The Influence of Police Reporting Styles on the Processing of Crime Related Information

Anita Eerland[1] and Tessa van Charldorp[2]*

[1] *Department of Communication Science, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands,*
[2] *Department of Languages, Literature, and Communication, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands*

Police records drawn up during or after a suspect's police interrogation play a crucial role in judicial systems and should therefore be factual representations of what occurred in the spoken interrogation. Within the judicial domain, however, little is known about how style of reporting (i.e., the specific language used) affects the interpretation of these facts. Furthermore, the relationship between police record 'quality' and variations in judgment of guilt, credibility or reliability has not been studied to date. In three studies, we investigated the influence of three commonly used recording styles (i.e., monolog, recontextualized and question-answer style) on judgments of guilt, credibility, and reliability in fictitious criminal cases. We hypothesized that participants would (1) find records in the question-answer style more credible and reliable than those in the monolog or recontextualized style, and (2) consider the recontextualized style to be the least credible and reliable. Experiment 1 showed that the Q&A style was perceived as more reliable than the other two styles. Experiment 2, a replication in which we also tested new hypotheses based on explorative analyses of Experiment 1, showed no effects of reporting style. To investigate whether the discrepancy in results was due to different scenarios, a third experiment that made use of multiple scenarios was conducted. We found effects of reporting style on perceived accuracy, imageability, and understandability. In sum, this study showed that factors as subtle as reporting style might impact the processing of information in contexts where only factual information should be taken into account.

Keywords: language comprehension, recording styles, judgments, police records, linguistic cues

INTRODUCTION

To understand language, people form mental representations of a described situation. This mental representation is known as a situation model (Johnson-Laird, 1983; Van Dijk and Kintsch, 1983; Morrow et al., 1987; Zwaan and Radvansky, 1998). Differences in the specific language used to convey a message are known to influence situation model construction. Perhaps the best-known example of research that supports this idea is the study by Loftus and Palmer (1974). In this study, participants first watched a video clip of a car crash and then answered questions about this clip. Some participants were asked to estimate the speed of the cars when they "hit" each other, others were asked to estimate the speed of the cars when they "smashed into" each other. Results showed, among other things, that the specific wording that was used to frame the question influenced the speed estimates of the participants. Participants that were asked the estimation question using the

word 'hit' thought the cars drove significantly slower than those who were asked the same question using "smashed into." One could argue that the difference between "hit" and "smashed into" is not a very subtle one. As a result, it is hardly surprising that the wording chosen to formulate a question impacts the answer given (i.e., the estimated speed of the cars). Later research, however, demonstrated that even more subtle differences in language use, known as linguistic cues, impact situation model construction and as a result influence how people think of a described situation (Givón, 1992; Magliano and Schleich, 2000). While there is a large variety of linguistic cues that could impact situation model construction, we will discuss two of these cues in more detail.

The first linguistic cue we will discuss is grammatical aspect. Grammatical aspect, more specifically the difference between the imperfective (e.g., *He was shooting a gun*) and the perfect/perfective aspect (e.g., *He had shot a gun/He shot a gun*), primes semantic knowledge associated with the described event, like location information (Carreiras et al., 1997; Ferretti et al., 2007; Ferreti et al., 2009; Anderson et al., 2013). For example, Ferretti et al. (2007) showed that participants were faster at naming a location after having read a verb conveyed in the imperfective aspect (e.g., *was skating*) than the perfective aspect (e.g., *skated*). Grammatical aspect is not only known to influence the construction of a situation model but also cognitive processes that rely on situation model information, like memory (Carreiras et al., 1997; Magliano and Schleich, 2000), problem solving (Salomon et al., 2013), and voting behavior (Fausey and Matlock, 2011).

Secondly, speech type, more specifically the use of direct speech (e.g., *The witness said: 'I saw the attacker entering the building'*) vs. indirect speech (e.g., *The witness said that she saw the attacker entering the building*), is also known to impact situation model construction (Yao and Scheepers, 2011; Yao et al., 2011; Stites et al., 2013; Eerland and Zwaan, 2018). For example, the use of direct speech results in a more vivid situation model and is as more perceptually engaging than indirect speech. In sum, the ways in which we formulate events matters with regard to how this information is processed and remembered. That can be problematic in contexts in which language is used as an objective means to record—on paper—what has taken place during a spoken interaction, like a police interrogation.

During or after a police interrogation, police records are constructed. These records—or written statements—play an important role in the judicial system as they can be used as evidence in court if they adhere to certain criteria. Therefore, it is important that police records contain information that is accurate and of high quality. The need for accuracy and quality of these reports has been discussed in the judicial and police context (Malsch et al., 2010; Gregory et al., 2011; Jansen, 2011).

However, while the importance of accurate police records seems obvious, guidance on *how* to write up police records has only recently received more attention, for example in the form of handbooks and training. Often, police officers have relative freedom as to how they produce a written document that is supposed to reflect the spoken interrogation (see De Keijser et al., 2012). In different countries or judicial systems, we also see different recording systems: whereas in some judicial systems the

recording of the police interrogation is typed up (or transcribed) verbatim afterwards (i.e., UK), other judicial systems require police officers to type up a police record while interrogating (i.e., the Netherlands).

Yu and Monas (2020) provide a brief overview of current literature on report writing in which they conclude that interview techniques and note taking are prioritized over actually *how to write* the police record. This finding can be confirmed by looking at various handbooks for police officers in which they are trained to interrogate and write a report (e.g., Schellingen and Scholten, 2014). There are exceptions where police officers are elaborately trained on how to write objectively, in a structured way, etc. (e.g., Reynolds, 2012; Miller and Whitehead, 2018). From the judicial and police training perspective there seems to be a focus on being accurate, objective and providing a step-by-step account (Reynolds, 2012), or on being accurate, objective, complete, concise and clear (Morley, 2008).

Although these instructions are helpful, in most cases they remain rather vague as to how to operationalize for example accuracy, conciseness, or objectivity in actual language. Furthermore, the impact of choosing certain linguistic characteristics over others is relatively unexplored and unattended to. For example, in an important study carried out by the Dutch police academy on judicial knowledge and police records, the author concluded with suggestions for improvement focusing on teaching police officers more judicial knowledge and judicial language (Jansen, 2011). Whereas, more judicial knowledge and language in a police record may be evaluated as qualitatively better according to judges and prosecutors—it tells us little about whether this actually improves the accuracy, comprehensibility, objectivity, or conciseness of the text. Furthermore, it seems to contradict the guidance from—for example—the Dutch law stating that a police record must be written as much as possible in the suspect's own words. Lastly, using more judicial language could possibly have other effects, such as on judgment.

Before we suggest how police records could be qualitatively improved, we need to have a much clearer understanding of how language use within the context of police records can affect judgments. In this study, rather than theorizing about what linguistic aspects could lead to what kinds of differences in quality or judgment, we take a bottom-up approach to see what kinds of linguistic aspects are prevalent in actual police records. Based on a corpus of 35 actual police records Van Charldorp (2011) found that there are three main linguistic styles used by Dutch police officers: the monolog style, the recontextualized style and the question-answer style. These styles make use of various linguistic constructions such as *perspective* and *visibility of the source*. Besides the restrictions that the computer format and the law provide, an officer is free to write up the suspect's story in either one of these three styles, or a combination thereof.

The *monolog style* is written from the perspective of the suspect (in first person, using direct speech). The questions asked or remarks made by the interrogator(s) do not appear in the record. This style is relatively informal, not too lengthy and comprehensible. An example is the following:

"I have heard and understood that I am not obliged to answer. I am 14 years old. I live with my father. My father's name is Steven Pinas. My mother passed away."

In the *recontextualized style* the officer's question is told from the perspective of the suspect using indirect speech. In other words, the "interrogator reworks his own questions, remarks, or suggestions into the narrative" (Komter, 2013) while still using the first-person perspective. By doing so, the officers "ensure their visibility" in the police records (Komter, 2013). This style is often lengthy, formal and somewhat complicated. An example is the following:

"You ask me which act I performed.
I state to you that I kept the garage door closed and I also helped pull Mervellino into the garage.
You ask me, if a knife was used.
I state to you that no knife was used. At least, not that I know."

In the *question-answer (Q&A) style* the police officers' questions and the suspect's answers are written up as such, generally written as "Q" and "A" or "Question: …" and "Answer: …". As a result, direct speech is used. The Q&A style is short, relatively informal and comprehensible. It most closely resembles the actual spoken interaction that occurred during the police interrogation. An example is the following:

"Question: In the living room we also encountered drug-resembling goods, whose are these?
Answer: I don't know, those are not mine.
Question: Have you ever seen your uncle with a fire weapon?
Answer: No."

As can be concluded from the description above, there are different dimensions that distinguish the reporting styles from each other. First of all, there is the number of *sources* that provide the information mentioned in the police record. Whereas, the recontextualized and the Q&A style include information from both the police officer and the suspect, the monolog style only states information provided by the suspect. Next, the styles differ when it comes to *perspective* taken in the record. The recontextualized and monolog style records only present the perspective of the suspect, whereas the Q&A style records represent the perspective of both the police officer and the suspect. One could derive a score from both dimensions as

indication of the record's representativeness of the interrogation that took place (see **Table 1**).

In this study we will explore how the above mentioned different linguistic reporting styles influence reader judgments concerning reliability (i.e., accuracy) and credibility (i.e., believability) of police records and the interrogated suspects. If linguistic style affects reader judgments, much clearer guidelines will be necessary for police officers on how to construct a written police record in order to most accurately institutionalize a suspect's spoken words.

Our predictions of how reporting style influences credibility and reliability judgments are based on (1) the representativeness score for each reporting style, and (2) how common each style is. Given that the Q&A style best represents the actual interrogation (see **Table 1**) and this format is commonly used in everyday discourse, we expected a Q&A style record to be perceived as more credible and reliable than records that use the recontextualized or monolog style. The recontextualized style has a higher representativeness score than the monolog style. However, because the recontextualized style is the most complex (and deviates the most from everyday language), we hypothesized that this style leads to the least credibility and reliability of the record. The credibility and reliability of the record reported in the monolog style is expected to be lower than that of the Q&A style but higher than that of the recontextualized style. Analyses regarding the credibility and reliability of the suspect will be exploratory. Our preregistration, materials, and data can be found on the following project page on the Open Science Framework: https://osf.io/fpgz5/.

EXPERIMENT 1

Methods

Sample

We aimed at 100 participants per condition. Therefore, we recruited 350 participants online through Amazon's Mechanical Turk (MTurk, http://www.mturk.com) and 352 completed the experiment (this can occur most likely due to technical issues involving the coordination of the platform we used for recruitment, MTurk, and the platform we used for running the experiment, Qualtrics). We excluded data from 28 participants because they had reading times <0.05 ms per word. This indicates that they could not have read the police report properly. We also excluded data from participants that did not report English as their native language ($n = 3$), indicated they found the report extremely difficult to understand ($n = 2$), found the report extremely difficult to understand and had reading times <0.05 ms per word ($n = 2$), or did not indicate their native language ($n = 1$). Because exclusion of these participants yielded unequal lists, we also removed the data from five last run participants in the recontextualized condition, and two last run participants in the Q&A condition. The remaining sample ($N = 309$; 103 participants per condition) had a mean age of 38.37 [$SD = 11.25$, range = 21–71, 172 (55.66%) females]. All participants were US residents and received $1 for their participation (that took ∼9 min).

TABLE 1 | Overview of the number of sources represented and the perspective offered by reports using various recording styles.

	Sources	Perspectives	Representativeness
Recontextualized	2 (officer, suspect)	1 (suspect)	3
Question-answer	2 (officer, suspect)	2 (officer, suspect)	4
Monolog	1 (suspect)	1 (suspect)	2

Representativeness is the sum of the number of sources and perspectives presented.

Materials and Procedure

We selected an authentic police report of the interrogation of a man being suspected of stealing a motorcycle for this experiment. We considered this case to be useful for our study, as the crime involved is moderately severe and the evidence presented could be interpreted as incriminating as well as exculpatory. This was done to prevent any ceiling or floor effects for the guilty judgments as these would make it more difficult to investigate how these judgments might be impacted by reporting style. Importantly, the suspect does not confess to the crime. The report was originally recorded in the recontextualized style in Dutch. We translated the report to English and we created two additional versions of the original report; one using the monolog style and one using the Q&A style. All three reports were checked by two native speakers of English.

Participants were randomly assigned to one of three reporting style conditions (i.e., monolog, Q&A, and recontextualized). After participants carefully read the police report we asked them (1) how easy or difficult it was to understand the police record (7-point scale, 1 = *extremely easy*, 7 = *extremely difficult*), (2) if they thought the suspect was guilty of the crime (stealing a motorcycle; *yes/no*), and (3) how confident they were about their judgment (7-point scale, 1 = *not at all*, 7 = *extremely*). In addition, participants indicated how credible (7-point scale, 1 = *extremely credible*, 7 = *extremely uncredible*) and reliable (7-point scale, 1 = *extremely reliable*, 7 = *extremely unreliable*) they thought the record and the suspect were. Finally, participants stated what they thought this study was about and provided some demographic information. We recorded the time people spent reading the report. This task was presented online in the Qualtrics survey research suite (http://www.qualtrics.com).

Results

To test whether recording style influenced credibility and reliability judgments, we conducted a one-way ANOVA with recording style as between subjects factor and credibility and reliability scores for the police record as dependent variables (see **Table 2**)[1]. We found that recording style impacted the perceived reliability of the police record [$F_{(2,306)} = 3.480$, $p = 0.03$, $\eta^2 = 0.022$] but not its credibility [$F_{(2,306)} = 2.775$, $p = 0.06$, $\eta^2 = 0.018$]. *Post-hoc* comparisons, using a Bonferroni correction, showed that police records written in the Q&A style were perceived as more reliable (as indicated by a lower score for unreliability) than those written in the recontextualized style. The perceived reliability of police records in the monolog style did not differ from that of records in the other two styles.

In addition, we conducted an ANOVA with recording style as between subjects factor and (1) understandability of the report, (2) reliability and (3) credibility scores for the suspect, (4) judgments of the suspect, and (5) judgment confidence as

[1]Please note that we preregistered the following: "For each participant we will use the two credibility scores (police record and suspect) and the two reliability scores (police record and suspect). We will perform a 2 (credibility) × 2 (reliability) ANOVA to examine if these scores differ across conditions." A 2 × 2 ANOVA, however, makes no sense, as we are interested in comparing these four outcome measures across the three reporting styles. Therefore, we deviated from our preregistered plan.

TABLE 2 | Mean (SE) scores per recording style in experiment 1 ($N = 309$).

Measures	Monolog (*n* = 103)	Question-answer (*n* = 103)	Recontextualized (*n* = 103)
Report			
Understandability	1.84 (0.11)	1.54 (0.09)	1.81 (0.11)
Credibility	2.35 (0.12)	2.15 (0.09)	2.52 (0.13)
Reliability	2.51 (0.13)	2.20 (0.09)[a]	2.65 (0.14)[b]
Suspect			
Credibility	2.99 (0.14)[a]	3.51 (0.16)[b]	3.60 (0.15)[b]
Reliability	3.13 (0.14)[a]	3.58 (0.16)	3.62 (0.14)[b]
Confidence	4.97 (0.14)	4.93 (0.12)	4.80 (0.13)

Confidence = reported confidence in judgements regarding suspect. Dependent variables were measured on a 7-point scale. Different superscripts indicate a significant (p < 0.05) difference.

TABLE 3 | Percentage of guilty judgments of the suspect per recording style in experiment 1 ($N = 309$).

	Recording Style			Total
	Monolog (*n* = 103)	Question-answer (*n* = 103)	Recontextualized (*n* = 103)	
Guilty judgments (%)	13.59	17.48	16.50	15.86

dependent variables. These exploratory analyses seem to suggest that recording style might also impact reliability [$F_{(2,306)} = 3.490$, $p = 0.03$, $\eta^2 = 0.022$] and credibility [$F_{(2,306)} = 4.774$, $p < 0.01$, $\eta^2 = 0.030$] judgments of the suspect (see **Table 2**). *Post-hoc* comparisons, using a Bonferroni correction, revealed that participants thought the suspect to be more reliable after reading the police record written in the monolog style than the recontextualized style. Perceived reliability of the suspect in these two conditions did not differ significantly from that in the Q&A condition. Regarding the perceived credibility of the suspect, we found a similar pattern. Participants considered the suspect to be more credible when they read the police record in the monolog style than in the Q&A or recontextualized style. Recording style did not seem to impact the perceived understandability of the report [$F_{(2,306)} = 2.489$, $p = 0.085$, $\eta^2 = 0.016$], or the confidence of participants regarding their judgment of the suspect [$F_{(2,306)} = 0.474$, $p = 0.623$, $\eta^2 = 0.003$]. An exploratory Chi square analysis suggested that reporting style did not impact the likelihood of a guilty judgment of the suspect ($\chi^2 = 0.63$, $p = 0.73$, Cramer's $V = 0.05$; see **Table 3**).

Discussion

Based on the number of sources represented and the number of perspectives presented we calculated a representativeness score for all three recording styles under investigation. We expected the recording style with the highest representativeness score, the Q&A style, to be perceived as more credible and reliable than police records that used either the recontextualized or the

monolog recording style. We expected the recontextualized style to be perceived as the least credible and reliable as this style is the most complex (and deviates the most from everyday language). Our hypothesis was partly supported by our data. We found that a Q&A recording style was perceived as more reliable, but not credible, than a recontextualized recording style and not a monolog recording style.

Although we expected the recontextualized style to be the most complex, our data suggest that recording style does not impact understandability. In other words, participants did not seem to perceive the recontextualized style to be more difficult to understand than the Q&A or monolog style. Recording style also does not seem to impact the perceived guilt of the suspect. Interestingly, our results seem to suggest that recording style impacts the perceived reliability and credibility of the suspect. As these analyses were exploratory in nature, we conducted a second experiment to test our newly generated hypotheses.

EXPERIMENT 2

Experiment 2 served as a conceptual replication study of Experiment 1. We used a comparable case (i.e., a robbery, no confession by the suspect, original report in the recontextualized style). Based on the results of Experiment 1, we hypothesized that recording style would impact the reliability of the police record with the record in the Q&A style perceived as more reliable than the record in the recontextualized style. We also expected the recording style to impact the reliability and credibility of the suspect, with the suspect being perceived as most reliable and credible when the police record was written in the monolog style. We did not expect recording style to impact understandability of the record, guilty judgments of the suspect, or confidence ratings with respect to this judgment.

Methods
Sample
We used the program G*Power (Faul et al., 2007) to conduct a power analysis based on the effect sizes found in Experiment 1. According to this power analysis we needed at least 495 participants (i.e., 165 per condition) to obtain statistical power at the recommended 0.80 level (Cohen, 1988). Therefore, we recruited 600 participants online through MTurk. Again, most likely to technical issues, 608 participants completed the experiment. We excluded data from 38 participants because they had reading times <0.05 ms per word. This indicates that they could not have read the police report properly. We also excluded data from participants that indicated they found the report extremely hard to understand ($n = 10$), or did not report English as their native language ($n = 12$). Because exclusion of these participants yielded unequal lists, we also removed the data from 13 last run participants in the recontextualized condition, and one last run participant in the Q&A condition. The remaining sample ($N = 534$; 178 per condition) had a mean age of 37.90 ($SD = 11.65$, range = 19–74, 279 [52.25%] females). All participants were US residents and received $1 for their participation (that took ~7 min).

Materials and Procedure
For this conceptual replication we selected another real police report of the interrogation of a suspect. This time, we selected a case in which a man was suspected of a robbery. Again, the suspect did not confess, and the information presented could be perceived as incriminating as well as exculpatory. The original report was written in the recontextualized style. We translated the original report from Dutch to English, and also created a monolog and Q&A style version of the translated original report. All versions were checked by two native speakers of English. We then followed the same procedure as in Experiment 1.

Results
To test whether recording style influenced credibility and reliability judgments, we conducted a one-way ANOVA with recording style as between subjects factor and credibility and reliability scores for the police record as well as the suspect as dependent variables (see **Table 4**). Contrary to our hypothesis and the results of Experiment 1, we found no impact of recording style on the perceived reliability of the police record [$F_{(2,531)} = 0.771$, $p = 0.46$, $\eta^2 = 0.003$], its perceived credibility [$F_{(2,531)} = 0.867$, $p = 0.42$, $\eta^2 = 0.003$], the perceived reliability of the suspect [$F_{(2,531)} = 0.468$, $p = 0.63$, $\eta^2 = 0.002$], and his perceived credibility [$F_{(2,531)} = 0.028, p = 0.97, \eta^2 < 0.001$]. As in Experiment 1, we did not find support for the idea that recording style influences the understandability of the record [$F_{(2,531)} = 0.637$, $p = 0.529$, $\eta^2 = 0.002$], guilty judgments ($\chi^2 = 0.76$, $p = 0.68$, Cramer's $V = 0.04$; see **Table 5**), or confidence regarding these judgments [$F_{(2,531)} = 1.230$, $p = 0.293$, $\eta^2 = 0.005$].

Discussion
Our conceptual replication of Experiment 1 yielded some interesting findings. Contrary to our expectations, we found no effects of recording style on reliability and credibility judgments of the police record or the suspect. The finding that recording style did not influence the understandability of the record, judgments regarding the guilt of the suspect, and the confidence with which these judgments were made confirmed the hypotheses generated through exploratory analyses of the data collected in Experiment 1.

In an attempt to explain why our results of Experiment 1 regarding the credibility and reliability of the police record and the suspect did not replicate, we looked more closely at the materials that we used. After all, we only used one scenario in each experiment. Although we controlled for some factors (e.g., type of crime, whether the suspect confessed or not, the style of the original police record), it might be that the scenarios we used differed in other ways. Any difference between our two scenarios might therefore (partly) explain why our experiments show different results. We looked specifically at the understandability of the case and the percentage of guilty judgments per experiment and over conditions (i.e., our results did not indicate a difference between conditions regarding the understandability and guilty judgments within experiments). It seems like the case used in Experiment 1 was easier to understand ($M = 1.73$, $SD = 1.04$) than the case used in Experiment 2 ($M = 2.96$, $SD = 1.48$). Also, we found far more guilty judgments in Experiment 2 (84.27%)

TABLE 4 | Mean (SE) scores per recording style in experiment 2 (N = 534).

Measures	Monolog (n = 178)	Question- answer (n = 178)	Recontextualized (n = 178)
Report			
Understandability	3.06 (0.12)	2.92 (0.11)	2.89 (0.11)
Credibility	2.66 (0.09)	2.50 (0.09)	2.52 (0.09)
Reliability	2.71 (0.10)	2.56 (0.09)	2.59 (0.09)
Suspect			
Credibility	3.50 (0.12)	3.53 (0.12)	3.53 (0.11)
Reliability	3.51 (0.11)	3.61 (0.12)	3.67 (0.12)
Confidence	5.77 (0.09)	5.58 (0.10)	5.61 (0.09)

Confidence = reported confidence in judgements regarding suspect. Dependent variables were measured on a 7-point scale.

TABLE 5 | Percentage of guilty judgments of the suspect per recording style in experiment 2 (N = 534).

	Recording Style			Total
	Monolog (n = 178)	Question- answer (n = 178)	Recontextualized (n = 178)	
Guilty judgments (%)	85.96	84.27	82.58	84.27

than in Experiment 1 (15.86%). It might thus be the case that recording style influences judgments (Experiment 1) but not when the case is somewhat more difficult to understand or when people are convinced the suspect is guilty (Experiment 2). Other studies have also shown that language effects may be overruled by other effects (e.g., order effects can overrule linguistic effects as was shown by Sherrill et al., 2015).

We considered our set of two experiments with mixed findings not strong enough to draw conclusions about the impact of reporting style on how people perceive a police record and a suspect. In addition, the fact that we only used one scenario in each experiment makes it difficult to generalize any result. Finally, in the two experiments so far, we found a strong correlation between credibility and reliability judgments for the record (0.90 in Experiment 1, 0.86 in Experiment 2) as well as the suspect (0.92 and 0.85, respectively). This raises the question whether we measured the same or different constructs. To address these issues, we conducted a third experiment.

EXPERIMENT 3

We conducted Experiment 3 to get a better understanding of *if* and *under what conditions* police reporting style impacts how people perceive a police record and a suspect. Our procedure for Experiment 3 deviated from that in Experiment 1 and 2 in several aspects. First, in Experiment 3 we used multiple scenarios instead of a single scenario (as was the case in Experiment 1 and 2). Second, we felt that—in retrospect—the questions regarding the reliability of the suspect and the credibility of the police

record might have been semantically odd. After all, participants could only judge whether they thought the suspect came across as believable (i.e., hence the question about credibility), and whether they thought the police record accurately reflected the interrogation (i.e., hence the question about reliability). Judging the believability of the record and/or the accuracy of the suspect seems odd and provided us with information that is difficult to interpret. Therefore, we decided to only include a credibility question for the suspect, and a reliability question for the record. Third, with the question about the reliability of the police record, we were interested in learning how well people thought the police record reflected the interrogation. We considered a question relating to the accuracy rather than the reliability of the police record to be more intuitive. Therefore, we decided to ask participants to judge the accuracy rather than the reliability of the police record. Asking about accuracy instead of reliability might make the difference with credibility more salient. We also asked participants about the likability of the suspect because judgments of credibility are known to be influenced by the likability of a person (e.g., Ohanian, 1990). Finally, we decided to also ask participants to rate the imageability of the described events. Imageability is known to be influenced by subtle linguistic differences (e.g., Carreiras et al., 1997; Magliano and Schleich, 2000; Yao and Scheepers, 2011; Yao et al., 2011; Stites et al., 2013) and might be one of the mechanisms through which language impacts cognitive processes. For example, information that is perceived as more vivid is remembered better and easier to retrieve from memory (Reyes et al., 1980). Adding a question about the imageability of events might be informative to the question *if* and *under what conditions* reporting style impacts information processing and guilty judgments.

Methods
Sample
According to an a priori power analysis (Faul et al., 2007) we needed at least 288 participants (i.e., 96 per condition) to obtain statistical power at the recommended 0.80 level (Cohen, 1988). In total, we recruited 375 participants online through MTurk (in several batches) and 376 completed the experiment. We excluded data from 68 participants because they had reading times <0.05 ms per word for at least one of the eight police reports. This indicates that they could not have read all reports properly. We also excluded data from participants that participated twice (due to the release of several batches, $n = 4$), did not report English as their native language ($n = 3$), or a combination of both ($n = 2$). The final sample ($N = 299$) involved 105 participants in the monolog condition, 98 in the recontextualized condition, and 96 in the Q&A condition. One participant did not provide any demographic information. The mean age of the remaining 298 participants was 37.35 [$SD = 10.85$, range = 19–71, 120 (40.27%) females]. All participants were US residents and received $3 for their participation (that took ∼30 min).

Materials and Procedure
We selected eight real police reports concerning various crimes of comparable severity (i.e., shoplifting (2×), street robbery, counterfeit money/robbery, domestic violence, threatening with

knife, stealing, attempted theft). In all cases a male suspect was brought to the police station for questioning where he actively denied being guilty of the crime. The reports of this interrogation contained information that could be perceived as incriminating as well as exculpatory. Some police records were based on authentic records. As in Experiment 1 and 2, we translated the original reports to English and we created two additional versions of each original report. Some police records were fictitious cases. All 24 reports were checked by two native speakers of English.

Experiment 3 had a mixed within-between subjects design with scenario as within subjects factor and reporting style as between subjects factor. That means that all participants were presented with the eight different scenarios, but that reporting style was consistent. We chose to present participants with eight scenarios in the same reporting style to make sure participants were not aware of the different reporting styles (and our interest in them).

Participants were randomly assigned to one of three reporting style conditions (i.e., monolog, Q&A, or recontextualized). Within each condition the eight scenarios were presented in random order to account for order effect. After participants carefully read a police report we asked them (1) how easy or difficult is was to understand the police record (7-point scale, 1 = *extremely easy*, 7 = *extremely difficult*), (2) how easy or difficult is was to imagine what happened (7-point scale, 1 = *extremely easy*, 7 = *extremely difficult*), (3) if they thought the suspect was guilty of the crime (*yes/no*), (4) how confident they were about their judgment (7-point scale, 1 = *not at all*, 7 = *extremely*), (5) how accurate they thought the report was (7-point scale, 1 = *extremely accurate*, 7 = *not accurate at all*), (6) how credible (believable) they thought the suspect was (7-point scale, 1 = *extremely credible*, 7 = *extremely uncredible*), and (7) how likable they thought the suspect was (7-point scale, 1 = *extremely likable*, 7 = *extremely unlikable*). Finally, participants stated what they thought this study was about and provided some demographic information. We recorded the time people spent reading the report. Again, this task was presented online in the Qualtrics survey research suite (http://www.qualtrics.com).

Results

To test whether reporting style influenced the imageability of the described crime, the understandability and accuracy of the police report, the credibility and likability of the suspect, or confidence regarding guilty judgments we used linear mixed models generated with SPSS (version 27). Compared to a repeated measures ANOVA, a linear mixed model is thought to reduce the chance of a Type I error (Quené and Van den Bergh, 2008). For all dependent measures we first estimated an intercept only model with a random intercept for *participant* and *scenario*. These models indicated that there was significant variance between participants regarding the imagebility of the described crime [$Var[u_{oj}] = 0.59$, $p < 0.001$], the understandability [$Var[u_{oj}] = 0.58$, $p < 0.001$] and accuracy [$Var[u_{oj}] = 1.04$, $p < 0.001$] of the police report, the credibility [$Var[u_{oj}] = 0.63$, $p < 0.001$] and likability [$Var[u_{oj}] = 0.44$, $p < 0.001$] of the suspect, and confidence regarding guilty judgments [$Var[u_{oj}] = 0.56$, $p < 0.001$]. There was also significant variance between scenarios

TABLE 6 | Linear mixed model results for all dependent measures in experiment 3.

Measures	Model 0		Model 1		Change in model fit p
	−2LL	Parameters	−2LL	Parameters	
Imageability	7,767.34	2	7760.57	4	0.034*
Report					
Understandability	7,536.35	2	7528.18	4	0.017*
Accuracy	8,076.94	2	8060.95	4	<0.001*
Suspect					
Credibility	9,114.85	2	9114.32	4	0.767
Likability	7,878.11	2	7877.69	4	0.814
Confidence	8,347.44	2	8347.03	4	0.812

Model 0 is the intercept only model. For Model 1 we added condition as fixed factor to the intercept only model.
*Significant at 0.05 level.

TABLE 7 | Estimated mean (SE) scores per recording style for experiment 3 (N = 299).

Measures	Recording style		
	Monolog	Question-answer	Recontextualized
Imageability	2.47 (0.19)[a]	2.16 (0.19)[b]	2.32 (0.19)
Report			
Understandability	2.35 (0.20)[a]	2.01 (0.20)[b]	2.22 (0.20)
Accuracy	3.21 (0.13)[a]	2.63 (0.13)[b]	2.78 (0.13)[b]
Suspect			
Credibility	4.41 (0.25)	4.50 (0.26)	4.47 (0.26)
Likability	4.74 (0.21)	4.70 (0.21)	4.67 (0.21)
Confidence	4.82 (0.17)	4.90 (0.17)	4.83 (0.17)
Guilty judgments*	0.62 (0.12)	0.65 (0.11)	0.65 (0.11)

Dependent variables were measured on a 7-point scale. Different superscripts in a row indicate a significant (p < 0.05) difference.
*Reported as the estimated proportion of guilty judgments.

regarding the imageability of the described crime [$Var[u_{1j}] = 0.23$, $p = 0.049$], the understandability of the police report [$Var[u_{1j}] = 0.26$, $p = 0.049$] and the credibility [$Var[u_{1j}] = 0.45$, $p = 0.049$] and likability [$Var[u_{1j}] = 0.29$, $p = 0.049$] of the suspect, but not regarding accuracy of the police report [$Var[u_{1j}] = 0.04$, $p = 0.076$] and confidence regarding guilty judgments [$Var[u_{1j}] = 0.16$, $p = 0.053$]. Following Barr et al. (2013) we decided to keep *participant* and *scenario* as random intercepts for all variables. We then added *condition* as fixed effect and compared for each variable separately the −2LL of this model that includes a fixed factor with the −2LL of the intercept only model. A decrease in −2LL indicates an increase in model fit. A significant increase in model fit suggests an effect of condition, and thus reporting style.

Table 6 shows how well the intercept only models fit our data and whether adding *condition* as a fixed effect significantly increases the fit of this model for each dependent measure. As can be seen, adding *condition* as a fixed effect did not improve the intercept only model for the credibility and the likability of the suspect, or confidence regarding guilty judgments. Reporting style did thus not influence these variables. The intercept only

model did, however, improve significantly by adding *condition* as a fixed effect for the remaining three dependent measures. Reporting style had a significant effect on the imageability of the described crime [$F_{(2,298.15)} = 3.423$, $p = 0.034$], and the accuracy [$F_{(2,298.68)} = 8.216$, $p < 0.001$] and understandability [$F_{(2,298.13)} = 4.144$, $p = 0.017$] of the police report.

Pairwise comparisons, for which we used the Šidák correction to correct for multiple comparisons, showed that participants indicated that it was easier for them to imagine the described crime when having read a report in the Q&A style than in the monolog style ($p = 0.028$). Also, participants indicated that reports written in the Q&A style were easier to understand than those written in the monolog style ($p = 0.013$). Finally, participants considered police reports written in the monolog style to be less accurate than those written in the Q&A ($p < 0.001$) or recontextualized style ($p = 0.011$). We found no other significant effects of reporting style (see **Table 7**).

In addition, we conducted a generalized linear mixed model (i.e., because our outcome variable is measured on a dichotomous instead of continuous scale) to test whether reporting style influenced judgments of guilt. Our intercept only model with a random intercept for *participant* and *scenario* correctly estimated 78.5% of the observations in our sample. This model indicated that there was significant variance between participants [$\mathrm{Var}[u_{oj}] = 0.64$, $p < 0.001$] but not between scenarios [$\mathrm{Var}[u_{1j}] = 1.87$, $p = 0.066$]. Adding *condition* as fixed effect resulted in a model with a predictive value of 78.5% which did not differ from that of the intercept only model. Comparing−2LL of the intercept only model (11049.85) and that of our model that included *condition* as fixed effect (11058.79) even suggests that adding *condition* decreased the model fit. Reporting style did thus not influence guilty judgments [$F_{(2,2.389)} = 0.485$, $p = 0.62$].

Discussion

We conducted Experiment 3 to gain a better understanding of the discrepancy in results between our first two experiments. To rule out that this discrepancy was caused by unintended differences between the scenarios that we used, we decided to use multiple scenarios in our third study. Our results showed that reporting style did influence the perceived accuracy of the report with the monolog style being perceived as less accurate than the Q&A style and recontextualized style. In our previous experiments, we did not ask participants to rate the accuracy of the report. Instead, we asked for its reliability which was impacted by recording style in Experiment 1 but not in Experiment 2. The results of Experiment 1 showed that a police report written in the recontextualized style was perceived as less reliable than that in the Q&A style. If accuracy and reliability tapped into the same construct, the results of Experiment 3 and Experiment 1 both suggest that the Q&A style is considered the most accurate/reliable. This is congruent with our hypothesis that a police report written in the Q&A style represents the actual interview better (i.e., reflected in a representativeness score) than a police report written in the monolog or recontextualized style.

In contrast to the findings of Experiment 1 and 2, we found a significant effect for reporting style on understandability. Police

reports written in the Q&A style were easier to understand than those written in the monolog style. This finding is consistent with Van Charldorp (2011) who concluded that out of the three main reporting styles the Q&A style is relatively informal and comprehensible. The fact that we found no significant difference in understandability between the recontextualized style and the monolog or Q&A style was surprising, given the complexity and rarity of the recontextualized style. After all, information that is presented in a way that deviates from our expectation (i.e., which is the case with presentations that we encounter less often) is more difficult to process (Zwaan, 1994).

Our finding that a described crime was easier to imagine after reading a police report in the Q&A style than in the monolog style fits with the result regarding understandability and supports the theory that mental model construction lies at the heart of language comprehension (Johnson-Laird, 1983; Van Dijk and Kintsch, 1983; Morrow et al., 1987; Zwaan and Radvansky, 1998). Information that is easier to imagine, is easier to understand, and also more likely to be remembered better (Reyes et al., 1980). Therefore, it is important that future research focuses on the impact of reporting style on memory and cognitive processes that rely on memory function (e.g., decision making).

Future research might also want to use alternative methods to measure the variables of interest. We were interested in very subtle effects of language use, yet our dependent variables were measured using a 7-point scale or a dichotomous scale. It might be that our method was not subtle enough to pick up on such subtle effects. This could also explain why we found no evidence that recording style influenced guilty judgments in all three experiments. An alternative measure of interest might be a think-aloud protocol. A think-aloud protocol—in which participants share their thoughts while reading police reports and answering questions—will provide useful information about how people process information.

CONCLUSION

So far, linguistic studies show that the written police record is often a selection of the actual interrogation that preceded it (Jönsson and Linell, 1991; Van Charldorp, 2011), and that transformations take place such that the written document becomes a structured, logical, chronological and neutrally told story of what happened (Van Charldorp, 2020). Processes of entextualisation, recontextualisation and decontextualisation across legal contexts have been elaborately discussed elsewhere (Heffer et al., 2013) showing, amongst many other things, that legal texts not only travel physically, but also across discursive spaces creating new contexts, interpretations and meaning. These types of transformations are not only relevant in the legal domain, but across many institutional settings where spoken interaction forms the basis of written documents. Such transformations, however, seem to be taken for granted in many studies. What the consequences are of very specific elements within this transformation process, has received very

little attention (however, see De Keijser et al., 2012). In this study we took a closer look at how different linguistic reporting styles influence reader judgments concerning reliability and credibility of police records and the interrogated suspects. We found that reporting style indeed influenced the processing of information. More specifically, reporting style impacted the perceived accuracy of the report, as well as the understandability and imageability of the described event.

In sum, our study showed that language is important and that subtle differences in language use might have unintended effects. Clearly more research is needed. Only when we better understand the impact of subtle differences in language use and the mechanisms through which language operates, we can design better guidelines for police officers on how to construct a written police record that does not—unintentionally—influence the course of justice.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in an online repository and can be found below: https://osf.io/fpgz5/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty Ethics Assessment Committee – Humanities, Utrecht University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AE and TC contributed to the planning of the studies, preparation of the materials, and wrote the manuscript. AE collected and analyzed the data. Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Anderson, S. E., Matlock, T., and Spivey, M. (2013). Grammatical aspect and temporal distance in motion descriptions. *Front. Psychol.* 4, 337. doi: 10.3389/fpsyg.2013.00337

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Memory Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Carreiras, M., Carriedo, N., Alonso, M. A., and Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory Cogn.* 25, 438–446. doi: 10.3758/BF03201120

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillsdale, NJ: Lawrence Erlbaum.

De Keijser, J., and Malsch, M., and Kranendonk, R., and de Gruijter, M. (2012). Written records of police interrogation: differential registration as determinant of statement credibility and interrogation quality. *Psychol. Crime Law* 7, 613–629. doi: 10.1080/1068316X.2010.526119

Eerland, A., and Zwaan, R. A. (2018). The influence of direct and indirect speech on source memory. *Collabra Psychol.* 4, 5. doi: 10.1525/collabra.123

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fausey, C. M., and Matlock, T. (2011). Can grammar win elections? *Polit. Psychol.* 32, 563–574. doi: 10.1111/j.1467-9221.2010.00802.x

Ferreti, T. R., Rohde, H., Kehler, A., and Crutchley, M. (2009). Verb Aspect, event structure, and coreferential processing. *J. Memory Lang.* 61, 191–205. doi: 10.1016/j.jml.2009.04.001

Ferretti, T. R., Kutas, M., and McRae, K. (2007). Verb aspect and the activation of event knowledge. *J. Exp. Psychol. Learn. Memory Cogn.* 33, 182–196. doi: 10.1037/0278-7393.33.1.182

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics* 30, 5–55. doi: 10.1515/ling.1992.30.1.5

Gregory, A. H., Compo, N. S., Vertefeuille, L., and Zambruski, G. (2011). A comparison of US police interviewers' notes with their subsequent reports. *J. Investig. Psychol. Offender Profiling* 8, 203–2015. doi: 10.1002/jip.139

Heffer, C., Rock, F., and Conley, J. (2013). *Legal-Lay Communication: Textual Travels in the Law.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199746842.001.0001

Jansen, T. G. (2011). *Juridische kennis en het proces-verbaal. Een onderzoek naar de kwaliteit en aandachtspunten voor verbetering.* Uitgave van de Politieonderwijsraad, Den Haag. Available online at: https://www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/PDF/81033.PDF. (accessed November 01, 2021).

Johnson-Laird, P. N. (1983). *Mental Models.* Cambridge, MA: Harvard University Press.

Jönsson, L., and Linell, P. (1991). Story generations: from dialogical interviews to written reports in police interrogations. *Text* 11, 419–440. doi: 10.1515/text.1.1991.11.3.419

Komter, M. (2013). "Travels of a suspect's statement," in *Legal-Lay Communication: Textual Travels in the Law,* eds C. Heffer, F. Rock, and J. Conley (Oxford: Oxford University Press), 126–146.

Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *J. Verbal Learn. Verbal Behav.* 13, 585–589. doi: 10.1016/S0022-5371(74)80011-3

Magliano, J. P., and Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Process.* 29, 83–112. doi: 10.1207/S15326950dp2902_1

Malsch, M., De Keijser, J., Kranendonk, P. R., and de Gruijter, M. (2010). Het verhoor op schrift of op band? De gevolgen van het 'verbaliseren' van verhoren voor het oordeel van de jurist. *Nederlands Juristenblad* 37, 1931–2407.

Miller, L. S., and Whitehead, J. T. (2018). *Report Writing for Criminal Justice Professionals, 6th Edn.* New York, NY: Routledge. doi: 10.4324/9781315267579

Morley, P. (2008). *Report Writing for Criminal Justice Professionals: Learn to Write and Interpret Police Reports.* West Berkshire: Kaplan Publishing.

Morrow, D. G., Greenspan, S. L., and Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *J. Memory Lang.* 26, 165–187. doi: 10.1016/0749-596X(87)90122-7

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *J. Advertising* 19, 39–52. doi: 10.1080/00913367.1990.10673191

Quené, H., and Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Memory Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002

Reyes, R. M., Thompson, W. C., and Bower, G. H. (1980). Judgmental biases resulting from differing availabilities of arguments. *J. Pers. Soc. Psychol.* 39, 2–12. doi: 10.1037/0022-3514.39.1.2

Reynolds, J. (2012). *Criminal Justice Report Writing.* Scotts Valley, CA: CreateSpace Independent Publishing Platform.

Salomon, M. M., Magliano, J. P., and Radvansky, G. A. (2013). Verb aspect and problem solving. *Cognition* 128, 134–139. doi: 10.1016/j.cognition.2013.03.012

Schellingen, R., and Scholten, N. (2014). *Verdachtenverhoor. Meer dan het stellen van vragen.* Mechelen: Kluwer Belgium

Sherrill, A. M., Eerland, A., Zwaan, R. A., and Magliano, J. P. (2015). Understanding how grammatical aspect influences legal judgment. *PLoS ONE.* 10:e0141181. doi: 10.1371/journal.pone.0141181

Stites, M. C., Luke, S. G., and Christianson, K. (2013). The psychologist said quickly, "Dialogue descriptions modulate reading speed!". *Memory Cogn.* 41, 137–151. doi: 10.3758/s13421-012-0248-7

Van Charldorp, T. C. (2011). *From police interrogation to police record.* Oisterwijk: Uitgeverij BOXPress.

Van Charldorp, T. C. (2020). "Reconstructing suspects' stories in various police record styles," in *The Discourse of Police Interviews*, ed M. Mason and F. Rock (Chicago, IL: The University of Chicago press), 329–348.

Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension.* New York, NY: Academic Press.

Yao, B., Belin, P., and Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *J. Cogn. Neurosci.* 23, 3146–3152. doi: 10.1162/jocn_a_00022

Yao, B., and Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition* 121, 447–453. doi: 10.1016/j.cognition.2011.08.007

Yu, H., and Monas, N. (2020). Recreating the scene: an investigation of police report writing. *J. Techn. Writ. Communic.* 50, 35–55. doi: 10.1177/0047281618812441

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *J. Exp. Psychol.Learn. Memory Cogn.* 20, 920–933. doi: 10.1037/0278-7393.20.4.920

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162