Routledge
Taylor & Francis Group

OPEN ACCESS   Check for updates

Accountability in legal decision-making

Enide Maegherman[a,b,c] (iD), Karl Ask[b] (iD), Robert Horselenberg[c] and
Peter J. van Koppen[a,d] (iD)

[a]*University College Maastricht, Maastricht, The Netherlands;* [b]*Department of Psychology,
Gothenburg University, Gothenburg, Sweden;* [c]*Department of Criminal Law and Criminology,
Faculty of Law, Maastricht University, Maastricht, The Netherlands;* [d]*Department of Criminal Law
and Criminology, Faculty of Law, VU University Amsterdam, Amsterdam, The Netherlands*

Having to explain a decision has often been found to have a positive effect on the quality of
a decision. We aimed to determine whether different accountability requirements for judges
(i.e., having to justify their decision or having to explicate their decision) affect evidence
use. Those requirements were compared to instructions based on the falsification principle
and a control condition. Participants ($N = 173$) decided on the defendant's guilt in a murder
case vignette and explained their decision according to the instructions. The explication and
falsification (but not the justification) instructions increased the use of exonerating
evidence. There was no significant difference between the groups in guilt perception. The
use of exonerating evidence was a significant positive predictor of acquittal rates. The
implications for the different forms of instructions in practice are positive, but suggest a
difference between the evidence considered and the evidence used to account for
the decision.

Key words: accountability; confirmation bias; falsification; legal decision-making.

Although the process of legal decision-making has been the subject of a variety of theoretical explanations as well as experimental research, the insight into how judges reach a final decision remains limited. Some aspects of the decision-making process are known, as they are prescribed by law. A requirement for judges in several countries is that they are, to a certain extent, required to account for their decision (e.g. Art. 359 and 360 Dutch Code of Criminal Procedure, DCCP; Mevis, 2019). In previous research on decision accountability, researchers have suggested that such a requirement could substantially alter the decision-making process (Lerner & Tetlock, 2003). In the current study, we used lay participants to investigate whether variations in the instruction on how to account for a decision affect the evidence considered and the decision made on the guilt of a defendant.

Reasoned judicial decisions

As a judge can almost never know for sure what exactly happened, an inherent leap is required for them to become sufficiently

Correspondence: Enide Maegherman, University College Maastricht, Maastricht University, PO Box
616, Maastricht, 6200 MD, The Netherlands. E-mail: enide.maegherman@maastrichtuniversity.nl

convinced about what happened based on the information provided in the evidence. One of the elusive aspects of legal decision-making is how that leap is made. The need and requirements for explaining a decision differ between the various legal systems, but one common expectation is that the explanation will provide some sort of insight into the judicial decision on guilt in criminal legal proceedings. The most important question to be answered by the judge is whether the suspect committed the crime they are accused of (Dreissen, 2007). In order to answer this question, judges in the Netherlands will first study the case file, which is likely to consist of mainly incriminating information (Crombag, 2017), and will then be presented with the prosecution's and the defence's arguments at trial (Verbaan, 2016). The reasoned decision should make it clear that the rules regulating the use of evidence were followed. It can also be seen as an explanation of why the judge was convinced beyond a reasonable doubt that the accused committed the crime (Dreissen, 2007).

Besides the requirements that the reasoned decision has to fulfil, there are several additional functions for why judges in the Netherlands must explain their decisions. Firstly, the explanation of their decision acts as justification for the punishment that follows for the convicted individual. Secondly, the reasoned decision is used to account for the decision to the general public. Thirdly, it informs the various parties involved in the legal proceedings. Lastly, the reasoned decision can serve as a potential quality control by other legal instances, such as the Supreme Court, although that rarely happens in practice (Dreissen, 2007; Verbaan, 2016). Furthermore, it has been argued by Gommer (2007) that there is a need for requiring an explanation due to the potential influence of thought-processes the decision-maker may not be aware of, such as biases. In theory, the explanation serves as a 'rational reconstruction' of what was considered by the judge for the decision (Gommer, 2007).

National differences in accountability requirements

Different legal systems incorporate different instructions on how a decision should be accounted for. Scholars have compared the content of the Dutch requirement to the German requirement for explaining a decision (Dreissen, 2007; Mevis, 2019). Although there is little difference between the codes of criminal procedure in the Netherlands and Germany on that issue, the literature on the explanation requirements makes it clear that the German system imposes stricter requirements on the judge (Dreissen, 2007; Mevis, 2019; Simmelink, 2001). Whereas the German instructions could be interpreted as requiring an *explication*, the Dutch instructions could be interpreted as requiring a *justification* of the decision. In the German system, the judge has to account for their selection and evaluation of evidence, and to pay attention to facts that indicate an alternative, but not accepted, version of events (Dreissen, 2007; Mevis, 2019). Furthermore, there are specific requirements of evidence evaluation. For instance, in cases of contradicting witness statements, the judge has to consider how both statements came about, as well as to explain the discrepancies between them. In the written decision, the judge will have to account for the grounds of his reliability judgement. Overall, the German judge is required to provide a more in-depth explanation of the decision than the Dutch judge. In doing so, the judge shows the decision was made by a professional with integrity rather than by a purely subjective individual (Mevis, 2019).

In the Netherlands, the requirements imposed on the judge to explain or motivate his decision are limited, due to the integrity and professionalism inherently expected of a judge (Mevis, 2019). The explanation provided by the judge does not have to be a reflection of the discussion or consideration that led to the decision. It suffices if the explanation contains arguments that, taken together, *justify* the decision that was rendered

(Reijntjes & Reijnjes-Wendenburg, 2018). The point of view that the selection and evaluation of evidence do not require motivation, with a few exceptions, is in stark contrast to the extensive requirements in the German system (Mevis, 2019). Although Article 360 of the DCCP requires that the judge explicitly accounts for why they consider certain evidence to be reliable, it is limited to evidence where the reliability is questionable (e.g. vulnerable or anonymous witnesses). Compared to the Dutch standards, the German judge has an extensive duty to motivate the decision – the written decision not only needs to include the proven fact and the evidence used, but also needs to explain the selection and evaluation of evidence (Simmelink, 2001). The Supreme Court of the Netherlands appears to be lenient in enforcing the rules regarding the reasoned decision provided by the judge (Dreissen, 2007). The review of the decision by the Supreme Court remains limited following a change in the DCCP in 2005; the judge now explicitly has to explain why their decision differs from the substantiated arguments raised by either the prosecution or the defence. Thereby, the extent of the reasoned decision becomes dependent on the points raised by one of the parties (Dreissen, 2007). The differences between the Netherlands and Germany in their requirements for the reasoned decision raise the question of how these differences affect judges' reasoning with evidence.

Impact of accountability on reasoned decisions

The need to account for the decision on guilt or innocence of the suspect thus appears to differ between legal systems. Researchers have identified several ways in which such accountability can affect the decision-making process (Lerner & Tetlock, 2003). A key aspect of accountability, which determines its effectiveness in reducing cognitive bias, is whether the requirement to account for a decision was known prior to making the decision. Prior accountability, as is the case for judges, is

thought to encourage exploratory reasoning and making an optimal judgement, whereas post-decisional accountability has been found to increase confirmatory and self-justifying reasoning (Lerner & Tetlock, 1999).

One of the frequently considered factors of accountability is the positive effect of having to explain the decision-making process (*process accountability*) versus having to explain the decision itself (*outcome accountability*; Tetlock, 1985). In light of the explanation required of judges in the Netherlands, it appears that their accountability is focused more on explaining the decision itself than on explaining the decision-making process that led to that decision. In fact, the Dutch Supreme Court has ruled that the reasoned decision does not have to reflect the evidence that was considered, but merely the evidence that the final decision could reasonably be based on. The reasoned decision is therefore not a valid reflection of the decision-making process but rather is focused on outcome accountability (Reijntjes & Reijnjes-Wendenburg, 2018).

Another factor that has been found to moderate the effects of accountability on the decision-making process is the audience to whom the decision needs to be accounted for. Researchers have found evidence that the accountable persons shift their opinion towards the perceived opinion of the audience (e.g. Pennington & Schlenker, 1999). However, research on multiple audiences is lacking (Hall et al., 2015). In the case of judges, the audience may hold a range of opinions. For instance, the decision will likely be read by the defendant and their relatives, but possibly also by the complainant and their relatives, as well as the public and other judges. Furthermore, the court of appeal may also read it.

Researchers investigating accountability have mainly focused on other areas of decision-making, and little research has been conducted into accountability in the context of legal decision-making. Tetlock (1983) investigated whether the influence of an initial

impression of guilt can be affected by prior accountability. He found that those who initially received evidence against the defendant first were more likely to find him guilty, but that this primacy effect was reduced by prior accountability. Therefore, prior accountability seems to be able to prevent an initial belief from biasing a decision on guilt, which has obvious positive implications for the requirement of judges to explain their decision.

Assessing quality in legal decision-making

Scholars have suggested that forcing judges to substantiate their decisions can enhance the accuracy of legal decision-making, by ensuring that the decision is not based on irrelevant information or speculation (Cohen, 2015). This notion, however, does not accommodate the intricate effects of accountability on decision-making as demonstrated by the psychological research reviewed above. The lack of understanding concerning the effects of accountability on legal decision-making may be due to the difficulty in assessing what constitutes a good decision in the legal context. Some elements of what the accountability literature considers to be important for a good decision can also be seen in the context of decision-making by judges, such as the notion of impartiality. However, in actual legal decision-making, an objective ground truth is often not available, which makes it difficult to determine the quality of the decision.

The quality of decision-making in general can also be related to underlying processes – for instance, the dual-process theory, in which System 1 is responsible for the fast, intuitive, perhaps biased decisions, whereas System 2 is responsible for the analytical and conscious decisions (Kahnemann, 2011). In the context of legal decision-making, routine, time constraints and the lack of feedback could also increase the risk of resorting to the heuristic thinking of System 1, adversely affecting the accuracy of decision-making (Kahnemann, 2011; Tay et al., 2016). One such bias thought

to be particularly relevant in legal proceedings is confirmation bias (Findley & Scott, 2006), the tendency to seek and interpret information in such a way that it confirms an initial belief, while paying disproportionately less attention to information that could contradict that belief (Nickerson, 1998). In the context of legal decision-making, an excessive focus on the guilt of the suspect could result in miscarriages of justice by insufficiently considering exonerating evidence.

A reasoned decision can give some insight into what was considered in making the decision. For instance, indicators of confirmation bias in a reasoned decision would suggest that the decision-making process may have deviated from its goal of determining the truth. As the determination of the truth is generally considered the aim of criminal proceedings (Cleiren, 2008; Crombag, 2017; De Keijser, 2017), written decisions containing indicators of confirmation bias could be considered as an arguably worse decision, as it would suggest a focus on an existing belief rather than finding the truth.

Based on previous research, the active consideration of alternative scenarios can mitigate the influence of a prior belief (O'Brien, 2009; Rassin, 2018). Therefore, explanations that consider alternative scenarios can be considered indicative of a less biased process of decision-making. Trying to disprove an existing idea, known as falsification, can also be considered an important process when trying to determine what most likely happened, as failed attempts at disproving a theory can act as support for the theory (Crombag et al., 2006). Falsification is closely related to the consideration of alternative scenarios. A scenario can be defined as a chronological or causal description of a central action (Van Koppen & Mackor, 2020). Evidence that disproves one scenario may confirm another scenario. Furthermore, trying to find a good alternative scenario for the available evidence can also be considered part of attempting falsification (Van Koppen & Mackor, 2020).

Although consideration of exonerating evidence and alternative scenarios remain indirect measures of the decision-making process, they can provide insight into whether the evidence considered for the decision, and thus the decision-making process, differs depending on the instruction given to account for the decision.

The current study

In the current study, we aimed to investigate whether prior instructions to account for a decision affect the legal decision-making process. In order to do so, participants were provided with one of four instructions before reading a vignette of a murder case and were then asked to make a reasoned decision on the guilt of the defendant. In the *justification* condition (based on the DCCP), participants were asked to mention evidence that supported their decision, while in the *explication* condition (based on the German Code of Criminal Procedure, GCCP) participants were asked to show that they had considered evidence both for and against their decision. In the *falsification* condition, participants were asked to describe the different possible versions of events and how they decided on the most plausible version by excluding the alternatives. The falsification condition was not based on a specific country. In the fourth condition, which was considered the *control* condition, participants only received the general instruction to explain their decision. After reporting and explaining their decision, participants were asked to rate the individual pieces of evidence in the case for how incriminating or exonerating they found them to be.

We formulated and pre-registered the following hypotheses:

> H1: Those in the justification condition were expected to use less exonerating evidence in their justification of the decision than those in the explication or the falsification condition, but more than those in the control condition.

> H2: Those in the justification condition were expected to consider fewer scenarios than those in the explication or falsification condition, while the control group was expected to consider fewer scenarios than those in the three experimental conditions.

> H3: (a) The justification condition was expected to have a higher conviction rate than the explication or falsification group, (b) but the control condition was expected to have a higher conviction rate than the three experimental conditions. (c) The average rating of guilt was also expected to be higher in the justification condition than in the explication and falsification conditions, (d) while the control condition was expected to have a higher rating of guilt than the three experimental conditions. This hypothesis was based on the idea that consideration of all evidence, including the exonerating evidence, as well as alternative scenarios would raise more doubt about the guilt of the suspect, and thus result in fewer convictions (Tenney et al., 2009).

> H4: The amount of exonerating evidence mentioned in the written decision was expected to be a significant negative predictor of the conviction rate of the suspect.

The pre-registration for the study can be found at: https://osf.io/fc962/?view_only=574 6019c60bf4a4e84ef103627a0a0e8

Method

Participants

Participants were recruited using Amazon MTurk, as well as via advertisements through social media. The survey platform Qualtrics was used for data collection. A power analysis was conducted in G*Power (v3.1; Faul et al., 2007). A medium effect size ($f = 0.25$) was estimated based on earlier research on excluding alternative scenarios, as the accountability literature offered no comparable studies that could be used to estimate effect size. Using a power of .8 and a .05 Type I error rate resulted

in a required sample size of 179 participants. To allow for potential exclusions, considerably more responses were collected. A large number of participants ($n = 366$) did not answer the control questions about the instructions correctly and were screened out from the survey at an early stage. Responses were also excluded for incorrect answers to the control questions about the case ($n = 33$) or to the attention checks ($n = 37$), and for open-ended answers that we suspected were not genuine (e.g. bots, duplicate responses; $n = 49$). Incorrect answers to the control questions and the attention checks had been pre-registered as exclusion criteria. One of the control questions was ultimately not used as an exclusion criterion because the answer was not clear enough from the vignette. Another eight participants were excluded because their rating for likelihood of guilt differed from the sample median by more than three times the absolute median deviation, and they were therefore thought to be outliers (Leys et al. 2013), which had also been pre-registered as an exclusion criterion. The final sample consisted of 173 participants. Participants received compensation through MTurk. Participants' mean age was 31 years old ($SD = 11.14$). The majority of participants (58%) were female. Ethical approval for this study was obtained from the ethical committee at Maastricht University.

Materials

Instructions

Participants were randomly assigned to one of four conditions, each of which received a different instruction on how to motivate the decision they had made (Appendix A). These were the *justification* condition (based on the DCCP), the *explication* condition (based on the GCCP), the *falsification* condition (based on the principle of falsification) and the *control* condition (in which participants were given only minimal instructions). The various instructions were constructed after consultation of the literature on the requirements for judges in the different countries to account for

their decision. They had also been pre-tested to ensure that they were understandable using multiple choice questions about the meaning of the instructions. The instructions were deemed understandable when all pre-test participants answered all multiple choice questions correctly within the number of attempts allowed, which differed according to the number of elements within the instruction. These questions were also used as control questions during the experiment so that participants who did not understand the instructions could be removed, as included in the explanation on the procedure.

Practice vignette

In order to familiarise participants with the instructions, they received a practice vignette and were asked to make a decision and motivate it. The case concerned a burglary, where a suspect had been charged for the crime, but was accusing someone else. The example was a simple task that allowed participants to practise applying the instructions.

Case vignette

Participants were then presented with a vignette of a fictional murder case (Appendix B). In the case, Emma Miller claimed to have found her husband James dead when she arrived home from seeing her friend. Emma was covered in blood when the police arrived. The case contained information about James having had an extra-marital affair, and that Emma knew about the affair. Emma was described as the main suspect based on the evidence against her. However, the case also contained a few indications of an alternative scenario, namely that James' mistress may be the perpetrator. The pieces of evidence in the case were pre-tested for the extent to which they were perceived as incriminating or exonerating. As intended, the case was perceived as indicating that Emma was guilty of killing James, with an average likelihood of guilt

rating of 69.1 (*SD* = 16.9) on a 0–100 scale in the pre-test (*N* = 71).

Measures

Case judgments

After writing their reasoned decision, participants were asked to rate how likely it is that the main suspect, Emma, killed James, using a visual analogue scale from 0 (*not at all likely*) to 100 (*very likely*). Following that rating, participants were asked whether or not they would convict Emma for murdering James by selecting one of two options (*acquit*/*convict*). After making their decision, participants then had to rate how confident they were about their conviction decision on a visual analogue scale from 0 (*not at all confident*) to 100 (*very confident*).

Valence ratings of evidence

Participants were asked how exonerating or incriminating they found individual pieces of evidence to be. They did so by using a visual analogue scale from 0 to 100, where 0 means completely exonerating and 100 means completely incriminating. In order not to influence participants' judgments in either direction, the starting position of the slider was set to 50 when participants were first presented with the scale. These ratings were not included in the hypotheses or used in the main analyses, but can be found in the Supplemental Materials. They were not included in the main analyses as it could not be determined whether participants used the evidence as predicted by the pre-test results.

Procedure

All participants completed the study online using Qualtrics. Participants were first welcomed to the study and provided with some information about the study, such as that they would have to judge the guilt of the defendant and have to explain their decision. They then provided informed consent before starting the study. Participants first filled in a short demographic section, including their age and educational background. In the next section, participants were randomised to one of the four experimental conditions and were given the instructions to explain their decision according to the condition. Multiple choice control questions about what participants were asked to do according to the instructions were included here to ensure that the instructions were correctly understood by participants. If participants did not answer all control questions correctly, they were directed back to the instructions and could then attempt the questions again. There was one control question for each element of the instruction, which resulted in two questions for the control condition, three questions for the justification and falsification condition, and four questions for the explication condition. Participants could attempt the questions twice in the control condition, three times in the justification and falsification condition, and four times in the explication condition. If, after the final attempt, they still did not answer all questions correctly, they were taken to the end of the survey and did not continue to the actual study.

The participants were then given a short practice vignette depicting a burglary case and were asked to decide on the guilt of the defendant and explain their decision, thereby familiarising themselves with the instructions. Participants were told that they were required to write a reasoned decision after reading the actual case vignette. In order to increase their sense of accountability, participants were told that their explanation would be reviewed by a panel of judges to determine how well they explained their decision according to the instructions. In the next section, participants were presented with the actual case vignette, and were asked to write a reasoned decision about the case. While writing their decision, participants were able to refer back to the case description, which was presented on the same page. In the final section, participants filled in the measures described above, first deciding

on the dichotomous verdict and the likelihood of guilt, then rating the valence of the evidence. Here, again, they could revisit the case description. Participants were thanked for taking part in the study and received further information about the aims of the study. The median response time was 35 minutes and eight seconds.

Coding of the reasoned decisions

The evidence that participants used in their decision was coded according to 21 different categories. A pre-test was done using the case ($N = 71$) prior to main data collection. Participants were asked to indicate the likelihood that the suspect was guilty on a scale from 0 to 100, and were then asked to rate the evidence in the case file on a scale from 0 (*exonerating*) to 100 *(incriminating)*. Based on the results of the pre-test, five categories of evidence were determined for the evidence in the case file: strongly incriminating (14 items), mildly incriminating (4 items), neutral (9 items), mildly exonerating (4 items) or strongly exonerating (3 items). The categories were determined based on the mean ratings of the evidence and their confidence intervals (CIs). The strongly exonerating category had a mean rating below 40, and the mildly exonerating category had a mean between 40 and 50 with an upper 95% CI bound no greater than 50. The neutral category had a mean between 40 and 60, with an upper 95% CI bound crossing 50. The mildly incriminating category had a mean between 50 and 60, and a lower 95% CI bound above 50. The strongly incriminating category had a mean rating above 60.

For each of the five categories determined by the pre-test, participants could use the evidence as incriminating, neutral or exonerating, resulting in a total of 15 pre-specified categories. Seeing as the 15 categories were used for the development of the material, these were also coded initially. However, as it was more informative to determine how the evidence was used, those 15 categories were then combined into the three categories that were used

to test the hypotheses, namely incriminating, neutral or exonerating evidence, as used by the participant. Using the evidence according to the coding categories of use of the evidence allowed for better incorporation of participants' interpretation of the evidence into the conclusions drawn based on the data.

However, during the coding, it became clear that participants were sometimes not specific about which evidence they referred to. For instance, several participants mentioned that Emma had a 'motive'. As there were several pieces of evidence that related to a motive, such as the fact that Emma knew James was having an affair and that she would benefit financially more from his death than from a divorce, we decided to code the mention of such evidence as 'unspecified'. Some participants also misremembered information that was provided in the case. For instance, they mentioned that DNA was found on the murder weapon, whereas the case only specified that fingerprints were found on the murder weapon. Therefore, an additional category was created for 'misremembered' evidence. For both the unspecified and misremembered categories, a distinction was also made for whether the evidence was used as incriminating, neutral or exonerating. Therefore, an additional six categories were created, resulting in a total of 21 categories to be coded. All the incriminating categories were combined into one incriminating category, which was used for the final analyses. The exonerating and neutral categories were also combined into one overall exonerating category and one overall neutral category.

All the responses were coded by one coder. The coder was trained using the coding of the evidence according to the 21 categories that were possible on the basis of the classification of the evidence. By doing so, the coders were not focused on the final categories, which served as a form of protection against potential bias in terms of the categories used for the final analyses. The coder was trained by the experimenter about the different categories,

Table 1. ICC for the coded evidence and scenarios in the written decisions.

		95% CI	
Category	ICC	lower bound	upper bound
Incriminating evidence	.949	.906	.972
Neutral evidence	.581	.340	.751
Exonerating evidence	.911	.841	.951
Scenarios implicating Emma	.816	.682	.896
Scenarios implicating James' mistress	.859	.752	.921
Scenarios implicating other perpetrators	.792	.645	.882

Note: ICC = intra-class correlation coefficient; CI = confidence interval.

and a few responses were coded to practice. A second coder, who was trained by the first coder, coded 21 of the responses (12%) in order to assess the inter-rater reliability of the coding. The intra-class correlation coefficient (ICC) for a two-way, random, single-measures, consistency analysis was conducted for the categories of exonerating, incriminating and neutral evidence, as used by the participants. The ICCs for the categories of evidence are reported in Table 1. As only the coded exonerating evidence by Coder 1 was used to test the study's hypotheses, the ICCs were considered good to excellent (Koo & Li, 2016).

Finally, the scenarios mentioned by the participants were coded according to the implicated perpetrator who committed the central action, namely killing James: Emma, James' mistress or 'other'. An intra-class correlation coefficient for a two-way, random, single-measures (consistency) analysis was conducted for each of the scenario categories (see Table 1). The ICCs for the scenario categories were good.

Design and analysis

The experiment included four independent conditions to which participants were randomly assigned. A number of dependent variables were used to test the hypotheses. A one-way between-groups analysis of variance (ANOVA), using the amount of exonerating evidence used in the reasoned decisions as dependent variable, had been planned to test H1. A similar one-way ANOVA, using the number of scenarios mentioned by participants as dependent variable, had been planned to test H2. A Pearson's chi-square analysis had been planned to determine whether the conditions differed in their decision to convict the participant, as predicted in H3. Furthermore, a one-way between-groups ANOVA, using the rating of likelihood of guilt as dependent variable, had been planned to test H3. Comparisons between conditions, individually or combined in accordance with the hypothesis, was done through planned contrasts for all further analyses. Lastly, a point-biserial correlation coefficient had been planned to determine whether the amount of evidence coded as exonerating in the written decisions was significantly associated with participants' decision to convict or acquit the main suspect (H4).

Results

Use of exonerating evidence (H1)

A one-way between-groups ANOVA was conducted to compare the amount of evidence coded as exonerating for each of the conditions (for means in each condition, see Table

Table 2. Descriptive statistics for the number of exonerating pieces of evidence mentioned in each of the conditions.

Condition	n	M	SD
Justification	45	1.31	1.84
Explication	40	1.90	1.77
Falsification	45	1.76	1.92
Control	43	0.79	1.26

Table 3. Number of scenarios considered per condition and type of scenario.

Condition	n	Mean number of scenarios			
		Emma	Mistress	Other	Total
Justification	45	1.00 (0.43)	0.42 (0.50)	0.29 (0.46)	1.71 (0.84)
Explication	40	1.00 (0.55)	0.70 (0.69)	0.28 (0.45)	1.98 (0.97)
Falsification	45	1.07 (0.45)	0.80 (0.59)	0.20 (0.40)	2.07 (1.05)
Control	43	1.00 (0.31)	0.30 (0.46)	0.21 (0.41)	1.51 (0.80)

Note: Standard deviations in parentheses.

2). As the assumption of homogeneity of variance was violated, a Welch ANOVA was conducted instead. A significant difference between the conditions was found, $F(3, 92.26) = 4.64$, $p = .005$, $\eta_p^2 = .131$, 90% CI [.026, .219]. A planned contrast was used to compare the mention of exonerating evidence in the justification condition to the combined falsification and explication conditions. There was no significant difference, $t(89.6) = 1.52$, $p = .132$, Hedges' $g = -0.28$, 95% CI [−0.64, 0.09]. Another planned contrast was conducted to compare the exonerating evidence mentioned in the justification condition to the exonerating evidence mentioned in the control condition. No significant difference was found, $t(78.1) = -1.55$, $p = .125$, Hedges' $g = 0.33$, 95% CI [−0.09, 0.75]. Hypothesis 1 was therefore not supported.

An additional exploratory contrast was conducted to test whether there was a significant difference between the combined explication and falsification conditions compared to the control condition for the mention of exonerating evidence. A significant difference was found, $t(114.08) = 3.73$, $p < .001$, Hedges' $g = 0.50$, 95% CI [0.24, 0.99], indicating that those in the combined explication and falsification conditions mentioned significantly more exonerating evidence than did those in the control condition.

Scenarios considered (H2)

An overview of the number of scenarios considered per condition and scenario type can be found in Table 3. A one-way between-groups ANOVA was conducted to compare the total number of scenarios that were mentioned in the written decisions across conditions. The analysis showed a significant effect of condition, $F(3, 169) = 3.26$, $p = .023$, $\eta_p^2 = .055$, 90% CI [.004, .106]. A planned contrast was conducted to contrast the control condition to the three experimental conditions combined. A significant difference was found, $t(169) = 2.50$, $p = .013$, Hedges' $g = -0.44$, 95% CI [−0.79, −0.10]. The justification condition was further contrasted to the explication and falsification condition, which were weighted together for the contrast due to the conceptual similarity. No significant difference was found, $t(169) = 1.82$, $p = .070$, Hedges' $g =$

Table 4. Likelihood of guilt ratings across conditions.

Condition	n	M	SD
Justification	45	68.53	23.99
Explication	40	66.58	23.11
Falsification	45	66.40	20.48
Control	43	74.23	19.06

−0.31, 95% CI [−0.04, 0.69]. Hence, Hypothesis 2 was partially supported.

Perception of guilt (H3)

Conviction rates (H3a and H3b)

A Pearson's chi-square analysis was used to determine whether the groups differed in their decisions on whether or not to convict the suspect. Although participants in the control condition were more likely to convict the main suspect (61.7%) than were participants in the justification (46.8%), explication (48.8%) and falsification (47.8%) conditions, there was no significant difference between the conditions, $\chi^2(3) = 4.33$, $p = .228$, $V = .123$. Hence, Hypotheses 3a and 3b were not supported.

Likelihood of guilt (H3c and H3d)

A one-way between-groups ANOVA was conducted to compare participants' ratings of the likelihood that Emma killed James across the conditions (see Table 4). No significant difference was found, $F(3, 169) = 1.22$, $p = .305$, $\eta_p^2 = .021$, 90% CI [.000, .054]. Thus, Hypotheses H3c and H3d were not supported.

Mention of exonerating evidence and conviction rates (H4)

A point-biserial correlation coefficient had been planned to determine whether the amount of evidence coded as exonerating in the decisions was significantly associated with the binary measure of whether or not the participant would convict the main suspect. However, the Shapiro–Wilk test showed that the assumption of normal distribution was violated for the amount of exonerating evidence mentioned,

$W(173) = .798$, $p < .001$. Therefore, a binary logistic regression was conducted instead. The amount of exonerating evidence mentioned by the participant was found to be a significant predictor of their decision on guilt, $\chi^2(1) = 50.35$, $p < .001$, odds ratio (OR) = 2.18, 95% CI [1.67, 2.85], showing that participants who mentioned more exonerating evidence were more likely to acquit the suspect. Thus, Hypothesis 4 was supported.

Discussion

In the current study, we aimed to determine whether detailed instructions to account for a legal decision influenced the evidence and scenarios considered by participants. Participants received instructions that were based on either the Dutch (justification) or the German (explication) Code of Criminal Procedure, based on the principle of falsification, or conveyed only general instructions to account for their decision (control). Although there was a significant difference between conditions for the amount of exonerating evidence they mentioned in their decisions, there was no significant difference between the justification condition and the explication and falsification condition. There was also no significant difference between the justification and control conditions, contrary to our expectations. The expectations were based on the idea that the focus on supporting the chosen scenario would result in less consideration of the exonerating evidence (e.g. O'Brien, 2009). The findings suggest a lack of influence from the justification instruction on the use of exonerating evidence. On the other hand, an exploratory analysis showed that the combined explication

and falsification conditions used significantly more exonerating evidence than the control condition. As the justification condition did not differ from the control condition, while the explication and falsification conditions did, it seems the instructions did have a significant effect on the consideration of exonerating evidence, as was expected on the basis of the accountability literature (Lerner & Tetlock, 2003; Tetlock, 1983). The observed pattern could be due to the same mechanism underlying our hypothesis, namely that the justification condition did not include the consideration of alternative scenarios, nor did it encourage the consideration of evidence not supporting the decision (O'Brien, 2009; Van Koppen & Mackor, 2020). Therefore, we had expected the justification condition to be biased towards the guilt of the suspect and the incriminating evidence. Although the difference was not significant between the justification and the other conditions, the fact that it did not differ from the control condition, while the explication and falsification conditions did differ from the control condition, suggests that participants in the justification condition may indeed have been primarily focused on the guilt of the suspect. Although only the explication condition was not contrasted with the control condition, the falsification and explication conditions differed significantly from the control condition, and the descriptive statistics indicate the highest average use of exonerating evidence in the explication condition. Subsequently, the specific request to include evidence beyond that which supports the decision, as in the German but not in the Dutch Code of Criminal Procedure, seems to have resulted in the consideration of more exonerating evidence.

It was expected that the justification condition would consider fewer scenarios than the explication or falsification condition, and that the control condition would consider fewer scenarios than the three experimental conditions. This hypothesis was based on the fact that the instructions in the justification condition did not articulate the need to consider other scenarios (Mevis, 2019) and the fact that the control instructions did not mention alternative scenarios at all. The hypothesis was partially supported, as the control condition did include fewer scenarios when contrasted with the combined justification, explication and falsification conditions, as well as according to the descriptive statistics. However, the justification condition did not differ from the explication and falsification conditions. Thus, while the detailed instructions in the experimental conditions do seem to have increased the consideration of alternative scenarios, the specific emphasis on alternative scenarios in the explication and falsification conditions was not sufficient to produce a further increase relative to the justification condition. It is therefore unclear whether the explicit instruction to consider alternative scenarios contributes to the use of alternative scenarios above and beyond the effect of providing detailed accountability instructions.

Contrary to our expectations, there was no significant difference in the perception of guilt, in either conviction rates or ratings of likelihood of guilt, between the different conditions. However, we did find, as predicted, that participants who mentioned more exonerating evidence were more likely to acquit the defendant. A possible interpretation of these findings is that, while accountability instructions effectively influenced the consideration of exonerating evidence, the effect was not sufficiently large for it to carry over and influence the global perception of the case. According to that reasoning, the consideration of exonerating evidence can be considered a mediator, which causally precedes the outcome variables (i.e. guilt perception and conviction decision). On methodological grounds, it is plausible that a manipulation exerts a stronger influence on the more proximal outcome (i.e. the mediator) than on the more distal outcome (i.e. the dependent variable). With limited statistical power (as in the current study) the proximal effect may be captured,

whereas the distal effect goes undetected. Future research should investigate the viability of such causal process models in the context of legal decision-making.

Another possible explanation for the latter finding is that participants in the control and justification conditions, despite not mentioning it in their reasoned decision to the same extent as those in the explication and falsification conditions, did take the exonerating evidence into consideration equally when making their decision. That account would be consistent with research showing that accountability may influence *how* the reasoning occurs rather than *what* the reasoning includes (Hall et al., 2015). It is also in line with the notion that written decisions may contain the evidence that the decision rests on, rather than the evidence that was actually considered in order to come to the decision (Reijntjes & Reijnjes-Wendenburg, 2018). In actual practice, however, the written decision also serves important communicative purposes (Verbaan, 2016). According to the European Court of Human Rights (2019), the reasoned decision is used to show the different parties that they have been heard, which should help them accept the decision that has been made. The reasoned decision is also important to enable parties to use their right to appeal. For those purposes, a discrepancy between what was considered and what is written down could result in a lack of understanding by the parties, and possibly impede the process of appeal. Therefore, decisions made under explication and falsification instructions seem to better serve some of the purposes of a reasoned decision.

In the current study, the communicative functions of the decision were not made explicit to the participants. Instead, they were told that their decision would be evaluated by a panel of experienced judges in terms of how well they followed the instructions. In previous studies on accountability, it has been suggested that the effect of accountability on decision-making is due to wanting to be viewed positively by others and to avoid receiving criticism (Simonson & Nye, 1992). The effect of accountability we observed may therefore differ from the effects in practice, where judicial reasoned decisions are evaluated, or observed, by several parties, including a court of appeal as well as the defence (Verbaan, 2016). Furthermore, personal consequences, such as receiving criticism, of the reasoned decision were also not included in the current study, although they are likely to affect judges in practice. Further research could determine whether specifying the audiences in accordance with the audience for real-life judges increases the influence of accountability on the decision-making process.

Limitations and future research

It should be taken into consideration that the current study cannot ascertain whether the need to account for a decision could counter the influence of a prior belief, as expressed in confirmation bias (e.g. Kassin et al., 2013; Nickerson, 1998). We intentionally did not measure the initial perception of guilt prior to participants writing their decision, as stating a hypothesis in itself can cause a preference for information supporting that hypothesis (O'Brien, 2009). We therefore anticipated that stating a hypothesis might cloud the effect of giving a reasoned decision. However, based on the pre-test, the case description used was biased towards the guilt of the suspect, which, based on previous research (e.g. Ask et al., 2008; Eerland & Rassin, 2012; Rassin et al., 2010), was expected to result in a bias towards incriminating evidence. As indicated by the conviction rates, a strong guilt bias was not observed, suggesting that the need to explain the decision may have countered the influence of the biased initial information across all conditions (including the control condition). That observation supports Gommer's (2007) argument that the requirement for an explanation in itself serves as a countermeasure against the potential influence of bias.

Another limitation of the study is that we cannot be sure how the written decision

provided by participants corresponds to their actual consideration of the evidence. Although their interpretation of the evidence was coded, it is particularly difficult to determine what weight participants assigned to the evidence. In most legal systems, there are limited legal rules regulating the weighing of evidence in order to become convinced and make a decision. For instance, there are no legal rules concerning how much weight should be attributed to a witness statement, or to no DNA match being found. There are, however, minimum rules of what evidence is required, such as the rule that one witness is not sufficient [DCCP, Art. 342 (2)], or that a confession cannot be the only evidence used for a decision [DCCP, Art. 341 (4)].

Furthermore, as participants were specifically asked to include certain elements in their reasoned decision, dependent on their condition, the extent to which the mention of the evidence means it was actually considered remains unclear. However, as the mention of exonerating evidence was a significant positive predictor of the decision to acquit, we can tentatively conclude that the inclusion of exonerating evidence also means participants attached value to the exonerating evidence. Further research into the weighing of evidence could also contribute to the understanding of reasoned decisions.

A further limitation of the current study is that no condition without instruction was included in the design. Although we intentionally did not include such a condition as it would not be a realistic representation of judicial decisions in practice, it also limits the conclusions that can be drawn on the basis of our findings. The mere requirement to explain a decision does not inform on the type of reasoning that was used. Despite the lack of ecological validity, a condition without instruction would have allowed us to compare the need to account for a decision to not having to explain a decision at all. It would therefore be advisable, in future research, to include a condition that requires no explanation at all, in addition

to the detailed instructions researched in the current study.

Lastly, a limitation of the current study relates to statistical power and precision. First, the sample was somewhat smaller than we had initially aimed for. Second, our power analysis was based on the detection of a medium-sized effect. Therefore, the current study was acceptably powered to detect only a medium-to-large effect size. As a consequence, we cannot exclude the existence of smaller effects that may nonetheless be of practical relevance. Furthermore, the observed effect sizes in the current study were not precisely estimated (as indicated by wide confidence intervals), which means that we may have over- or underestimated the true underlying relationships. The accumulation of further empirical evidence is necessary to establish the robustness and validity of our statistical conclusions. Similarly, although using real judges would likely have been more informative, we would not have been able to reach the required sample size. If, in future research, a large-enough sample of real judges can be included, that would have considerable added value to the findings in the current study. Using judges would greatly increase the ecological validity, as well as provide insights into the effect of training and experience on the weighing of evidence, both on its own and in combination with other factors, such as the instructions studied in the current study.

Conclusions

Overall, the findings of the current study have positive implications for the requirement imposed on judges to explain their decision in a written decision. Our findings indicate that variations in the instructions as to how (mock) judges should explain their decisions can influence the type of evidence considered when making the decisions. We did not find evidence, however, that instructions focusing primarily on incriminating evidence (as dictated by the DCCP) negatively affect the final

decision on guilt. The lack of statistical power, however, prevents us from concluding that the effectiveness of instructions based on the principles of justification, explication and falsification does not differ. Although the explication and falsification instructions led to an increased use of exonerating evidence compared to control instructions, this did not translate into differences in the final decision on guilt. This finding suggests that a key component of the GCCP – the requirement of judges to explicate their decision and consider alternative scenarios – may improve the transparency and thoroughness of reasoned decisions.

Ethical standards

Declaration of conflicts of interest

Enide Maegherman has declared no conflicts of interest

Karl Ask has declared no conflicts of interest

Robert Horselenberg has declared no conflicts of interest

Peter van Koppen has declared no conflicts of interest

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee of Maastricht University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study

Acknowledgements

Supplemental material

Supplemental data for this article can be accessed online at https://dx.doi.org/10.1080/13218719.2021.1904452.

ORCID

Enide Maegherman [iD] http://orcid.org/0000-0001-9872-0708

Karl Ask [iD] http://orcid.org/0000-0001-5093-5902

Peter J. van Koppen [iD] http://orcid.org/0000-0001-5781-6869

Data availability statement

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

References

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology*, *22*(9), 1245–1259. https://doi.org/10.1002/acp.1432

Cleiren, C. P. M. (2008). Het monopolie op de strafrechtelijke waarheidsvinding [The monopoly on truth-finding in criminal law. *Strafblad*, *3*, 272–285.

Cohen, M. (2015). When judges have reasons not to give reasons: Comparative law approach. *Washington and Lee Law Review*, *72*, 483–572.

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

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

De Keijser, J. W. (2017). *Als de waarheid eraan moet geloven: Alledaagse bedreigingen voor waarheidsvinding in het strafprocess* [When the truth is taken down: Everyday threats for truth-finding in criminal law procedure]. Boom Juridisch.

Dreissen, W. H. B. (2007). *Bewijsmotivering in strafzaken* [Reasoned decisions in criminal law cases]. Boom Juridisch.

Eerland, A., & Rassin, E. (2012). Biased evaluation of incriminating and exonerating (non)-evidence. *Psychology, Crime & Law*, *18*(4), 351–358. https://doi.org/10.1080/1068316X.2010.493889

European Court of Human Rights. (2019). Guide on article 6 of the European convention on human rights: Right to a fair trial (criminal limb). Retrieved from: https://www.echr.coe.int/documents/guide_art_6_criminal_eng.pdf

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*(2), 175–191. https://doi.org/10.3758/bf03193146

Findley, K. A., & Scott, M. S. (2006). Multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, *2*, 291–398.

Gommer, H. (2007). Onbewuste denkprocessen maken motivering tot noodzaak [Subconcious thoughtprocesses necessitate reasoned decisions]. *Trema*, *4*, 127–134.

Hall, A. T., Frink, D. D., & Buckley, M. R. (2015). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, *38*(2), 204–224. https://doi.org/10.1002/job.2052

Kahnemann, D. (2011). *Thinking, fast and slow*. Straus and Giroux.

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*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

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

Lerner, J. S., Tetlock, P. E. (2003). Bridging individual, interpersonal, and institutional approaches to judgment and decision making: The impact of accountability on cognitive bias. In S. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (Cambridge Series on Judgment and Decision Making, pp. 431–457). Cambridge University Press. https://doi:10.1017/CBO9780511609978.015

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

Mevis, P. A. M. (2019). Modernisering van het strafprocesrecht op z'n Duits [Modernisation of criminal procedural law in the German way]. *Delikt & Delinkwent*, *7*, 40–47.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(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*(4), 315–334. https://doi.org/10.1037/a0017881

Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin*, *25*(9), 1067–1081. https://doi.org/10.1177/01461672992512001

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*(2), 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*(3), 231–246. https://doi.org/10.1002/jip.126

Reijntjes, J. M., & Reijnjes-Wendenburg, C. (2018). De Bewijsconstructie [The evidence construction]. In *Handboek Strafzaken 34: Bewijs* [EPUB version]. Wolters Kluwer. Retrieved from: https://www.navigator.nl

Simmelink, J. B. H. M. (2001). Bewijsrecht en bewijsmotivering [Evidence law and reasoned decisions]. In M. S. Groenhuijsen & G. Knigge (Ed.), *Het onderzoek ter zitting* (pp. 397–453). Rijksuniversiteit Groningen.

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*(3), 416–446. https://doi.org/10.1016/0749-5978(92)90020-8

Tay, S. W., Ryan, P., & Ryan, C. A. (2016). Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. *Canadian Medical Education Journal*, *7*(2), e97–103. https://doi.org/10.36834/cmej.36777

Tenney, E. R., Cleary, H. M., & Spellman, B. A. (2009). Unpacking the doubt in "beyond a reasonable doubt": Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology*, *31*(1), 1–8. https://doi.org/10.1080/01973530802659687

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, *46*(4), 285–292. https://doi.org/10.2307/3033716

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, *48*(3), 227–236. https://doi.org/10.2307/3033683

Van Koppen, P. J., & Mackor, A. R. (2020). A scenario-approach to the Simonshaven case. *Topics in Cognitive Science*, *12*(4), 1132–1151. https://doi.org/10.1111/tops.12429

Verbaan, J. H. J. (2016). *Straf(proces)recht begrepen [Criminal (procedural) law understood]*. Boom Juridisch.

Appendix A: instructions

Justification condition

Please explain your decision on the guilt of the defendant.

- The decision should rest on the evidence that you mention in your verdict.
- Your verdict should include facts and circumstances that give reasons for your decision.
- If your decision differs from explicitly substantiated points raised by either the prosecution or defence**, give reasons for this.

**Points that the prosecution or defence provide evidence to support or prove the truth of.

Explication condition

Please explain your decision on the guilt of the defendant.

- Your verdict should specify what relevant facts are deemed to be proven or not proven.
- Demonstrate that you considered and evaluated all relevant facts and circumstances both for and against your belief in judging the likelihood of the defendant's guilt.
- Explain any obvious alternative scenarios that are equally consistent with the facts as the scenario you decided on.
- Explain how you determined the weight of the individual pieces of evidence you considered.

Falsification condition

Please explain your decision on the guilt of the defendant.

- Your verdict should describe the different possible versions of the events that you considered.
- Use the available evidence to explain how you excluded alternative scenarios.
- Explain how the evidence supports your decision to convict or acquit the defendant.

Control condition

Please explain your decision on the guilt of the defendant.

- Describe how you came to your decision.
- Your verdict should refer to the available evidence.

Appendix B: case vignette

On Monday the 23rd of January 2017, Emma Miller, James Miller's wife, found her deceased husband lying on his back on the bed in the bedroom of their suburban home.

Upon her discovery, Emma called the emergency services and told the operator about what had just happened. After being informed by the operator, the police immediately rushed to the scene of the crime. When they arrived at the Miller home they found Emma covered in blood sitting next to her dead husband's body. It immediately became clear that James had multiple stab wounds in his chest.

Emma Miller was interviewed by the police. She claims to have left the house around 19:30 to visit her friend Catherine Hughes. Since James was visiting his parents and therefore not at home when she left, Emma claims she locked the front door to the house. Emma arrived at her friend's house around 20:30 but the police consider it suspicious that it took Emma an hour to get to her friend's house while this trip should normally only take her 30 min. According to Emma, she stopped by her office on the way, but this could not be confirmed. Emma claims to have left Catherine's house around 21:50 and arrived back home around 22:15. When she arrived home, she noticed the front door was unlocked. When she called James' name but did not get a response, she decided to go look for him. This is when she found James dead on the bed they shared.

The police immediately start a large-scale investigation to clarify what happened to James Miller. Various pieces of forensic evidence were found during the investigation of the crime scene. The Technical Criminal Investigation Department found the victim lying in a pool of blood. Furthermore, they found bloody fingerprints on the edge of the bed that turned out to belong to Emma. They also found traces of blood on the wall behind the bed. On the pillows, they found both long brown and long blonde hairs. Emma's DNA was found at the crime scene and on James' body. DNA from an unidentified woman was found on the door handle of the bedroom door. In the bathroom sink, it was clear that Emma had washed her hands. The sink contained traces of James' blood and there was a bloody fingerprint on the tap. The fingerprint was Emma's. The police believe she was trying to wash away traces of evidence.

An autopsy of the victim's body confirmed that the stab wounds in the chest had been the cause of death. The stab wounds seemed to have been caused by a right-handed person, but the medical examiner was not certain about this. Time of death was between 19.30 and 20.30. It seems as if James had had sexual contact with a woman shortly before he died.

In order to find out who might have had a motive to kill James, the police start interviewing friends and family. Amongst the interviewees were two of James' friends: John Taylor and Paul Baker. John stated that James told him a few months ago that he was having an affair. Paul confirmed John's story and stated that James also told him about the affair, but about a week before James told John. Neither of the witnesses could confirm who the mysterious mistress is, but both testify that they had seen him talking to a brunette on Thursday January 19th in the bar where they always play darts. Judging from how James was communicating with her and gently touching her, they were under the impression that their friend and the unknown woman were intimate with each other. The only thing John and Paul can confirm is that the woman was not Emma, as she has blonde hair, not brown. Eventually, police were unable to track James' presumed mistress down. According to John Taylor, James had been planning on ending the affair as soon as possible, because he could tell Emma was very suspicious. John thought James might have planned to meet his mistress on Monday night, as James had said he could not meet at the bar that night.

After finding out about James' mistress, the police now suspect Emma has killed her husband out of anger over the affair. Friends of Emma told the police that she had previously threatened to hurt James if he cheated on her. A few days ago, she had told one of her friends that she thought James was

cheating on her and that she was looking for proof. Emma showed almost no emotion when talking to the police about James' death. According to the prenuptial agreement, Emma and James would divide their possessions equally in case of divorce, but if one of them died, the other would get everything. In order to get a clear timeline of the events that night, they decide to interview Emma's friend Catherine about the fatal night. Catherine confirms Emma's story completely and states that her friend was with her that night at the times indicated by Emma. However, through further investigation, the police find out that Catherine owes Emma a large sum of money and now believe that this is a valid reason for providing Emma with a false alibi for that night. Emma also received a parking ticket at 21:15 while being parked outside of her friend's house.

The police also interviewed the neighbours. One of them claimed to have heard screaming coming from the Millers' home somewhere between 15:00 and 17:00 on that specific Monday. According to this neighbour, it seemed as if a woman and a man had a fight, but she could not say if it were Emma and James she heard screaming. Emma was at home in the afternoon, which was confirmed by witnesses who saw her car in the driveway at 16.30. However, Emma denies fighting with James. Another neighbour heard someone arrive at the Millers' house that Monday evening between 19:45 and 20:00. He heard the front door close around that time but claims he did not hear the doorbell.

Another neighbour living a block further down, said that he saw a woman walking down the street that Monday evening. The woman was wearing a hood, so her face and hair were covered, but the neighbour claims the woman had Emma's posture. In his opinion, the woman seemed very nervous. This neighbour saw how this woman stopped briefly at one of the trash cans in the street and then disappeared from sight. Based on this story, the police searched the trash cans in the street. They found a carpet knife (as pictured below) which was covered in blood.

Based on DNA from the blood traces found on the knife, the Technical Criminal Investigation Department confirmed that this must be the knife used to kill James. Emma's fingerprints were found on the knife, and Emma is right-handed. There were also fingerprints which did not match James or Emma on the knife. It is not known who these prints belonged to.

In sum, the police believe Emma has killed her husband, based on the forensic evidence found at the scene, the fingerprints on the carpet knife, the witness testimonies from James' friends and the neighbours, her shaky alibi and the fact that she was found covered in blood at the crime scene. Emma is also the only person with a motive for killing James: the fact that he was having an affair. Emma's defence lawyer argues that the police should continue to try and find James' mistress.