attention to my opinion
3. The interviewee was attentive to me
4. The interviewee was interested in my point of view
5. The interviewee was honest with me
6. The interviewee respects my knowledge
7. The interviewee can generally be trusted to keep the word
8. I can trust the interviewee to keep their word to me
9. The interviewee did the job with skill
10. The interviewee performed expertly
11. The interviewee made effort to do a good job
12. The interviewee acted like a professional
13. We have our culture in common
14. The interviewee and I share one ethnicity
15. The Interviewee shares my culture
16. We worked well as a team
17. The communication went smoothly
18. The interviewee and I got along well

Questionnaire Purpose and Confidence Cues
Please fill in the following questions based on your expectations and training so far.

1. How confident are you in your ability to determine whether or not a suspect is guilty or innocent before you interview a suspect on a 1 (very much confidence) to 5 (no confidence) scale?

2. How confident are you in your ability to detect whether a suspect is telling the truth during an interview on a 1 (very much confidence) to 5 (no confidence) scale?

3. How confident are you in your ability to detect whether a suspect is guilty or innocent after an interview on a 1 (very much confidence) to 5 (no confidence) scale?

4. Please rate the following statements to indicate the extent to which you feel they reflect the purpose of a suspect interview in your typical experience

	Never the purpose	Occasionally the purpose	Sometimes the purpose	Often the purpose	Very often the purpose	Always the purpose
To gather new evidence						

To test existing evidence					
To get a confession					
To confirm guilt					
To confirm innocence or rule someone out of an investigation					

Plausibility of the Suspect's Story

Please answer the following questions on the 5-point Likert scale with the response options:

1=strongly agree

2=agree

3=neither agree nor disagree

4=disagree

5=strongly disagree

1. The suspect appears to be trustworthy.
2. The suspect appears to be likeable.
3. The suspect appears to be friendly.
4. The suspect appears to be an honest person.
5. The suspect seems like they care about others.
6. The suspect appears to be empathic.
7. The suspect seems like they would act ethically.

Q98 To what extent do you agree/disagree with the following statements?

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I believe the suspect´s narrative could be true (1)	o	o	o	o	o	o	o
The suspect´s narrative was plausible (2)	o	o	o	o	o	o	o
The suspect´s narrative seems to be true (3)	o	o	o	o	o	o	o
It was easy to follow the narrative from beginning to end (4)	o	o	o	o	o	o	o

I could understand the story the suspect provided (5)	o	o	o	o	o	o	o
If I were telling this narrative, I would have organized it in the same way (6)	o	o	o	o	o	o	o
The information presented in the suspect´s narrative was consistent (7)	o	o	o	o	o	o	o
All of the facts in the suspect´s narrative agreed with each other (8)	o	o	o	o	o	o	o

There was no important informatio n missing from the suspect´s narrative (9)	o	o	o	o	o	o	o
There were "no holes" in the suspect´s narrative (10)	o	o	o	o	o	o	o
I think the suspect covered all aspects of the case in their account (11)	o	o	o	o	o	o	o

Q107 The "consistency" of a narrative refers to the extent to which a narrative does not contradict itself or contradict other things you know to be true or false.

How would you rate the suspect´s narrative in terms of "consistency"?

○ Strongly inconsistent (1)

○ Inconsistent (2)

○ Somewhat inconsistent (3)

○ Neither consistent nor inconsistent (4)

○ Somewhat consistent (5)

○ Consistent (6)

○ **Very consistent (7)**

Intention to Further Investigate
Q110 To what extent do you agree/disagree with the following statements?

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
If I was the investigating police officer, I would suggest my supervisor to continue the investigation on this suspect (1)	○	○	○	○	○	○	○
I think this person should be further investigated (2)	○	○	○	○	○	○	○

I believe that a second investigative interview would contribute to the progress in this case (3)   ○ ○ ○ ○ ○ ○ ○

I think the suspect should not be released without further questioning (4)   ○ ○ ○ ○ ○ ○ ○

There would be a reason continuing to investigate this individual (5)   ○ ○ ○ ○ ○ ○ ○

I do think that another interview with this suspect would be necessary (6)   ○ ○ ○ ○ ○ ○ ○

Questions on Immersion

To what extent do you agree / disagree with the following statements?

	Strongly disagree	**Somewhat disagree**	**Neither agree nor disagree**	**Somewhat agree**	**Strongly agree**
I took the task seriously					
I could imagine myself as an interviewer					
The interview felt real to me					
The suspect played his/her part well					

Open Text Box

Do you have anything left to say to the researchers?

Appendix B

Case Description

Background information about the case and list of allegations

Imagine yourself being 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 30/03/2023 the police arrested a woman named Mrs Brown for dealing drugs. The woman was caught selling different types of drugs in the park in your town. The women dealt with Opiates (Heroin, morphine), Hallucinogens (LSD), and Marijuana. She was arrested at 4.30 pm 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 30/03/2023 is listed here:

- An old woman with schizophrenia saw someone that looked like the suspect together with Mrs Brown in the park, 5 minutes before and after Mrs Brown dealt the drugs.
- A 12-year-old child 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.

(When Guilt Presumption condition: letter A)

Your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually **committed** the crime.

(When Innocence Presumption condition: letter B)

Your supervisor reminds you that four out of five (80%) people interviewed as suspects for this crime actually **did not commit** the crime.

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 suspects 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 suspect's responses, and afterwards, we will ask you questions about your experience in the interview.

Appendix C

Interview Script

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

Interviewee: I don't really have a lot to say. I don't even know this woman and now I am suspected of being involved in a crime with her. I don't really know what to say.

Interviewer: Can you tell me what you did on the afternoon of the 30th of March 2023?

Interviewee: It was a Thursday. On Thursday afternoons I am usually working. I am a taxi driver in my town, and my work shift is always from 2 pm until 8 pm. I remember that this day was a really busy day, and I had a lot of clients.

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

Interviewee: I am sorry. I don't remember the clients I had that day, because there were so many. But I can ask my boss if he still has the list of clients I drove on that day.

Interviewer: Do you remember anything else?

Interviewee: No, sorry. You could explain why you are questioning me because at this point I don't know why I am being suspected of the crime.

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

Interviewee: Mhh... so ... sometimes during my break, which is usually at around 4 pm, I go in the park and walk around a bit. You can ask my colleagues, I have done that since I started working at the taxi company, even when it is a busy day. I am sorry, I forgot to mention that earlier. Sometimes, I feel like talking with other people in the park, so I just ask them how their day was. Most of them are quite friendly, so I have a short chat with them while walking around in the park. Maybe this Mrs Brown could have been one of the people I talked to.

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: Probably this woman was a client of mine. Probably she wanted me to drive her to the park, and then I decided to use my break to talk to her in the park. I mean, to be honest, I don't remember all the people I drive, even if I have good conversations with them. There are just too many to remember them all. I am sorry.

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

Interviewee: My boss does not arrange my clients for me but I schedule my own clients. I have the same phone number for my work and my private stuff. A lot of people contact me outside of my work shift. Most of them are desperately trying to get a taxi, as there are not a lot of taxi drivers around our town. So, when I start my shift in the afternoon, I already have a lot of phone calls and messages from clients wanting me to drive at a certain time and to a certain location.

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: As I already mentioned, I like to chat with people. I am a social person and a lot of people know me. I mean, as a taxi driver, you get around quite a lot in town and the people just know you. Many people get to know you, but the problem is that I am really bad at remembering their names and faces. I mean, I might have talked with this woman at some point. However, I am certain that I am not involved in any drug business with her, that is how much I can tell you.

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

Interviewee: I am telling you, you could basically find traces of drugs in every taxi driving around the city. I mean it is not like all my clients are saints. They certainly like to have fun. Just because there were traces of drugs in my car, it doesn't mean that they belong to me. I would never risk my job by transporting drugs in my taxi. But you said that I drove this woman, this Mrs Brown to the park. I mean if she dealt drugs in the park, she already had drugs on her and this is the reason why you found traces of drugs in my taxi. There is no reason to assume the drugs were from me.

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

Appendix D

Debrief

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 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.

Britt Hudepohl: b.a.e.hudepohl@student.utwente.nl

Sina Akkermann: s.akkermann@student.utwente.nl

Lisa Zeppelzauer: l.s.zeppelzauer@student.utwente.nl

Project Supervisor: Dr. Steven Watson; s.j.watson@student.utwente.nl