

# Witnesses' failure to detect covert manipulations in their written statements

## Citation for published version (APA):

Sagana, A., Sauerland, M., & Merckelbach, H. (2017). Witnesses' failure to detect covert manipulations in their written statements. *Journal of Investigative Psychology and Offender Profiling*, 14(3), 320–331. <https://doi.org/10.1002/jip.1479>

## Document status and date:

Published: 01/10/2017

## DOI:

[10.1002/jip.1479](https://doi.org/10.1002/jip.1479)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

**RESEARCH ARTICLE**

# Witnesses' failure to detect covert manipulations in their written statements

Anna Sagana  | Melanie Sauerland | Harald Merckelbach

Forensic Psychology Section, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands

**Correspondence**

Anna Sagana, Forensic Psychology Section, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands.

Email: [anna.sagana@maastrichtuniversity.nl](mailto:anna.sagana@maastrichtuniversity.nl)

**Abstract**

Law enforcement agencies and legal professionals often have to rely on witness statements. Undetected errors in witnesses' statements, however, could impede the accurate reconstruction of a crime and lead to the incrimination of innocent suspects. Here, we examined whether witnesses can detect manipulations in their written statements. We expect that writing a statement could provide a good means for discrimination between what is truly recalled and what is an error. This is because writing allows to monitor and control the previously produced information. In 3 experiments, participants watched a mock crime film and subsequently provided a written statement of what they had witnessed. Following a delay of several minutes (Experiment 1), 48 hr (Experiment 2), or 1 month (Experiment 3), participants were exposed to and interviewed about their testimony. Unknown to them, they were confronted with statements, which included 4 secretly manipulated details. Participants' missed a substantial number of manipulations in their written statements. Importantly, the detection rates varied as a function of delay (Experiments 1 and 2: 74–89%; Experiment 3: 36%–52%). Detection rates also varied as a function of the type of details that were manipulated. Our findings indicate that writing a statement comes with limited benefits in witnesses' ability to detect errors in their statements.

**KEYWORDS**

choice blindness, interrogation protocol, misinformation, source monitoring, witness statements

## 1 | INTRODUCTION

When investigating crimes, police sometimes resort to techniques such as changing or omitting details from witnesses' statements to elicit a confession (Pearse & Gudjonsson, 1999). However, few studies to date have attempted to directly manipulate participants' statements (Christianson, Engelberg, & Gustafson, 2007; van Bergen, Horselenberg,

Merckelbach, Jelicic, & Beckers, 2010). In fact, much of the current literature on the misinformation effect employs the typical paradigm in which the misleading information is provided via another person's narrative of the event. One disadvantage of this approach is that the misleading information is often not tailored to the memory of the participant. As a consequence, participants may fail to detect the misinformation because it refers to a detail that was never encoded in the first place (Lindsay, 1993; McCloskey & Zaragoza, 1985). Furthermore, detecting misinformation in others' accounts is substantially different from detecting changes in one's own statement. Participants often report the misleading information because they are more trusting to the external source than to their own memory (Mazzoni, 1998; Wright & Livingston-Raper, 2002). They assume the source of the misleading information to be more valid than the information they remember as part of the original event. Additionally, providing a statement requires witnesses to deliberate and evaluate the strength of their memory in an attempt to decide which details they should report (Goldsmith, Koriat, & Weinberg-Eliezer, 2002). Thus, to report a detail, the representation of the source should be reasonably strong and meet a certain degree of conviction. For these reasons, it could be argued that in the case of their own testimony, witnesses should be able to detect manipulations in their statements.

Findings by Christianson et al. (Experiment 1; 2007), however, suggest that witnesses accept unreported details as their own testimony, even when they are instructed to search actively for errors. In their study, participants were questioned about a mock crime. Whilst transcribing the statements, the experimenter made six changes to the original report. Participants were then asked to inspect the statements for mistakes either by listening to the experimenter reading the statement or by reading the statement out loud themselves. Participants' performance was far from ideal. Though the ability to detect unreported details was above chance level when participants read the statement themselves (detection rate: 71.6%), the detection rate was as low as 15% when participants listened to the statement being read to them. Likewise, van Bergen et al. (2010), who examined the acceptance of misinformation in one's transcribed statement, found that individuals accept not only changes in the previously reported information but also additions of completely new elements (detection rates ranging from 27.5%–55.0%). Furthermore, the longer time interval between providing a statement and presenting the transcribed version (1 day vs. 2 weeks) promoted the acceptance of misinformation.

Although the findings regarding the ability of participants to detect changes in their testimony appear to be robust, there is a certain lack of attempts aiming to extend the findings beyond verbal and later transcribed statements. For instance, studies on the accuracy of autobiographical memories written down as diary entries suggest that participants tend to accept manipulated diary fragments as genuine autobiographical memories (Merckelbach, Wessel, & Horselenberg, 1997). However, these results were obtained in the context of diaries where participants write down diary entries about memorable, yet mundane activities. Hence, the acceptance of manipulated fragments might reflect the liberal recognition thresholds that people rely on in this setting. On the other hand, there are findings indicating that written recall is associated with the creation of fewer false memories than verbal recall (Kellogg, 2001; Smith & Hunt, 1998). That is because writing is a relatively slow process where the ideas have to be stored whereas the graphemic representations for spelling words are activated (Kellogg, 2007), allowing for monitoring and control over the previously produced information (Grabowski, 2007). Moreover, the visual record of the written recall should offer retrieval cues that benefit long-term performance (Kellogg, 2007). For these reasons, writing a statement should provide effective means to discriminate between originally recalled and externally derived information (e.g., information provided by the interviewer). Therefore, more conceptual replications are necessary before we can postulate that the effect extends beyond transcribed statements in the eyewitness context.

With these considerations in mind, the aim of the present study was to examine to what extent participants detect manipulations in their written statements for a variety of retention intervals. The consideration of various time intervals stems from the observation that in real cases, witnesses may be confronted with their statements in court, often months or even years after providing the statement. Extensive research indicates that the longer the time interval between misinformation and the target event, the stronger is the effect of the misinformation (Belli, Windschitl, McCarthy, & Winfrey, 1992; Frost, Ingraham, & Wilson, 2002; Higham, 1998; Loftus, 2005; Underwood & Pezdek, 1998). According to the source monitoring framework (SMF; for a review, see Johnson, Hashtroudi, & Lindsay, 1993), the increased acceptance of misinformation for longer time intervals can be explained by the fact that the cues

that allowed participants to successfully discriminate determine the origin of the information (original recollection vs. inserted misinformation) fade over time. Hence, source misattributions become more likely. Drawing from these findings, we expected that participants would be less able to detect manipulations in their statements when the intervals between giving testimony and the presentation of the manipulated statement are longer than when the intervals are shorter. Accordingly, we performed three experiments in which participants first provided a written statement and were later interviewed about it. During the interview, participants were confronted with statements that included up to four covertly manipulated details. The procedure was similar across experiments except that the confrontation with the manipulated statement occurred several minutes (Experiment 1), 48 hr (Experiment 2), or 1 month (Experiment 3) after giving testimony.

The failure of witnesses to detect manipulations in their statements can have serious consequences for legal decision making. Undetected manipulations in official legal witness documents could compromise the accurate reconstruction of a criminal act. Such undetected errors could cause ambiguity about the sequence of events or the key features of the perpetrator, affect other pieces of evidence, and possibly lead to an innocent suspect being incriminated. The undetected errors could bias the police by blinding them to evidence inconsistent with these errors whilst heightening their awareness towards cues that support the erroneous statement. Consequently, important physical evidence may be lost, police investigation could be delayed, and the discovered evidence could be corrupted (for similar findings regarding confessions, see Hasel & Kassin, 2009). The judges and the jury, who often assume independence of evidence, might be disproportionately affected in their decision making by the intertwined, corrupted evidence. Finally, when presented in court, the manipulated statement could alternate the true memory of the witness, bolstering the negative consequences of the error. If written statements could provide good means for discrimination between what is truly recalled and what is an error, investigators might want to reconsider the procedure by which they obtain statements.

## 2 | METHOD

The testing materials and procedure were largely analogous across experiments, with the exception of the retention interval (henceforth delay) between providing the statement and being presented with the manipulated version. Delays varied between several minutes (i.e., <30 min, Experiment 1), 48 hr (Experiment 2), and 1 month (Experiment 3). Note that the studies were planned and executed sequentially, with knowledge of the prior results when performing the later experiments. All three studies were approved by the standing ethical board of the faculty.

### 2.1 | Participants

In Experiments 1 and 2, participants were 29 and 40 (22 men,  $M_{\text{age}} = 21.2$  years,  $SD_{\text{age}} = 3.2$ , age range: 18–41) students of German ( $n = 35$ ) and Dutch ( $n = 34$ ) origin. Participants were Bachelor students at the faculty of Psychology and Neuroscience (92.8%) or other faculties (7.2%). In Experiment 3, participants were 32 Germans (6 men,  $M_{\text{age}} = 21.1$  years,  $SD_{\text{age}} = 2.9$ , age range: 18–30). All were Bachelor students at the faculty of Psychology and Neuroscience (93.5%), except for two individuals who were working in the private sector. Participation was voluntary in exchange for course credit or a gift certificate. All participants were naïve to the purpose of the study and were tested individually in their native language.

### 2.2 | Materials

#### 2.2.1 | Stimulus film

A film fragment depicting a staged assault attempt was displayed on a computer screen with a resolution of 768 × 1024 pixels. The duration of the fragment was 2 min 43 s and comprised the following plot: A young red-haired woman with a fuller figure (i.e., the victim) walks down the road with two friends (bystanders) who are pushing their

bicycles. The bystanders are a slim, tall woman with brown hair and a young, tall man with brown hair. All three arrive at a bus station where a young man with a fuller figure and black hair (i.e., the perpetrator) is sitting on a bench. After the victim tells her friends that the bus she has to catch should arrive within minutes, the bystanders ride away. On their departure, the perpetrator approaches the victim and invites her out for a drink. Her negative response turns the perpetrator violent. He grabs and then pushes the victim into the bus stop. The victim manages to escape and runs away entering a park behind the bus stop. However, the perpetrator catches up with her and here the fragment ends.

### 2.2.2 | Manipulations

The four performed manipulations referred to persons, actions, and location details. Our a priori selection of the details was based on a pilot study with 21 participants (4 men,  $M_{age} = 27.0$ ,  $SD_{age} = 11.7$ , age range: 21–53) in which we looked at the frequency with which event details were provided in participants' free recall. These selected details and their report frequency can be found in Table 1. Because not all participants reported the preselected details in their free reports, the a priori selection of the details to be manipulated resulted in loss of data points. We decided not to deviate from the selected items (i.e., by manipulating different ones ad hoc) because this would have introduced variation in the manipulated details, complicating interpretation of the results. We also did not add details into participants' statements.

### 2.2.3 | Posttest questionnaire

The posttest questionnaire aimed to assess whether participants had noticed the manipulations in their statements but were reluctant to reveal them during the interview. First, participants were asked three open-ended questions as to whether they had noticed any problems or anything strange during the experiment. Then, they were misinformed that the study employed two conditions: an experimental condition where some details in participants' statements had been manipulated and a control condition where no manipulations had been made. Participants had to indicate in which condition they thought they had been assigned to. If they thought they had been in the experimental condition, they had to specify which details they believed had been manipulated. If these responses were correct, they were counted as retrospective detection.

## 2.3 | Design

We employed a within-subjects design with detail type (person, action, and location) as the independent variable. Concurrent and retrospective detections served as the dependent variables. Concurrent detection occurs when concerns about the integrity of statement details are raised at the time of presentation (i.e., whilst reading the statement or during the interview). Retrospective detection includes instances of suspicion that are expressed after the interview in the posttest questionnaire. Overall detection rate includes concurrent and retrospective detections. If participants indicated manipulations that had not occurred, no detection was assumed. Finally, the memorability of the details was assessed by coding the accuracy (accurate vs. inaccurate) of the later to-be-manipulated details in participants' reports.

**TABLE 1** Frequencies of reported details based on a pilot study

	Manipulated detail	Alteration category	M (%)	95% CI
1	The hair of the female bystander were <i>red</i>	Person	43.0	24.2–63.5
2	The figure of the victim was <i>normal</i>	Person	66.0	45.4–82.8
3	Perpetrator <i>pushed</i> the victim to the <i>ground</i>	Action	71.0	50.0–86.2
4	The surrounding was a <i>parking lot</i>	Location	33.0	17.2–54.6
5	The bystanders <i>walked away</i> with their bikes	Action	80.0	60.0–92.9

Note. Detail 5 was manipulated only in Experiment 3, in exchange of detail 4. CI = confidence interval.

## 2.4 | Procedure

The experiment consisted of two parts. In part 1, participants provided a witness statement. In part 2, they were interviewed about 10 details in their statements, four of which (at a maximum) had been manipulated. The purpose of the interview was to examine whether participants would raise concerns about the accuracy of the statement when their attention was pointed to specific details. A cover story indicating that the study investigated the link between personality and memory was used to mask the real purpose of the study.

During the first part, after signing the informed consent form, participants watched the stimulus film and then filled out personality questionnaires as a filler task. Next to allowing for a time interval between the film and preparing the statement, the filler task served to strengthen the credibility of our cover story. No time limit was imposed for the filler task. Hereafter, participants of all three Experiments were given unlimited time to provide a free report about the witnessed crime. They were instructed to follow the chronological sequence of the events and to describe all people involved in the film. Furthermore, they were discouraged from guessing about details they could not remember (Gabbert, Hope, & Fisher, 2009; Hastie, Landsman, & Loftus, 1978). Participants typed their statement in a text file, saved it in a designated computer folder, and then printed the statement. This procedure was followed to give participants the impression that they had control over their statement. Finally, participants either continued with a working memory computer task (Experiment 1) or were thanked and dismissed (Experiment 2 or 3).

In Experiment 1, whilst participants were busy with the computer task, the experimenter covertly accessed the digital file with their statements from another computer and manipulated the critical details as described above. Then she or he printed and saved the manipulated version of the statement in the same location and with the same name as the original one. In Experiments 2 and 3, the same procedure was followed after participants had left the lab.

Part 2 took place on completion of the computer task (<30 min) for participants in Experiment 1, and participants returned to the lab 48 hr later in Experiment 2 and 1 month ( $M_{\text{days}} = 28.7$ ,  $SD_{\text{days}} = 1.6$ ) later in Experiment 3. In Experiment 3, we allowed a deviation of 2 days to accommodate participants' classes and weekends.

The interview started with the experimenter instructing participants to open their statement on the computer screen and handing them the manipulated printed version. Participants were first asked to read their statement as a reminder. Subsequently, the experimenter inquired about the four manipulated (e.g., "In your statement you mention that the victim was of *normal* weight. Can you estimate her weight in kg?") and six nonmanipulated details (e.g., "According to your description the perpetrator started talking to the victim. What was the content of the conversation?" and "You mention that the victim wore a pink coloured jacket. Were there any prints on the jacket?"). Similar to the four manipulated details, the six nonmanipulated details referred to persons, actions, and location details. If participants had not mentioned the detail in their original statement, no manipulation was performed. However, to keep the length and the content of the interview constant across participants, the interviewer formulated a question related to the nonreported detail in general terms (e.g., "Could you estimate the height and the weight of the victim?"). If participants, at any point of the interview, raised concerns about the manipulated details (e.g., "Well, she was a bit overweight"), the experimenter first referred them to their statement. If they insisted (e.g., "This is a mistake"), then the experimenter kept a neutral position (e.g., "If you feel there is a mistake, I will make a note and check it later. Shall we continue with the interview?"), noted the detail as detected, and continued with the interview. All instances of suspicion (major and minor), if correct, were counted as concurrent detection. After the interview, participants were asked to fill out the posttest questionnaire and were thanked and debriefed.

## 3 | RESULTS

### 3.1 | Detection rates

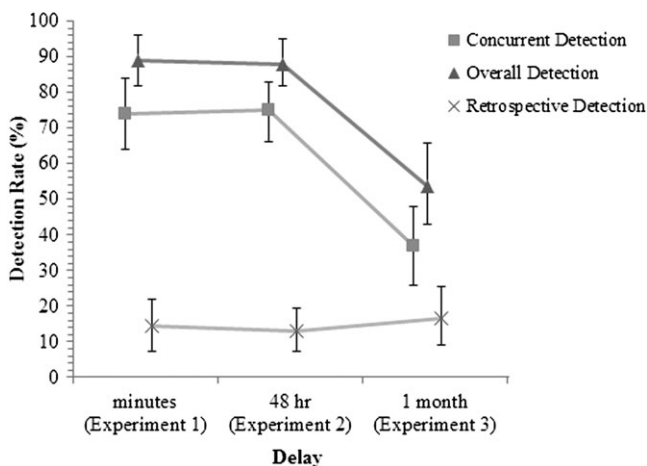
As many participants did not report all the to-be-manipulated details in their original statements, we changed fewer details than initially planned. In Experiment 1, of the  $29 \times 4 = 116$  planned manipulations, we could only change 82

details. Of these, 61 (74.4%) were detected concurrently and 12 (14.6%) additional trials were detected retrospectively, leading to 89.0% overall detection rate. In Experiment 2, we performed 108 of 160 planned manipulations. Eighty-one (75.0%) were detected concurrently, and 14 (13.0%) additional trials were detected retrospectively, leading to an 88.0% overall detection rate. In Experiment 3, we performed 78 of 128 planned manipulations. Twenty-nine (37.2%) were detected concurrently, and 13 (16.7%) additional trials were detected retrospectively, leading to 53.8% overall detection rate. The mean detection rate across studies can be found in Figure 1 (see also Table 2). Though overall participants' ability to detect manipulations in their written statements was higher than the typical findings in this line of research (Christianson et al., 2007; van Bergen et al., 2010), participants missed a substantial number of manipulations in their written statements, particularly for the longest delay.

The total number of performed manipulations (range: 1–4) did not affect the detection rates, all  $F_s(3, 29-40) \leq 1.01$ ,  $p_s \geq .401$ . Furthermore, participants seldom indicated manipulations that had not occurred. Across all three experiments, only 23 nonmanipulated details were suspected falsely to be altered. If participants' objections (detections) to the manipulated details reflected memory updating or imprecise writing in the original statements, one should expect that participants would be equally likely to raise suspicion about nonmanipulated details as well. That was not the case.

To further examine the effect of delay on detection rates, we analysed the data of all three experiments directly. For the purpose of this analysis we included only the three manipulated details that remained constant across all experiments (see Table 1). A standard least square regression analysis would not be appropriate for this design because it includes categorical within-subject response data. Rather, a regression methodology known as generalised estimating equation (GEE) is in order. The GEE model provides an appropriate alternative as it accounts for the correlated residuals via the specification of the working correlation matrix. This method enables the incorporation of the repeated observations for all performed manipulations, leading to a total of  $N = 192$  observations.

Two separate GEE analyses were performed with concurrent and overall detection rates (detected vs. nondetected) as dependent variables and the delay and the three different manipulated details as predictors. Note that the use of the manipulated details as predictor is exploratory. Our analysis confirmed that delay significantly affected both concurrent, Wald  $\chi^2(2, N = 192) = 11.43$ ,  $p = .003$ , and overall detection rates, Wald  $\chi^2(2, N = 192) = 8.97$ ,  $p = .011$ . Specifically, the 1-month delay resulted in significantly lower detection rates, compared to the shorter delays (i.e., minutes and 48 hr). The results of this statistical analysis can be found in Table 3. Interestingly, which specific detail was manipulated had an effect on both concurrent, Wald  $\chi^2(2, N = 192) = 11.34$ ,  $p = .003$ , and overall detection, Wald  $\chi^2(2, N = 192) = 10.83$ ,  $p = .004$ . The detailed results of this analysis can also be found in Table 3. Inspection of



**FIGURE 1** Proportion of concurrent, retrospective, and overall detection rates as a function of delay. Error bars represent (bootstrapped) 95% confidence intervals

**TABLE 2** Mean concurrent and retrospective detection rates and (bootstrapped) 95% confidence intervals across three experiments

	Order	Manipulated detail	n	Concurrent		Retrospective		Overall	
				M (%)	95% CI	M (%)	95% CI	M (%)	95% CI
Experiment 1 (minute delay)	1	The hair of the female bystander was <i>red</i>	25	80.0	55.5–94.1	6.67	4.5–25.0	86.6	66.6–94.7
	2	The figure of the victim was <i>normal</i>	15	55.5	32.0–80.0	22.2	5.6–42.9	77.7	57.1–94.1
	3	Perpetrator <i>pushed</i> the victim to the ground	24	88.0	73.6–96.5	12.0	3.6–27.0	100	n/a
	4	The surrounding was a <i>parking lot</i>	18	70.8	53.3–89.4	16.6	4.4–34.5	87.5	72.2–96.1
	Overall		116	74.4	64.7–84.0	14.6	7.3–21.9	89.0	82.1–95.9
Experiment 2 (48-hr delay)	1	The hair of the female bystander was <i>red</i>	34	57.8	33.3–80.9	10.5	4.2–27.4	68.4	47.3–87.5
	2	The figure of the victim was <i>normal</i>	19	52.1	30.7–72.2	26.1	8.0–45.0	78.2	60.8–94.7
	3	Perpetrator <i>pushed</i> the victim to the ground	32	88.2	77.1–97.1	8.8	2.6–19.3	97.0	89.1–97.5
	4	The surrounding was a <i>parking lot</i>	23	87.5	75.0–96.7	9.3	2.8–21.0	96.8	89.1–97.5
	Overall		108	75.0	66.7–83.2	13.0	7.4–19.4	88.0	81.7–94.1
Experiment 3 (1-month delay)	1	The hair of the female bystander was <i>red</i>	25	30.0	9.0–62.4	20.0	7.7–50.0	50.0	19.3–82.3
	3	The figure of the victim was <i>normal</i>	10	47.5	26.6–70.0	26.0	9.1–45.8	73.9	54.5–90.9
	4	Perpetrator <i>pushed</i> the victim to the ground	20	52.0	30.3–70.8	20.0	5.2–36.7	72.0	52.0–89.4
	2	The bystanders <i>walked</i> away with their bikes	23	10.0	4.1–25.0	0.0	n/a	10.0	4.0–26.6
	Overall		78	37.2	26.2–48.1	16.7	9.0–25.6	53.8	42.5–65.1

Note. The *ns* refer to the total number of manipulations performed across participants. In Experiment 3, the order of the manipulations differs because they followed the chronological sequence of the events. CI = confidence interval.

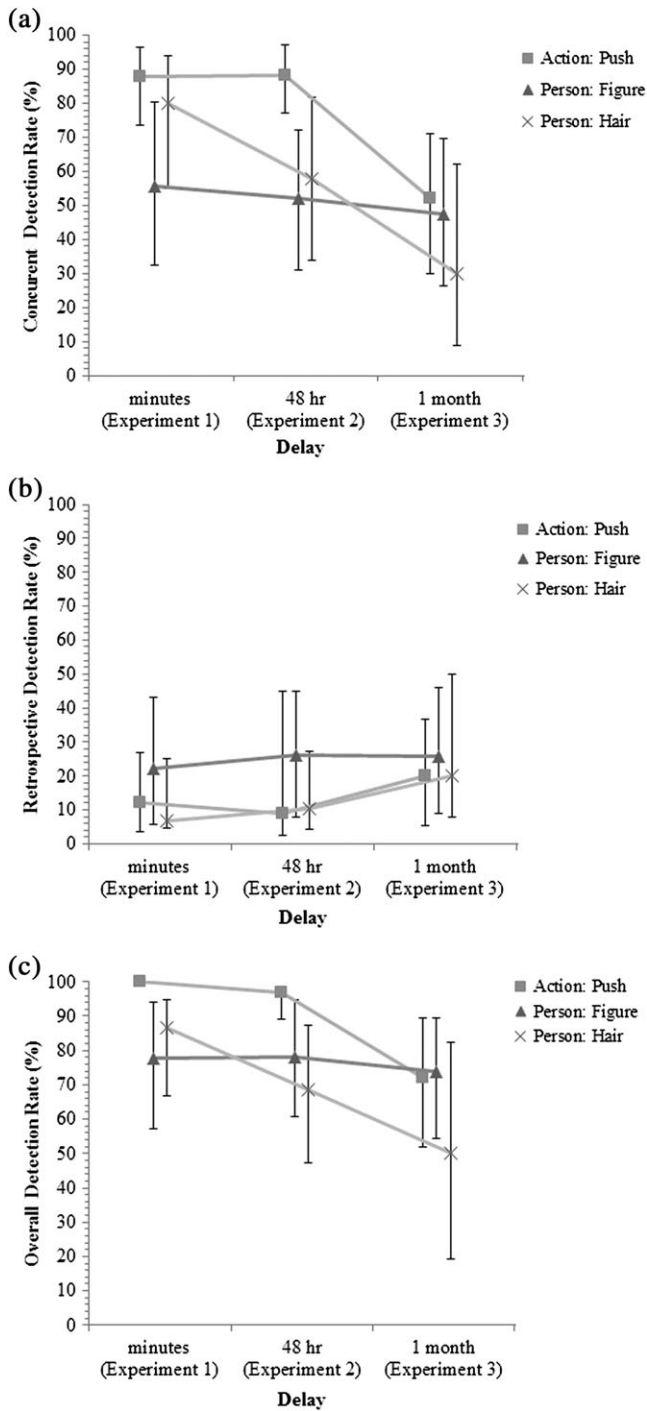
**TABLE 3** Fit for predictors for generalized estimating equation models investigating whether the delay between giving statement and being presented with manipulations and the type of the manipulated detail affect concurrent and overall detection rates

	Concurrent detection					Overall detection				
	b	SE	Wald $\chi^2$	p	OR	b	SE	Wald $\chi^2$	p	OR
Delay (base = min)										
48 hr	0.345	.398	0.75	.387	1.41	0.527	.537	0.95	.328	1.69
1 month	1.356	.427	10.04	.002	3.89	1.500	.543	7.63	.006	4.48
Detail (base = person: hair)										
Person: figure	0.166	.414	0.16	.689	1.18	-0.517	.486	1.13	.287	0.60
Action: push	-1.015	.403	6.34	.012	0.36	-1.561	.484	10.39	.001	0.21
Intercept	-0.838	.386	4.72	.030	0.43	-1.494	.479	9.737	.002	0.22

Note. *df* = 1. SE = standard error; OR = odds ratio.

the plotted detection rates per manipulated detail (Figure 2), reveals a sudden drop in concurrent and overall detection of the detail regarding the action of the perpetrator for the longest delay. The detection rates for the manipulation regarding the hair of the bystander exhibit a linear, yet nonsignificant, decline whereas detection rates for the manipulation about the figure of the victim seem to be entirely unaffected by the retention interval. Note, however, that the inclusion of the interaction term of the retention interval and the manipulated detail in the model did not improve the fit and yielded nonsignificant results.





**FIGURE 2** Proportion of concurrent (panel a), retrospective (panel b), and overall (panel c) detection rates for the three manipulated details that remained constant across Experiments as a function of delay. Error bars represent (bootstrapped) 95% confidence intervals

The low number of retrospective detections allows only a descriptive approach to the data. Visual inspection of the mean detection rates (Figures 1 and 2) reveals that retrospective detection remained stable across the different delays. Though this might reflect differences in memory and decision making between the instances where a manipulation is detected concurrently or retrospectively, given the limited number of the observations, we cannot draw conclusions about the effect of delay on retrospective detection.

### 3.2 | Accuracy of witness memory

Recall accuracy of the later to-be-manipulated details was high in all three Experiments (Experiment 1: 97.5%; Experiment 2: 97.1%; and Experiment 3: 98.8%). Of the six inaccurately recalled details that were manipulated, five were detected concurrently and one was detected retrospectively. These findings suggest that the ability to detect manipulations was not confined by the ability to accurately reproduce the witnessed details and are consistent with earlier work on eyewitness identification decisions (Sagana, Sauerland, & Merckelbach, 2014a, 2014b, 2015).

## 4 | DISCUSSION

In a series of three studies, we examined the ability of eyewitnesses to detect manipulations introduced to their testimony. We showed that eyewitnesses miss manipulations in the specifics of a statement they had written earlier and that the magnitude of the effect was larger for longer (1 month; Experiment 3) than for shorter delays (minutes or 48 hr; Experiments 1 and 2). For the longer delay, the results were most striking with only 36%–52% of the manipulations being detected, compared with the 74%–89% detections for shorter delays (see Figure 1). Additionally, the overall high levels of recall accuracy (97.3%–98.7%) for the to-be-manipulated details imply that even information that has been encoded and recalled accurately is susceptible to covert manipulations.

Our findings replicate and expand on earlier work in that witnesses can miss errors and manipulations in their written and spoken statements (Christianson et al., 2007; van Bergen et al., 2010). Though direct comparisons across studies are unwarranted, on a descriptive level, our study yielded higher detection rates for longer delays than those reported in the studies of Christianson et al. (self-read condition: 71.6%) and van Bergen et al. (2010; day: 55.0%; 2 weeks: 27.5%). This observation supports the idea that writing allows for strong monitoring and control of the produced information (Grabowski, 2007) and thus increases the chances of detecting manipulations in one's statement. Nevertheless, our results also suggest that typing a statement may not be sufficient for filtering out manipulations in one's statement after long delays (see Figures 1 and 2). The latter finding is in line with studies on the misinformation effect, which indicate increased acceptance of misinformation for longer compared with shorter retention intervals (Belli et al., 1992; Frost et al., 2002; Higham, 1998; Loftus, 2005; Underwood & Pezdek, 1998). Our findings are also consistent with recent studies indicating that eyewitnesses can be unaware of covert changes in the outcome of their identification decisions (Sagana, Sauerland, & Merckelbach, 2013, 2014b).

From a theoretical stance, our result can be explained by the SMF (for a review, see Johnson et al., 1993). The SMF postulates that misinformation like effects are due to individuals' struggle to discriminate between externally and internally derived information. This struggle results in source monitoring errors. Accordingly, the failure to detect changes in one's statement could be the result of source monitoring errors. The SMF also proposes that source monitoring errors can occur because of disruptions either in memory (e.g., insufficient encoding and forgetting) or in decision making (e.g., due to heuristics and social influences). Applying the SMF logic to our findings, we assume that because writing allows for increased input control (Grabowski, 2007) and generates additional visual records between the event and the recall (Kellogg, 2007), written statements should generate strong memory cues. In turn, these cues should protect participants against committing source monitoring errors when judging whether the derived information has been generated externally (i.e., due to the manipulation) or internally (i.e., due to recollection process) and result in increased detection rates. Indeed, looking at the detection rates for the shorter delays, this seems to be

the case. However, the SMF also suggests that long retention intervals should increase source monitoring errors. The cues that allowed participants to successfully discriminate the source of the information fade over time, leaving participants unable to determine the origin of the information, hence increasing misinformation acceptance. As a result, they misattribute the source and accept the change in their statement as their own recollection. The finding that concurrent and overall detection rates decreased for the longer as compared with the shorter delays is in line with this interpretation. The retrospective detection rate that remained stable across delays constitutes an exception to this pattern of findings. This is likely attributable to the small number of detections in retrospect or due to ceiling effects in detection rates in Experiments 1 and 2. Determining whether the decrease in concurrent and overall detection rates results from disruptions in memory (e.g., forgetting the content of the original statement), decision making (e.g., reliance on biases or social influences) or an interplay of the two is beyond the scope of the present investigation. Nevertheless, considering previous results showing that participants can be fairly accurate at remembering their original choice when their outcome of this choice was manipulated (Sagana et al., 2014a), a disrupted decision-making process when determining the source of the information may be a more likely explanation. For example, van Bergen et al.'s (2010) results indicate that participants who distrust their memory are more likely to accept misinformation than participants with positive views about their memory.

Interestingly, the current results suggest that the type of the manipulated detail might be of importance for the ability of witnesses to detect errors in their statements. As illustrated in Figure 2, detection rates for the manipulation concerning the figure of the victim remained moderate yet stable across the different delays, whereas the manipulation regarding the hair of the bystander exhibited a linear decline. In contrast, the detection rate of the manipulation regarding the action of the perpetrator to push the victim to the ground, which is the most noticed manipulation for the shorter delays, showed a sudden drop for the longer delay. Drawing from Heider's attributional approach (Heider, 1958), witnesses should focus their attention on actions rather than the descriptions of targets and location details. In line with this idea are findings indicating that witnesses do not only report more actions than descriptive details but are also more likely to accurately remember actions than descriptive details (Clifford & Scott, 1978; Ibabe & Sporer, 2004). Moreover, the negative influence of newly acquired (mis)information on memory is weaker for details with a strong causal and temporal context (Miguelés & García-Bajos, 2007; Myers, O'Brien, Balota, & Toyofuku, 1984). Whereas descriptions of persons and locations can be independent pieces of information that may not be causally related to the event or to other descriptions, details referring to actions are interconnected as one action can lead to another. This could explain the remarkably high detection rate of the manipulation concerning the action of the perpetrator for the shorter delays. However, if the causal, spatial, and temporal links to the action detail are disrupted, then the memory superiority should dissipate (Burke, Heuer, & Reisberg, 1992) and the detection advantage for details regarding actions should be undermined. One way to disrupt the aforementioned links is to increase the retention interval as increased delays weaken the memory. This appears to be the case for the 1-month delay where the greatest decline in detection rates refers to the action detail: The longer delay appears to have disrupted the memory of the event. Undoubtedly, additional factors may moderate the effect as details are unlikely to differ only in one aspect. They differ in valence, memorability, and semantic relatedness, all of which can influence memory (Dolcos, LaBar, & Cabeza, 2005; Jacques & Levine, 2007; LaBar & Cabeza, 2006; Talmi & Moscovitch, 2004). Therefore, on the basis of the present exploratory findings, we cannot draw firm conclusions about the effect of specific detail types on detection. Yet our results suggest that this topic requires systematic investigation.

Turning to the practical relevance of the present work, our findings are pertinent to law enforcement agencies and legal professionals. Mistaken or deliberate manipulations even of minor details can impede the accurate reconstruction of an event and possibly incriminate innocent suspects. Importantly, the present work speaks to the risk of long delays between police investigation and court trials. Considering that even under the *speedy trial act*, as applied in the U.S. judicial system, trials begin at least 90–120 days after the arrest of the suspect (Shermer, Rose, & Hoffman, 2011), the increasing acceptance of manipulations with longer delays is of concern. The current findings suggest that by the time witnesses would be cross-examined in court, a large proportion of witnesses would fail to realise that their statement has been tampered with. Furthermore, our work indicates that writing a statement comes

with limited benefits in witnesses' ability to spot mistaken or deliberate manipulations in a statement. Finally, the issues emerging from our line of work directly appeal to the importance for triers of fact to check for the consistency between witness statements and video recordings of witness interviews.

In conclusion, in a series of three experiments, we showed that eyewitnesses can be unaware of errors introduced in their written statements, with the effect being more pronounced for longer than shorter delays. Future research should look systematically into the factors that underlie poor detection of manipulations. The current findings, along with earlier work on the fallibility of autobiographical memory (Barclay & Wellman, 1986; Merckelbach et al., 1997), show how easy it is to compromise the quality of eyewitness statements and call for further research in this area.

## ACKNOWLEDGEMENTS

The authors acknowledge the valuable suggestions and assistance of Alana Krix and Eveline Ritzen in designing the present experiments. The authors would also like to thank Annika Horter, Felix Lippe, and Dyllian Scholten for their help in data collection.

## REFERENCES

- Barclay, C. R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language*, 25, 93–103. doi:10.1016/0749-596X(86)90023-9
- Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 356–367. doi:10.1037/0278-7393.18.2.356
- van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology*, 24, 885–896. doi:10.1002/acp.1595
- Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, 20, 277–290. doi:10.3758/BF03199665
- Christianson, S. A., Engelberg, E., & Gustafson, A. (2007). Recognition of previous eyewitness testimony from an altered interrogation protocol: Potential effects of distortions. *Psychology, Crime & Law*, 13, 583–589. doi:10.1080/10683160601095230
- Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, 63, 352–359. doi:10.1037/0021-9010.63.3.352
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2626–2631. doi:10.1073/pnas.0409848102
- Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognised over time: A source monitoring account. *Memory*, 10, 179–185. doi:10.1080/09658210143000317
- Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, 33, 298–307. doi:10.1007/s10979-008-9146-8
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General*, 131, 73–95.
- Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (Vol. 20). (pp. 165–179). Amsterdam: Elsevier.
- Hasel, L. E., & Kassin, S. M. (2009). On the presumption of evidentiary independence: Can confessions corrupt eyewitness identifications? *Psychological Science*, 20, 122–126. doi:10.1111/j.1467-9280.2008.02262.x
- Hastie, R., Landsman, R., & Loftus, E. F. (1978). Eyewitness testimony: The dangers of guessing. *Jurimetrics Journal*, 19, 1–8.
- Heider, F. (1958). *The psychology of interpersonal relations*. New Jersey: Lawrence Erlbaum Associates.
- Higham, P. A. (1998). Believing details known to have been suggested. *British Journal of Psychology*, 89, 265–283. doi:10.1111/j.2044-8295.1998.tb02684.x
- Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology*, 18, 711–726. doi:10.1002/acp.1025
- Jacques, P. L. S., & Levine, B. (2007). Ageing and autobiographical memory for emotional and neutral events. *Memory*, 15, 129–144. doi:10.1080/09658210601119762

- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28. doi:10.1037/0033-2909.114.1.3
- Kellogg, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 913–919. doi:10.1037/0278-7393.27.4.913
- Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *The American Journal of Psychology*, 120, 415–428. doi:10.2307/20445412
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7, 54–64. doi:10.1038/nrn1825
- Lindsay, D. S. (1993). Eyewitness suggestibility. *Current Directions in Psychological Science*, 2, 86–89.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366. doi:10.1101/lm.94705
- Mazzoni, G. (1998). Memory suggestibility and metacognition in child eyewitness testimony: The roles of source monitoring and self-efficacy. *European Journal of Psychology of Education*, 13, 43–60.
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1–16. doi:10.1037/0096-3445.114.1.1
- Merckelbach, H., Wessel, I., & Horselenberg, R. (1997). The accuracy of autobiographical memory: A replication of Barclay & Wellman (1986). *Behavioural and Cognitive Psychotherapy*, 25, 103–111. doi:10.1017/s1352465800018336
- Migueles, M., & García-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, 21, 1157–1172. doi:10.1002/acp.1323
- Myers, J. L., O'Brien, E. J., Balota, D. A., & Toyofuku, M. L. (1984). Memory search without interference: The role of integration. *Cognitive Psychology*, 16, 217–242. doi:10.1016/0010-0285(84)90008-2
- Pearse, J., & Gudjonsson, G. H. (1999). Measuring influential police interviewing tactics: A factor analytic approach. *Legal and Criminological Psychology*, 4, 221–238. doi:10.1348/135532599167860
- Sagana, A., Sauerland, M., & Merckelbach, H. (2013). Witnesses' blindness for their own facial recognition decisions: A field study. *Behavioral Sciences & the Law*, 31, 624–636. doi:10.1002/bsl.2082
- Sagana, A., Sauerland, M., & Merckelbach, H. (2014a). Memory impairment is not sufficient for choice blindness to occur. *Frontiers in Psychology*, 5, 449. doi:10.3389/fpsyg.2014.00449
- Sagana, A., Sauerland, M., & Merckelbach, H. (2014b). 'This is the person you selected': Eyewitnesses' blindness for their own facial recognition decisions. *Applied Cognitive Psychology*, 28, 753–764. doi:10.1002/acp.3062
- Sagana, A., Sauerland, M., & Merckelbach, H. (2015). The effect of choice reversals on blindness for identification decisions. *Psychology, Crime & Law*, 22, 303–314. doi:10.1080/1068316x.2015.1085984
- Shermer, L. O. N., Rose, K. C., & Hoffman, A. (2011). Perceptions and credibility: Understanding the nuances of eyewitness testimony. *Journal of Contemporary Criminal Justice*, 27, 183–203. doi:10.1177/1043986211405886
- Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, 5, 710–715. doi:10.3758/BF03208850
- Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, 32, 742–751. doi:10.3758/BF03195864
- Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review*, 5, 449–453. doi:10.3758/BF03208820
- Wright, D. B., & Livingston-Raper, D. (2002). Memory distortion and dissociation: Exploring the relationship in a non-clinical sample. *Journal of Trauma & Dissociation*, 3, 97–109. doi:10.1300/J229v03n03\_06

**How to cite this article:** Sagana A, Sauerland M, Merckelbach H. Witnesses' failure to detect covert manipulations in their written statements. *J Investig Psychol Offender Profil.* 2017;14:320–331. <https://doi.org/10.1002/jip.1479>