Science and Justice 56 (2016) 123–128

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Bloodstain pattern classification: Accuracy, effect of contextual information and the role of analyst characteristics

Nikola K.P. Osborne [a,b], Michael C. Taylor [b], Matthew Healey [c], Rachel Zajac [a,*]

[a] Department of Psychology, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[b] The Institute of Environmental Science and Research (ESR), Christchurch Science Centre, PO Box 29181, Christchurch 8540, New Zealand
[c] Department of Women's and Children's Health, University of Otago, PO Box 56, Dunedin 9054, New Zealand

ARTICLE INFO

Article history:
Received 20 September 2015
Received in revised form 23 December 2015
Accepted 28 December 2015

Keywords:
Bloodstain pattern analysis
Contextual bias
Individual characteristics
Cognitive science
Bayesian analysis

ABSTRACT

It is becoming increasingly apparent that contextual information can exert a considerable influence on decisions about forensic evidence. Here, we explored accuracy and contextual influence in bloodstain pattern classification, and how these variables might relate to analyst characteristics. Thirty-nine bloodstain pattern analysts with varying degrees of experience each completed measures of compliance, decision-making style, and need for closure. Analysts then examined a bloodstain pattern without any additional contextual information, and allocated votes to listed pattern types according to favoured and less favoured classifications. Next, if they believed it would assist with their classification, analysts could request items of contextual information – from commonly encountered sources of information in bloodstain pattern analysis – and update their vote allocation. We calculated a shift score for each item of contextual information based on vote reallocation. Almost all forms of contextual information influenced decision-making, with medical findings leading to the highest shift scores. Although there was a small positive association between shift scores and the degree to which analysts displayed an intuitive decision-making style, shift scores did not vary meaningfully as a function of experience or the other characteristics measured. Almost all of the erroneous classifications were made by novice analysts.

1. Introduction

Forensic science is currently at a crossroads. The reliability of conclusions about forensic evidence, and the methods used to reach those conclusions, are firmly under the microscope. This juncture is largely due to the much talked about National Research Council (NRC) report [1] into the state of forensic science. This report highlighted the need for known error rates in forensic science, and recommended that forensic disciplines acknowledge the role of the examiner in the interpretation process. In particular, the report recognised the need for research into the effect of contextual information on the interpretation process [1]. Context effects in forensic science are commonly referred to as contextual bias—a term that typically describes the unconscious influence of irrelevant information on judgements.

Prior to the publication of the NRC report [1], research into the performance of forensic experts was sparse and had primarily focussed on fingerprint evidence [2,3]. This research highlighted a high degree of subjectivity in fingerprint interpretation, showing that fingerprint decisions are vulnerable to bias. Now, many other forensic disciplines have faced the same scrutiny, with investigations into contextual bias being carried out in forensic odontology [4,5], handwriting examination [6], forensic anthropology [7], shoeprint examination [8], bullet comparison [9], DNA interpretation [10], and bloodstain pattern analysis [11,12]. The general consensus from this research is that forensic interpretations are vulnerable to contextual bias—a finding that is not surprising to psychological scientists, who have long investigated these basic human decision-making processes [e.g.,13,14].

In response to this growing body of literature, forensic laboratories around the world are developing ways to minimise the potential for contextual bias [6,15,16]. In Australia, for example, the Victoria Police Forensic Services Department has introduced a system of contextual information management for handwriting examinations [6]. Here, a designated context manager removes all irrelevant contextual details before passing the document on to be examined. Consequently, there is minimal chance for irrelevant contextual details to cloud judgement. Implementing this type of bias-minimising procedure would also be relatively uncomplicated for many other forensic disciplines, such as fingerprint interpretation, shoeprint examination, and DNA interpretation. The interpretation of such evidence requires minimal to no additional contextual information, and most or all contextual details can be removed.

Not all forensic disciplines, however, are presented with such a straightforward solution to eliminating the negative effects of context.

* Corresponding author.
E-mail addresses: nosborne@psy.otago.ac.nz (N.K.P. Osborne), michael.taylor@esr.cri.nz (M.C. Taylor), matthew.healey@otago.ac.nz (M. Healey), rachelz@psy.otago.ac.nz (R. Zajac).

http://dx.doi.org/10.1016/j.scijus.2015.12.005
1355-0306/

In bloodstain pattern analysis (BPA) for example, such an approach is more complex because much of the contextual information encountered seems both unavoidable and necessary for a complete analysis. BPA is largely a pattern recognition task in which interpretation of the size, shape, and distribution of bloodstains can provide valuable information in a crime scene investigation [17]. For example, features in a bloodstain pattern can indicate the mechanism for deposition, such as a blunt force impact or expiration from the mouth. This analysis can help to piece together the events of the crime and, in some cases, might help to distinguish between, for example, homicide and suicide, or self-defence and murder. In addition to classifying bloodstain patterns, analysts are sometimes required to reconstruct the crime events, in which case elements of the entire scene might contribute to their conclusions. So although context elimination might be recommended in other forensic disciplines, such an approach may not be possible in BPA.

Despite the complexities of managing contextual information in BPA, moves towards implementation seem prudent given the potential for bias to occur. In a recent study, experienced bloodstain pattern analysts made judgements about the classification of bloodstain patterns on ridged non-absorbent [11] and fabric surfaces [12]. Case scenarios presented alongside each bloodstain pattern were formulated to suggest how the bloodstaining occurred. This information either suggested the correct pattern type (positive biasing information), suggested the incorrect pattern type (negative biasing information), or was neutral. Relative to the neutral context, analysts were more often correct with the positive biasing information, and more often incorrect with the negative biasing information—findings consistent with confirmation bias [13].

To know more about the potential for bias in BPA, it is crucial that we understand how analysts use contextual information, and the degree to which this information influences their decision-making. Although Taylor et al.'s studies [11,12] are an important first step towards understanding context effects and reliability in BPA, there are still several unanswered questions. First, because each analyst was presented with different bloodstain pattern targets, we do not know if different analysts will reach the same conclusion when presented with the same pattern. Second, although the case scenarios in Taylor et al.'s studies contained information from various sources (e.g., medical findings, witness statements, police investigator's theory), it is not clear which of these sources exerted the greatest influence on analysts' decisions—data that would be crucial to the development of contextual information management systems in BPA. Finally, we do not know whether some analysts are more vulnerable to context effects than others, and whether or not training and experience in BPA plays a role in the degree to which context effects might emerge.

Dror [18,19] proposes that a cognitive profile representing cognitive abilities that underpin specific forensic tasks (e.g., visual attention, perceiving and comparing visual features) could help to identify forensic examiners who are the best suited to particular jobs. It is conceivable that such profiles could also consider a person's vulnerability to context effects, thereby enhancing endeavours to reduce the risk of contextual bias in forensic science. In the present study, we chose to assess three variables that could be associated with context effects in forensic decision-making: 1) need for closure (NFC; i.e., the extent to which people will be driven to reach any conclusion to avoid confusion and ambiguity [20]); 2) general decision-making style (GDMS; [21]); and 3) compliance (i.e., the extent to which people obey or conform with instructions when they would rather not [22]). These three variables were chosen based on their relevance to forensic decision-making and the availability of validated scales for their measurement.

2. Method

2.1. Participants

Study participants were attendees at a workshop titled "How do we reach conclusions about pattern classification in BPA?" at the International Association of Bloodstain Pattern Analysts (IABPA) Training Conference in San Diego, 2013. Participation in the workshop was voluntary and free of charge. The participants comprised analysts from Australasia (New Zealand and Australia), North America, Asia, and Europe. The participants were informed that, as part of the workshop, the researchers would be collecting data that may be published in a peer-reviewed journal. All 44 workshop attendees completed the experimental procedure. For data analysis, however, we excluded those participants who had no formal training in BPA ($n = 5$), giving us a final sample of 39 bloodstain pattern analysts. We considered the analysts who had advanced BPA training and experience presenting BPA testimony in court as experts ($n = 23$). The remaining analysts were considered as novices ($n = 16$).

2.2. Materials

2.2.1. Analyst characteristic measures

2.2.1.1. Need for closure (NFC) scale. We used a brief, 15-item version of the NFC scale [20], created and validated by Roets and Van Hiel [23]. This scale measures the respondents' NFC as it relates to five subscales: order, predictability, decisiveness, ambiguity, and closed-mindedness. The respondents are presented with statements such as "I don't like situations that are uncertain" and "I enjoy having a clear and structured mode of life," and are required to indicate their agreement on a 6-point Likert Scale (1 = completely disagree, 6 = completely agree). Roets and Van Hiel [23] recommend that researchers use the abridged scale to calculate a total NFC score, rather than separate subscale scores.

2.2.1.2. General decision-making style (GDMS). We used a 25-item scale developed and validated by Scott and Bruce [21] to measure decision-making style as it relates to five constructs: rational, avoidant, intuitive, dependent, and spontaneous. The respondents are required to rate statements such as "I double-check my information sources to be sure I have the right facts before making decisions" and "I generally make decisions that feel right to me" on a 5-point Likert Scale (1 = strongly disagree, 5 = strongly agree).

2.2.1.3. Compliance scale. To measure compliance, we used a 20-item questionnaire, developed and validated by Gudjonsson [22]. The respondents answer "true" or "false" to statements such as "I give in easily when I am pressured" and "I try hard to do what is expected of me." The scale also consists of reverse-score statements such as "I am not too concerned about what people think of me." The total number of "true" responses to non-reverse-scored statements and the number of "false" responses to the reverse-scored statements are combined to give a total score.

2.2.2. Pattern classification task

2.2.2.1. Bloodstain pattern target. We used a bloodstain pattern presented in a colour photograph (22.5 cm × 25 cm) for the classification task. The photograph (Fig. 1) was obtained courtesy of Taylor et al. [11]. The pattern was cast-off spatter, created in the laboratory by swinging a blood-soaked wrench. The analysts were informed that the photograph was of a bloodstain found on a vertical section of wall, where the bottom of the wall was at floor level, indicating that the bloodstain was approximately 30 to 40 cm from the ground. A scale was provided within the photograph.

2.2.2.2. Contextual information. We compiled items of contextual information, said to relate to the bloodstain pattern target, from six commonly encountered information sources in BPA (see Table 1). Because the bloodstain pattern was created in the laboratory, all of the information was fictitious.

2.2.2.3. Response format. The analysts classified the pattern by allocating 10 points to the pattern type(s) which supported their opinion, giving

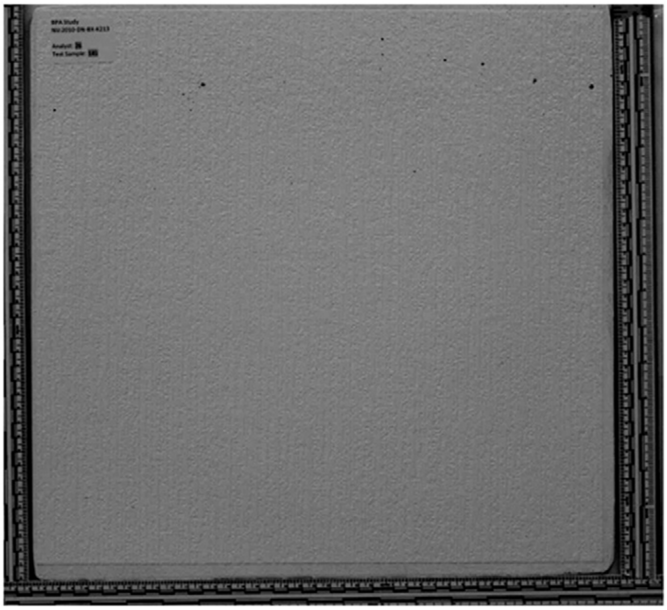

Fig. 1. The target pattern for the classification task (analysts saw this target pattern in colour).

more or fewer votes to reflect favoured and less favoured opinions. The pattern types that could be voted for were: cast-off, drip stain/trail, expiration pattern, projected (e.g., arterial), gunshot spatter (forward or backward), swipe, blunt force impact, saturation, wipe, transfer pattern. The analysts were given the following instructions:

You have 10 votes to use at each step. Allocate your votes according to your opinion. Give more votes to your favoured option(s), and fewer votes to your less favoured option(s). Assign zero votes to options that you have ruled out. At the end of each step, your votes should combine to total 10.

Table 1
Contextual information for pattern classification task.

Context item	Details
Police briefing	The deceased is Mr. Robert Harold Sing. A work colleague, Mr. Simon Peters, has gone to Mr. Sing's home after he failed to show for his 7 am work shift, and has found his body in the back yard. Mr. Peters called police, who launched a homicide investigation. A scene examination revealed bloodstaining on the walls and floor in the hallway leading to Mr. Sing's bedroom, and also on the west facing bedroom wall.
Medical findings	An autopsy shows that Mr. Sing died of a combination of blunt force trauma to the head and subdural haemorrhaging. Injuries to the eyes and nose were consistent with an object impacting Mr. Sing's face with considerable force. It is likely that he was alive for several hours after the injuries were sustained.
Other bloodstain patterns	An analysis of other bloodstains in the hallway and bedroom show patterns consistent with drip stains, leading from the bedroom to the hallway and outdoor area where the body was discovered.
Other forensic evidence	A bloodstained fingerprint pattern was analysed and found to match to Mr. Sing's brother in-law Mr. Jacob Walters. Several shoeprints found in the hallway of Mr. Sing's home, and on the path leading to the back yard matched to that of a woman Mr. Sing had been in a relationship with, Ms. Jessica Patel. A partial bloodied palm print was found on the west facing bedroom wall above the blood panel that is under analysis; however, police were not able to match it to any known person. A search of the homes of Mr. Jacob and Ms. Patel did not uncover any further evidence.
DNA statement	The blood on the wall panel being analysed matches to that of Mr. Sing [deceased].
Witness statement	Neighbours heard an argument between a man and a woman around 11 pm on the night before Mr. Sing was found dead. About 2 am that night they awoke to loud bangs and another intense argument between two men before a car was heard speeding away. The neighbours did not recognise the voices.

2.3. Procedure

The analysts completed the analyst characteristic measures before the pattern classification task.

The analysts were first presented with the target pattern in the absence of any contextual information, and asked to allocate their votes according to the instructions outlined above. Then, if they believed it would assist with their analysis, analysts could request one item of contextual information at a time from the six listed items. The order in which context items were listed was counterbalanced across participants.

After receiving each requested context item, the analysts were able to revise their allocation of votes on a new response sheet. This process was repeated until either: 1) an analyst was satisfied that no further contextual information would assist with his or her analysis, or 2) an analyst had requested all six items of contextual information.

In addition to their final vote allocation, analysts were asked to write down their final pattern classification—the pattern type or types that that they would include in their BPA report. We used this written classification to determine analysts' accuracy.

2.4. Data analysis

Analyses were conducted under an empirical Bayesian framework [24–28], primarily because this approach provides for robust handling of small sample sizes. All Bayesian tests performed are alternatives to established frequentist tests. The main difference in interpretation is the replacement of classical *p*-values and corresponding confidence intervals with 95% highest density intervals (HDI).[1] All computations were carried out in the statistical software package R [29], using various functions within the Bayesian First Aid package [30].

3. Results

3.1. Classification performance

We determined classification performance according to three main factors: overall accuracy, requests for contextual information, and opinion update (i.e., vote shift).

3.1.1. Overall accuracy

In addition to making their final vote allocation, recall that the analysts were asked to write down their final pattern classification. We categorised the accuracy of this classification according to whether it was definitively accurate, conservatively accurate, incorrect, or inconclusive. A definitively accurate response was assigned if analysts recorded only one pattern classification and it was the ground-truth pattern type (i.e., “cast-off”). A conservatively accurate response was assigned if analysts included cast-off among multiple classifications, or classified the pattern as “spatter stains”.[2]

If analysts recorded a classification that did not include “cast-off” or “spatter stains” then the response was considered incorrect. If analysts did not respond at all, or they indicated that there was not enough detail in the pattern to make a decision, their response was scored as inconclusive. The distribution of accuracy scores, and resulting proportion analyses according to experience level, are displayed in Table 2. Cohen's *d*

[1] In the current context, HDIs are presented as the probability that the estimated parameter is zero (e.g., percent chance that the mean difference = 0).

[2] *Spatter* describes bloodstains that occur as a result of force(s) additional to gravity; this could include blood being released from a moving object, as in a *cast-off* pattern (SWGSTAIN, 2009).

indicates the effect size (small, medium or large) of the difference between the two means being observed [31].

When examining the proportion of analysts in each accuracy category according to experience level, expert analysts represented the greatest proportion of analysts who were definitively (MD $= -0.13$, $d = 0.36$, 19% MD $= 0$) and conservatively accurate (MD $= -0.26$, $d = 0.76$, 3.5% MD $= 0$). Novice analysts represented the greatest proportion of analysts who were incorrect (MD $= 0.36$, $d = 1.57$, 0% MD $= 0$) or inconclusive (MD $= 0.05$, $d = 0.22$, 11% MD $= 0$).

3.1.2. Requests for contextual information

We were interested in four factors associated with the analysts' requests for contextual information: 1) the number of analysts who requested contextual information; 2) the average number of context items requested; 3) which items of contextual information were requested; and 4) the order in which the information was requested.

All analysts requested at least one item of contextual information ($M = 4.85$, SD $= 1.32$), and 19 (49%) requested all six. There was no meaningful difference in the mean number of items requested between novice ($M = 4.87$, SD $= 1.34$) and expert ($M = 4.63$, SD $= 1.38$) analysts. To determine the chronology of information acquisition, we calculated a request-order score for each of the six context items. For each analyst, the first item requested was given a score of six, the next item requested was given a score of five, and so on. A score of zero was given to items not requested. The mean request-order score for each item and the number of analysts requesting each item are displayed in Table 3.

3.1.3. Opinion update: shift score

Recall that on receiving each new item of contextual information, analysts could reallocate – or shift – their votes to reflect their updated opinion. Our next step was to consider how the contextual information contributed to analysts' propensity to shift their votes. To answer this question, we calculated the number of votes reallocated across all pattern types as a result of each context item (i.e., shift score; Table 3). For example, an analyst who had allocated five votes to “cast-off” and five votes to “impact” and then, after receiving the medical information, allocated seven votes to “cast-off” and three votes to “impact” would receive a shift score of two as a function of the medical information (i.e., two votes from impact were reallocated to cast-off). If the analyst's distribution of votes had remained the same, the shift score for that context item would be zero.

In addition to determining the amount of shift as a function of each item of contextual information, we calculated an overall shift score: the number of votes reallocated, across all pattern types, between the analysts' initial (i.e., without context) allocation and their final allocation. The mean overall shift score was 3.25 (SD $= 2.15$, $d = 1.54$ large, 0.0% MD $= 0$ [30]). The novice analysts had higher shift scores ($M = 3.81$, SD $= 1.76$) than the expert analysts ($M = 2.91$, SD $= 2.37$; 8.7% MD $= 0$). Only four analysts did not shift any votes across the task—all were in the expert group.

Using a Bayesian proportion test [30] we examined the relationship between the number of items requested and overall accuracy. Relative to the other analysts, analysts who were definitively accurate had selected fewer items of contextual information (MD $= -2.18$, SD $= 0.06$, $d = 0.71$ large, 0.0% MD $= 0$). Similarly, relative to the other analysts, analysts who were incorrect had requested more items of contextual information (MD $= 1.03$, SD $= -0.68$, $d = 1.03$ large, 0.0% MD $= 0$). When comparing the analysts' shift scores by overall accuracy, we found that the analysts who classified the pattern incorrectly had higher shift scores than those in the other overall accuracy categories (MD $= 1.53$, SD $= 0.44$, $d = 0.79$ large, 4.9% MD $= 0$). In all of the other relations examined, there was a greater than 10% chance that the mean difference was zero.

3.1.4. Instances of include-after-exclude and exclude-after-include decisions

When an analyst allocated zero votes to a given pattern type, this indicated that the analyst was ruling out that pattern type as a possible mechanism for the deposited stains. We were interested to know how the provision of contextual information was related to analysts' propensity to; a) rule out pattern types that had previously been allocated votes (i.e., exclude-after-include), and b) allocate votes to pattern types that had previously been ruled out (i.e., include-after-exclude; Table 4). Of the 1870 vote allocation decisions made across the study, there were 54 instances where analysts ruled out pattern types that they had previously allocated votes to, with almost half of these changes occurring as a function of the medical findings. There were six instances where analysts included a pattern type that they had previously excluded as a possible mechanism. Again, half of these changes occurred after the analysts requested the medical findings.

3.2. Analyst characteristics as predictors of classification performance

Next, we explored whether any of the analyst characteristic variables (mean scores presented in Appendix 1) were associated with the use of contextual information (i.e., number of items requested and mean shift score). We examined this using Bayesian correlations via the Bayesian First Aid packages in R [30]. There were no meaningful associations between the number of items requested and analysts' characteristic measure scores; where meaningful implies a less than 10% chance that the relationship was zero. Shift scores were positively associated with an intuitive decision-making style ($r = 0.23$, 8.1% $r = 0$). That is, higher intuition scores were associated with higher shift scores. There were no meaningful associations between the shift scores and any of the other individual characteristic measures.

4. Discussion

Bloodstain patterns can be critical evidence in crime scene investigations. As in many other forensic disciplines, however, the bloodstain interpretation process is vulnerable to contextual bias [11,12]. In the present study, we examined the consistency of classifications between analysts, the role of contextual information in BPA, and the role of analyst characteristics in decisions about bloodstain pattern evidence.

Although around 70% of the analysts made correct classifications, 20% made errors and 10% were not prepared to make a classification. When we compare our findings to decisions about cast-off patterns in Taylor et al. [11], the rate of correct classifications is similar (69%), however, the analysts in that study made fewer errors (14% incorrect classifications). We note that in Taylor et al.'s research [11], all analysts had received advanced BPA training and had a minimum of 5 years

Table 2
Accuracy scores and proportion analyses as a function of number and percentage of expert and novice analysts.

Accuracy score	n (%) expert	n (%) novice	n (%) total	Cohen's d	MD (SD)	Prob HDI = 0
Definitive accuracy	9 (39.1%)	4 (25%)	13 (33.3%)	0.36^{S}	−0.13 (0.14)	19%
Conservative accuracy	11 (47.8%)	3 (18.8%)	14 (35.9%)	0.76^{M}	−0.26 (0.14)	3.5%*
Incorrect	1 (4.3%)	7 (43.8%)	8 (20.5%)	1.57^{L}	0.36 (0.12)	0%*
Inconclusive	2 (8.7%)	2 (12.5%)	4 (10.3%)	0.22^{S}	0.05 (0.11)	11%

Cohen's d effect sizes: ^{S}small; ^{M}medium; ^{L}large.
* Less than 5% chance that mean difference = 0.

Table 3
Shift score, request-order score, percentage of analysts requesting each context item, percentage of cases that led to shift, and 95% HDI for likelihood of shift.

Context item	Shift score (SD)	Request-order score (SD)	% analysts requesting context item	% cases context item led to shift	Context more likely to facilitate shift than not (95% HDI)
Medical findings	1.78 (1.49)	4.31 (1.54)	92.3	90.0	Yes
DNA statement	0.86 (1.50)	2.62 (1.87)	74.4	63.0	Yes
Police briefing	0.83 (1.65)	2.38 (1.74)	74.4	60.4	Yes
Other bloodstain patterns	0.79 (1.75)	5.64 (0.67)	97.4	58.8	Yes
Other forensic evidence	0.43 (1.17)	2.64 (1.93)	71.8	52.0	Yes
Witness statement	0.00 (0.00)	0.92 (1.22)	48.7	0	No

casework experience. When considering only the expert analysts in our study, our error rate was much lower (4.3%) than that of Taylor et al.'s [11]: all but one of the incorrect classifications in the current study was made by a novice analyst. The comparatively high rate of errors (43.8%) made by novice analysts emphasises the importance of training and experience to increase reliability. Expertise alone, however, is not enough to prevent errors in pattern classification.

We note that all analysts in our sample had elected to participate in a workshop about decision-making and cognitive factors in BPA, and were therefore likely aware of – and open-minded about – the potential negative effects of context. That said, analysts clearly considered contextual information to be important to their decisions. All of the participants requested contextual information from at least one source; on average, they requested information from five sources. The most frequently requested items of information were the other bloodstain patterns and the medical findings. Interestingly – because it is likely to be the first source of information encountered in practice – the police briefing was not typically requested until several other items had already been viewed.

How did the analysts' opinions about the pattern evolve with each piece of contextual information? Before we consider the answer to this question, it is important to note here that the analysts' task in the current study was pattern classification—the recognition of a pattern according to stain characteristics. Because completing the task did not require any information other than the bloodstain pattern itself, any update in opinion as a function of the contextual information could be thought of as contextual bias. In the context of this study, however, we believe that it is more appropriate to consider opinion update as an effect of context rather than as bias. The distinction here is subtle, but important. Although both terms describe the influence of external factors on the perception of a stimulus, we refer to bias as an unconscious influence, and to the effect of context as a conscious influence. So although bias is strictly an effect of context, an effect of context in BPA does not necessarily constitute bias. In fact, by virtue of allowing analysts to request information and subsequently update their opinion, analysts in this study likely thought that they should be making an explicit attempt to use the contextual information in this way.

Regardless of the manner in which contextual information influences opinion, however, a situation in which one piece of information has influenced another has important implications if both pieces of information are presented to fact-finders as independent evidence. This situation is often referred to as double counting [32]. One way to conceptualise double counting is to consider the following question: if the BPA relies on information from other sources of information about the crime, then how do we weigh up what it is adding to an investigation? Knowledge about the various ways in which contextual items can influence analysts' decisions – and the extent to which this occurs – is therefore crucial.

There were only four analysts in our study (10%) whose conclusions were not influenced by context; that is, these analysts did not reallocate any votes across the entire classification task. Medical information led to the largest shift scores and the greatest number of "exclude-after-include" and "include-after-exclude" decisions. These findings are perhaps not surprising given that: 1) in practice, medical findings are likely to be perceived as a reliable source of information, and 2) there is direct link between a blood-letting injury and the deposition of bloodstains. Indeed, on closer analysis of explicit decisions to exclude or include a given pattern type (i.e., vote shifts to and from zero), the most commonly excluded pattern types in response to the medical information were gunshot spatter, expirated, arterial, and blunt force impact, all of which are pattern types in which the injury mechanism is inherent in the term used to describe the pattern.

Although witness statements did not significantly influence analysts' opinions, it would be premature to conclude that these are not a potential source of bias in practice. Witness statements often form the basis for a hypothesis to be tested. In fact, they may be the only information available to analysts when they perform their initial analysis. However, we know that reports from eyewitnesses are also a particularly fallible form of evidence [32,33]. Because information obtained early in an investigation can influence subsequent interpretations [34], the risks associated with witness statements deserve further exploration.

We recognise that contextual information cannot be easily categorised into independent items. For example, assessing the relevance of medical findings to the deposition of blood is only possible if analysts know that the source of the blood corresponds to the subject of the medical findings. This cross-information dependence could explain why DNA led to the second largest shift scores, when in principle whose blood it is should have no bearing on the classification of a pattern.

Our findings clearly demonstrate that contextual information can influence decision-making, but can the characteristics of an individual analyst predict the degree to which this occurs? Interestingly, the degree of opinion change was only associated with one of our analyst characteristic measures: intuitive decision-making style. Intuition describes a gut instinct or feeling [21]. By nature, intuitive decision-makers tend to rely on subjective cues (e.g., "this decision feels right") rather than objective information [21]. If those analysts with high intuitive decision-making scores were less likely to complete the task in an objective way, then it stands to reason that they might be more likely to consider sources of information outside of the pattern itself.

Even when we controlled for intuitive decision-making scores, however, there was still an overall effect of contextual information on classifications. Overall then, our findings demonstrated that propensity towards an opinion update due to contextual information was not a function of the cognitive factors or decision-making styles under investigation. Consequently, although a "cognitive profile" suggested by Dror [18,19] might be useful to determine the suitability of a candidate for particular forensic tasks, from our findings we cannot determine the

Table 4
Number of exclude-after-include and include-after-exclude decisions.

Context item	Exclude-after-include	Include-after-exclude
Medical findings	24	3
DNA statement	8	1
Police briefing	5	2
Other bloodstain patterns	15	0
Other forensic evidence	2	0
Witness statement	0	0
Total changes (1870)	54 (31 analysts)	6 (5 analysts)

types of people who are more prone or less prone to change their opinion as a result of contextual information. Instead, our findings suggest that context effects should be addressed at a discipline level, rather than targeting analysts who fit a certain profile.

What do our findings mean for the management of contextual information in BPA? To answer this question, we must consider how the use of – and need for – contextual information might vary depending the bloodstain pattern analyst's task. A bloodstain pattern analyst's task can range from pattern classification, to mechanism determination, to crime scene reconstruction. The boundaries between these components, however, are frequently blurred. This blurring is accentuated by the fact that the terms used to classify patterns often describe the mechanistic cause of the pattern, and therefore form part of a reconstruction theory. As noted earlier, our participants were required to engage only in pattern classification, in which all of the information needed to make a decision is inherent in the pattern itself.

In crime scene reconstruction, however, certain contextual information may be highly relevant and necessary for a complete analysis. In this way, linear sequential unmasking – obtaining contextual information in a step-by-step fashion after making an initial context-free interpretation [35,36], much like the method used in this study – holds considerable promise as a way of managing the effects of context in BPA. Specifically, this method allows analysts to transparently incorporate relevant contextual information into their scene reconstruction. Critically, this method also allows analysts to cognitively “retrace their steps” in the event that any item of contextual information is found to be unreliable or erroneous. Given that medical findings resulted in the greatest effect of context in this study, it may be particularly important for analysts to make their initial observations and classifications in the absence of this information where possible. Further research is required to understand when and how contextual information can be safely integrated at each stage of BPA.

The potential for bias or context effects in BPA should not overshadow the value that bloodstain pattern evidence can provide in crime scene investigations. We encourage analysts to thoroughly document both their analysis process and their use of contextual information to inform their conclusions. In doing so, analysts can be aware of – and accountable for – the role of contextual information in their analyses, with a view to increasing the reliability of their classification decisions.

5. Acknowledgements

Funding for this study was provided by the Marsden Fund Council (from Government funding administered by the Royal Society of New Zealand), the University of Otago, and the Institute of Environmental Science and Research Core Funding. Statistical analyses were made possible with support from New Zealand eScience Infrastructure (NeSI). The authors gratefully acknowledge Andrew Mills for his contribution to data input.

Appendix 1. Mean individual characteristic measure scores.

Individual characteristic measures and subscales	*M* score (SD)
Need for closure (maximum 6)	3.83 (0.54)
General decision-making style (maximum 5)	
Rational	4.14 (0.47)
Dependent	3.46 (0.61)
Intuitive	3.24 (0.67)
Avoidant	2.34 (0.77)
Spontaneous	2.30 (0.62)
Compliance (maximum 15)	9.28 (3.32)

References

[1] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academy of Sciences, Washington, DC, 2009.
[2] I.E. Dror, D. Charlton, A.E. Péron, Contextual information renders experts vulnerable to making erroneous identifications, Forensic Sci. Int. 156 (2006) 74–78.
[3] I.E. Dror, D. Charlton, Why experts make errors, Journal of Forensic Identification 56 (2006) 600–616.
[4] N.K.P. Osborne, S. Woods, J. Kieser, R. Zajac, Does contextual information bias bitemark comparisons? Sci. Justice 54 (2014) 267–273.
[5] M. Page, J. Taylor, M. Blenkin, Context effects and observer bias—implications for forensic odontology, J. Forensic Sci. 57 (2012) 108–112.
[6] B. Found, J. Ganas, The management of domain irrelevant context information in forensic handwriting examination casework, Sci. Justice 53 (2013) 154–158.
[7] S. Nakhaeizadeh, I.E. Dror, R.M. Morgan, Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias, Sci. Justice 54 (2014) 208–214.
[8] J.H. Kerstholt, R. Paashuis, M. Sjerps, Shoe print examinations: effects of expectation, complexity and experience, Forensic Sci. Int. 165 (2007) 30–34.
[9] J.H. Kerstholt, A. Eikelboom, T. Dijkman, R.D. Stoel, R. Hermsen, B. van Leuven, Does suggestive information cause a confirmation bias in bullet comparisons? Forensic Sci. Int. 198 (2010) 138–142.
[10] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, Sci. Justice 51 (2011) 204–208.
[11] M.C. Taylor, T.L. Laber, P.E. Kish, G. Owens, N.K.P. Osborne, The Reliability of Pattern Classification in Bloodstain Pattern Analysis, Part 1: Bloodstain Patterns on Rigid Non Absorbent Surfaces, (in press).
[12] M.C. Taylor, T.L. Laber, P.E. Kish, G. Owens, N.K.P. Osborne, The Reliability of Pattern Classification in Bloodstain Pattern Analysis, Part 2: Bloodstain Patterns on Fabric Surfaces, (Under revision).
[13] R.S. Nickerson, Confirmation bias: a ubiquitous phenomenon in many guises, Rev. Gen. Psychol. 2 (1998) 175.
[14] J.S.B.T. Evans, Bias in human reasoning: causes and consequences, Essays in Cognitive Psychology, Lawrence Erlbaum Associates, Hillside, NJ, England, 1989.
[15] E.J. Mattijssen, R.D. Stoel, W. Kerkhoff, Minimizing contextual bias in forensic firearms examinations, Wiley Encyclopedia of Forensic Science, (2015).
[16] R.D. Stoel, C.E. Berger, W. Kerkhoff, E.A. Mattijssen, I.E. Dror, M. Hickman, K. Strom, Minimizing contextual bias in forensic casework, in: K.J. Strom, M.J. Hickman (Eds.), Forensic Science and the Administration of Justice: Critical Issues and Directions, SAGE Publishing, California, USA 2014, pp. 67–96.
[17] T. Bevel, R.M. Gardner, Bloodstain Pattern Analysis: With an Introduction to Crime Scene Reconstruction, third ed. CRC Press, Boca Raton, 2008.
[18] I.E. Dror, Practical solutions to cognitive and human factors challenges in forensic science, Forensic Sci. Policy Manage. 4 (2013) 1–9.
[19] I.E. Dror, Cognitive neuroscience in forensic science: understanding and utilizing the human element, Philos. Trans. R. Soc. B 370 (2015) 1–8.
[20] D.M. Webster, A.W. Kruglanski, Individual differences in need for cognitive closure, J. Pers. Soc. Psychol. 67 (1994) 1049–1062.
[21] S.G. Scott, R.A. Bruce, Decision-making style: the development and assessment of a new measure, Educ. Psychol. Meas. 55 (1995) 818–831.
[22] G.H. Gudjonsson, Compliance in an interrogative situation: a new scale, Personal. Individ. Differ. 10 (1989) 535–540.
[23] A. Roets, A. Van Hiel, Item selection and validation of a brief, 15-item version of the need for closure scale, Personal. Individ. Differ. 50 (2011) 90–94.
[24] J.K. Kruschke, Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Academic Press, 2014.
[25] J.K. Kruschke, Bayesian estimation supersedes the t test, J. Exp. Psychol. Gen. 142 (2013) 573.
[26] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, third ed. CRC Press, Boca Raton, FL, 2013.
[27] M.D. Lee, E.-J. Wagenmakers, Bayesian Cognitive Modeling: A Practical Course, Cambridge University Press, 2014.
[28] R. Wetzels, E.-J. Wagenmakers, A default Bayesian hypothesis test for correlations and partial correlations, Psychon. Bull. Rev. 19 (2012) 1057–1064.
[29] R Core Team, R: a language and environment for statistical computing. Vienna, Austria; 2014, URL http://www.R-project.org , (2015).
[30] R. Bååth, Bayesian first aid: a package that implements Bayesian alternatives to the classical *.test functions in R, Proceedings of UseR! 2014-the International R User Conference, 2014.
[31] J. Cohen, A power primer, Psychol. Bull. 112 (1992) 155.
[32] D. Davis, E.F. Loftus, Internal and external sources of misinformation in adult witness memory, in: M.P. Toglia, J.D. Read, D.F. Ross, R.C.L. Lindsay (Eds.),Handbook of Eyewitness Psychology (Vol 1). Memory for Events, NJ: Erlbaum, Mahwah 2007, pp. 195–237.
[33] M.L. Howe, L.M. Knott, The fallibility of memory in judicial processes: lessons from the past and their modern consequences, Memory 23 (2015) 633–656.
[34] J.H. Kerstholt, A.R. Eikelboom, Effects of prior interpretation on situation assessment in crime analysis, J. Behav. Decis. Mak. 20 (2007) 455–465.
[35] D.E. Krane, S. Ford, J.R. Gilder, K. Inman, A. Jamieson, R. Koppl, I.L. Kornfield, M. 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.
[36] I.E. Dror, W.C. Thompson, C.A. Meissner, I.L. Kornfield, D.E. Krane, M. Saks, M. Risinger, Letter to the editor — context management toolbox: a linear sequential unmasking (LSU) approach for minimising cognitive bias in forensic decision-making, J. Forensic Sci. 60 (2015) 1–2.