Implementing context management in forensic casework:

Minimizing contextual bias in firearms examination

E.J.A.T. Mattijssen[a], W. Kerkhoff[a], C.E.H. Berger[a,b], I.E. Dror[c], R.D. Stoel[a]

[a] Netherlands Forensic Institute, PO Box 24044, 2490 AA The Hague, The Netherlands
[b] Leiden University, Institute for Criminal Law and Criminology, PO Box 9520, 2300 RA Leiden, The Netherlands
[c] University College London (UCL) and Cognitive Consultants International (CCI-HQ)

Managing contextual information in forensic casework aims to minimize the task-irrelevant information while maximizing the task-relevant information that reaches the examiner. Any particular forensic domain will require its own context management system and this paper describes the design and implementation of context management for minimizing bias in forensic firearms and ammunition examination. Guided by a categorization of different sources of contextual information, a flow-chart was constructed that specifies the process of casework examinations and context management. Despite these measures, another examiner may need to be involved when context management fails and the risk of bias is real. Application of this context management system does not make a subjective examination objective, but can limit the risks of bias with a minimal investment of time and resources.

1. Introduction

Traditional forensic identification sciences, which have been accepted in court for over a hundred years, are now criticized for their lack of scientific rigor (e.g. [1-5]). Part of the criticism focuses on the biasing effect of domain-irrelevant information on the judgment of experts in forensic casework[1]. Such concerns were originally raised long ago [7-10], but were mainly ignored. The change in the forensic domain came only with the presentation of empirical evidence for the existence and importance of cognitive problems [11-13].

It has been argued that forensic scientists should acknowledge the risks of contextual and other biases, and minimize their effects by implementing appropriate methods and procedures for forensic casework [3, 14-17]. With the growing acceptance that these issues are real and relevant, it seems appropriate to take steps to deal with contextual bias. However, while much has been written and is being written on contextual bias in forensic science (e.g. [2, 3, 15, 18-26]), there is a need to develop and provide sufficiently detailed guidelines.

Blind procedures such as (linear) sequential unmasking [15, 27, 28], evidence line-ups [14], and the 'case managers' model [3, 17] have been proposed. Sequential unmasking minimizes bias by '*preventing analysts from knowing the profile of submitted references (i.e., known samples) when interpreting testing results from evidentiary (i.e., unknown or questioned) samples.*' [27]. While sequential unmasking is a very powerful methodology, it can only be implemented for specific types of evidence. For these types of evidence the features can be defined and measured prior to examining the reference material and prior to the comparison. This is the case for domains such as DNA and fingermark examinations, but much more challenging for areas such as toolmark examination, firearm examination, and handwriting analysis, where in current practice it is not as straightforward to objectively define and record all features of the evidence.

Risinger's 'case managers model' distinguishes a case manager who is fully informed of the facts of the case, and an analyst who is 'blind' to irrelevant information [3,

---

[1] In contrast to earlier publications we will use the terms 'task-relevant' and 'task-irrelevant' in the remainder of this article. It is our opinion that when managing information to minimize contextual bias this usually focuses on a specific task, e.g. the comparison of markings in two bullets. Information which is irrelevant for this specific task (comparison) might be relevant for another task (e.g. shooting scene reconstruction) within the same forensic domain [6].

15-17]. Page et al. propose to apply a combination of the case managers model and sequential unmasking in forensic odontology [29]. Found et al. describe the management of contextual information in forensic handwriting examination casework [30].

Although there is a call for the development of appropriate methods and procedures to minimize the effects of (contextual) bias, others question the need for this. As a recent example, Langenburg et al. claim that only a small proportion (2%) of the casework from the Minnesota Bureau of Criminal Apprehension Latent Print Unit, from 2009/2010, consisted of cases where there was a high level of interaction between the examiner and the investigator(s) or prosecutor and which had a high amount of contextual information [31]. However, serious concerns have been raised about the methodology of this study, which call the findings into question [32]. Champod voiced the concern that examiners will be deprived completely from external information and that the need for fundamental research in various forensic fields will be neglected by prioritizing research on the presence of contextual bias [33]. Context management should take these concerns into account ensuring that it is efficient and does not lead to depriving the examiner of relevant information.

Indeed, when implementing context management, the implemented system should be practical and efficient while at the same time enabling an effective use of task-relevant contextual information [17]. In a reply to Champod [33], Berger and Stoel [34] stress that context management should both decrease the amount of task-irrelevant information and increase the amount of task-relevant information. They agree that implementing procedures to deal with contextual information does not require the research agenda to be switched from fundamental issues to the existence of bias in every single forensic discipline.

This article describes the development and implementation of an efficient context management procedure to deal with contextual information within forensic firearms and ammunition examination. Context management [30, 35] aims to optimize the flow of information to and from a forensic examiner in a case, by minimizing as much as possible the exposure to task-irrelevant information while maximizing the role of and focus on task-relevant information. The specific procedure to be applied in practice depends on the type of contextual information. We will use the classification of contextual information as proposed by Stoel, Berger, Kerkhoff, Mattijssen & Dror [35] in order to decide on the appropriate context management procedure.

The structure of this article is as follows: We will first give a short overview of the classification of contextual information, and the corresponding context management procedures as proposed by Stoel et al. [35]. Then we will describe the consecutive steps that

were taken, and choices that were made during the implementation of the context management procedure in the Firearms section of the Netherlands Forensic Institute (NFI). The project confirmed that the implementation of context management was as feasible as expected; it is currently being implemented for other disciplines within the institute, wherever applicable.

We want to make explicit that we do not quantify whether, how, and how much bias by contextual information is an issue in forensic casework in practice. While these questions are important and interesting from an academic perspective, we took the position that we cannot ignore the large amount of findings from cognitive research outside of forensic science (e.g. [11-13]) providing convincing evidence for the existence of contextual influences. Therefore, it is prudent to perform forensic examinations in such a way that the risk of contextual bias is minimized, while not leading, of course, to inefficient approaches that result in much longer lead times. As Dror stated: "For forensic science to successfully take on the issue of contextual bias, it is important that one correctly considers the risks, that measures are taken when needed, and that they are proportionate and appropriate" [36].

Procedures that deal with rare occurrences should not make the everyday work of examiners (too) inefficient, as implied by Helsloot and Groenendaal [37]. During the design of context management in the Firearms section we have taken into consideration both the positive effects of minimizing bias in forensic casework as well as the negative impact on e.g. lead times. Given current best practice, this has resulted in a system that both maximizes benefits in overall casework and minimizes resources and efforts.

2. Levels of contextual information, and context management

2.1 Levels of contextual information

Stoel et al. [35] classify sources of contextual information into four levels ordered by their proximity to the information in the trace: the trace itself (Level 1), the reference material (Level 2), the case information (Level 3), and the 'base rate' information (Level 4)[2].

Level 1 contains contextual information inherent to the examined questioned material and can usually not easily be separated from it. This information is coming from

---

[2] Dror et al. [6, 28, 38] added a fifth level to the taxonomy of Stoel et al. [35] containing cultural and organizational factors (Level 5). This level is not discussed in this paper.

(physical) features of the questioned material, some of which are relevant and some of which may not be relevant for the examination.

Whenever questioned material is analyzed simultaneously with reference material of a known source (be it a suspect, a firearm, etc.), the perception of the relevant features of the questioned material may become partly dependent on what the examiner has seen in the reference material. Usually the forensic question is whether the questioned material comes from the same source as the reference material. The comparison therefore depends on both the questioned material and the reference material. However, the perception of the features of the questioned material should not be affected by the reference material, since the relevance of the reference material to the case is the very thing at stake. The reference material itself is denoted as Level 2 contextual information when analyzing the questioned material.

Level 3 contains case information in the broadest sense. That is, all information (both oral, written, and non-verbal information) that concerns the case.

Level 4 contains information that is not specific to the case, it includes information such as base rate. Base rate information is organization and discipline specific information about the outcomes of previous cases that can create an expectation about obtaining inculpatory evidence prior to any examination in the present case.

2.2 Managing levels of contextual information

Minimizing bias may require a different approach for each type of contextual information. Exposure to Level 1 contextual information is generally difficult to control since this type of contextual information is inherent to the evidential material. Even so, approaches do exist to control the exposure to Level 1 contextual information, for instance by removing task-irrelevant features from the questioned material. For example, in a signature comparison, only the signature on the questioned document can be given to the examiner by using a physical (or digital) overlay. If there is any doubt about the effect of managing level 1 information, a sequential procedure can be applied where the examination is first applied *without* Level 1 contextual information, and then *with* the specific Level 1 contextual information. Managing Level 1 contextual information as described above constitutes Level 1 context management.

Exposure to Level 2 contextual information is relatively easy to manage for disciplines like DNA and fingerprint examination. Since the reference material can have a biasing effect on the perception of the features of the questioned material, it should not be

given to the examiner before or during the analysis of the questioned material. This sequential analysis of the questioned material and the reference material, and the comparison of the two, has been termed 'sequential unmasking' for forensic DNA examination [27], but applies to most forensic disciplines. The expert can only proceed to a next step after having finished and documented the current step. Sequential unmasking is now a standard procedure in some forensic DNA laboratories [39] and is relatively easy to implement for evidence types where the questioned material can be examined without knowledge of the reference material. This requires that the examiner is capable of defining (and measuring) the relevant features based on the questioned material only.

While the relevant features are well-defined for DNA and fingerprints, this does not hold for cartridge case and bullet comparisons. Examiners in this field typically compare striations in the questioned bullet and in the reference bullets with a comparison microscope, explicitly looking for agreement or disagreement in the form of matching or non-matching striations. In current best practice it is not straightforward to objectively define features based on the questioned material only. Striations in bullets for example vary in continuous dimensions (width, depth and length) while for DNA the relevant features are categorical and more restricted in their possible appearances and easier to define. Analyzing each separate striation on for example the position in a bullet, width, depth and length is very difficult without the use of upcoming 3D measurements and is extremely time consuming [40]. Besides this, a lot of the invested effort will turn out to be redundant because variations in striations are expected anyway, also between fired bullets from the same firearm. Without considering the reference material is it is not clear which of the features of the numerous striations are reproducible and thus important for the actual comparison. To illustrate this point a comparison of one area of two bullets is shown in Figure 1 where some matching striations are visible at the bottom while the features of the other striations seem to vary between the two bullets. Reference samples might show that the matching striations at the bottom of Figure 1 are reproducible while the other striations show a lot of variability from fired bullet to fired bullet.

FIGURE 1

Level 3 contextual information is the sum of information about the case at hand, transmitted by, for instance, interaction with colleagues, the police, and legal practitioners. In the NFI's practice, a large part of this information originates from the documents describing the case when it enters the laboratory. The general solution to control exposure to Level 3 contextual information is straightforward: optimize the information flow to minimize the risks of the forensic examiner receiving task-irrelevant information, or too little task-relevant case information. This requires a careful separation of the task-relevant from the task-irrelevant information in the case by a knowledgeable context manager[3]. The context manager and examiner in any given case must necessarily be two different persons, which parallels the case manager approach of Risinger et al. [3] (see also [30, 35]).

Deciding on what is task-relevant or task-irrelevant information for an examination and the interpretation of its results can be a difficult task and depends on the specific discipline and case. Therefore, development and implementation of context management in everyday casework may require quite some thought. To ensure that the examiner performing the comparisons receives all of the needed task-relevant and none of the task-irrelevant information for the comparisons, some case information will have to be removed while other information might have to be added to the case file. Removing task-irrelevant information and adding task-relevant contextual information constitutes Level 3 context management.

Finally, base rate information based on previous cases may result in a continuing expectation that the evidence under consideration will be inculpatory or exculpatory, depending on the examiner's experience from previous comparable casework. Minimizing the effect of base rate information constitutes Level 4 context management. One approach, implemented by the NFI's Firearms section, is to add fake cases with conclusions opposite to those in the case flow to balance the kind of evidence most commonly encountered [17,41]. Adding numerous fake cases to the case flow would increase the caseload of the examiners, and for this reason the number of added fake cases will have to be kept low in practice. When fake cases are entered into the main case flow the examiners performing the

---

[3] We introduce the term 'context manager' instead of using the term 'case manager' as proposed by Risinger et al. [3] to emphasize the role of this person. The context manager's role is to ensure a correct management of the contextual information when starting work on a case, after which his responsibility ends. The term case manager implies a larger role after the initial context management which differs from the proposed context management in this paper.

comparisons should not be able to recognize the cases as fake. To be able to do this, fake cases should be very realistic in appearance. It should be noted that adding only a few fake cases to the main case flow just slightly lowers the actual base rate. But when examiners are made aware that each case could be a fake case, for which the correct conclusion is not in agreement with the expected conclusion based on the overall base rate, this awareness might be sufficient to alter the examiner's perception of the base rate.

3 Implementing context management

This section describes the steps taken to minimize the influence of context effects during design and implementation of context management at the Firearms section. Broadly speaking, those steps start with developing an overview of the possible instances of exposure to contextual information during casework, followed by choosing context management actions, and end with ensuring the quality of the context management. All examples given will be for the field of forensic firearms examination but can be translated to other forensic domains.

3.1 Identifying exposure to contextual information

For efficient context management it is important to know during which steps of the examination process examiners might be exposed to contextual information. This can be visualized in a flow-chart of all steps that are taken from the moment of the evidential material entering the laboratory until the report is sent to the applicant. The resulting flow-charts from different institutes, departments, and even sections may differ due to variations in routing systems.

For each step in the examination process the applicable level(s) of contextual information should be identified. During some process steps the examiner will not be exposed to contextual information while during other steps the examiner will be exposed to multiple levels of contextual information. For instance, during the comparison step the examiner could be exposed to all four levels of contextual information. Information about the number of cartridges in a cartridge-clip, the brand and type of the cartridges, cartridge cases and the bullets, and striking marks (starters) in the fired cartridge cases or bullets constitute Level 1 contextual information. Since current practice in firearm and ammunition examinations does not yet enable the examiner to objectively analyze striations or striation patterns in bullets and cartridge cases in a straightforward way, the firearms examiner will

examine the markings in the questioned material and the reference material simultaneously, which exposes the examiner to Level 2 contextual information. The examiner will read Level 3 contextual information in the description of the case, stating for instance that eyewitnesses report having seen two shooters. Results from earlier comparable cases might already have an effect on the expectations of the examiner (Level 4) before the examination starts.

The flow-chart developed for the Firearms section (before the implementation of context management) is shown in Figure 2. In this figure it is shown that the findings, interpretation, conclusion on paper, and the draft report (i.e. the completed report) are reviewed by another examiner. During peer review the examiner can be exposed to all levels of contextual information.

FIGURE 2

The Firearms section of the NFI, performs three different main types of examination:

1) general comparisons (including comparisons of questioned ammunition parts to each other and to reference ammunition fired from questioned firearms);
2) comparisons for police shooting incidents;
3) technical examinations, ballistics, and reconstructions.

Managing all levels of contextual information for these three types of examinations will take a lot of time. Based on an assessment of the possibilities of controlling the exposure to different levels of contextual information, the effect on workflow, and the associated costs, choices have been made about which levels of contextual information will be managed for the different types of examinations.

To illustrate the different levels of contextual information to manage for each type of examination we introduce a case description of a shooting incident which is present in the submitted case file:

"A shooting incident has occurred where one man shot twice towards two police officers driving in their patrol car. The police officers exited the car and together fired a total of three shots towards the man. The man dropped his firearm on the ground and ran

away before he could be apprehended. At the crime scene three standard law enforcement cartridge cases, two additional cartridge cases and the dropped firearm were recovered. Afterwards two bullets were recovered from the patrol car. The police officers state that each of them fired at least one shot, but due to emotional stress at the time they are not sure who fired one round and who fired two rounds"

Depending on the question posed about this shooting incident various levels of contextual information are managed, which will be discussed in the following sections.

3.1.1 General comparisons

For a general comparison, where only the dropped firearm, the two non-law enforcement cartridge cases and the two bullets are submitted, Level 3 contextual information is managed. The written information can easily be managed in this case by removing the case description from the case file. In this case example the information that two police officers saw one man shooting and saw him drop the firearm can be regarded as a source of potentially biasing and task-irrelevant information when the question is whether the two cartridge cases were fired from the firearm. The (perception of the) base rate is slightly altered by introducing fake cases in the case flow (Level 4). For such a case, Level 1 and 2 contextual information cannot be practically managed currently. For Level 1 this is because, for example, the striations and available starters cannot be compared separately from the bullets in which they are present. Finding agreement within one land engraved area might cause the expectation that the striations in the following land engraved area will also correspond. Furthermore, the results from the examined cartridge cases might - when giving evidence for one source (the firearm) - influence the subsequent examination of the bullets from the same case[4]. Managing this possible source of bias would result in a logistical challenge, as it might require more examiners than available. The Level 2 context information cannot be managed since the questioned material cannot be examined without knowledge of the reference material.

---

[4] This form of bias differs slightly from the original explanation of Level 1 contextual information. The findings from the comparison of the cartridge cases affects the comparison of the bullets. This type of bias is also related to a form of base rate (Level 4) where in most cases the received cartridge cases and bullets are fired from the same firearm. Note that while we focus on task-irrelevance and task-relevance, these terms should be used pragmatically. When tasks are too narrowly defined a resulting context management system will become too inefficient for everyday practice.

3.1.2. Comparisons for police shooting incidents

For a comparison resulting from police shooting incidents where additionally to the general comparison example the two service firearms and the three law enforcement cartridge cases are submitted, Level 1 contextual information is partially managed in addition to Level 3 and 4 contextual information. In this case example the information that each police officer shot at least once can be regarded as a source of potentially biasing and task-irrelevant information when the question is which of the three cartridge cases was fired from which law enforcement firearm.

   In the Netherlands, every police officer is equipped with the same brand and model of firearm, the same brand and type of ammunition, and a standardized number of cartridges in the cartridge-clip. Because of this, the remaining number of cartridges still present in the submitted two cartridge-clips will give direct information about which officer shot twice. When the manufacturing lots (stamped in the bottom of the cartridge, as part of the so-called headstamp) of the cartridges within the two cartridge-clips differ and match those of the questioned cartridge cases, this will give additional information on which cartridge case was shot from which firearm. While this information is surely relevant when answering the question which cartridge case was fired from which firearm, it should not be taken into account while comparing and interpreting the actual markings on the cartridge cases[5]. Exposure to this information (Level 1) is avoided by letting one examiner prepare the case for a second examiner. The first examiner creates the test fires (reference cartridge cases), using both manufacturing lots for each firearm and provides only these and the three questioned cartridge cases from the incident to a second examiner who performs the comparisons in the absence of any case information. This type of comparison is a so-called 'blind comparison'. A blind comparison differs from a general comparison in that the second examiner, performing the comparisons, will not receive the 'context managed' case file and the possibly available related task-relevant evidence.

3.1.3. technical examinations, ballistics, and reconstructions

While for comparison cases the emphasis generally lies on the removal of contextual information, this is usually not the case for technical examinations, ballistic examinations,

---

[5] This information is domain-relevant and might be taken into account for the overall interpretation of the evidence, but should not be part of the specific examination task, the comparison of the markings in the cartridge cases.

and reconstructions. To be able to investigate these cases, as much information as possible about the circumstances at the time of the incident is required to, for example, set up experiments that are as meaningful as possible for the particular case. When the question for the case example would be to reconstruct the bullet trajectories based on the bullet holes in the patrol car, the firearms examiner would probably be exposed to all case information. To ensure that the comparisons within this case are as unbiased as possible, the comparisons will be performed again by a second examiner as a blind comparison in the absence of any case information.

3.2 Use of contextual information

In cases where context management results in the removal of contextual information prior to examination this does not mean that this information is irrelevant for the case as a whole. While this information may be irrelevant for performing the task of the comparison itself, it might change the interpretation of the findings. As an example, information that a questioned firearm was stored in a rain gutter in-between two possibly related shooting incidents might explain differences in the striations in the fired bullets. Rain might have caused the barrel to corrode, changing the bearing surface of the barrel, and thus the resulting striations. The evidential strength of the differences in striations in the fired bullets from the two shooting scenes might change based on this contextual case information. Because of this possible effect, the removed information will always be made available to the examiner after the examination and interpretation of the markings. When the removed information changes the previous interpretation of the findings, the role that the information plays will become transparent. The former interpretation (*without* contextual information) and the adjusted interpretation (*with* contextual information) will both be described, clearly showing the effect of the information.

Depending on the level of contextual information to manage, the most appropriate moment for management must be identified. Level 3 information will mostly be managed at an early stage, before an examiner starts working on the case, while Level 1 and 2 contextual information will probably be managed at a later stage (during case preparation), since the questioned and reference material will actually have to be handled by an examiner.

3.3 How to manage contextual information: removing nothing unless, or everything except

When setting up context management the main focus of the NFI Firearms section was on the management of Level 3 contextual information for general comparisons. This type of comparisons calls for a source level interpretation, with a fixed question about the source of the ammunition parts and standard working procedures. For these examinations contextual information needs to be removed, while for examinations towards activity level questions, such as ballistics investigation and reconstructions additional contextual information is usually necessary. To manage Level 3 contextual information for source level comparisons, two distinct approaches have been tried by the firearms section:

1) The removal of only the potentially biasing information (removing *nothing, unless*);
2) The removal of all verbal and written information except for information that is necessary for the examination, comparison, and interpretation at source level (removing *everything, except*).

To remove only the potentially biasing information (Approach 1), a list of criteria for biasing information was made which contains the following criteria for general firearm comparisons:

- Information about the number of suspects, used firearms, and number of shots;
- Information about the source of various evidential material;
- Information about the relationship between the suspect and the victim;
- Information that directly (or indirectly) addresses the answer to a particular question, e.g. an eyewitness stating the brand of the firearm;
- Information about the past behavior of the suspect, or whether there is a confession.

To remove all verbal and written information except for information that is necessary for the examiner to be able to do his job (Approach 2), a list of the necessary information was made which contained the following criteria for general firearm comparisons:

- Information about the presence of blood on the evidential material;
- Information about other examinations which have to be performed on the same evidential material;
- Information about the conservation of the evidential material up until this moment;
- Information about materials that have been hit by a bullet.

The first - intuitively very appealing - approach of removing only the potentially biasing information has been tried for approximately half a year. At the start of this period the criteria and the procedure of removing contextual information were explained to all ten examiners with the help of case description examples. After this introduction all examiners were asked to manage the contextual information in newly submitted cases based on a rotation schedule where one examiner was responsible for the cases submitted in one week, followed by the next examiner during the next week. In this way each examiner was responsible for context management for at least two separate weeks. During this time a total of approximately 250-300 cases (± 25-30 per examiner) were managed. Examiners who examined the cases after the contextual information was managed were asked to provide feedback to the context manager on the application of the criteria when they disagreed. Afterwards, the case descriptions of ten randomly selected cases were provided as a test to all of the examiners and they were asked to state whether there was potentially biasing information that should be removed and if so, which (parts of) sentences they would remove from the description based on the criteria. For only three of the ten case descriptions the examiners unanimously agreed whether or not there was potentially biasing information present. For the other seven case descriptions two to four examiners disagreed with the other examiners. The results of the next step of deciding which information should be removed according to the criteria were even less consistent. The amount of removed information varied a lot between examiners and when discussing these differences it was not always possible to find a consensus. An example of such a sentence is: "A firearm was used to shoot at the front door of a house". Some examiners thought this was not potentially biasing, because shooting does require a firearm, while other examiners argued it was potentially biasing, because the sentence referred to 'a' [one] firearm. Figure 3 provides an example of the diversity seen in removing contextual information from one case description by ten different examiners. The marked (parts of) sentences represent the information that should be removed according to the examiners when applying the criteria.

The first approach of only removing the potentially biasing information was chosen for reasons of expected efficiency and intuitive appeal. Removing information takes time, so when choosing to do this only when it is absolutely necessary the invested time is minimized. During the project, approximately one-fourth of the cases were found to contain potentially biasing contextual information, according to our criteria. When only removing information from these cases, a minimum amount of time needs to be invested in the other three-fourth of the cases. On the other hand, setting-up a complete list of the criteria for

potentially biasing information and implementing these criteria in such a way that every examiner applies them in the same way turned out to be very difficult. This approach resulted in a lot of variation between examiners when considering the removal of biasing context information. The time spent on discussions resulting from different interpretations of the criteria outweighed the timed saved by only removing contextual information when it is potentially biasing.

FIGURE 3

The second approach of removing all verbal and written information (unless necessary for the examiner to be able to do his job) has been tried after the first approach and showed far more consistent results. At the start, the criteria and the procedure of removing contextual information and providing the necessary information were explained to all ten examiners. A context management group of three examiners was created to manage the contextual information in newly submitted cases based on a weekly rotation schedule. When context management of a case is unsuccessful, in the sense that part of the description (containing potentially biasing information) is still present in the case file, or when it turns out that not all necessary information from the case description was provided, the context manager is informed through a feedback form. The implementation of Approach 2 was not concluded with a test similar to that conducted after Approach 1. We will therefore provide an overview of the feedback forms received after implementation of Approach 2. After the initial familiarization with the approach, context management turned out to be unsuccessful approximately once every forty cases. In approximately half of these cases contextual information was still present somewhere in the case file, and in the other half the context manager had neglected to provide necessary information which was present in the case description.

     The second approach of removing all verbal and written information (unless task-relevant) ensures a straightforward, fast and objective procedure when compared to the removal of only the potentially biasing contextual information. When following this procedure every case will need to be managed, which will result in additional time that will have to be invested for all cases without potentially biasing information. However, the decrease in time spent on discussing the interpretation of criteria for removal resulted in an

overall lower time investment compared to the first approach, because the criteria for Approach 2 were apparently less complicated to apply. In current practice following Approach 2, the removal of information takes about 5-10 minutes for every case. This is more than reasonable in relation to the total time invested on a case (many hours rather than minutes).

When context management is unsuccessful because available necessary, task-relevant information was not provided to the examiner performing the examination, this might cause contamination issues, unnecessary exposure to for example blood, or a loss of criminalistic overview for the examiner, resulting in misleading conclusions or missed opportunities. Possibilities for further forensic examinations might be overlooked and in the worst case lost because the examiner was unaware of their importance to the case. Until now (after more than three years) no irreversible problems have arisen from this approach. To minimize these risks and to ensure that the benefits of context management outweigh the potential downsides, the context manager and the examiner should both be experts in the specific forensic discipline.

3.4 What to do when the examiner was exposed to task-irrelevant contextual information?

After implementation of context management, task-irrelevant information might still reach the examiner in two distinct situations:

1) The contextual information was not managed correctly;
2) The examiner was exposed to task-irrelevant information in spite of correct initial context management.

When the contextual information was not managed successfully (Situation 1), in the sense that part of the description (containing potentially biasing information) is still present in the case file, the examiner could redo the context management and give the complete case to another examiner or could prepare the case for another examiner who will then only receive the questioned and reference material to compare and the propositions to consider, the earlier mentioned blind comparison. The latter procedure could of course be applied to every case, but might not be suitable for every team of examiners. When in practice one examiner is responsible for the complete examination process (pre-assessment, examination,

interpretation, and report writing) this procedure is dependent on the availability of a third examiner. The changes in logistics for each case will probably result in longer lead times.

An examiner can receive additional contextual information during the examination (Situation 2), through, for instance, communication with the applicant. When dealing with this new information it is important to note that as long as the pre-assessment, examination and interpretation are not completed, the received information could influence the examiner. When the new information arises, the examiner should again choose to let a second examiner perform a blind comparison. To be able to recognize exposure to new potentially biasing information, all examiners should be familiar with the context management procedure and the applicable criteria. Because this is a self-report measure in our system, the observed variations in interpretation of the criteria (see Section 3.3, Approach 1) can influence the reproducibility. In the firearms section of the NFI, contact with the applicant for general comparison cases is very rare so this will not have a large influence on the reliability of the system. When contact with the applicant is more frequent, a logbook of the conversations could be used. This enables the peer reviewer to double check the exposure to additional potentially biasing information. Alternatively, communication with the applicant could be done by another examiner than the one performing the examinations.

We note that the proposed procedure of a blind comparison is only possible when there are sufficient examiners available. In small groups this procedure will be more challenging. Problems might be overcome by letting another (related) group of examiners perform the initial management of contextual information. When, due to a limited number of examiners or repeated unsuccessful context management, it is not possible to perform a blind comparison, the result could still be reported but should be accompanied by a statement that no measures have been taken in this specific case to control potential contextual bias. Based on this information, the trier-of-fact will have to decide the overall added value of the examination's conclusions for their final decision in the case.

3.5 Dealing with unavoidable, but task-irrelevant information

Even though contextual information from the case file itself can be successfully managed, examiners could still be exposed to media reports. When there is a steady case flow this will not be a real problem when minimizing bias, because when contextual information is removed most cases are rather similar with respect to the number of ammunition parts and used firearms in the Netherlands. Although rare in the Netherlands, for cases which deviated from average, such as larger shootings with assault rifles and police shooting incidents, it

will be difficult to ensure that the examiners are not exposed to any contextual information through the media.

In some countries one examiner is responsible for the evaluation of different evidence types in one case (e.g. firearms, toolmarks and fingerprints). In the Netherlands this is currently not the case. A firearms examiner will only evaluate the firearm related evidence. This ensures that potentially biasing information from the results of one type of examination will not influence the subsequent examination by the same examiner. For institutes where one examiner is responsible for different types of examination this would result in a potential threat to the reliability of a context management system.

Discussing the findings from different forensic disciplines with colleagues might expose the examiner to task-irrelevant information. To minimize bias the examination should be performed before discussing the findings with colleagues or the examination should performed by another examiner as a blind comparison.

3.6 Quality control

To ensure that context management is performed correctly, different checks are built into the system. Before starting the examination the examiner will check whether all potentially biasing information is removed from the case file. After the comparisons and the interpretation of the results, the examiner will receive the removed contextual information and will check whether all necessary information from the case description was provided. During peer review of the entire case, where in our firearms section a second examiner reviews on paper the findings, interpretation, conclusion, and the draft report, the reviewer will also check whether context management was correct. If during either one of these checks it is shown that possibly biasing task-irrelevant information was still present before or during examination, a (second) blind comparison will be performed. When, after examination, it turns out that task-relevant information was mistakenly removed this will be reported back to the context manager, to improve the quality of context management in subsequent cases. When the previously removed information affects the interpretation of the results this will be reassessed.

3.7 Flow-chart

The project at the NFI's Firearms section resulted in context management for the three main types of examination mentioned. In Figure 4 a flow-chart shows the steps in the

implemented context management system. After initial classification of the case as a general comparison, a comparison for police shooting incidents, or a technical examination, ballistic examination or reconstruction, the context management steps and checks are depicted.

For a general comparison case, the context manager removes the case description and provides all necessary information, after which another examiner checks whether all potentially biasing information was removed before starting the examination of the case. If all contextual information - except for possibly available necessary information - was removed correctly, the second examiner will start the examination. If context management was not successful, this examiner will prepare the case for a third examiner as a blind comparison. After examination, the second examiner checks whether the necessary information available in the case description was provided by reviewing the original (unmanaged) case file. Then the findings, interpretation, conclusion, and the complete draft report will be peer reviewed. The examiner performing this task will again check whether context management was successful. If this was not the case and the markings are perceived as ambiguous, another examiner will be asked to perform an additional blind comparison.

For a police shooting incident the examiner prepares the case for a blind comparison. A second examiner, just receiving the fired ammunition parts and the test fires from the submitted firearms, will perform the examination, ensuring that the examiner cannot be biased by e.g. the number of cartridges left in the cartridge-clip and the lot number of the cartridges. The first examiner receives the findings, interpretation, and conclusion from the examiner performing the blind comparison and writes the complete draft report. During peer review it is checked whether a blind comparison was performed. If this has not happened the examination has to be redone by performing a blind comparison.

For technical examinations, ballistic examinations, and reconstructions, the emphasis generally lies on the gathering of information about the case, to be able to perform the experiments in the most informative way. When a comparison of markings is part of such a case, the evidence will be prepared for a blind comparison by the examiner. During peer review it is again checked whether a blind comparison was performed. If this has not happened the examination has to be redone by performing a blind comparison.

FIGURE 4

3.8 Dealing with the examiner

Implementing context management requires the cooperation of the examiners working with the procedures. The results of published studies on the risk of contextual bias in the forensic field are not all consistent[6] and some examiners might disagree with the necessity of such a system and might even feel that their level of expertise or even their integrity is in doubt. Such responses derive from a lack of understanding of cognitive effects. Since the cooperation of the examiner is critical for the success of context management to minimize bias, it is important to provide some education on cognitive effects to all forensic examiners.

The NFI, as well as the Netherlands National Police have been provided with such training on a number of occasions (such training has also been provided to forensic examiners in e.g. the United States, United Kingdom, Canada, Finland, and Australia). This illustrates the growing recognition that human examiners are the main instrument in many forensic domains. And that therefore their training must include how cognition relates to their work.

Forensic examiners who received cognitive training have for the most part been very supportive and receptive of new context management procedures to minimize potential bias.

3.9 A first step

The implemented context management system is a first step in ensuring that task-irrelevant contextual information will not influence the comparisons and interpretations by forensic firearm examiners. Other forms of context management could be applied when possible, but the implementation of context management for every additional level of contextual information will make the procedure more complex and time-consuming. When implementing context management this should be taken into account. It is expected that contextual bias mainly plays a role when traces are ambiguous [2,35]. An assessment of the costs compared to the benefits can result in a workable compromise.

The implemented system mainly focuses on the removal of contextual information for the benefit of comparisons at source level interpretation. When the examination request asks for an interpretation towards activity level such as with reconstructions, context

---

[6] An overview of studies on contextual bias in the forensic sciences can be found in Refs. [14, 42, 43]. The few studies in the forensic literature do not provide a clear answer. Some conclude that it is not a big issue [19-21, 31]. In contrast, other studies do find effects of contextual information [32].

management will also help in gathering the needed additional information, ensuring the optimal use of contextual information in casework.

5 Discussion and conclusion

We described the theoretical, psychological, and forensic aspects of contextual bias in forensic science, and we proposed a taxonomy of different sources of contextual irrelevant biasing information that can aid in the development of context management earlier [35]. Context management is a general approach for minimizing the impact of task-irrelevant information, and maximizing the role of task-relevant information. In this paper we applied that theory in the practice of firearm examinations with interpretation at source level. We described the development of context management in the Firearms section of the NFI. Context management can be implemented following the steps outlined in this paper.

The two main approaches that are described in this paper are first, the removal of only the potentially biasing information, and second, the removal of all verbal and written information except for information that is necessary for the examiner to be able to do his job. The first approach was thought of to be less time consuming, since only cases containing possibly biasing contextual information have to be managed. However, because this approach is more susceptible to differences in interpretation of the removal criteria by the examiners, resulting discussions made it more time-consuming. The second approach resulted in a more straightforward, faster, and more objective procedure to minimize contextual bias. Depending on the type of examination performed, the number of examiners available, the preferred level of quality assurance, and the attitude of the examiners, different choices and compromises can be made, resulting in different procedures.

During the implementation of context management some examiners stated that minimizing bias due to contextual information makes the examination more objective. This argument is - although appealing - not correct. Minimizing bias does not make the examination more objective, but it ensures that the subjective examination will be less susceptible to bias. The subjective nature of the examination is the very reason why procedures like context management should be implemented, to decrease the potential for bias.

Making an examination more objective would require the possibility to measure the relevant features before comparison and to base the examination and interpretation on quantitative data. Implementing context management may result in an increased awareness

of the issues of contextual bias, and as a result an awareness of the need of more objective, evidence-based methods. To be able to work towards (linear) sequential unmasking, to negate the potential bias caused by the examination of the reference material before or simultaneously with the trace (Level 2), one needs to define what features to extract from the trace. If one can objectively measure features of the trace and reference material, algorithms and methods exist to compare those features, and provide a quantitative (and objective) assessment of the evidential strength. Of course, even with more quantitative and objective methods, human judgment will still play an important role in forensic science. For instance, representativeness of the background and training data and sampling decisions, but also choosing features and comparison algorithms, will always require (subjective) human judgment. The risk of being affected by external factors increases with increasing subjectivity [44], but 'subjective' does certainly not imply 'unreliable', or 'not valid'.

References

[1]   I.E. Dror, R. Rosenthal, Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts, J Forensic Sci 53 (2008) 900-903.

[2]   I.E. Dror, D. Charlton, A.E. Péron, Contextual information renders experts vulnerable to making erroneous identifications, Forensic Science International 156 (2006) 74-78.

[3]   M.D. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion, California Law Review 90 (2002) 1-56.

[4]   M.J. Saks, D.M. Risinger, R. Rosenthal, W.C. Thompson, Context effects in forensic science: A review and application of the science of science to crime laboratory practice in the United States, Science and Justice 43 (2003) 77-90.

[5]   The National Research Council (U.S.).Committee on Identifying the Needs of the Forensic Sciences in the United States. Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington DC, 2009.

[6]   W.C.Thompson. Determining the Proper Evidentiary Basis for an Expert Opinion: What Do Experts Need to Know and When Do They Know Too Much, in: C.Robertson and A.Kesselheim (Eds.), Blinding as a Solution to Bias in Biomedical Science and the Courts: A Multidisciplinary Approach, Elsevier, New York, 2015, forthcoming.

[7]   C.J.F. Böttcher, Theory and Practice of forensic handwriting comparison, Tijdschrift voor Strafrecht LXIII (1954) 77-131.

[8]   W.E.Hagan. A treatise on disputed handwriting and the determination of genuine from forged signatures, Banks & Brothers, New York, 1894.

[9]   L. Miller, Bias among forensic document examiners: a need for procedural change, Journal of Police Science and Administration 12 (1984) 407-410.

[10]   L. Miller, Procedural bias in forensic examination of hair, Law and Human Behavior 11 (1987) 157-163.

[11]   R.E. Nisbett, T.D. Wilson, Telling more than we can know: verbal reports on mental processes, Psychological Review 84 (1977) 231-259.

[12] R.E.Nisbett, L.Ross. Human Inference: Strategies and Shortcomings in Social Judgement, Prentice Hall, 1980.

[13] T. Tversky, D. Kahneman, Judgment Under Uncertainty: Heuristics and Biases, Science 1124 (1974) 1128-1130.

[14] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: Problems, perspectives, and proposed solutions, Journal of Applied Research in Memory and Cognition 2 (2013) 42-52.

[15] W.C. Thompson, Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation, Law, Probability and Risk 8 (2009) 257-276.

[16] W.C. Thompson, What role should investigative facts play in the evaluation of scientific evidence?, Australian Journal of Forensic Sciences 43 (2010) 123-134.

[17] I.E. Dror, Practical Solutions to Cognitive and Human Human Factor Challenges in Forensic Science, Forensic Science Policy & Management: An International Journal 4 (2013) 105-113.

[18] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, Science and Justice 51 (2011) 204-208.

[19] L.J. Hall, E. Player, Will the introduction of an emotional context affect fingerprint analysis and decision-making?, Forensic Science International 181 (2008) 36-39.

[20] J. Kerstholt, A. Eikelboom, T. Dijkman, R. Stoel, R. Hermsen, B. van Leuven, Does suggestive information cause a confirmation bias in bullet comparisons?, Forensic Science International 198 (2010) 138-142.

[21] J.H. Kerstholt, R. Paashuis, M. Sjerps, Shoe print examinations: Effects of expectation, complexity and experience, Forensic Science International 165 (2007) 30-34.

[22] G. Langenburg, C. Champod, P. Wertheim, Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons*, J Forensic Sci 54 (2009) 571-582.

[23] N.K.P. Osborne, S. Woods, J. Kieser, R. Zajac, Does contextual information bias bitemark comparisons?, Science and Justice 54 (2014) 267-273.

[24] B. Schiffer, C. Champod, The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation, Forensic Science International 167 (2007) 116-120.

[25] R.B. Stacey, A report on the erroneous fingerprint individualization in the Madrid train bombing case, Journal of Forensic Identification 54 (2004) 706-718.

[26] W.C. Thompson, S.A. Cole, Lessons from the Brandon Mayfield case, The Champion 29 (2005) 42-44.

[27] D.E. Krane, S. Ford, J.R. Gilder, K. Inman, A. Jamieson, R. Koppl, I.L. Kornfield, D. Michael Risinger, N. Rudin, M.S. Taylor, W.C. Thompson, Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation, J Forensic Sci 53 (2008) 1006-1007.

[28] I.E. Dror, W.C. Thompson, C.A. Meissner, I. Kornfield, D. Krane, M. Saks, M. Risinger, Letter to the Editor - Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making, J Forensic Sci 60 (2015) 1111-1112.

[29] M. Page, J. Taylor, M. Blenkin, Context Effects and Observer Bias: Complications for Forensic Odontology, J Forensic Sci 57 (2012) 108-112.

[30] B. Found, J. Ganas, The management of domain irrelevant context information in forensic handwriting examination casework, Science and Justice 53 (2013) 154-158.

[31] G. Langenburg, F. Bochet, S. Ford, A Report of Statistics from Latent Print Casework, Forensic Science Policy & Management: An International Journal 5 (2014) 15-37.

[32] J. Kukucka, The Journey or the Destination? Disentangling Process and Outcome in Forensic Identification, Forensic Science Policy & Management: An International Journal 5 (2014) 112-114.

[33] C. Champod, Research focused mainly on bias will paralyse forensic science, Science and Justice 54 (2014) 107-109.

[34] C. Berger, R. Stoel, Letter to the Editor, Science and Justice 54 (2014) 510-511.

[35] R.Stoel, C.Berger, W.Kerkhoff, E.J.A.T.Mattijssen, I.Dror. Minimizing Contextual Bias in Forensic Casework, in: K.J.Strom and M.J.Hickman (Eds.), Forensic Science

and the Administration of Justice: Critical Issues and Directions, SAGE Publications, Inc., 2014, pp. 67-86.

[36] I.E. Dror, Letter to the Editor - Combating Bias: The Next Step in Fighting Cognitive and Psychological Contamination, J Forensic Sci 57 (2012) 276-277.

[37] I. Helsloot, J. Groenendaal, Naturalistic Decision Making in Forensic Science: Toward a Better Understanding of Decision Making by Forensic Team Leaders, J Forensic Sci 56 (2011) 890-897.

[38] I.Dror. Cognitive neuroscience in forensic science: Understanding and utilising the human element. Philosophical Transactions of the Royal Society. 2015, In Press.

[39] A.J. Meulenbroek, T.J.P. de Blaeij, A.D. Kloosterman, Richtlijnen borgen onbevooroordeeld DNA-onderzoek Stapsgewijze benadering voorkomt 'post hoc target shifting, Expertise en Recht 6 (2009) 119-129.

[40] M. Baiker, I. Keereweer, R. Pieterman, E. Vermeij, J. van der Weerd, P. Zoon, Quantitative comparison of striated toolmarks, Forensic Science International 242 (2014) 186-199.

[41] W. Kerkhoff, R.D. Stoel, C.E.H. Berger, E.J.A.T. Mattijssen, R. Hermsen, N. Smits, H.J.J. Hardy, Design and results of an exploratory double blind testing program in firearms examination, Science & Justice (2015).

[42] I.E. Dror, S.A. Cole, The vision in 'blind' justice: Expert perception, judgment, and visual cognition in forensic pattern recognition, Psychonomic Bulletin & Review 17 (2010) 161-167.

[43] M.D. Risinger, The NAS Report on Forensic Science: A Glass Nine-Tenths Full (This is About the Other Tenth), Jurimetrics 50 (2009) 21-34.

[44] I.E. Dror, The ambition to be scientific: Human expert performance and objectivity, Science and Justice 53 (2013) 81-82.
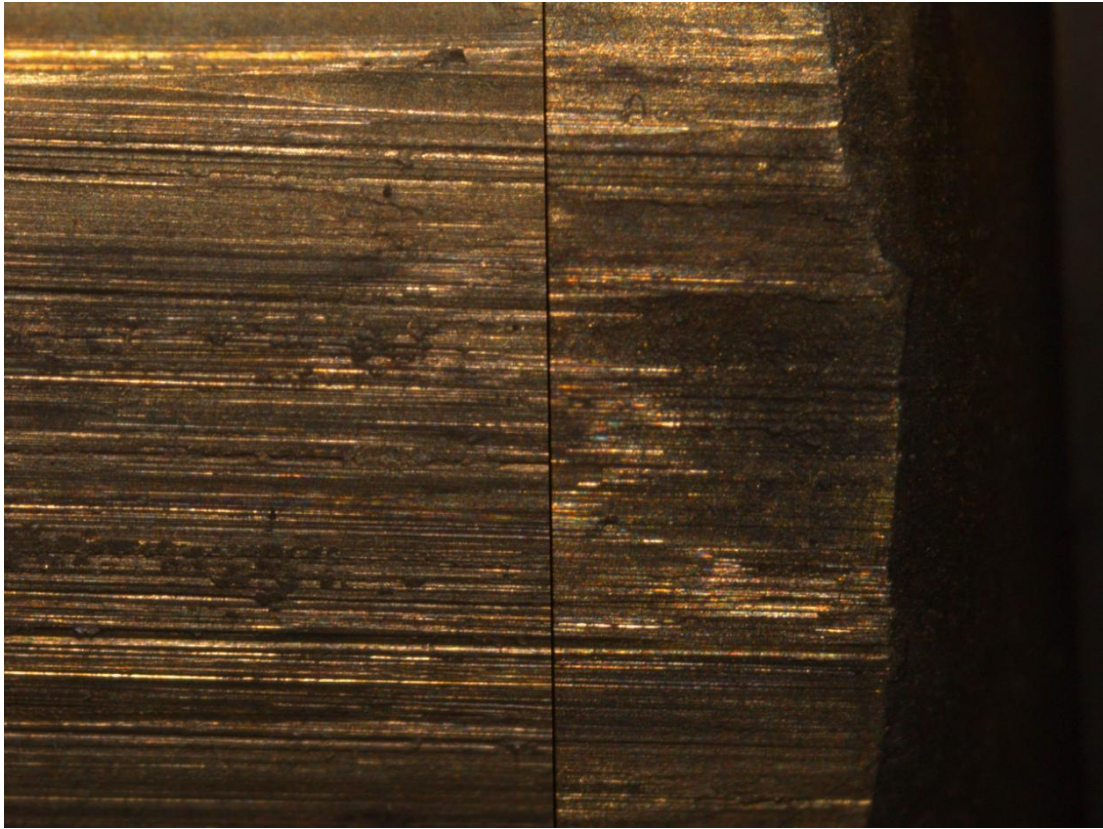
Figure 1 Example of a comparison of one area of two bullets where some matching striations are visible at the bottom while the other striations seem to vary between the two bullets.
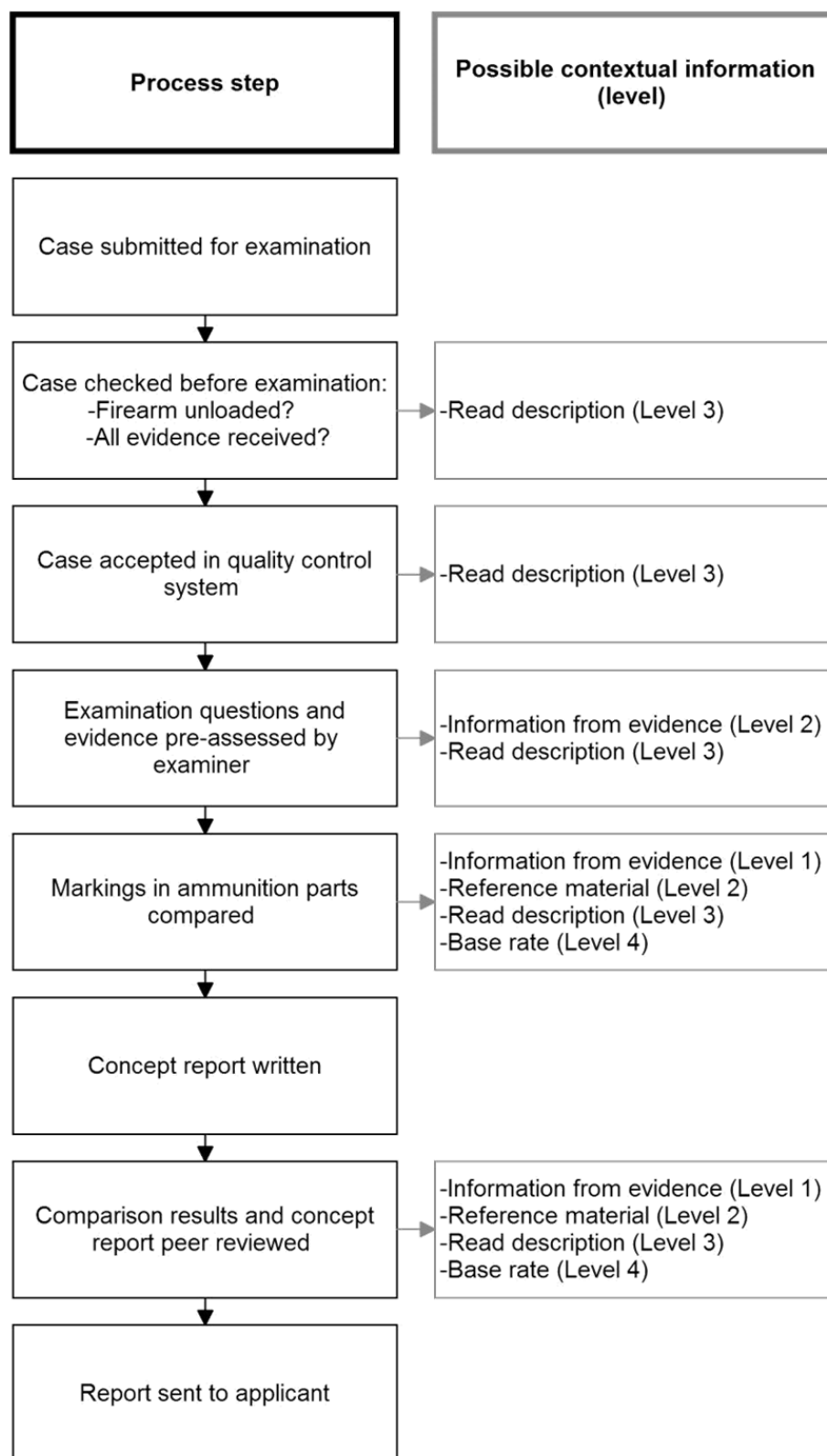
Figure 2 Flow-chart of casework for a Firearms section. The possible sources and corresponding levels of contextual information are given for every step of the process.

Figure 3 An example of the diversity seen in removing contextual information from one case description by ten different examiners. The marked (parts of) sentences represent the information that should be removed according to the examiners when applying the criteria.
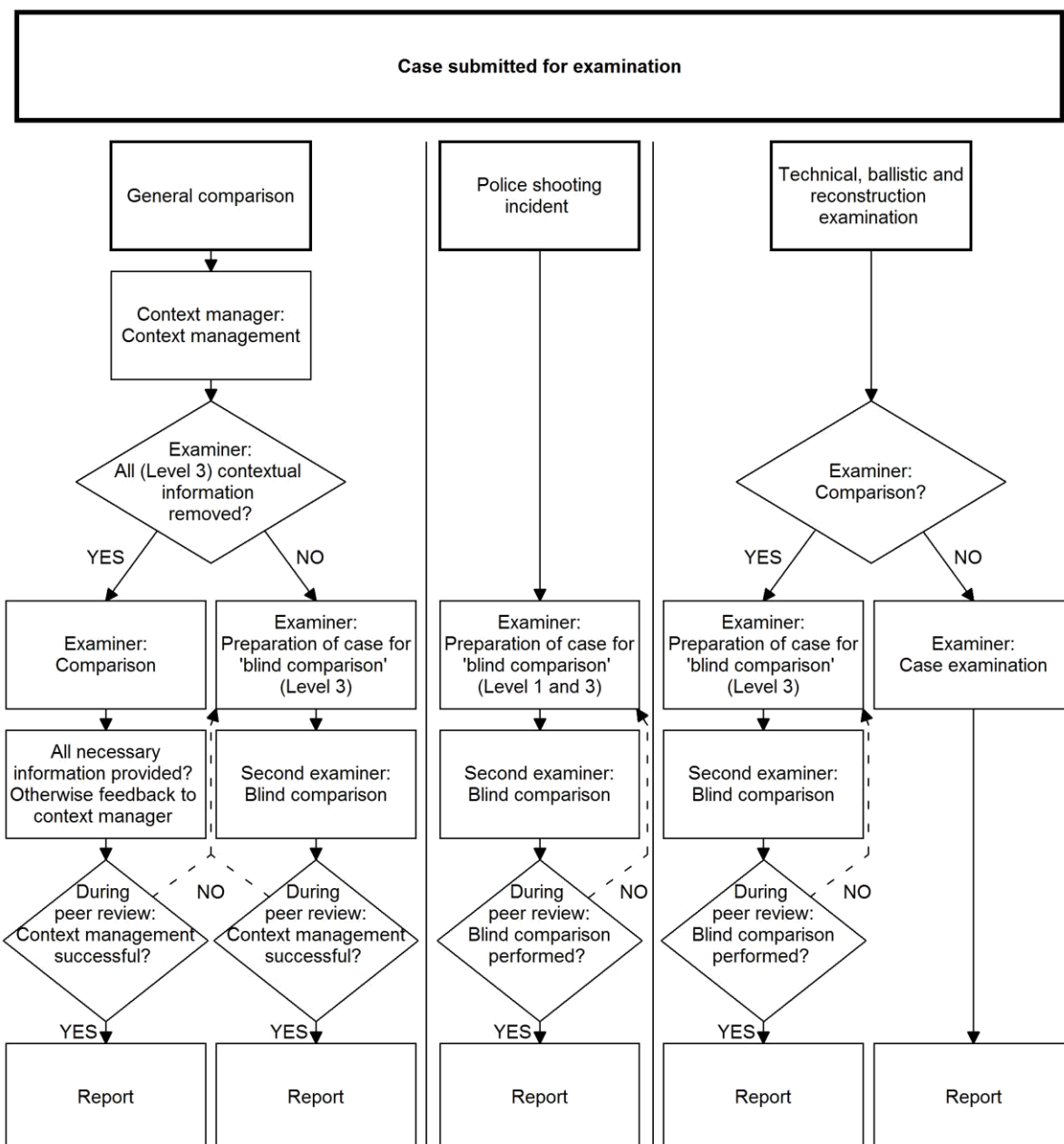
Figure 4 Flow-chart of the three different routes of context management based on the three main types of examination. Within the flow-chart the levels of contextual information which are (partially) managed are shown.