WILEY

Test of the analysis of competing hypotheses in legal decision-making

Enide Maegherman[1,2] ⓘ | Karl Ask[2] | Robert Horselenberg[1] | Peter J. van Koppen[1,3]

[1]Department of Criminal Law and Criminology, Faculty of Law, Maastricht University, Maastricht, The Netherlands

[2]Department of Psychology, Gothenburg University, Gothenburg, Sweden

[3]Department of Criminal Law and Criminology, Faculty of Law, VU University Amsterdam, Amsterdam, The Netherlands

Correspondence
Enide Maegherman, Department of Criminal Law and Criminology, Faculty of Law, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands.
Email: enide.maegherman@ maastrichtuniversity.nl

Summary

The analysis of competing hypotheses (ACH) has been suggested to be a method that can protect against confirmation bias in the context of intelligence analysis. In the current study, we aimed to determine whether ACH could counter confirmation bias in the reasoning with evidence in the context of criminal law proceedings. Law students (N = 191) received information about the ACH method or general information about biases. They were given a case vignette with a main suspect and a list of 24 questions, 6 of which they could ask about the case. Half of the questions related to incriminating information, whereas the other half related to exonerating information. Contrary to our expectations, participants in both conditions favoured questions relating to exonerating information and rated the exonerating evidence as being more important for their decision. Despite the lack of bias observed, it seems participants failed to properly apply the ACH method.

KEYWORDS
ACH, confirmation bias, evidence, legal decision-making

1 | INTRODUCTION

In criminal law proceedings, the judges or juries have to determine whether suspects are guilty of the offence they are accused of. Such a decision requires the decision maker (hereafter referred to as judge) to reason from the available evidence to a final verdict. To do so, the judge has to decide what has been proven to be true. That process can be compared to testing hypotheses about what happened; the guilty scenario presented by the prosecution, and possibly the alternative scenario of innocence most likely presented by the defence. These can then be tested using the available evidence. Several theories exist about how a decision is made in legal proceedings (e.g. Pennington & Hastie, 1988; Simon, 2004; Van Koppen & Mackor, 2019; Wagenaar, van Koppen, & Crombag, 1993). The reasoning processes undertaken by decision makers in criminal law proceedings are inherently cognitive processes, and subsequently, can be vulnerable to cognitive biases. Such cognitive biases can also lead to

miscarriages of justice (Bandes, 2006; Martin, 2002). In the present study, we investigated whether a structured analysis technique taken from the field of intelligence analysis, namely the analysis of competing hypotheses (ACH), can reduce the effect of cognitive biases.

An important aspect of determining whether something is true, addressed by Popper (2005), is the need for attempted falsification rather than merely confirmation of the theory to be tested. Falsification refers to searching for facts that disconfirm the existing hypothesis. Popper (2005) pointed out that, even with evidence supporting the theory, a theory can only be maintained until evidence that disconfirms it is found. The fact that a theory has not been disproven yet does not mean it must be true. In short, when trying to determine whether something is true, one must look not only at evidence supporting that, but must foremost try to find evidence that falsifies the rule. Failed serious attempts at falsification increase the likelihood of the hypothesis being true (Crombag, van Koppen, & Wagenaar, 2006).

1.1. | Confirmation bias

Although attempting falsification might seem like a simple process, it is complicated by a common cognitive bias, namely confirmation bias (Wason, 1968). When confirmation bias occurs, one focuses excessively on finding evidence that confirms an initial hypothesis, while paying disproportionately little attention to evidence that contradicts that hypothesis (Nickerson, 1998). That hypothesis can then be maintained, even when faced with contradicting evidence, which is known as *belief perseverance*. Belief perseverance has also been observed in some prosecutors who, in exoneration cases, have maintained that the convicted individual was guilty despite overwhelming evidence to the contrary (Burke, 2007). The biased interpretation and selection of evidence which constitute confirmation bias consequently help maintain the belief. For instance, confirmation bias could affect the interpretation of evidence in such a way that it becomes more supportive of the existing hypothesis (Kassin, Dror, & Kukucka, 2013). It has been argued that confirmation bias is particularly relevant in legal proceedings (Findley & Scott, 2006). In addition, in some jurisdictions (e.g. the Netherlands), the case file presented to judges is likely to be biased towards the guilt of the suspect, which has been argued to make judges prone to confirmation bias (Crombag, 2017).

Schünemann and Bandilla (1989) describe a test of the influence of incriminating information presented to judges prior to trial on judges' decisions. One group received pre-trial information that was mainly incriminating for the suspect, while the other group received information that was less incriminating prior to trial. All judges then received identical records of the subsequent proceedings. In the first group, 82% of judges chose to convict the suspect, whereas only 53% of judges in the second group chose to convict the suspect. Based on these findings, it appears that judges' decisions are strongly affected by their initial belief. As discussed, according to Popper's argumentation (2005), judges should also try to falsify their hypothesis. According to Wagenaar et al. (1993), the simplest way to try to falsify the criminal charge would be to investigate what happened according to the suspect. However, in line with research on confirmation bias (Ask & Granhag, 2005; Jonas, Schulz-Hardt, Frey, & Thelen, 2001; Wason, 1968), one would expect that there will be a preference for evidence that confirms the initial hypothesis, presumably the charge, and a subsequent failure to attempt falsification of that hypothesis.

In light of the potential detrimental effect of confirmation bias, it is paramount to examine the relation between confirmation bias and the process of falsification. Mynatt, Doherty, and Tweney (1977) used a computer-controlled environment to this end. Participants were asked to formulate a hypothesis, after which they were asked to choose between an environment where they could make observations that would likely confirm their hypothesis, or an environment where they could test alternative hypotheses. Participants showed a strong preference for the former environment. Participants who did receive falsifying information used that information to reject incorrect hypotheses.

Other researchers have found that participants presented with a criminal case file show a preference for evidence that supports their existing beliefs. For instance, Rassin, Eerland, and Kuijpers (2010) presented students with a case file and asked them to decide whether the suspect was innocent or guilty. Participants could then choose from 20 further investigative measures, half of which were directed at obtaining further evidence against the suspect, and half which were aimed at obtaining exonerating information. The initial decision was found to predict the further information sought, with participants choosing investigative measures fitting their initial decision. Similar findings were presented by O'Brien (2009), who found that participants who formulated a hypothesis about the likely culprit in a mock crime investigation showed more confirmation bias than those who did not formulate a hypothesis or who formulated both a hypothesis and a counterhypothesis. Interestingly, participants who were asked to write down three alternative suspects did not show less bias than those who only identified one primary suspect. Therefore, whether and how the formulation of multiple alternative hypotheses can reduce bias requires further investigation.

The consideration of alternative scenarios in legal decision-making has also been investigated by Rassin (2018). In his study, some participants were presented with an alternative scenario in a criminal case whereas others were additionally instructed to indicate how well each piece of evidence fitted with the main and the alternative hypothesis, respectively. They used pen and paper to do so, which means it was an active task rather than simply having them consider the evidence. Participants were asked to rate police findings and to decide whether or not to convict the main suspect. Rassin (2018) found that the pen-and-paper tool counteracted an excessive focus on evidence confirming the main hypothesis in both lay people and criminal justice professionals. The pen-and-paper tool group also had a lower conviction rate than the alternative scenario group. When combined with the findings by O'Brien (2009), it therefore seems that only formulating alternative scenarios may not be sufficient; one should actively consider the alternative scenarios in order to counteract the influence of confirmation bias.

1.2. | Analysis of competing hypotheses

A potential tool to structure the consideration of alternative scenarios is the ACH. ACH is a procedure which was originally designed to help intelligence analysts avoid common pitfalls (Heuer, 1999). It requires careful weighing of alternative explanations. These are compared against each other, thus preventing the analyst from deciding on the first solution that seems satisfactory. One of the pitfalls targeted by ACH is the concern that focusing on trying to confirm one hypothesis may cause analysts to fail to recognise that much of the evidence is also consistent with other explanations or conclusions that have not yet been refuted (Heuer, 1999). It is evident that that problem shows similarities with confirmation bias. It would therefore be useful to determine whether the ACH method can also be used in other areas where confirmation bias is likely to play a role, such as decision-making by judges.

The ACH method consists of eight steps, as summarised below on the basis of Heuer's (1999) description. The first step requires the

identification of potential hypotheses; that can be accomplished in collaboration with colleagues who hold differing views. The second step is to make a list of significant evidence, noting not only the presence of evidence but also the absence. For each hypothesis, it should be considered what you would expect to see, and not see, if it were true. During the third step, a matrix is prepared with hypotheses across the top, and evidence down the side. The matrix is then used to analyse how each piece of evidence relates to each hypothesis. In that step, the analyst works across the matrix, examining how consistent each piece of evidence is with each of the hypotheses. In the corresponding cell of the matrix, it should be noted whether the evidence is consistent, inconsistent or irrelevant to the hypothesis. The matrix aids the determination of the diagnosticity of the evidence. For instance, if a piece of evidence is consistent with all hypotheses, its diagnostic value is very low to non-existent. In the fourth step, the matrix should be refined. It may be possible that some hypotheses should be added, or more distinctions should be made to include other alternatives. Evidence which has no diagnostic value should be removed from the matrix.

During the fifth step, tentative conclusions can be drawn about the likelihood of each hypothesis. It is important to try to disprove hypotheses rather than trying to prove them. The hypothesis which has the fewest inconsistent pieces of evidences is probably the most likely. However, that does not mean that the hypothesis with the most consistent pieces of evidence is the most likely. The matrix only serves as a tool that causes giving more consideration and more analysis of hypotheses that would otherwise have been thought to be unlikely. During the sixth step, it should be considered how sensitive the conclusion is to a few critical items of evidence, and what the consequences would be if that evidence were misleading or subject to a different interpretation. Additional investigations may be needed. Subsequently, conclusions of the analysis are drawn during the seventh step. The relative likelihood of all hypotheses should be discussed. It is important to proceed by eliminating hypotheses, rather than by confirming them. During the final step, analysts should identify possible future information that may change the outcome of the analysis (Heuer, 1999).

Because it was initially developed for use in intelligence settings, it is worth studying whether the ACH procedure can also be used in the consideration of evidence in judicial proceedings. If found to be effective, the procedure would then offer a systematic approach to decision-making, which could reduce the influence of biases. To date, ACH has not yet been the subject of many studies (Puvathingal & Hantula, 2012). In one study, Lehner, Adelman, Cheikes, and Brown (2008) found that ACH mitigated confirmation bias for participants who did not have experience working in intelligence analysis. That manifested itself in weighing of the evidence rather than in evidence interpretation. Dhami, Belton, and Mandel (2019) found mixed evidence for the effectiveness of ACH in reducing confirmation bias in trained analysts instructed to use the ACH method, as well as a lack of adherence to the method. The current study examined whether informing people about the ACH method can reduce the influence of an initial belief by increasing the attempts made to falsify the original hypothesis in the context of a criminal case.

1.3. | The current study

Based on existing theory and previous research, in the current study, we examined the relation between an initial belief, attempts at falsification, and the final decision using a mock criminal case. Furthermore, it was tested whether training in ACH can affect the extent of attempted falsification. In order to do so, participants were either trained in the ACH procedure, or merely received general information on biases. Participants were then provided with a description of a case, in which a guilty scenario for the main suspect was obvious, and an alternative scenario was subtly implied. After participants had indicated their initial belief about the case they were given the opportunity to receive answers to a number of investigative questions. The questions related to information with potential to either confirm or disconfirm their initial belief. Answers to the disconfirming questions could potentially either support the alternative scenario or give evidence that contradicted the scenario that the main suspect was guilty (e.g. providing an alibi, finding traces of an alternative suspect). After obtaining additional information based on the selected questions, participants were again asked to indicate their belief about the case by, for instance, rating the likelihood that the main suspect was guilty.

We expected that those who were instructed in the ACH procedure would show more attempts at falsification than those in the control condition, by choosing more questions that could potentially disconfirm the main suspect's guilt (H1). Those in the ACH condition were subsequently expected to have a lower final rating of likelihood of the suspect's guilt than those in the control condition (H2).

2 | METHODS

2.1 | Participants

Participants were recruited through advertisement at various law faculties in the Netherlands and Belgium. Of the original sample of 361 participants, 170 participants were excluded for reasons related to failing the attention or control checks, or having a background in something other than law (for a detailed overview of the exclusion of participants, see Data S1). Participants could choose to take the survey in Dutch or in English. The final sample included 191 participants, of which 154 chose to take the survey in Dutch. Participants' average age was 22.8 years (SD = 3.5). The majority of the participants were female (68.6%). Out of the final sample, 71.7% completed the survey online. A sensitivity power analysis based on the final sample size, a power of .80, and a .05 significance level was conducted using the G*power software (v3.1; Faul, Erdfelder, Lang, & Buchner, 2007). According to the power analysis, the collected sample size was acceptably powered to detect an effect size of d = .36.

2.2 | Materials

2.2.1 | ACH and biases information

In the ACH condition, participants were given a short explanation of the ACH procedure. The explanation focused on producing several hypotheses and using the matrix, as these are thought to be the key components of the method relevant to decision-making. In addition, participants were given a short description of confirmation bias. To ensure that the only difference between the conditions was the explanation of ACH, participants in the control group also received the information on confirmation bias.

2.2.2 | Case description

A short fictional case description about a criminal investigation was used. In the case, a man named Jasper Kostal had been murdered. The main suspect was his wife, Sabine. The case contained information that appeared incriminating for Sabine (e.g. she was covered in blood, her fingerprints were on the knife). However, an alternative scenario was also subtly implied in the case description. Specifically, there was evidence that Jasper had been having an affair. The case was pretested by asking participants to rate the likelihood that Sabine was guilty of killing Jasper on a scale from 0 (*not at all likely*) to 100 (*very likely*). As the case had to be perceived as incriminating against Sabine, it was adapted in response to the pretest ratings. The case was also edited for authenticity by an experienced court clerk. In the final pretest ($N = 20$), the average rating for the likelihood that Sabine was guilty of killing Jasper was 73.4 ($SD = 19.3$).

2.2.3 | Investigative questions

Participants were presented with a list of 24 investigative questions. Of these questions, 12 could potentially provide answers that confirmed the obvious scenario of guilt, whereas the other 12 questions could potentially disconfirm the obvious scenario of guilt and support a scenario of innocence. Answers to the disconfirming questions could indicate that the that the obvious scenario could not be true, provided alternative explanations for the evidence found, or provided support for the alternative scenario. Participants were asked to choose six questions in total, and were returned to the list of questions after having received the answer to each question. In order to ensure that participants perceived the questions to be confirming or disconfirming for their hypothesis as intended, participants were asked to report why they had chosen to ask each question using open-ended answers. Although we had initially planned to use the questions selected by participants as a measure of falsification attempts, it soon became clear during coding that this was not feasible. Based on participants' responses, their interpretation of the question did not always match the classification we had given it. For example, for the question asking whether Sabine is right-handed, some participants emphasised that it

would have been exonerating for Sabine if she had turned out to be left-handed, whereas others emphasised that Sabine's right-handedness fits with the hypothesis that she was the perpetrator. It thus became clear that the selected questions could not easily be classified as confirming or disconfirming without knowing what outcome the participant had anticipated. Unfortunately, participants' open-ended answers did not always offer insight into their expectations, as several answers were of the type 'it's important to know', 'I was wondering', or even 'just checking'. It was therefore decided to use an alternative measure of falsification (see below).

2.2.4 | Importance ratings

At the end of the study, participants were asked to rate the importance for their final decision on guilt of each piece of evidence that they had received as answers to the investigative questions, using a scale from 0 (*not at all important*) to 100 (*very important*). As these ratings related to the answers to the investigative questions, rather than to the questions themselves, the interpretation of these ratings was much more straightforward and thus more suitable than question selection as a measure of preference for falsifying information. In testing H1, it was therefore expected that participants in the ACH condition would have a higher average importance rating for the evidence that falsified the main suspect's guilt than would those in the control condition.

2.3 | Procedure

Initially, the study was administered in person as we considered also using the side notes made by participants. Some of these participants came to the lab, whereas others were tested in a classroom setting. However, data collection in person proved to be extremely difficult, and the notes participants made had little additional value. We noticed that participants were often not making notes, or merely writing down a few keywords of the case rather than using a structured approach as instructed by the ACH instruction. Data collection was therefore continued online. The survey platform Qualtrics was used. Several questions were added to ensure the quality of the gathered data. These included four control questions about the case, two control questions about the ACH instruction, and three attention checks throughout the procedure.

After providing informed consent, participants either read only general information about biases (control condition) or read general information about biases *and* received an explanation of the ACH procedure (ACH condition). In the online study, participants in the ACH condition then also received control questions related to the ACH manipulation (e.g. a true or false question asking whether the main focus was on evidence supporting the hypothesis). They were told how many questions they answered correctly and were then given the option to revisit the ACH instruction. Participants were then given the case file. After having read the case file, participants were asked to rate the likelihood that Sabine was guilty (0 = *not at all likely*, 100 = *very likely)*, whether or not they would convict her for the

TABLE 1 Medians and Mann–Whitney U analyses for type of question across conditions

| Type of question | Condition | | Total | Mann–Whitney U | p | r |
	ACH	Control				
Confirming questions	2.0	2.0	2.0	4,289.50	.476	.052
Disconfirming questions	4.0	4.0	4.0	4,289.50	.476	.052

Abbreviation: ACH, analysis of competing hypotheses.

murder of her husband,[1] and how confident they felt about their conviction decision (0 = *not at all confident*, 100 = *very confident)*. These answers were used as a measurement of their initial belief. Furthermore, participants were asked to write down their hypothesis or hypotheses about what happened. For these, the number of scenarios formulated were counted, as well as the perpetrators named in the scenarios.[2] Participants were then given the list of confirming and disconfirming investigative questions.

All participants selected six investigative questions and received answers to those questions. Participants were again asked to rate the likelihood of the suspect being guilty, whether or not they would convict the suspect, and how confident they felt about their conviction decision. Finally, participants were then asked to rate the importance for their final decision on guilt of the evidence they had received as answers to the investigative questions.

Participants took an average of 1 hr and 47 min to complete the study (*SD* = 8 hr and 58 min). The median response time was 28 min and 7 s.[3]

3 | RESULTS

3.1 | Language

Participants could choose to complete the survey in either Dutch or in English. The Dutch and English groups did not differ for the measures of likelihood of guilt, the number of confirming or disconfirming questions selected, or importance ratings. Language also did not interact with condition for any of the above measures (detailed results for the analyses involving language can be found in Data S1). It was therefore decided that language was not a confounding factor and it was not included in subsequent analyses. The data from the Dutch and the English groups were combined for all further analyses.

3.2 | Question selection

Based on Shapiro–Wilk tests, it was concluded that the data for the dependent variables were not normally distributed (detailed outcomes for the analyses can be found in Data S1). It was therefore decided to conduct non-parametric test for the main analyses.

In order to determine whether those instructed in ACH showed a stronger preference for the questions which were potentially disconfirming for the suspect's guilt, the questions chosen were compared between the ACH and the control condition using Mann–Whitney

U tests (Table 1). Participants could choose a total of six questions. The number of disconfirming and confirming questions was calculated for each participant. Participants in the ACH condition and the control condition did not differ significantly in the number of selected questions that were confirming of guilt. The ACH group also did not differ from the control group for the number of disconfirming questions chosen. A Wilcoxon signed rank test indicated that participants overall chose significantly more questions that were disconfirming than questions that were confirming for the suspect's guilt, $z = -9.05$, $p < .001$, $r = .463$. The median for the type of chosen questions for all participants can be found in Table 1.

3.3 | Importance ratings

The importance ratings for the exonerating evidence was averaged for each participant. The same was done for the importance ratings for the incriminating evidence. Participants in the ACH condition did not differ significantly from participants in the control condition in their importance ratings for the exonerating evidence. There was also no significant difference between the ACH group and the control group in the average rating for incriminating evidence (Table 2). Overall, participants rated the exonerating evidence as significantly more important than the incriminating evidence, $z = -7.13$, $p < .001$, $r = .376$. The median importance ratings for the incriminating and exonerating evidence overall can be found in Table 2. As there was no evidence that participants in the ACH condition showed a stronger preference for disconfirming evidence than did control participants, we found no support for H1.

3.4 | Ratings of guilt

Participants were asked to rate the likelihood of the main suspect, Sabine, being guilty immediately after reading the case (Time 1) and again after they had received answers to the investigative questions (Time 2). Participants in the ACH condition did not differ significantly from participants in the control condition for the rating of likelihood of Sabine being guilty at Time 1. For the likelihood rating at Time 2, participants in the ACH condition also did not differ significantly from participants in the control condition (Table 3). This finding indicates that there was no support for H2. Overall, the ratings for likelihood of guilt decreased significantly between Time 1 and Time 2, $z = -10.27$, $p < .001$, $r = .526$. Median ratings for likelihood of suspect guilt for all participants can be found in Table 3.

TABLE 2 Median importance ratings and Mann–Whitney U analyses

Type of evidence	Condition		Total	Mann–Whitney U	p	r
	ACH	Control				
Exonerating evidence	74.0	71.0	72.7	3,896.50	.202	.093
Incriminating evidence	56.3	54.0	55.3	3,623.50	.830	.016

Abbreviation: ACH, analysis of competing hypotheses.

TABLE 3 Median ratings for likelihood of suspect guilt and Mann–Whitney U analyses

Time point	Condition		Total	Mann–Whitney U	p	r
	ACH	Control				
Time 1	70.0	69.0	70.0	4,281.00	.479	.051
Time 2	40.0	45.0	40.0	4,381.00	.657	.032

Abbreviation: ACH, analysis of competing hypotheses.

3.5 | Compliance with the ACH instructions

Participants in the ACH condition were asked whether they had made an ACH matrix, how helpful they found it, and how difficult they found it after giving the importance ratings. Of the participants in the ACH condition, 38% reported having used the matrix. Participants are therefore thought not to have complied with the ACH instructions. The average helpfulness rating for those who reported using the matrix was 57.1 (SD = 18.8) on a scale from 0 (*not helpful at all*) to 100 (*very helpful).* The average difficulty rating was 46.3 (SD = 28.5) on a scale from 0 (*not difficult at all*) to 100 (*very difficult*).

4 | DISCUSSION

The current study was intended to evaluate whether ACH could protect against confirmation bias in the context of a criminal case. This question proved difficult to answer as, contrary to our expectations, participants in the current study did not show confirmation bias in either condition. Participants in both conditions seemed to favour questions that were disconfirming for the guilt of the main suspect in the case. Similarly, they also rated the exonerating evidence as more important for their decision than the incriminating evidence. Their perception of guilt of the main suspect also decreased between time 1 and time 2, as did the conviction rates for the suspect. Moreover, participants in the ACH condition did not differ significantly from the participants in the control condition in their initial ratings of likelihood of guilt. Participants in both groups had a relatively high rating of guilt, and the majority of participants chose to convict Sabine after only having read the case file. Subsequently, they also did not differ between conditions for the questions they chose to ask, or in their importance rating of the exonerating and incriminating evidence. Therefore, it appears there was no difference between those who received the ACH instructions and those who did not receive the ACH instructions in terms of perception of guilt, nor in their search for further information.

A significant difference was, however, found for the ratings of guilt and conviction rates at Time 2 compared to Time 1. The perception of guilt decreased after participants received answers to the chosen questions. That finding is in line with the fact that participants also chose more questions that we had originally coded as disconfirming of the suspect's guilt than questions that were considered confirming of the suspect's guilt. Similarly, participants overall rated the exonerating evidence as more important for their decision than the incriminating evidence. The results therefore support the conclusion that participants were not biased towards the scenario of the main suspect and updated their belief in the main suspect's guilt appropriately in response to the exonerating evidence received.

There are several potential explanations for these findings. Firstly, the lack of confirmation bias found in all participants should be considered. Confirmation bias was expected based on the fact that participants received an initial case file that was biased towards guilt of the main suspect (O'Brien, 2009). Even though the case file also contained the suggestion of an alternative scenario, preliminary testing of the material still indicated that it was biased towards the guilt of Sabine, with an average rating of guilt of 75 (out of 100). Furthermore, the initial ratings of guilt in the main study also supported that the case file initially created a belief in guilt. Therefore, it seems that the lack of confirmation bias was not due to a lack of an initial impression of guilt. It may be so that the manipulation failed to create confirmation bias as, despite the impression of guilt, participants were not sufficiently invested in the case. A lack of investment could also have resulted in less cognitive dissonance experienced in response to contradicting information, and subsequently, no preference for supporting information in order to achieve or maintain consonance (Ask, Reinhard, Marksteiner, & Granhag, 2011; Festinger, 1957). Obtaining such an investment from participants would likely require a task that allowed them to be more cognitively or emotionally involved.

Nevertheless, a preference for incriminating information following an initial impression of guilt has been observed in several studies (e.g. Ask & Granhag, 2005; Marksteiner, Ask, Reinhard, & Granhag, 2011; O'Brien, 2009). Rassin (2018) also found that participants showed tunnel vision in response to an incriminating case file. The findings of the current study are thus not in line with previous research, as the initial impression of guilt did not cause a preference

for incriminating information. In previous studies, where information that contradicted the suggested guilty scenario was included, such information was found to be insufficient to protect against confirmation bias (e.g. Marksteiner et al., 2011; O'Brien, 2009). However, the more explicit suggestion of an alternative scenario in the current study, which clearly implicated an alternative perpetrator, may have countered confirmation bias more than the exonerating information included in previous studies. Although Rassin (2018) similarly included an explicit alternative perpetrator, based on the material used by Ask and Granhag (2005), he did not find students to be less influenced by the hypothesis of the main suspect than students who did not receive information about an alternative suspect. A possible explanation for that differing finding, thought to be due to the sample of participants, is described below.

Although the lack of confirmation bias and the appropriate interest in the exonerating materials are, of course, encouraging findings, they were also unexpected based on earlier research. One possible explanation that has also been offered by Maegherman, Ask, Horselenberg, and Van Koppen (2020) is that the lack of confirmation bias might be due to the sample of law students that was used. In several previous studies which observed confirmation bias, the participants included college students (O'Brien, 2009), police officers (Ask & Granhag, 2005) or police trainees (Marksteiner et al., 2011). Law students were chosen as the population for the current study due to the context of the material and the fact that finding real judges to participate in research has proven very difficult. Law students were thought to have an affinity with the reasoning required for the current study, which was not expected from a community sample or psychology students. However, the element that may have been underestimated is that law students are also the lawyers of the future. A possible implication of the findings might therefore be that law students have a keener eye for alternative scenarios, and might have a greater interest in the exonerating evidence, than expected. A similar finding was reported by Rassin et al. (2010), who also used law students as participants. In their study, participants chose mainly exonerating investigations in response to an incriminating case file, although the amount of incriminating investigations increased in response to more evidence and increased severity of the case. One question that remains is whether the critical stance observed in law students in experimental research will ultimately also result in increasingly critical judges in practice, or whether external factors can threaten their openness to alternative explanations.

Another potential explanation for the lack of observed confirmation bias in the current study could be that participants in the control condition were also warned about biases prior to reviewing the case file. However, that seems unlikely to provide a sufficient explanation for the lack of bias when considering the existing evidence for a bias blind spot (Pronin, Lin, & Ross, 2002). The bias blind spot has recently also been documented by Kukucka, Kassin, Zapf, and Dror (2017) in an international survey of forensic examiners. Examiners showed a tendency to acknowledge bias in other domains than their own, as well as in other examiners but not in themselves. Therefore, it seems unlikely that merely informing participants about biases would have protected against confirmation bias in the current study.

While neither of the groups showed confirmation bias, so that ACH could not counter confirmation bias, there were also exploratory findings suggesting that ACH was not properly used by participants. For instance, based on the notes participants made, it was clear that they were mainly making a list of evidence rather than using the ACH matrix as instructed. Secondly, when participants were asked whether they had used the matrix, the majority responded they had not used the matrix. Of those who did, the average rating for how helpful they found the matrix was only just over the midpoint, whereas the difficulty level they reported was just under the midpoint. Therefore, the intended use of ACH did not seem to be adopted by most participants in the ACH condition.

There are several possible explanations related to the ACH method itself as to why ACH did not work or was not used. One of these is the lack of training that participants received. Participants were given a simplified description of the ACH method, which had been pre-tested for clarity. They were also given an example case on which they could practice. They were then asked two control questions about the instructions. Although participants who answered those questions wrong were excluded, we could not check whether they had actually practiced on the example case. Therefore, the instructions participants received may not have been applied by participants. One potential implication is therefore that ACH may require in-depth training before it can be used at all, let alone effectively. Dhami et al. (2019) also found that the majority of intelligence analysts in their study who had been trained and instructed to use ACH did not follow one or several of the steps prescribed by the technique.

Lastly, ACH has also received criticisms for various reasons in more recent research. For instance, Dhami et al. (2019) pointed out that the criteria used in ACH are vague, which can make the judgement process by the analysts unreliable. It is unclear which criteria should be used to determine whether evidence is consistent or inconsistent with a scenario. They also found that those trained in ACH were less consistent in evidence assessment and in their final conclusions, compared to earlier decisions, than those who had not been trained in ACH. Criticisms of ACH have also been voiced in more general terms about structured analytic techniques, which ACH falls under (Chang, Berdini, Mandel, & Tetlock, 2017). The lack of research supporting the effectiveness of these techniques has also repeatedly been used to question their use in practice (Chang et al., 2017; Dhami et al., 2019). Several variations of the ACH method have been suggested in an attempt to improve the original ACH method. These include for instance the argumentation-based ACH (Murukannaiah, Kalia, Telang, & Singh, 2015) and a collaborative, web-based, version of ACH (Convertino, Billman, Pirolli, Massar, & Shrager, 2008). However, further research is needed to determine the effectiveness of both the original and adapted ACH methods in a variety of contexts where confirmation bias exists.

Based on the growing number of studies in which the use of the ACH technique has been questioned, it should be considered whether adapting the ACH method would be sufficient to address its purpose of countering confirmation bias, or whether it would be more beneficial to aim to develop an alternative training method or technique. In

order to do so, it would be useful to further research the ACH method and the identified problems. For instance, perhaps the use of ACH is dependent on the context in combination with extensive training, thereby limiting its applicability. Extensions of the already complex method may also further limit its use, whereas simplifications may result in a lack of application by those trained, which seems to be supported by the findings of the current study. If, in further research, the effectiveness of ACH continues to be found to be problematic, a novel method, based on the core principle of falsification, but designed to be widely applicable, may be a more fruitful area of research.

Beyond the criticisms of the ACH method, limitations of the current research include the fact that the study was conducted online for a portion of the participants. In order to protect against this limitation, a number of attention checks were included throughout the study. These included directed queries (i.e. answering the question in a specified way) and memory tests (Abbey & Meloy, 2017). Furthermore, participants also had to answer control questions about the ACH instruction and about the case. Participants also had the option to review the case at several points during the study, so there was no limit on their access to the necessary material due to the study being online.

Another limitation of the study is that we did not know what participants' expectations were when they chose an investigative question. In other words, we could only judge what they were testing by inference from the question they asked, which may not be a valid measurement of what they tried to find out. In future research, it would be beneficial to ensure that participants also have to indicate what outcome they are expecting.

5 | CONCLUSION

Findings of the current study did not show evidence of confirmation bias among participants in the control condition or the ACH condition, thereby preventing us from conclusively testing whether ACH could protect against confirmation bias. Despite that difficulty, it was nevertheless clear that participants did not adequately make use of the ACH method. In order to validly determine whether ACH can prevent against confirmation bias in the context of criminal proceedings, further research is needed, although it may be advisable to use an adaptation of the ACH method.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Enide Maegherman https://orcid.org/0000-0001-9872-0708

ENDNOTES

[1] The conviction ratings were analysed and can be found in Data S1.

[2] The exploratory analysis of the formulated scenarios can be found in Data S1.

[3] We suspect the large mean and *SD* can be explained by participants not closing the tab immediately after they completed the survey. The longest response time was 4.5 days.

REFERENCES

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Ask, K., & Granhag, P.-A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*, 43–63. https://doi.org/10.1002/jip.19

Ask, K., Reinhard, M.-A., Marksteiner, T., & Granhag, P. A. (2011). Elasticity in evaluations of criminal evidence: Exploring the role of cognitive dissonance. *Legal and Criminological Psychology*, *16*, 289–306. https://doi.org/10.1348/135532510X510153

Bandes, S. (2006). Loyalty to one's convictions: The prosecutor and tunnel vision. *Howard Law Journal*, *49*, 475–494.

Burke, A. (2007). Neutralizing cognitive bias: An invitation to prosecutors. *New York University Journal of Law Liberty*, *2*, 512–530.

Chang, W., Berdini, E., Mandel, D. R., & Tetlock, P. E. (2017). Restructuring structured analytic techniques in intelligence. *Intelligence and National Security*, *33*, 337–356. https://doi.org/10.1080/02684527.2017.1400230

Convertino, G., Billman, D., Pirolli, P., Massar, J. P., & Shrager, J. (2008). The CACHE study: Group effects in computer-supported collaborative analysis. *Computer Supported Cooperative Work (CSCW)*, *17*, 353–393.

Crombag, H. F. M. (2017). Invloeden op rechterlijke beslissingen [Influences on judicial decisions]. In P. J. van Koppen, J. W. de Keijser, R. Horselenberg, & M. Jelicic (Eds.), *Routes van het recht [Routes of the law]* (pp. 385–397). Hague: Boom Juridisch.

Crombag, H. F. M., van Koppen, P. J., & Wagenaar, W. A. (2006). *Dubieuze zaken: De psychologie van strafrechtelijk bewijs* [Dubious cases: Psychology of evidence in criminal law]. Olympus (Original work published 1992).

Dhami, M. K., Belton, I. K., & Mandel, D. R. (2019). The "analysis of competing hypotheses" in intelligence analysis. *Applied Cognitive Psychology*, *33*, 1080–1090. https://doi.org/10.1002/acp.3550

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/bf03193146

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.

Findley, K. A., & Scott, M. S. (2006). Multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, *2006*, 291–398.

Heuer, R. J. (1999). Analysis of competing hypotheses. In *Psychology of intelligence analysis* (pp. 95–110). Washington, DC: CQ Press. https://doi.org/10.1007/s10606-008-9080-9

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An

expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*, 557–571. https://doi.org/10.1037//0022-3514.80.4.557

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*, 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: A global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition*, *6*, 452–459. https://doi.org/10.1016/j.jarmac.2017.09.001

Lehner, P. E., Adelman, L., Cheikes, B. A., & Brown, M. J. (2008). Confirmation bias in complex analyses. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *38*, 584–592.

Maegherman, E., Ask, K., Horselenberg, R. & Van Koppen, P. J. (2020). *Law and order effects: On cognitive dissonance and belief perseverance* (Unpublished manuscript). Department of Criminal Law and Criminology, Maastricht University.

Marksteiner, T., Ask, K., Reinhard, M.-A., & Granhag, P.-A. (2011). Asymmetrical scepticism towards criminal evidence: The role of goal- and belief-consistency. *Applied Cognitive Psychology*, *25*, 541–547. https://doi.org/10.1002/acp.1719

Martin, D. L. (2002). Lessons about justice from the laboratory of wrongful convictions: Tunnel vision, the construction of guilt and informer evidence. *UMKC Law Review*, *70*, 847–864.

Murukannaiah, P. K., Kalia, A. K., Telang, P. R., & Singh, M. P. (2015). Resolving goal conflicts via argumentation-based analysis of competing hypotheses. *IEEE 23rd International Requirements Engineering Conference (RE)* (pp. 156–165). Ottawa, ON: IEEE. https://doi.org/10.1109/re.2015.7320418

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85–95. https://doi.org/10.1080/00335557743000053

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. https://doi.org/10.1037/1089-2680.2.2.175

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*, 315–334. https://doi.org/10.1037/a0017881

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 521–533. https://doi.org/10.1037/0278-7393.14.3.521

Popper, K. (2005). *The logic of scientific discovery*. England: Routledge (Original work published 1959).

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*, 369–381. https://doi.org/10.1177/0146167202286008

Puvathingal, B. J., & Hantula, D. A. (2012). Revisiting the psychology of intelligence analysis: From rational actors to adaptive thinkers. *American Psychologist*, *67*, 199–210. https://doi.org/10.1037/a0026640

Rassin, E. (2018). Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *Journal of Investigative Psychology and Offender Profiling*, *15*, 227–233. https://doi.org/10.1002/jip.1504

Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: An analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling*, *7*, 231–246. https://doi.org/10.1002/jip.126

Schünemann, B., & Bandilla, W. (1989). Perseverance in courtroom decisions. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 181–192). Berlin: Springer.

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *The University of Chicago Law Review*, *71*, 511–586.

Van Koppen, P. J., & Mackor, A. R. (2019). A scenario approach to the simonshaven case. *Topics in Cognitive Science*. Advance online publication. https://doi.org/10.1111/tops.12429

Wagenaar, W.-A., van Koppen, P.J. &, Crombag, H.F.M. (1993) *Anchored narratives: The psychology of criminal evidence*. Hertfordshire: Harvester Wheatsheaf.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281. https://doi.org/10.1080/14640746808400161

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.