Minimizing Bias in Forensic Handwriting Examinations

L.C. Alewijnse,[1] E.J.A.T. Mattijssen[2] and R.D. Stoel[3]

Abstract: *The purpose of this paper is to contribute to the increasing awareness about the potential bias on the interpretation and conclusions of forensic handwriting examiners (FHEs) by contextual information. We briefly provide the reader with an overview of relevant types of bias, the difficulties associated with studying bias, the sources of bias and their potential influence on the decision making process in casework, and solutions to minimize bias in casework. We propose that the limitations of published studies on bias need to be recognized and that their conclusions must be interpreted with care. Instead of discussing whether bias is an issue in casework, the forensic handwriting community should actually focus on how bias can be minimized in practice. As some authors have already shown (e.g., Found & Ganas, 2014), it is relatively easy to implement context information management procedures in practice. By introducing appropriate procedures to minimize bias, not only forensic handwriting examination will be improved, it will also increase the acceptability of the provided evidence during court hearings.*

Reference: Alewijnse, L.C., Mattijssen, E.J.A.T., Stoel, R.D. (2015) Minimizing Bias in Forensic Handwriting Examinations. J. Forensic Document Examination, Vol 25, pp. 17-26.

Key Words: Contextual bias, Confirmation bias, Base rate bias, Selection bias, Content information management

1. Introduction

Forensic handwriting examination is regarded as a difficult field of expertise that relies on human observations, analysis and expert judgment. Dror (2013) argued that with increasing subjectivity, the risk of being affected by external factors increases. This is particularly true for difficult or ambiguous questioned material encountered during handwriting examinations. Much has been written on bias in forensic science, but only a few papers focus specifically on forensic handwriting examination. Experiments on the occurrence of bias have been performed with mixed results with extensive discussion afterwards. Some papers have been written on methods to reduce the likelihood and potential effects of bias. It is important to understand the importance of bias in casework ant to take measures to minimize this bias.

1 Netherlands Forensic Institute.
Email: alewijnse@nfi.minvenj.nl
2 NFI. Email: e.mattijssen@nfi.minvenj.nl
3 NFI. Email: r.stoel@nfi.minvenj.nl

We will begin by discussing several types of bias that might occur in casework, and provide an overview of literature on bias in forensic handwriting examination. Secondly, we describe an experiment that was conducted in 2012. We include this experiment not for the results, but to show that it is extremely difficult to perform experiments to prove bias does, or does not, exist. Finally, we discuss sources of bias and provide suggestions which will help to minimize potential bias in forensic casework.

Bias can be defined as an inclination that inhibits impartial judgment. Bias can occur in any phase of the handwriting examination; during the phases of selection, pre-assessment, comparison, interpretation and/or reporting the findings. Individual differences between FHEs in susceptibility to bias make the matter even more complex. Additionally, some types of evidence are more vulnerable to bias than others (Kassin, Dror & Kukucka, 2013).

Confirmation bias is probably the most relevant type of bias for the FHE. Confirmation bias is the tendency of looking for confirming evidence for a

hypothesis that is regarded to be true while disregarding conflicting evidence. Alternative terms associated with *confirmation bias* are *experimenter bias* or tunnel vision. Two subcategories of confirmation bias are *selection bias* and *contextual bias*.

The selection of reference writing provides room for selection bias as it involves subjectivity in the decisions about what evidence to collect and to examine. In an ideal setting, the conditions of selecting the examination material are clearly defined and protocols are followed. But in practice, there are no generally accepted international standard operating procedures for the collection of reference material. For example, in the literature, the number of recommended reference signatures ranges from a minimum of five, to ten, to as many as possible known signature samples. This number differs among FHEs and can also be case dependant.

Contextual bias is the tendency for a conclusion to be influenced by extraneous information. This type of bias results from domain-or task-irrelevant context information. Domain or task specific extraneous information can be introduced through police case descriptions or from results of examinations performed by other forensic disciplines (Risinger et al., 2002). The potentially biasing information does not always have to be very clear. It can range from a case description that mentions that the suspect admits to have written the threatening letter to more subtle forms such as the content of the writing itself.

2. Publications on bias

More than a hundred years ago, Hagan (1894) addressed the issue of bias and irrelevant factors in *A Treatise on Disputed Handwriting and the Determination of Genuine from Forged Signatures*. Hagan suggested that the FHE should not know the the outcome or results the party employing the FHE wants. He points out that, "Where the expert has no knowledge of the moral evidence or aspects of the case […] there is nothing to mislead him, or influence the forming of an opinion." Although he did not use the actual term *bias*, Hagan was clearly referring to contextual bias.

Sixty years later, Böttcher (1954) addressed blind testing for document examiners in his paper, "Theory and practice of forensic handwriting comparison," which was written in Dutch. Ninety years after Hagan's publication, Miller (1984) published a paper on bias among forensic document examiners. He proposed implementing procedural changes that would minimize bias that could lead to erroneous decisions. The experimental setup of this study and its results were extensively criticized by the FHE community (see Stoel, Dror & Miller, 2013 and Found, 2014). Miller's concerns were not accepted nor adopted. The forensic science community did not even consider it as a serious topic for discussion. The reason for this was most likely the result of the belief that Miller's research somehow attacked the discipline as not being a scientific one.

At the beginning of this century there was renewed attention on bias within the forensic community, (Dror et al. 2006) The authors used the misidentification of Brandon Mayfield based on a fingerprint found at the crime scene of the Madrid Bombing case as context information in an experiment. Their study showed that with new context provided in a case, four out of the five fingerprint examiners provided different conclusions on previously examined evidence, from which three examiners changed their conclusion from identification to exclusion.

Furthermore, the 2009 National Academy of Science report, "Strengthening Forensic Sciences in the United States: A Path Forward," stressed that research on the sources and effects of bias in forensic science disciplines is necessary. The example that is given points towards studying the effects of contextual bias in forensic practice to determine whether and to what extent the results of forensic analyses are influenced by knowledge, regarding the background of the suspect and the investigator's theory of the case. Additionally the report expresses the need to develop detailed protocols for performing forensic examinations.

Various authors described bias in forensic science from a cognitive perspective (e.g., Risinger et al., 2002, and Saks et al. 2003). The reader is also referred to a more recent paper by Kassin, Dror and Kukucka (2013) which provides a psychological perspective on confirmation bias, an overview of studies on forensic confirmation bias, and proposes best practices to reduce bias.

When focusing on bias in forensic handwriting examination there is an interesting contribution

on the subject by Sulner (2014). He writes about cognitive bias in forensic handwriting examinations and the way in which decision-making and expert testimony is susceptible to biasing influences. In Sulner's paper, the impact of confirmation bias is illustrated with case studies. Another relevant study for FHEs is presented by Kukucka and Kassin (2012). The authors demonstrated that knowledge on a withdrawn confession of a suspect affects the conclusions of FHEs: they were more likely to give erroneous conclusions.

3. Experiment on contextual bias

To explore the influence of context information on forensic handwriting examination, we set up an experiment. The purpose of the experiment was to study the effects of contextual bias. The main question of the experiment was To determine if FHEs would draw similar conclusions on the same writings when confronted with different case descriptions.

3.1 Participants

A group of FHEs was approached to participate by sending a request on the docexam email list in October 2012. Twenty-four FHEs from Australia, Canada, the UK and the US were interested in participating in this experiment. They were not made aware of the main purpose of the study, i.e., to investigate bias.

Special attention was given to the experience of the participants. All but two of the FHEs were court-qualified. The two non-qualified FHEs worked for two and three years in the field and were still in training at the time. The overall experience ranged from two to 40 years, with an average of 14 years. The number of cases that the court-qualified FHEs handle per year ranged from 15 to over 200 cases, with an average of 73 cases a year.

3.2 Procedure

This study focused on signature examinations. We wanted to introduce the bias ourselves. To minimize the problem that bias could unintentionally result from the provided evidence itself we chose to ask the FHEs to compare signatures instead of complete texts. A signature is believed to contain less information about the content of the writing than text could.

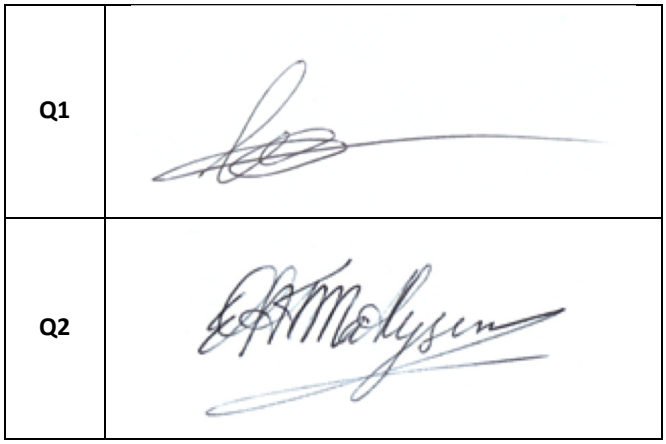

Figure 1. Questioned signatures:The two questioned signatures, Q1 and Q2, used in this experiment. The signatures are both simulations.

Because the effect of context information is most likely to occur when the writing is ambiguous, two simulated signatures from the study by Mattijssen, Van den Heuvel and Stoel (2011) were chosen. From the results of that study, with twelve participating FHEs, we could deduce that these signatures might be difficult to examine. Six of the FHEs gave an inconclusive for the first simulation, and for the second simulation, six conclusions were incorrect: both providing indications for ambiguous evidence. The two selected signatures are shown in Figure 1. They are further referred to as *questioned signatures 1* and *2* or *Q1* and *Q2*.

Each of the questioned signatures was included in a separate mock case. Each mock case consisted of one of the two scenarios. The scenarios are case descriptions with strong task-irrelevant information. Within one scenario the FHEs would be confronted with either strengthening, i.e., where the case description suggests that the questioned signature is a simulation, which is true, or opposing information, i.e., where the case description suggests that the questioned signature is authentic when it was a simulation.

1. The first scenario involved a loan agreement between a woman and her ex-husband. The strengthening information suggested that the signature of the woman was simulated by the ex-husband to get the loan without her knowing. The opposing information stated that a woman refuses to pay back a loan, indicating that the signature is authentic. The case description mentioned that the bank confirms the authenticity of the signature.

	Context information	N	Correct (Inc)	Incorrect (Inc)	Correct	Incorrect
Q1	Opposing	10	10 (6)	0 (0)	100%	0%
	Strengthening	11	10 (8)	1 (1)	91%	9%
Q2	Opposing	10	10 (2)	0 (0)	100%	0%
	Strengthening	12	8 (3)	4 (0)	67%	33%

Table 1. Results: The experiment contained two questioned signatures, with signature origin opposing and strengthening case descriptions. From all participants (N=24), only 21 completed both examinations. One participant only examined Q2. The number between brackets is the number of participants who gave an inconclusive. The percentages of correct and incorrect conclusions are shown In the last two columns, and are based on the conclusions disregarding the inconclusive opinions. (N = the number of participants, Inc = Inconclusive)

2. The second scenario involved a vehicle sales agreement. The strengthening information was given by the claim of the buyer: he has never seen an agreement or signed it but has to pay for the car. The opposing information claimed that three reliable witnesses can confirm a man signing a sales agreement for a car.

The FHEs were requested to provide conclusions on the questioned signatures for each of the two mock cases. They were requested to read the case description to learn which signature at the bottom of the contract was questioned. The FHEs were then asked to give a categorical conclusion (authentic or simulation) and in case of significantly limiting factors in the examination, they could check the box for inconclusive. *Inconclusive* is used synonymously with *no conclusion* when the examiner is at the zero point on the confidence scale (McAlexander, Beck & Dick, 1991).

3.3 Results of the experiment

The results on the authenticity assessment of signature Q1 and Q2 (Table 1) show that no errors were made when opposing context information was given. Incorrect conclusions were only given when strengthening information was provided. For Q1, this incorrect conclusion was marked as inconclusive. However, for Q2, four errors were made. None of these conclusions were marked as inconclusive.

Additionally, it is worth mentioning that the number of inconclusive scores for Q1 was relatively high in comparison to those for Q2. Both these observations correspond to the scoring by FHEs in the previous study. Again, the ambiguity of the origin of the Q1 signature seems to present itself in the high inconclusive scores, and of the Q2 signature in the incorrect conclusions.

The results for Q2 are more complex. The relative high percentage of incorrect conclusions when given strengthening information (33%) contrasts to the low percentage of incorrect conclusions when given opposing information (0%). Four out of twelve FHEs gave incorrect conclusions when strengthening context information was given, i.e., the case description suggested that the questioned signature was a simulation, as it was.

3.4 Interpretation of the results

We set up a simple assessment of the possible outcomes we could have obtained. There are several possibilities for evidence of presence or absence of bias which are explored in more detail.

First, let us take a look at what the results would be in a no contextual bias situation. We then expect equally proportionate percentages for opposing and strengthening case information. In Table 2 the most extreme outcomes are shown by 100% correct or 100% incorrect rates for both opposing and strengthening information. These similar percentages for opposing and strengthening case information show that the conclusions were not affected by the provided contextual information. The correct or incorrect rates do not have to be 100%. Without contextual bias the expected percentages are expected to be the same when given either opposing or strengthening information. Intermediate percentages (e.g., 80%-20%)

Context information	Correct	Incorrect
Opposing	100%	0%
Strengthening	100%	0%
Opposing	0%	100%
Strengthening	0%	100%
Opposing	80%	20%
Strengthening	80%	20%

Table 2. Confirmation of bias

correct-incorrect rates can also be expected in case of absence of bias. These percentages reflect that neither opposing nor strengthening information affected the FHE's conclusions.

Now, let us take a look at what the results would look like under strong contextual bias. Evidence for bias will be reflected by disproportionate incorrect rates for opposing and strengthening information. Table 3 shows the most extreme configurations of evidence for bias. In the first part, all participants who read strengthening case information gave correct conclusions and all participants who read opposing case information gave incorrect conclusions. Intermediate percentages (e.g., 80%-20%) will also indicate evidence of bias, but here the influence seems to be less strong.

3.5 Discussion

This experiment was initially set up as a within-subject design; each participant had to judge the same signature twice over a period of time, with different context information. In a former study on simulation quality, twelve FHEs had already given conclusions on a large set of questioned signatures (including the two simulated signatures used for this study). All FHEs from this previous study were again approached to participate in the new experiment. This is a common approach in studies on context bias, and is described by Dror, Charlton and Péron (2006), who emphasize the difficulty of conducting proper scientific research on bias in forensic science. Unfortunately, only three of the twelve former FHEs responded to the new call, which made a within design undesirable. Because of

Context information	Correct	Incorrect
Opposing	0%	100%
Strengthening	100%	0%
Opposing	20%	80%
Strengthening	80%	20%

Table 3. Strong evidence for bias

this we decided to change the design and continue with a new group of participants who had never seen the signatures before.

In hindsight, neutral context information (e.g., no information at all) could have been included in the experimental setup, next to the opposing and strengthening information. The data resulting from this neutral context information situation could have served as the control group. The drawback would be that the group sizes would have become even smaller than they were now.

When evaluating the results of the study, it is important to keep in mind two factors: sample size and effect size. When the contextual information provided within both scenarios induces bias to a certain extent, the probability of proving this with a given study is dependent on the sample size. With a large number of participants even a small difference between groups can be demonstrated, while this potential decreases with a decreasing number of participants. Especially when the expected biasing effect becomes smaller, a small number of participants might be insufficient to find an existing effect in the data (imagine Table 3 with a total sample size of four in each conditional). This brings us to the effect size. When the provided information induces a strong bias, less participants will be needed to find this in the data. Kukucka & Kassin (2012) demonstrated that in their study, the manipulation of case information apparently had an effect on the conclusions drawn by FHEs. In our experiment, however, we honestly did not know whether the manipulation of the provided context information was strong enough to affect the decisions of FHEs and, if so, how much. When the biasing effect would turn out to be small, the number of participants in this experiment (N=24), is not enough. A so-called

statistical power analysis (see Cohen (1988); details not shown here), shows us that for the detection of small effects we would have needed approximately 600 participants. For the detection of a large effect, 40 participants are required in this situation. In other words, the number of participants in this study was not sufficient. To draw conclusions from any experiment, sample size should be appropriate for the expected effect.

The initial purpose of the experiment was to study the effects of contextual bias. The current experiment could not provide (strong) evidence for effects of bias by introducing task-irrelevant case information on FHE conclusions regarding signature authenticity, but was also insufficient in proving that bias does not affect the examination process. Taking into account the limitations of the study design we would like to stress that one must be very careful when drawing conclusions from experiments such as this one, about whether bias exists or not. Even if the results would follow the pattern of the upper part of Table 3, i.e., the strongest indication of bias, study design (e.g., sample size, representativeness of the signatures and the experts, the context information, the unrealistic test environment) would impede a straightforward generalization of the results to casework.

4. Minimizing bias

After increased attention on bias in the literature e.g., Found (2015), the importance of bias in casework currently is acknowledged by only a few in the forensic handwriting community. Under the realistic expectation that irrelevant information will bias forensic casework, methods and ready-to-implement procedures have been proposed in the recent years to minimize the influence of bias during forensic casework. These have been proposed both within the field of forensic handwriting examination, as well as in other forensic fields where visual comparison and human decision making play an important role (Krane et al., 2009; Found & Ganas, 2013; Stoel et al. 2014; Mattijssen, Stoel & Kerkhoff, 2015; Dror et al. 2015; Mattijssen et al., in press). At this time the general focus lies on minimizing the influence of context information. To provide a framework to think about the possible sources of context information with which FHEs might be confronted during forensic casework, Stoel, et al. (2014) suggested a taxonomy including four levels:

- Level 1: the trace;
- Level 2: the reference material;
- Level 3: the case information;
- Level 4: 'base rate' information;

Subsequently, Dror et al. (2015) added a fifth level with regard to organizational and cultural factors.

When looking at the field of forensic handwriting examination, all of the described levels might play a role. We would like to address the first four levels and provide possible solutions to minimize their influence on forensic casework through e.g., context information management (CIM, Mattijssen et al., in press; see also Stoel et al. 2014, Found & Ganas, 2013, and Risinger et al., 2002).

Content information management has become a part of the standard operating procedure in some forensic laboratories with respect to forensic handwriting examination, such as at the Victoria Police Forensic Services Department in Australia and at the Netherlands Forensic Institute in The Netherlands (Found & Ganas, 2013). Both laboratories have proposed procedures where the first person of contact is the department chief, a specially trained evidence control unit, or a content information manger (CIM) that prepares the case for further examination by an FHE. Found and Ganas (2013) describe that all essential case information is noted on a Document Unit Casework Information Sheet (DUCIS) and a Context Manager (CM) is responsible for passing on only the domain-relevant information to the FHE, while withholding the domain irrelevant information. The handwriting department at the Netherlands Forensic Institute has introduced CIM in 2014. Here, all written and verbal case information is removed during the intake, except for the information that is deemed necessary for the FHE to perform the examination in a proper way. A list of criteria is composed with the sort of information that is critical for the FHE to known. After the CIM has managed the information during the intake, he/she cannot perform the actual examination nor peer review the case. When the FHE has drawn his/her conclusion, and the case is peer reviewed, the original case description becomes

available before the findings are reported. In this way, potential errors introduced by the CIM in withholding relevant information are avoided.

4.1. Level 1: The trace

The first level contains the context information that is inherent to the examined questioned material itself. For example, the addressee or content of a threatening letter or the content of a contract of which only the source of the signature is questioned. The FHE is asked to provide a conclusion on the source of the writing. The FHE is expected to compare the characteristics from the written text or the signature with the provided reference material. He/she is not supposed to take the meaning of the text itself into account. This poses a problem as both the meaning of the text and the characteristics to compare are both part of the same object.

Although the meaning of the text is an inherent part of the evidential material it might be possible to minimize bias. The Fischhof Method (Sulner, 2014), of upside-down comparison offers the FHE a means of avoiding undesirable observer effects from the written text. This inverted image comparison can be applied to handwritten texts as well as to signatures. Another method is to use an overlay. When a signature examination is requested, an overlay might be used to make the additional text of the contract unavailable to the FHE. An alternative would be to scan or copy the text of the threatening letter at high resolution, after which a selection of non-biasing words could be provided to the FHE in a random order, highly decreasing the perception of the meaning of the text. An often heard argument against using only a selection of the material as described above is that the examiner loses important information. While this may well be a good argument, we believe it should be made explicit exactly what the examiner thinks he/she is missing. Future studies could focus on counter effects of such CIM procedures in practice.

4.2. Level 2: The reference material

The second level contains contextual information coming from the reference material. When examining the questioned and reference material simultaneously, perceived characteristics from the reference material could be superimposed on the questioned material. An FHE might unconsciously start looking for characteristics in the questioned material which are seen in the reference material. When traces are ambiguous this might result in the FHE perceiving characteristics which might not have been observed if the questioned material was examined before examining the reference material. Even more problematic, the FHE might even perceive characteristics which aren't there to begin with.

Selection bias, as explained in the introduction, might also be regarded to attribute to bias resulting from the reference material. Depending on the selection of the appropriate reference material, and in particular when only small quantities of writing are available, the visible characteristics might differ between samples. A strong selection bias regarding the reference material could thus influence the results of the examination.

Sequential unmasking is a procedure that could be introduced to minimize contextual bias by the reference material. First implemented within the field of forensic DNA examination, this procedure states that the questioned material should be examined first and its features should be described. Sequentially the reference material should be examined and described. Next, the features can be compared. When new features are added, this must be accounted for and documented as such. Following this procedure will prevent the process of post-hoc target shifting (Krane et al., 2008 and Thompson, 2009). Implementing (linear) sequential unmasking (Krane et al., 2008; Dror et al., 2015) in forensic handwriting examination requires a detailed description of the questioned material and the handwriting features which are expected to be seen in the reference writing. These features should be described as objectively as possible. Consensus about the important features for examination is a prerequisite.

Due to the subjectivity associated with noting the handwriting features and the amount of work in practice, it is more efficient to take a look at the nature of the reference writing before the comparison phase, i.e., does the general writing style match and is it contemporaneous? This can easily be done by the Context Manager (CM), CIM or the Evidence and Quality Control Officer (EQC), proposed respectively

by Found and Ganas (2013), Mattijssen et al. (in press) and Saks et al. (2003).

4.3. Level 3: The case information

The third level contains all types of received written and non-verbal context information, such as case descriptions, eyewitness statements, communication with the police, the prosecutor, or colleagues. The bias that was introduced in our experiment (i.e., the extraneous case information) also belongs to this level. CIM will help to minimize bias resulting from these sources. For example, a CIM can remove the additional domain- or task-irrelevant information from the case file before an FHE starts his/her examination. In that way the FHE will not see the information and can thus not be influenced by it. There could of course be domain or task relevant information in the case file which is necessary for the FHE, such as temporal information about when the questioned and/or reference material has supposedly been written. A list of domain-relevant information will help to ensure that the FHE performing the examination will receive the information necessary for the handwriting examination.

A drawback of this might be that a second person needs to be instructed to recognize the relevant information in the case file. Depending on the specificity of the list of domain-relevant information, this person should at least have some knowledge about forensic handwriting examination. When an FHE performs the role of a CIM, this will unable him or her to perform the subsequent examination. Especially in smaller groups of examiners this might be a practical problem, but might be overcome by training colleagues from other disciplines to manage the context information.

4.4. Level 4 Base rate information

The fourth level contains the information coming from the results of previous cases. This can be interpreted as an expectancy of the outcome of the new case based on the results of former, comparable cases. When for example 99 out of 100 former cases were concluded with identification, the FHE might expect this new case to result in identification as well. This source of bias might be hard to minimize in practice. Preferably the proportion identification exclusion is 50 - 50. When the identification : exclusion ratio diverges from the preferred ratio, an option to minimize base rate bias might be by introducing false cases. A similar approach of introducing false cases has been implemented in the field of forensic firearms examination at the NFI (Kerkhoff et al., 2015). Adding only a few false cases to the case flow and making the FHEs aware that this is being done, might already have a profound effect on the mind-set of the FHEs. Through the false cases it is also possible to provide feedback on errors in actual casework, which is normally not possible. In this way false cases will not only help to minimize bias by changing the base rate but will also provide a possibility for quality management in a casework environment, in comparison to the existing proficiency tests.

4.5 Peer review

During peer review the second FHE could be biased when this FHE would receive the interpretation and conclusions of the first FHE. When peer review is carried out independently (blind) of the initial examination, the results from the first FHE cannot influence those of the second. An FHE who does not know what conclusion the first FHE has given cannot be influenced by it.

4.6 Evidence lineup

The evidence lineup fairly regularly shows up in the literature as an approach to minimize the biasing effect of irrelevant context information. The evidence lineup is a procedure whereby the disputed material is compared with a lineup of reference material. The FHE knows what constitutes as the questioned material, and must compare this with several pieces of reference material. Of this reference material, one or more will originate from the suspect, other writings will come from other sources that resemble the questioned material to a certain extent. These added writings are also called *fillers* or *foil specimens*.

Superficially, the evidence lineup seems to provide a means to minimize bias, but a critical assessment might prove otherwise. Much of the literature on the use of evidence lineups can be traced back to Saks et al. (2003). After Saks et al. (2003), the importance of the evidence lineup was stressed in a vacuum without giving a detailed description of how to set up and

perform the lineup, and how the results should be interpreted.

Although it seems like a watertight procedure, an important shortcoming is that it is hard to establish appropriate criteria for the choice of fillers. It is precisely the choice of the fillers which largely determines the outcome. If one uses a random selection of handwriting the task is likely to be too easy for the FHE, and the lineup will add nothing to minimize bias. When simulated handwritings are used as fillers, the task will become more complex, but what might actually be tested is how well the person did in simulating the handwriting (i.e., simulating quality). An interesting option may be to let a second FHE select a sample of handwriting that resembles the questioned handwriting from a large database. But then this may only prove how well the first FHE performs at selecting similar handwritings. It is, therefore, not clear what the added value of the evidence lineup is, as it offers no guarantee for a better assessment of the probative value of the evidence. The evidence lineup is an attempt to solve several problems at once, but fails because there are too many variables varying simultaneously. Testing the evidence and the examiner could better be done separately.

5. Conclusion and discussion

In this paper we have described confirmation bias with two related subcategories; selection bias and contextual bias. Focusing on sources of bias, we have discussed various sources following the taxonomy proposed by Stoel et al. (2014). Some countermeasures to minimize bias in casework have been proposed for each of the levels, such as using an overlay (Level 1), introducing sequential unmasking (Level 2), implementing context information management (Level 3) and adding false cases to the normal case flow (Level 4).

Implementing any of these countermeasures will result in additional time that has to be invested in casework. To justify this additional time effort it is important to acknowledge that task-irrelevant contextual information can have a profound influence on casework. Our own study and several others have shown that studying the influence of bias on casework is extremely difficult and the results might raise more questions than provide answers. Instead of focusing on finding proof for bias in forensic casework through scientific studies, bias should be acknowledged as a human factor that requires effective countermeasures.

Implementing context information management has been shown to be feasible in practice, Found & Ganas, (2014). The handwriting department at the Netherlands Forensic Institute also introduced such a procedure. The main reason for this has been the pursuit of impartiality. When FHEs are unaware of the circumstances of the case, these circumstances cannot influence the examination. Especially for a difficult and relatively subjective discipline such as forensic handwriting examination, with the FHE as the main instrument of analysis, it is extremely important to ensure that the examination is as unbiased as possible. This is also being recognized in court hearings where FHEs are increasingly being questioned concerning the subjective nature of their examinations. When countermeasures have been implemented the FHE can easily show that he/she could not have been influenced by the task-irrelevant contextual information, because the FHE was not aware of this information during examination. Even though the debate about the existence and impact of bias is still on-going within the handwriting community, implementing countermeasures will demonstrate, at the very least, to the trier-of-fact that the issue of bias is taken seriously.

6. References

Böttcher, C.J.F. (1954). Theory and practice of forensic handwriting comparison. *Tijdschrift voor Strafrecht, 63*, 77-131.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), L. Erlbaum Associates.

Dror, I.E., Charlton, D. & Péron, A.E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International, 156*(1), 74-78.

Dror, I.E. (2013). Practical solutions to cognitive and human factor challenges in forensic science. *Forensic Science Policy & Management, 4*, 1-9.

Dror, I.E., Thompson, W.C. Meissner, C.A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Linear sequential unmasking context management toolbox: A Linear Sequential Unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Science, 60* (4), 1111-1112.

Dror, I.E. (2015). Cognitive neuroscience in forensic science: Understanding and

utilizing the human element. *Philosophical Transactions of the Royal Society B, 370*, 1674.

Found, B. & Ganas, J. (2013). The management of domain irrelevant context information in forensic science handwriting examination casework. *Science and Justice, 53*(2), 154–158.

Found, B. (2015). Deciphering the human condition: the rise of cognitive forensics. *Australian Journal of Forensic Sciences, 47*(4), 386-401.

Hagan, W.E. (1894). *A treatise on disputed handwriting and the determination of genuine from forged signatures.* New York, Albany, N. Y.: Banks & Brothers.

Kassin, S.M., Dror, I.E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*(1), 42–52.

Kerkhoff, W., Stoel, R.D., Berger, C.E.H., Mattijssen, E.J.A.T., Hermsen, R., Smits, N., & Hardy, H.J.J. (2015). Design and results of an exploratory double blind testing program in firearms examination. *Science and Justice, 55*, 514-519.

Krane, D.E., Ford, S., Gilder, J., Inman, K. Jamieson, A., Koppl., R., Kornfield, I., Risginger, D.M., Rudin, N., Taylor, M.S., & Thompson, W.C. (2008). Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Science, 53*, 1006-1007.

Kukucka, J., & Kassin, S. M. (2012). Do confessions taint juror perceptions of handwriting evidence? Paper presented at the Annual Meeting of the American Psychology-Law Society San Juan, Puerto Rico, March 14–17.

Mattijssen, E.J.A.T., Van den Heuvel, C.E., & Stoel, R.D. (2011). The relation between signature complexity and the perceived quality of signature simulations. Proceedings of the 15th IGS Conference in Cancun, Mexico.

Mattijssen, E.J.A.T., Stoel, R.D. & Kerkhoff, W. (2015). Minimizing Contextual Bias in Forensic Firearms Examinations. In A. Jamieson & A.A. Moenssens (Ed.) *Wiley Encyclopedia of Forensic Science*. Chichester: John Wiley. doi: 10.1002/9780470061589.fsa1117.

Mattijssen, E.J.A.T., Kerkhoff, W., Berger, C.E.H., Dror, I.E. & Stoel, R.D. (in press). Implementing context information management in forensic casework: Minimizing contextual bias in firearms examination. *Science and Justice*.

McAlexander, T.V., Beck, J., & Dick, R. (1991). The Standardization of Handwriting Opinion Terminology. *Journal of Forensic Science, 36*(2), 311–319.

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

The National Research Counsil (U.S.), Committee on identifying the needs of the forensic sciences in the United States (2009). *Strengthening forensic science in the United States: A path forward.* Washington DC: The National Academies Press.

Risinger, Saks, Thompson & Rosenthal (2002). The Daubert/kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion. *California Law Review, 90*, 1-56.

Saks, M.J., Risinger, D.M., Rosenthal, R., & Thompson, W.C. (2003). 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, 2*, 77-90.

Stoel, R.D., Berger, C.E.H., Kerkhoff, W., Mattijssen, E.J.A.T., & Dror, I.E. (2014). Minimizing contextual bias in forensic casework. In: Strom, K.J. & Hickman, M.J. (Eds). *Forensic science and the administration of justice*. Sage Publications Inc, 67-86.

Stoel, R.D., Dror, I.E., & Miller, L.S. (2014). Bias among forensic document examiners: Still a need for procedural changes. *Australian Journal of Forensic Sciences, 46*(1).

Sulner, A. (2014). Handwriting: Cognitive bias. *Wiley Encyclopedia of Forensic Science.* Chichester, UK: John Wiley & Sons Ltd.

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