233 East Main Street
Suite C
Bridgeport, WV 26330
(304) 933-3162

Final Technical Report

DoJ/NIJ grant #2010-DN-BX-K270

Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments

Peter Fraser-Mackenzie[1,2], Itiel Dror[1,3], Kasey Wertheim[4]

[1] Cognitive Consultants International (CCI), London, U.K.
[2] School of Management, University of Southampton, Southampton, U.K.
[3] Institute of Cognitive Neuroscience, University College London (UCL), London, U.K.
[4] Complete Consultants Worldwide (CCW), Bridgeport, WV.

Corresponding Author:
Dr Peter A.F. Fraser-Mackenzie
p.fraser-mackenzie@soton.ac.uk
+44 (0) 23 8059 7430

*Supported by a research grant from the National Institute of Justice (NIJ) on "Forensic Decision Making on Suitability for Identification Judgments" (Grant Number 2010-DN-BX-K270). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agency.

Table of Contents

Abstract

We examined forensic fingerprint examiners' suitability determinations of latent fingerprints comparing situations in which the latent is assessed solo (in isolation) versus situations in which it is presented alongside a comparison (matching or non-matching) exemplar print. The presence of a non-matching comparison exemplar led examiners to be more inclined to draw the conclusion that the latent was suitable for comparison compared to when the latent was presented solo. This effect persisted even when the latent presented was highly unsuitable for comparison. The presence of a matching comparison exemplar led examiners to be less likely to decide that the latent was suitable and more likely to decide the latent was questionable compared to solo analysis. This effect persisted even when the latent presented was highly suitable, suggesting a strong main effect. Knowledge of another examiner's previous determination that the latent was unsuitable was found to increase the likelihood that the examiner would conclude that the latent was unsuitable. However, knowledge of a previous "suitable" determination by another examiner did not increase the likelihood of a "suitable" conclusion by examiners. The finding that effects were weaker, although not entirely removed, in those with IAI certification suggests that training may be an appropriate route for reducing the effect of contextual influence and bias in suitability determinations. It was also shown that a latent prints that were previous classed as "unsuitable" in a non-biasing context tended to still be judged to be "unsuitable" by examiners that were presented with the latent in a strongly biasing context (a major case in which a previous examiner was purported to have made an Individualization).

Keywords

Forensic Decision-Making; Contextual Effects; Fingerprints; Suitability; IAI Certification.

1. Introduction

In the initial "Analysis" stage of ACE-V (Analysis, Comparison, Evaluation & Verification), a latent is assessed in isolation for its suitability for the comparison process. Suitable prints are then compared to potential matching exemplars from ten prints in order to attempt to identify the source of the latent print. The reliability of this suitability assessment plays an important part in the ACE-V process as a whole. If the isolated judgment of suitability is not a reliable indication of the actual identification potential, the identification may be delayed, precluded or errors may occur. This study explores the reliability and biasability of suitability judgments using a controlled, covert, study of forensic latent fingerprint examiners.

In this study, we investigate the question of whether suitability judgments may differ in the presence of a comparison exemplar print due its impact as contextual information on the cognitive processing of visual information and judgment. A long history of research into human cognition reveals that visual judgment processing relies extensively on both goal-directed attention and automatic attention direction from contextual information [1-9]. Thus, in addition to attention being directed by the goals of the searcher, attention can also be directed automatically as a result of contextual cues. Cue priming can increase the saliency of search targets [5,10,11] or can guide attention towards certain information and away from others [12]. For example, having been exposed to a particular feature previously (e.g. features such as color or spatial frequency), the subsequent search for similar features is generally found to be more efficient [13]. As a result, the true extent to which we are in command of where the "mind's eye" is directed is a major question in cognitive psychology [9]. On this basis, we expect the presence of a comparison exemplar to have a considerable effect on the visual attention and search behavior of examiners and that this has the potential to affect the reliability and biasability of suitability conclusions.

For these reasons, Dror [14] has recommended that the ACE be conducted (and documented) linearly (i.e. sequentially), and with each phase independent from each other. Although such practices are not common, they have been implemented by the US Federal Bureau of Investigation (FBI), the

Netherlands Forensic Institute (NFI) and the Dutch National Police. For example, the revised Standard Operating Procedures (SOPs) of the FBI "include some steps to avoid bias: examiners must complete and document analysis of the latent fingerprint before looking at any known fingerprint" and "instructs examiners conducting analysis of a latent fingerprint to analyze it for evidence of distortion, determine whether it is 'of value,' and document the data used during analysis" (p.27) [15].

However, initial latent analysis in isolation may lack the benefit of direction guided by the comparison exemplar. Therefore Dror [14] suggests that examiners may be allowed to return and revisit the analysis stage, but they must document and justify it. Indeed, the Office of the Inspector General (OIG) [15] clearly takes this cognitively informed approach on board, citing this approach in its report: "a solution to bias may be requiring initial analysis of the latent fingerprint in isolation from the known fingerprints, but also permitting, with clear and detailed documentation, some 're-analysis' of the latent print after comparison" (p. 28). A recent Expert Group set up by the National Institute of Standards and Technology (NIST) [16] has reached similar conclusions and has recommended that: "Modifications to the results of any stage of latent print analysis (e.g., feature selection, utility assessment, discrepancy interpretation) after seeing a known exemplar should be viewed with caution. Such modifications should be specifically documented as having occurred after comparison had begun" (Recommendation 3.2, see NIST [16]).

Furthermore, Dror recommends that examiners be restricted to the extent that such re-analysis be allowed, e.g., that "clear" features during analysis not be changed, but "ambiguous" ones can benefit from hindsight cognitive attention (for details, see Dror, 2009 [14]). A similar approach has been adopted by Langenburg and Champod [17].

The reliability of the Analysis stage of ACE-V has not received as much attention from research as the assessment of the reliability of subsequent stages, such as evaluation (e.g., Ulery, Hicklin, Buscaglia and Roberts [18]) and there is a lack of data and literature in this area. One exception was a recent study investigating feature selection in which the examiner searches for valid, usable, minutiae. The more valid minutiae there are, the greater the information available to undertake the comparison

and the latent has more evidential strength. The study by Dror, Champod, Langenburg, Charlton, Hunt and Rosenthal [19] found that the number of minutiae observed by examiners can vary between examiners (inter-examiner inconsistency) as well as within examiners assessing the same latent at different times (intra-examiner inconsistency). This variation in the number of minutiae indicates some initial evidence that suitability judgments may vary between individuals.

While the cognitive literature indicates that suitability judgments could differ in the presence of a comparison exemplar due to the impact of contextual information on the cognitive processing of visual information, very little work has investigated this important topic. One exception was Dror et al [19] who investigated the presence of a matching comparison print on the number of minutiae observed by examiners. The presence of a matching comparison print was found to affect this feature selection by reducing the number of minutiae observed by examiners [19]. It was suggested that the matching comparison print provided contextual information that guided the visual search attention of the examiners in a way that limited the scope of their search for minutiae, or changed their thresholds. The presence of a non-matching comparison print on feature selection was not tested in their study. Nevertheless, the finding that the presence of a matching comparison print can result in a different number of minutiae observed compared to solo analysis, indicates that suitability judgments may be susceptible to contextual bias. A previous study on contextual bias found that the Analysis stage was relatively robust to contextual bias, but they used a very different sort of context [20]. Rather than manipulating the existence and type of the exemplar comparison print, they manipulated contextual information (for details, see Schiffer and Champod [20]).

For the first time, we examine whether the suitability analysis conclusion is affected by contextual factors. Specifically, our main aim is to determine whether the presence of a matching or non-matching comparison exemplar results in different suitability conclusions compared to suitability conclusions made when the suitability determination is undertaken in isolation. Other cognitive and contextual influences may further affect suitability determination, but were not the object of this current study.

1.1 Latent Clarity and Biasability

Previous studies have demonstrated that latent-to-exemplar comparison conclusions (i.e. "Individualization", "Exclusion", "Inconclusive") can be affected by biasing contextual factors [21-23]. These effects appear to be strongest when the comparison is more difficult to judge, such as when the latent is of poorer quality due to noise or distortion [21,24]. Therefore, it is possible that suitability conclusions are reliable in clear cut cases, regardless of whether the comparison print is present versus when it is absent, but are unreliable and biasable when the latent is of poor quality and more difficult to judge. In these cases, we would expect greater reliance on contextual cues to help deal with this visual complexity. Therefore, we would expect to observe a greater contextual biasing effect on suitability judgment as a result of the presence of a comparison exemplar. For this reason, we aim to study cases in which the suitability of the latent is clear and relatively simple to judge, versus cases in which the latent suitability is more difficult to judge.

1.2 Inferred Suitability Conclusions

In both experiment 1 and 2 we shall be comparing the suitability conclusions in both a solo suitability determination task and a latent-exemplar comparison task. We employed an inferred suitability conclusions measure to compare suitability determinations between these different tasks. Suitability conclusions where the examiner has analyzed the latent in isolation (solo) provide a direct and clean indication of an examiner's view of the suitability as they can judge a latent to be "Suitable", "Unsuitable" or "Questionable". However, during latent-to-exemplar comparisons such suitability judgment is not provided. Nevertheless, we can infer the suitability judgment from their comparison conclusions. If the examiner decides that a latent is a "Match" or "Non-match", then we can infer that the examiner considers the latent to be "Suitable" for comparison; otherwise they should not draw this conclusion. If the examiner draws an "Inconclusive" conclusion regarding the latent to exemplar comparison, it could be that the examiner may consider the latent suitability to be "Questionable". Finally, the examiner can draw the conclusion that the latent is "Unsuitable" for comparison with the exemplar, directly indicating their suitability assessment.

Using this framework we can compare the suitability conclusions made by examiners when undertaking solo suitability assessment with inferred suitability conclusions drawn from the examiners' comparison conclusion decisions. Table 1 summarizes the inferred suitability categories on the basis of solo suitability conclusions and latent-to-exemplar comparison conclusions.

Table 1
The Inferred suitability conclusion categories derived from solo suitability assessment or latent-to-exemplar comparison conclusions

Inferred Suitability Conclusion	Solos Suitability Assessment Conclusion	Latent-to-Exemplar Comparison Conclusions
Suitable	Suitable	Match/Non-match
Questionable	Questionable	Inconclusive
Unsuitable	Unsuitable	Unsuitable

1.3 Examiner Qualifications

This study will also examine whether possessing International Association for Identification (IAI) Certified Latent Print Examiner (CLPE) status has any bearing on the performance and biasability of examiners. Given the automaticity of contextual effects on perception, we are doubtful that CLPE qualified examiners will be immune to the biasing effects of the comparison prints on inferred suitability conclusions. Nevertheless, the extent of bias may be reduced if the training of these individuals is effective. This study will allow us to draw conclusions as to whether examiner certification has any measurable effect on suitability analysis determinations.

1.4 Exogenous Contextual Biases

We also seek to examine the effect of exogenous contextual biases on suitability assessment, i.e. biases arising outside of the latent or exemplar being examined. In experiment 2, we explore whether examiners' solo suitability judgments may be biased by the knowledge of another examiner's purported suitability determination. In experiment 3 we assess whether suitability judgments continue to be correlated with the underlying latent suitability in a strongly biasing context (identification made by another examiner in a major case).

2. Experiments

This study is important as it provides the opportunity to examine whether, in latent-to-exemplar examinations, the judgment of suitability for identification is subject to revision in the presence of a record print target. If the Analysis stage is affected by the presence of a comparison print we have reasonable evidence supporting the proposals for a linear process in ACE-V [14]. It is generally understood that the process cannot, practically, be absolutely linear, if only to allow for correction of missed data during Analysis. However, it is important to understand the degree to which Analysis varies with and without exposure to a comparison exemplar presented alongside the latent.

Experiment 1 investigates the effect of matching or non-matching comparison prints on suitability determinations regarding latent prints from a range of predetermined suitability classes (from highly suitable to highly unsuitable) compared to solo suitability determination. We examine if suitability determinations depend on the presence of a matching or non-matching comparison exemplar presented alongside the latent being assessed for suitability and we expect such effects to be most prominent in more borderline cases (i.e. when suitability class is hard to determine, near the decision threshold, rather than being highly suitable or highly unsuitable). This main study tests the linearity of the ACE-V process. We assess the impact that IAI certification has on any observed effects.

Experiment 2 investigates whether solo suitability assessments may be biased by the knowledge of a previous examiner's suitability determination. We want to examine if this contextual information biases the likelihood of a suitable or unsuitable determination. Again, we assess whether this biasing effect is mediated whether examiners are IAI certified or not.

Finally, assuming that examiners' suitability assessments are not random and do align with the underlying suitability class of the latent, we also assess whether this persists even in a strong bias scenario. Experiment 3 assesses whether the prior determined suitability class of the latent does predict examiners' suitability conclusions given the knowledge that the case is a major case and previous examiner has already made an identification.

3. Experiment 1

Inferred Suitability Conclusions in Solo and Pairwise Comparison Task Types

3.1 Method

In this study we compare the inferred suitability conclusions of examiners when undertaking solo-suitability assessment versus when they are undertaking latent-to-exemplar comparisons. We introduced latent prints that differed in their previously determined degree of suitability from clear cut cases (i.e. "Highly Suitable" or "Highly Unsuitable") to those that were slightly less clear cut (referred to as simply "Suitable" or "Unsuitable") to borderline cases that were difficult to determine with respect to suitability (referred to as "Inconclusive"). We examine whether the inferred suitability conclusions may differ depending on the presence of a comparison print (i.e. when undertaking latent to exemplar comparison) but expect that this effect may only be strongly observable when the latent suitability is more difficult to judge (i.e. in the "Inconclusive" predetermined suitability class compared to the "Highly suitable" or "Highly unsuitable" classes). In total, 54% the participants were IAI certified.

3.1.1 Materials

We developed a dataset specifically for this task. In total, 16 donors provided 6,400 latent prints (400 from each donor) and 16 known exemplar sets. The latent prints were classified into five predetermined suitability classes ("Highly Suitable", "Suitable", "Inconclusive", "Unsuitable" and "Highly Unsuitable") based on the suitability determinations of four IAI Certified Latent Print Examiners (CLPEs). Latent prints were securely presented to the CLPEs in Complete Consultant's Worldwide's (CCW) Web-based Remote Examination (WebREx) client/server software system. The system allows the CLPEs to grade each latent print in terms of its suitability for individualization purposes. CLPE graders were not shown image(s) of the matching exemplar in order to make their determination of "Suitable", "Unsuitable", or "Unsure". Highly Suitable latent prints are defined as ones that are graded as "Suitable" to be able to be positively identified, if a clear (i.e. also suitable) impression of the matching friction ridge skin source were made available for comparison, by all 4 CLPEs. Highly Unsuitable latent prints are ones that are graded as "Unsuitable" by all 4 CLPEs.

Suitable and Unsuitable latent prints are graded as such by 3 out of 4 CLPEs, where the 4th reached a determination other than the other 3 CLPEs (e.g. 3 "Suitable" versus 1 "Unsuitable" or "Unsure"; or 3 "Unsuitable" versus 1 "Suitable" or "Unsure"). Inconclusive latent prints are graded as "Suitable" or "Unsuitable" by 2 CLPEs when the other 2 CLPEs reached another determination, or where there are 3 or more "Unsure" determinations. Appendix 1 shows the full 81 combinations of CLPE determinations and the associated predetermined suitability class. Note that while this methodology enables us to broadly categorize the latent prints into our different predetermined suitability classes, we can expect a degree of variation in suitability determinations between the four CLPEs that derives from their training, experience, etc. and not purely from the latent suitability. For this reason, as we explain in the results section, we need to control for this potential variation in the latent suitability over and above our predetermined suitability class. From this dataset, 640 latent print images and source record print images were used. The 640 images are composed of 320 latent prints predetermined to be highly suitable, and 80 latent prints from the other predetermined suitability categories. More highly suitable latent prints were chosen to keep examiners from tiring of complex image examination.

3.1.2 Participants

In total, 24 expert latent print examiners took part in the study. All examiners were experienced latent print examiners that have qualified as expert witnesses in U.S. courts. Thirteen of the participants were IAI certified (CLPEs) with a mean of 13.0 years' experience and standard deviation of 6.32 and the remaining eleven were not IAI certified (Non-CLPEs) with a mean of 12.1 years' experience and standard deviation of 7.96.

3.1.3 Design

The dependent variable was the inferred suitability conclusion made by the participants: "Suitable", "Unsuitable" or "Questionable" (see Table 1). There were two independent variables. The first was the Predetermined Suitability Class (explained in the material section) that consisted of five levels: Highly Suitable, Highly Unsuitable, Suitable, Unsuitable and Inconclusive. The second independent

variable was the Task Type: solo suitability task (Solo), latent-to-exemplar comparison task with a matching comparison exemplar (Match Pair) and latent-to-exemplar comparisons with a non-matching comparison exemplar (Non-Match Pair). A further independent variable was the IAI certification status of the examiner (CLPE vs. Non-CLPE)

3.1.4 Procedure

Data was collecting covertly, meaning that the study was administered in the midst of participants who believed they were conducting casework In the solo suitability task (Solo), each examiner was presented with at least 125 latent print images selected randomly without replacement from our 640 latent dataset and instructed to analyze them and declare each to be Suitable or Unsuitable for identification or Questionable. In the latent-to-exemplar comparison task, each examiner was presented with at least 180 latent-print images (again selected randomly without replacement) from those not used in the solo task for that examiner alongside an exemplar. Half of these latent prints were presented with a matching exemplar (Match Pair) and in the other half latent prints were not matching (Non-Match Pair). Accordingly, while each examiner was presented with a particular latent print only once, a latent could be presented solo or with an exemplar to different examiners. Participants were instructed to conduct an examination to reach one of the standard conclusions of Individualization (i.e. Match), Exclusion (i.e. Non-Match), Inconclusive, or to conclude the latent was Unsuitable for any firm conclusion. The inferred suitability conclusion variable was determined for all three tasks (see table 1). Note that this equates to the actual suitability determination in the solo task and is derived in the comparison task.

3.2 Results

Figure 1 shows the percentage of "Suitable" (A), "Unsuitable" (B), and "Questionable" (C) conclusions by our examiner participants in solo and matching/non-matching pair comparison tasks for different classes of latent defined by their predetermined suitability. As we should expect, Figure 1A shows that a very high percentage of "Highly Suitable" class latent prints were judged to be "Suitable" by the examiners and a very low percentage of the "Highly Unsuitable" class latent prints

were judged to be "Suitable". Similarly, as expected Figure 1B shows that a very high percentage of "Highly Unsuitable" class latent prints were judged to be "Unsuitable" by examiners and only a small percentage of "Highly Suitable" class latent prints were judge to be "Unsuitable". Indeed, it would be surprising if we did not observe this correspondence between the predetermined suitability of the latent and the participant's suitability conclusions.

However, in different task types (x-axis) these percentage rates of each conclusion appear to differ and these differences appear to be greatest when the predetermined suitability class (shown by the different lines) was more borderline (e.g. Inconclusive).

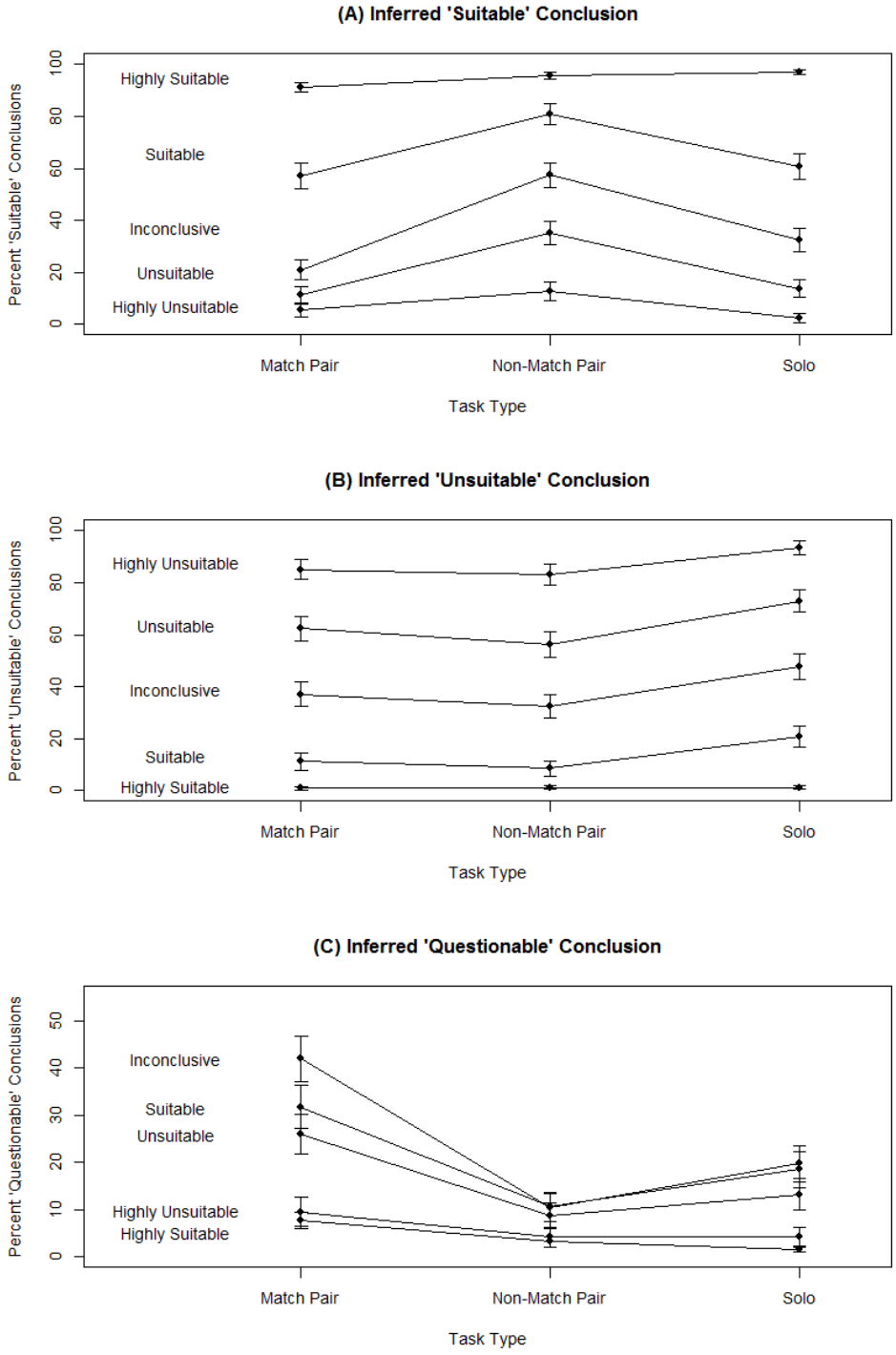

Figure 1. Results from Experiment 1. The percentage of the total count of each of the three inferred suitability conclusions for each predetermined latent suitability category in each task type. *The predetermined suitability class is indicated by the different lines, the task type is indicated on the x-axis and the percentage of each inferred suitability conclusion is shown on the y-axis. Hence, the percentages of inferred "Suitable" (shown in A), "Unsuitable" (B), and Questionable (C)*

conclusions for a particular predetermined suitability class (e.g. "Highly Suitable") for a particular task type (e.g. "Match Pair") will sum to 1. Note, also, that the lines serve only as a visual aid to indicate the degree of discontinuity in inferred suitability determinations between the task types and do not reflect data points between task three type categories.

Spearman's rank correlation coefficient between the predetermined suitability class and the examiner conclusions was substantial ρ=.734. Important to this experiment, however, is not the strength of this correlation. Rather, we are interested in whether or not this correspondence differs within individuals in the presence or absence of a comparison exemplar. Indeed, when we compare the Spearman's rank correlation coefficient between the examiners' inferred suitability conclusions and the predetermined suitability class of the latent, in the solo assessment, task we observe a correlation of ρ=.795. In contrast, in the presence of a matching comparison exemplar, this correlation reduces to ρ=.755, which is found to be significant[1] using a z-test based on a Fisher r-to-z transformation of the coefficients, $z(5977) = 3.84$, $p < .001$. Furthermore, in the presence of a non-matching comparison exemplar this correlation reduces to ρ=.650, which is again found to be significantly lower compared to the solo assessment correlation, $z(5976) = 11.78$, $p < .001$. While these basic correlational descriptive statistics suggest that some effect on suitability judgments is occurring in the presence/absence of a comparison exemplar, in order to establish exactly where the effects lie, we need to employ a more sophisticated statistical approach.

As discussed in the materials section, we need to control for the potential variation among latent prints over and above our predetermined suitability class. We also need to control for the potential variation between examiners in their criteria for judging suitability. As found by Dror et al [19] we expect to that some examiners may generally tend draw more "suitable"/"unsuitable"/"questionable" conclusions than others. Therefore, we employed a mixed-effects modeling approach that will correctly control for this possible variation within particular examiners and within particular latent prints across trials that would otherwise undermine the assumption of observation independence. This

[1] Throughout this report we use the term "significance" in the statistical sense used in hypothesis testing whereby the null hypothesis is rejected at a 5% alpha level.

is achieved by including two random factor regressors (an Examiner ID and a Latent ID variable) in addition to our main (fixed) effects (Task Type and Predetermined Suitability Class). The analyses were run in R using the lme4 package [25]. Each suitability conclusion was modeled separately with the first model predicting the likelihood of a "Suitable" decision, the second model predicting the likelihood of "Questionable" decision and the third modeling the likelihood of an "Unsuitable" decision.

We can use this modeling technique to assess whether the presence/absence of a matching/non-matching comparison exemplar (Task Type) has an effect on the likelihood of each suitability conclusion having properly controlled for the variation between different examiners and latent prints. Secondly, we can assess whether any effect of the comparison exemplar on suitability conclusions depends on the predetermined suitability class of the latent. This is because we expect that there will be a stronger effect of a comparison exemplar on suitability judgments when the latent is more difficult to judge (i.e. when the latent is not Highly Suitable/Unsuitable). This hypothesis would be confirmed if there was a significant Prior Determined Suitability × Task Type interaction in the direction we expect. Indeed, Figure 1 does appear to show that there is greater differential in the effect of task type on the likelihood of each of the suitability conclusions as long as the predetermined suitability class is not Highly Suitable or Highly Unsuitable.

Table 2 summarizes the results of these models[2] where the log-likelihood ratio tests indicate the significance of each factor in predicting the likelihood of each suitability conclusion. Table 2 shows that we observe a significant effect of task type on the likelihood of each of the three suitability conclusions. This significant main effect of task type indicates that there is a different likelihood of each inferred suitability conclusion depending on the task being undertaken (i.e. whether the latent is assessed solo or with a matching or non-matching comparison print). However, this effect appears to depend on the predetermined suitability class of the latent.

[2] Akin to an ANOVA for linear regression models, the log-likelihood ratio tests in this table show the significance of main effects by comparing nested models with each term added sequentially to the previous model.

Table 2

Results from Experiment 1. The main effects of predetermined latent suitability and task type on the predicted likelihood of each decision (suitable, questionable or unsuitable) based on log likelihood ratio (LLR) tests.

Factor	*ΔAIC*	χ^2 *difference*	*Df*	*p*
Likelihood of "Suitable" Conclusion				
Prior Determined Suitability	-982.3	990.34	4	< .001
Task Type	-366.9	370.92	2	< .001
Prior Determined Suitability × Task Type	-66.7	85.10	8	< .001
Likelihood of "Questionable" Conclusion				
Prior Determined Suitability	-373.2	381.27	4	< .001
Task Type	-290.6	294.56	2	< .001
Prior Determined Suitability × Task Type	-14.4	30.12	8	< .001
Likelihood of "Unsuitable" Conclusion				
Prior Determined Suitability	-998.9	996.94	4	< .001
Task Type	-151.6	155.56	2	< .001
Prior Determined Suitability × Task Type	3.6	12.39	8	.135

ΔAIC indicates the change in *Akaike* information criterion as a result of including the factor in that row. $\chi 2$ difference indicates the chi squared statistic for the change in log-likelihood as a result of including the factor in that row (LRR test). Df indicates the degrees of freedom for that LLR test. p indicates the statistical significance of the LLR test.

This is because we also observe significant Prior Determined Suitability × Task Type interactions for "Suitable" and "Questionable" conclusions. This interaction indicates that the effect of the task type depends on the predetermined suitability of the latent. From Figure 1 it seems that that these interactions are due to a greater effect of the task type on suitability judgments when the suitability of the latent is more difficult to judge (i.e. in borderline cases such as the "Inconclusive" class of latent prints) compared to cases in which the latent is clearly either suitable or unsuitable (i.e. such as in the Highly Suitable/Highly Unsuitable classes).

3.2.1 Borderline Suitable/Unsuitable Latent Prints

In order to examine these interaction effects in more detail, we followed up with Bonferroni adjusted pairwise Wilcoxon rank sum tests on different predetermined suitability classes. We first examined the borderline latent print class used in the study; the "Inconclusive" predetermined suitability class. Table 3 shows these results by indicating the count (in brackets) and percentage of each of the conclusions made by all the examiners under each of the task types for the "Inconclusive" (borderline) class of latent. This class of latent would be the most difficult to judge and therefore was expected to be the most affected by the presence of comparison exemplar. It was found that examiners were 2.73

times more likely to decide that a borderline latent was suitable when it was paired with a non-matching exemplar compared to when it was paired with a matching exemplar ($p < .001$). They were 1.76 times more likely to decide a borderline latent was suitable when it was paired with a non-matching exemplar than when it was examined solo ($p < .001$). The examiners were also more likely to conclude that borderline latent prints were suitable when examining them solo than when examining the latent with a matching exemplar ($p < .001$).

Table 3
Results from Experiment 1. The percentage of the total count of each inferred suitability conclusion drawn in each task type for borderline class latent prints (i.e. predetermined suitability = "Inconclusive"). The total count is shown in brackets.

Conclusion	Task Type		
	Solo	Matching Pair	Non-Matching Pair
Suitable	32.5% (135)	21.0% (87)	57.3% (238)
Questionable	19.8% (82)	41.9% (174)	10.4% (43)
Unsuitable	47.7% (198)	37.1% (154)	32.3% (134)
Total	100% (415)	100% (415)	100% (415)

Examiners were 37.5% more likely to decide that a borderline latent was unsuitable when examining the latent solo that when examining it with a comparison exemplar ($p < .001$). There was no significant difference between matching and non-matching comparison tasks in the likelihood of "Unsuitable" conclusions regarding the borderline latent prints ($p = .435$).

It was also found that examiners were 4.03 times more likely to decide that a borderline latent was questionable when it was presented with a matching exemplar compared to when it was paired with a non-matching exemplar ($p < .001$), and twice as likely compared to when presented solo (cf. with a matching exemplar, $p < .001$). Solo assessment resulted in 1.91 as many "Questionable" conclusions of borderline latent prints compared to when presented with a non-matching exemplar ($p < .001$).

3.2.2 Non-borderline, Highly Unsuitable Class, Latent Prints

In a similar fashion to Table 3 for the borderline cases, Table 4 shows the results of the "Highly Unsuitable" class and Table 5 shows the results of the "Highly Suitable" class. As with the borderline

class latent prints, examiners were more likely to conclude that a highly unsuitable latent was suitable for comparison when it was presented with a non-matching comparison exemplar compared to both solo ($p < .001$) and when presented with a matching comparison print ($p = .002$). There was no significant difference in the rate of "Suitable" conclusions ($p = .137$) between solo assessment and when the latent was presented with a matching exemplar for these "Highly unsuitable" latent prints.

There were significantly more "Unsuitable" conclusions drawn when the latent was assessed solo versus when it was analyzed in the presence of a comparison exemplar ($p < .001$) for "Highly unsuitability" latent prints. Finally, the same proportion of "Questionable" conclusions were drawn in the solo assessment task when analyzed with a non-matching comparison exemplar ($p > .999$), but twice as many were observed in the presence of a matching comparison exemplar ($p = .019$ vs. non-matching and $p = .018$ vs. solo) for "Highly Unsuitable" class latent prints.

3.2.3 Non-borderline, Highly Suitable, Latent Prints

For the "Highly Suitable" class, the increased likelihood of "Questionable" conclusions in the presence of a matching exemplar compared to solo analysis ($p < .001$) or in the presence of a non-matching exemplar ($p < .001$) persisted. There was no significant difference in the rate of "Suitable" conclusions regarding "Highly Suitable" latent prints between non-matching comparison and solo analysis tasks ($p = .15$). However, more "Suitable" conclusions were drawn after solo assessment that when analyzed in the presence of a non-matching comparison exemplar even for this "Highly Suitable" class ($p < .001$ cf. both).

Table 4
Results from Experiment 1. The percentage of the total count of each inferred suitability conclusion drawn in each task type for non-borderline unsuitable class latent prints (predetermined suitability = “Highly Unsuitable”). The total count is shown in brackets.

Conclusion	Task Type		
	Solo	Matching Pair	Non-Matching Pair
Suitable	2.4% (8)	5.4% (18)	12.9% (43)
Questionable	4.2% (14)	9.6% (32)	4.2% (14)
Unsuitable	93.4% (313)	85.1% (285)	82.9% (277)
Total	100% (335)	100% (335)	100% (334)

Table 5
Results from Experiment 1. The percentage of the total count of each inferred suitability conclusion drawn in each task type non-borderline suitable class latent prints (predetermined suitability = "Highly Suitable"). The total count is shown in brackets.

Conclusion	Task Type		
	Solo	Matching Pair	Non-Matching Pair
Suitable	97.0% (1863)	91.1% (875)	95.6% (918)
Questionable	1.7% (32)	7.8% (75)	3.2% (31)
Unsuitable	1.3% (25)	1.0% (10)	1.1% (11)
Total	100% (1920)	100% (960)	100% (960)

3.2.4 IAI Qualifications

There was no significant difference between CLPE qualified and Non-CLPE qualified examiners in the number of years' experience, $t(19) = -0.306$, $p = .763$. To test for the effect of IAI certification, we test whether the addition of an additional factor for IAI certification significantly improved the fit of the linear mixed models, again using the log-likelihood ratio test approach. The results showed that IAI certification was found to play a mediating role in the previously found interaction between task type and prior determined suitability on inferred suitability conclusions. The Prior Determined Suitability × Task Type interactions × IAI Qualification interaction was found to be a significant effect in the linear mixed model for the likelihood of a "Suitable" decision ($\Delta AIC = -48.9$, $\chi2(10) = 68.9$, $p < .001$), the likelihood of an "Unsuitable" decision ($\Delta AIC = -33.3$, $\chi2(10) = 53.3$, $p < .001$) and the likelihood of a "Questionable" decision ($\Delta AIC = -3$, $\chi2(10) = 22.94$, $p = .011$).

Table 6
Results from Experiment 1. The Spearman's correlation coefficients between the predetermined suitability class of the latent and IAI certified (CLPEs) versus non IAI certified (Non-CLPEs) examiners' inferred suitability conclusions in solo suitability, matching pair comparison, and non-matching pair comparison task types.

	Overall	Solo	Matching Pair	Non-Matching Pair
CLPEs	.753	.783	.801	.676
Non-CLPEs	.708	.808	.698	.613

Table 6 shows the Spearman's rank correlation coefficients between the predetermined suitability class of the latent and the examiners' inferred suitability judgments. Overall, the inferred suitability

conclusions of CLPEs tended to have a greater correlation with the underlying predetermined suitability class of the latent, $z(7776) = 4.28$, $p < .001$. The results also showed that the inferred suitability conclusions of both CLPEs and non-CLPEs demonstrated worse correlations with the predetermined latent suitability class in the presence of a non-matching exemplar compared to the solo suitability task ($z(3023) = 6.17$, $p < .001$ for CLPE and $z(2471) = 10.07$, $p < .001$ for Non-CLPE).

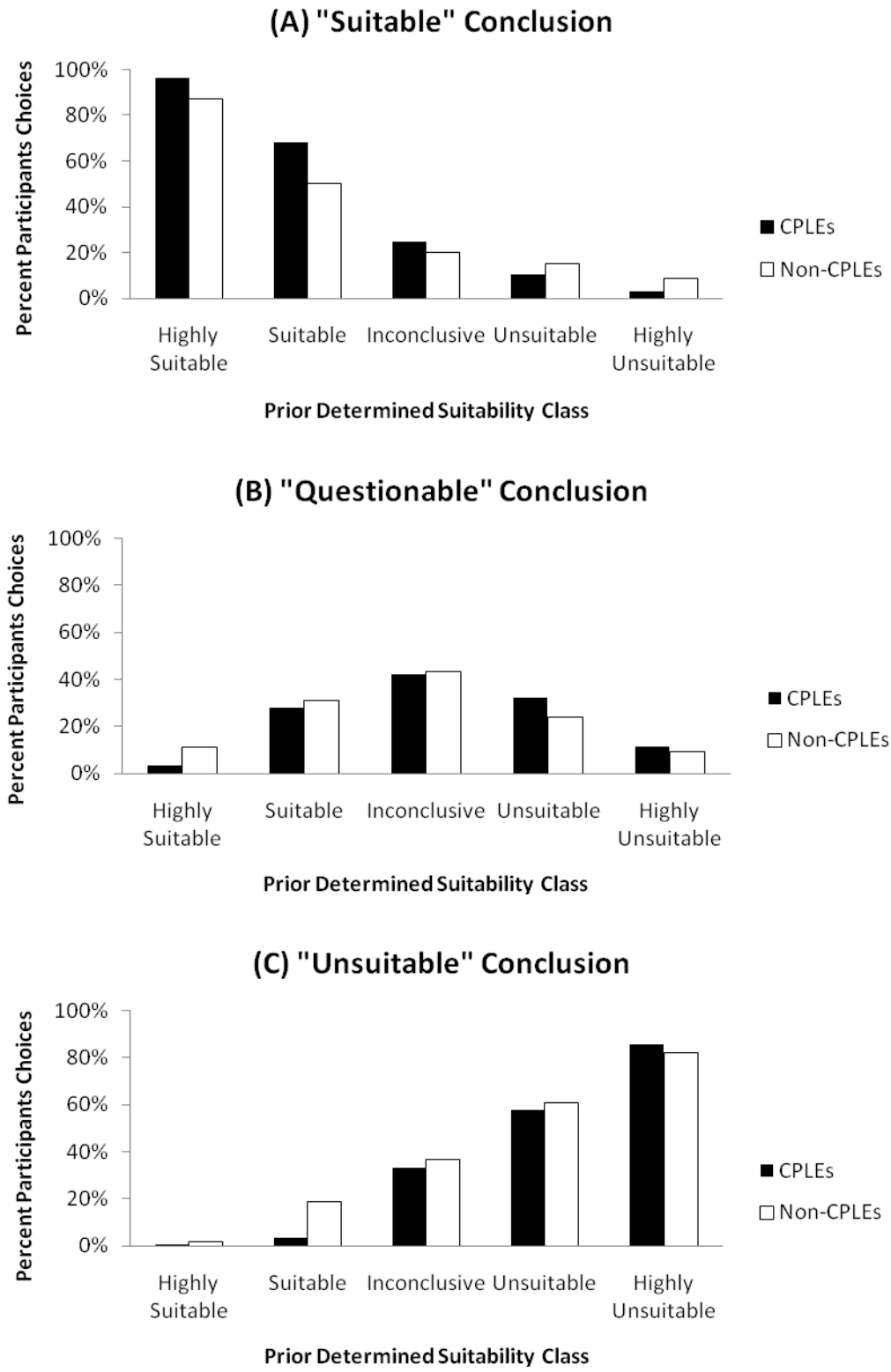

Figure 2. The percentage of CLPEs and Non-CPLEs with a "Suitable" (A), "Questionable" (B) and "Unsuitable" (C) inferred suitability determination regarding latent prints that varies in the prior determined suitability class when undertaking a matching pair comparison task. *The total of the percentages shown in A, B and C for each prior determined suitability class within each certification group will sum to 100%.*

However, the main difference between CLPEs and Non-CLPEs appears to be in the presence of a matching comparison exemplar. While the inferred suitability conclusions of Non-CLPEs were found

to be less correlated with the predetermined suitability class of the latent in the presence of a matching exemplar, $z(2415) = 6.27$, $p < .001$, the CLPEs appeared to be unaffected in both solo and matching comparison tasks demonstrating similar correlations, $z(3060) = 1.32$, $p = .187$.

In order to examine this effect of the IAI certification during matching pair comparisons we, again, employed Bonferroni adjusted pairwise Wilcoxon rank sum tests. Table 7 shows the percentage of "Suitable", "Questionable" and "Unsuitable" conclusions for the five different predetermined suitability classes in the matching pair comparison task. The results show that the suitability conclusions for CLPEs and Non-CLPEs are reasonably similar in the borderline "Inconclusive" class of latent. The main differences appear at the extremes, i.e. "Highly Suitable" and "Highly Unsuitable" classes. When presented with a "Highly Unsuitable" class of latent alongside a matching exemplar, the Non-CLPEs were significantly less likely to decide that a latent was unsuitable ($p < .001$) or questionable ($p < .001$) and more likely to decide the print was suitable for comparison than CLPEs ($p < .001$).

Importantly, all these Non-CLPEs that decided the latent was suitable for comparison did not make the correct "Individualization" conclusion, instead concluding that the latent prints *did not* match, despite both images coming from the same source. While five "Highly Unsuitable" latent prints were also judged to be "Suitable" by the CLPEs, one examiner did manage to draw the correct conclusion that the pair did in fact match despite the highly unsuitable class of the latent being examined. For the "Highly Suitable" class of latent prints, the pattern reversed; the Non-CLPEs were less likely to judge that the matching pairs were suitable for comparison than CLPE examiners ($p < .001$) and more likely instead to conclude that the latent prints were "Unsuitable" or "Questionable".

4. Experiment 2

Extra Suitability Information on Solo Suitability Assessment

4.1 Method

In this experiment, we examined the role of knowledge regarding another examiner's conclusion on solo suitability determination. Examiners undertook solo suitability assessment of latent prints as in

the previous study, however this time the examiners were also shown a text display above the latent print indicating a previous examiner's suitability determination. This extra suitability information was manipulated to determine whether it biases the suitability judgments of examiners.

4.1.1 Design

The dependent variable was the inferred suitability conclusion made by the participants: "Suitable", "Unsuitable" or "Questionable" (see Table 1). There were two independent variables. The first was the prior determined suitability category explained in the materials section above that consisted of three levels: Suitable, Unsuitable and Inconclusive. The second independent variable was a purported suitability information bias in the form of text displayed at the top of the screen indicating a previous examiner's suitability determination of the latent being examined (either "Suitable" or "Unsuitable"). While an "Inconclusive" text display would be a possible further condition, this was unlikely to provide any biasing effect and would, therefore, provide little additional value to this particular experiment. Again we assessed impact of IAI certification status of the examiners on this potentially biasing information.

4.1.2 Procedure

As in experiment 1, the trials were presented covertly. The Each examiner was presented with 30 latent prints in each category ("Inconclusive", "Suitable" and "Unsuitable"). Fifteen of the latent prints were presented with the word "Suitable" displayed above the latent print and the other half of the latent prints were presented with the word "Unsuitable" displayed. As the repeated presentation of "Highly Unsuitable" latent prints with a "Suitable" previous examiner determinations or "Highly Suitable" latent print with "Unsuitable" previous examiner determinations might seem suspicious to the participants, we decided not to include these conditions in the design. Nevertheless, in order to preserve the base rate prevalence of each prior determined suitability class, highly suitable and highly unsuitable latent prints were also presented, but only with the congruent "Suitable" and "Unsuitable" text display. This unbalancing data was ignored in the subsequent analysis.

4.2 Results

Table 7 shows the results of the linear mixed modeling. The results indicate that there is no effect of the suitability information bias on the likelihood of "Suitable" choices or "Questionable" choices. However, there was a significant effect of the suitability information bar on the likelihood of "Unsuitable" conclusions by the examiners. As before, there was a significant effect of the prior determined suitability class on the examiners' suitability conclusions, however, there were no significant interactions.

Table 8 appears to show this effect whereby there was a greater likelihood of "Unsuitable" conclusions by examiners when the text above the latent being assessed (purported suitability bias) stated that a previous examiner had found the latent to be "Unsuitable" compared to when the text stated a previous "Suitable" determination. There was no significant effect of CLPE qualification on the likelihood of "Suitable" conclusions (ΔAIC = 8.6, $\chi2(6) = 3.39$, $p = .759$), "Unsuitable" conclusions (ΔAIC = 10.1, $\chi2(6) = 1.91$, $p = .928$) or "Questionable" conclusions (ΔAIC = 2.4, $\chi2(6) = 9.52$, $p = .146$).

Table 7

Results from Experiment 2. The main effects of prior determined latent suitability and the presence of suitability information on the predicted likelihood of each suitability decision (suitable, questionable or unsuitable) based on log likelihood ratio (LLR) tests.

	ΔAIC	χ^2 *difference*	*Df*	*p*
Likelihood of "Suitable" Conclusion				
Prior Determined Suitability	-198.7	202.64	2	< .001
Suitability Information Bias	0.9	1.02	1	.312
Prior Determined Suitability × Task Type	2.4	1.65	2	.439
Likelihood of "Questionable" Conclusion				
Prior Determined Suitability	0.6	3.42	2	.181
Suitability Information Bias	-0.9	2.92	1	.087
Prior Determined Suitability × Task Type	3.6	0.48	2	.786
Likelihood of "Unsuitable" Conclusion				
Prior Determined Suitability	-169.2	173.14	2	< .001
Suitability Information Bias	-6.3	8.28	1	.004
Prior Determined Suitability × Task Type	2.2	1.86	2	.395

ΔAIC indicates the change in *Akaike* information criterion as a result of including the factor in that row. χ2 difference indicates the chi squared statistic for the change in log-likelihood as a result of including the factor in that row (LRR test). Df indicates the degrees of freedom for that LLR test. p indicates the statistical significance of the LLR test.

Table 8
Results from Experiment 2. The percentage of the total count of each inferred suitability conclusion drawn under different predetermined latent classes during solo suitability assessment and with information regarding a previous examiner's purported suitability assessment presented was a text display above the latent being assessed. The count of the conclusions is shown in brackets.

	Prior Determined Suitability					
Examiners'	Suitable		Inconclusive		Unsuitable	
Suitability	Suitability Bias Text Display		Suitability Bias Text Display		Suitability Bias Text Display	
Conclusion	"Suitable"	"Unsuitable"	"Suitable"	"Unsuitable"	"Suitable"	"Unsuitable"
Suitable	63.4% (223)	59.4% (209)	32.6% (100)	32.7% (100)	10.1% (31)	10.2% (31)
Questionable	20.2% (71)	17.6% (62)	17.6% (54)	16.0% (49)	16.9% (52)	12.8% (39)
Unsuitable	16.5% (58)	23.0% (81)	49.8% (153)	51.3% (157)	73.1% (225)	77.0% (235)
Total	100% (352)	100% (352)	100% (308)	100% (306)	100% (308)	100% (305)

5. Experiment 3

Solo Suitability Assessment in a Serious Case

5.1 Method

The final experiment examined suitability assessments in a serious case with information regarding a previous examiner's decision. Examiners were asked to undertake analysis of a latent and comparison print purported to have been found to be a match by a previous examiner in the context of a major case.

5.1.1 Design

The 24 examiners were split into three groups, the first were presented with a "Highly Suitable" latent, the second with a "Suitable" latent and the third with a "Inconclusive" latent. Examiners were asked to provide one of six conclusions: "Latent Unsuitable", "Ten print Unsuitable", "Common Area Unsuitable", "Inconclusive", "Individualization" or "Exclusion". The latent and exemplar always matched. We assessed whether the examiner's suitability determinations would still correlate with the prior determined suitability class of the latent despite the strong biasing context to suggest a match.

5.1.2 Procedure

Participants were sent an email containing images files of the latent and exemplar as attachments stating the following, "I need you to examine this pair of major case prints and make a decision. They have been identified by another examiner, and we need a second opinion. Please do the comparison

and decide whether: (1) Latent unsuitable, (2) Ten print unsuitable, (3) Common area unsuitable, (4) Inconclusive, (5) Individualization, (6) Exclusion". Examiners could also leave a comment regarding their conclusion.

5.2 Results

Table 9 shows the twenty four examiners' conclusions depending on the prior determined suitability condition. As shown, the majority of examiners acted in accordance with the prior determined suitability class of the latent and did not conclude that the latent was "Suitable" for comparison. Only three examiners drew the conclusion that the latent was "Suitable" for comparison; two concluding that the pair matched and one concluding the pair did not match. Whether or not the two individualizations and one exclusion judgment were significantly outside of the rates observed in non-serious cases would require a larger sample of data for analysis. Nevertheless, using logistic regression analysis, we can conclude that prior determined suitability does impact on the likelihood of the conclusion that the latent, exemplar, or common area was unsuitable, $\chi 2(2) = 8.19$, $p=.017$. These results indicates that prior determined suitability was a significant factor in the examiners' conclusions despite the biasing context of a major case in which a previous examiner had decided the pair matched.

Table 9
Results from Experiment 3. The counts of examiner conclusions under the three prior determined suitability conditions.

Examiner Conclusion	Inferred Suitability	Prior Determined Suitability		
		Highly Unsuitable	Unsuitable	Inconclusive
Latent Unsuitable	Unsuitable	7	3	1
Tenprint Unsuitable	Unsuitable	0	1	1
Common Area Unsuitable	Unsuitable	1	0	1
Inconclusive	Questionable	0	3	3
Individualization	Suitable	0	0	2
Exclusion	Suitable	0	1	0

6. Discussion

The results of Experiment 1 showed that, while there is a substantial correlation between examiners' inferred suitability judgments and the underlying suitability of the latent (measured in this study by predetermined suitability), the suitability judgments of examiners is still influenced by the presence of a matching and non-matching comparison exemplar. The effects of a comparison exemplar on suitability judgments were strongest when the latent is more difficult to judge, i.e. borderline suitable/unsuitable.

The main effect of the presence of a non-matching comparison exemplar was that examiners were more inclined to draw the conclusion that the latent was suitable compared to when the latent was presented solo. This effect persisted even when the latent was in the "Highly unsuitable" class of latent. One possible explanation for this is that the presence of a non-matching exemplar directed examiners' attention towards the differences between the latent and the exemplar. It may have been easier to identify some differences between the latent and the exemplar than to identify usable minutiae in the solo assessment task. Using these observed differences they may have found enough evidence to decide that the latent did not match the comparison exemplar.

However, the question remains as to whether these decisions were correct or not. On the one hand, it seems possible that some latent prints may be generally unsuitable for comparison in most cases, but in cases in which the comparison is clearly very different, the examiner may be able to form the correct judgment. Thus the suitability of a latent may be relative, depending on the exemplar with which it is being compared. However, the danger is if the presence of a comparison exemplar highlights erroneous differences in areas of noise or distortion in the latent. In such cases it is possible for the examiner to conclude "Exclusion" purely on the basis of noise when the correct conclusion is that the comparison is inconclusive. More research is required to determine whether the increased "Suitable" conclusions are errors arising from the non-matching comparison exemplar highlighting noisy differences between the two, or whether the conclusions are correct and that latent was *relatively* suitable given the particular comparison print being compared.

The presence of a matching exemplar was also shown to have an effect on suitability judgments. The main effect seemed to be that examiners were less likely to decide the latent was "Suitable" and were more likely to decide the latent was "Questionable" in the presence of a matching comparison exemplar compared to when it was assessed solo. This effect persisted even in the "Highly suitable" latent class suggesting a strong main effect. This finding fits with the results found by Dror et al [19] whereby the presence of a matching comparison exemplar reduced the number of minutiae observed by examiners. Assuming the same effect is occurring here, if examiners fail to find an adequate number of minutiae due to the presence of the matching comparison exemplar, they may be more inclined to conclude that the latent is "Questionable". It may also indicate a tendency for solo suitability judgments to overestimate the suitability of latent compared to the practical suitability when undertaking latent-to-exemplar comparisons.

The analysis of examiners' qualifications indicates that there was a difference between IAI certified (CLPEs) and non-IAI certified examiners (Non-CLPEs) in the effect of a comparison exemplar on suitability determination. In our study, the CLPE group tended to demonstrate greater correlation with the predetermined suitability than Non-CLPEs. However, both groups showed worse correlation with the underlying predetermined suitability when presented with a non-matching exemplar compared to their solo assessment. The main difference between the groups arose in the presence of a matching exemplar. When the latent being examined was highly unsuitable but presented alongside a matching exemplar, the Non-CLPEs were more likely to conclude that the latent was suitable for a comparison but that the pair did not match compared to CLPEs. When the latent being examined was from the "Highly suitable" class, the Non-CLPEs were more likely to conclude that the pair was unsuitable or questionable compared to CLPEs. An important question is whether there are any confounds between these groups that could also explain the difference over and above IAI certification. For example, while we did control for the number of years of experience, we did not control for differences in gender, age, education, etc. Indeed, it is possible that IAI certification reflects other factors, such as more dedicated and motivated examiners who opt to undertake the certification – hence, it may not be the IAI certification per se.

The results of Experiment 2 showed that examiners could be biased towards the "Unsuitable" conclusion given the knowledge that another examiner has concluded that the latent is unsuitable. Given that this effect was not observed in the likelihood of a "Suitable" conclusion, the results suggest that examiners are more prone to bias by the suggestion that a latent is unsuitable than they are to the suggestion that it is suitable. IAI certification was not found to mediate this effect.

The results of Experiment 3 indicate that even in strongly biasing scenarios (examinations in which another examiner had purportedly made an identification and the case is a major case) the underlying suitability of the latent still plays an important role in examiners' suitability determinations. While the results are promising, a further, more in-depth, study of the effects of case seriousness suitability determinations would be a valuable contribution to the literature.

7. Conclusion

The finding that suitability determination conclusions can be vary depending on the presence/absence of a matching/non-matching comparison exemplar indicates that ACE process should be well structured and linear in nature (with some exception, see [14]). Such a linear, sequential, approach to exposing forensic examiners to information has been suggested in DNA [26]. The results are consistent with the cognitive psychological literature indicating that individuals' attentional processes can be automatically directed by contextual information – in this case; the presence of a matching or non-matching exemplar. We observed a lower correspondence between the predetermined latent suitability and the examiners' inferred suitability judgments in the presence of a comparison print. Also as expected, the greatest effect of the presence of a comparison print occurred when the latent being compared was more difficult to judge (when the latent was on the borderline between being suitable and unsuitable).

The finding that examiners can be biased towards the conclusion that latent is unsuitable by the purported conclusion of another examiner, but not biased towards the conclusion that the latent is suitable, tells us a little about the decision thresholds for examiners in terms of false positives and

false negatives. This finding is consistent with other studies of forensic examiners' views of error types [27].

The finding from Experiment 3 was that the suitability determination of examiners in highly biasing scenarios is consistent with the underlying suitability of the latent. This shows the relative robustness of the Analysis stage in ACE, and is consistent with Schiffer and Champod [20].

Finally, the finding that examiners with IAI certification qualification appeared to be less affected by the contextual effect of comparison prints on suitability judgments indicates some initial evidence that certain types of training could help to reduce errors in suitability judgments. However, even IAI certified examiners were not entirely immune to the effects indicating that there is still room for improvement in the training of forensic examiners.

References

[1] Jonides, J. Voluntary versus automatic control over the mind's eye's movement. In Attention and Performance IX (Long, J.B. and Baddeley, A.D., eds) (1981) pp. 187–203 Lawrence Erlbaum Associates

[2] Wolfe, J.M. Cave, K.R. Franzel, S.L. Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance 15(3) (1989) 419–433.

[3] Folk, C.L. Remington, R.W. Johnston, J.C. Involuntary covert orienting is contingent on attentional control settings. Journal of Experimental Psychology: Human Perception and Performance 18(4) (1992) 1030–1044

[4] Desimone, R. Duncan, J. Neural mechanisms of selective visual attention. Annual Review of Neuroscience 18 (1995) 193–2227

[5] Egeth, H.E. Yantis, S. Visual attention: control, representation, and time course. Annual Review of Psychology 48 (1997) 269–297

[6] Itti, L. Koch, C.A. saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40(10-12) (2000) 1489–1506

[7] Kastner, S. Ungerleider, L.G. Mechanisms of visual attention in the human cortex. Annual Review of Neuroscience 23 (2000) 315–341.

[8] Corbetta, M. Shulman, G.L. Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience 3(3) (2002) 201–215

[9] Awh, E. Belopolsky, A.V. Theeuwes, J. Top-down versus bottom-up attentional control: a failed theoretical dichotomy. Trends in Cognitive Science 16(8) (2012) 437-443

[10] Julesz, B. Texton Gradients: The texton theory revisited. Biological Cybernetics 54(4-5) (1986) 234-251

[11] Moraglia, G. Display organization and the detection of horizontal line segments. Perception and Psychophysics 45(3) (1989) 265-272.

[12] Wolfe, J.M. Butcher, S.J. Lee, C. Hyle, M. Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. Journal of Experimental Psychology: Human Perception and Performance 29(2) (2003) 483-502

[13] Maljkovic, V. Nakayama, K.. Priming of pop-out: I. Role of features. Memory & Cognition 22(6) (1994) 657-672.

[14] Dror, I.E. How can Francis Bacon help forensic science? The four idols of human biases. Jurimetrics: The Journal of Law, Science, and Technology 50(1) (2009) 93-110.

[15] Office of the Inspector General. A review of the FBI's progress in responding to the recommendations in the office of the inspector general report on the fingerprint misidentification in the Brandon Mayfield case. http://www.justice.gov/oig/special/s1105.pdf (2011)

[16] National Institute of Standards and Technology. Expert Working Group on Human Factors in Latent Print Analysis. Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach. U.S. Department of Commerce, National Institute of Standards and Technology. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=910745 (2012).

[17] Langenburg, G. Champod, C. The GYRO System—A Recommended Approach to More Transparent Documentation. Journal of Forensic Identification 61(4) (2011) 373 – 384.

[18] Ulery, B.T. Hicklin, R.A. Buscaglia, J. Roberts, M.A. Repeatability and reproducibility of decisions by latent fingerprint examiners. PLoS One 7(3) (2012) e32800

[19] Dror, I.E. Champod, C. Langenburg, G. Charlton, D. Hunt, H. Rosenthal R. Cognitive issues in fingerprint analysis: Inter-and intra-expert consistency and the effect of a 'target' comparison. Forensic Science International 208 (2011) 10-17.
[20] Schiffer, B. Champod, C. The potential (negative) influence of observational biases at the analysis stage of finger mark individualization. Forensic Science International 167 (2007) 116–120.
[21] Dror, I.E. Charlton, D. Why experts make errors. Journal of Forensic Identification 56(4) (2006) 600-616.
[22] Dror, I.E. Rosenthal, R. Meta-analytically quantifying the reliability and biasability of forensic experts. Journal of Forensic Sciences 53(4) (2008) 900-903.
[23] Dror, I.E. Wertheim, K. Fraser-Mackenzie, P. Walajtys, J. The impact of human-technology cooperation and distributed cognition in forensic science: Biasing effects of AFIS contextual information on human experts. Journal of Forensic Sciences 57(2) (2012) 343-352.
[24] Dror, I.E. Peron, A. Hind, S. Charlton, D. When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. Applied Cognitive Psychology 19(6) (2005) 799-809.
[25] Bates, D. M. Fitting linear mixed models in R. R News 5 (2005) 27-30.
[26] Krane, D., Ford, S., Gilder, J., Inman, K., Jamieson, A., Koppl, R., Kornfield, I., Risinger, D., Rudin, N., Taylor, M., & Thompson. W.C. Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation. Journal of Forensic Science 53(4) (2008) 1006-7.
[27] Charlton, D. Fraser-Mackenzie, P. Dror, I.E. Emotional experiences and motivating factors associated with fingerprint analysis. Journal of Forensics Sciences 55(2) (2010) 385-393.

Appendix

Table A1

The predetermined suitability class based on all the possible combinations of suitability determinations made by the four CPLEs

CLPE 1	CLPE 2	CLPE 3	CLPE 4	Predetermined Suitability Class
Suitable	Suitable	Suitable	Suitable	Highly Suitable
Unsure	Suitable	Suitable	Suitable	Suitable
Unsuitable	Suitable	Suitable	Suitable	Suitable
Suitable	Unsure	Suitable	Suitable	Suitable
Unsure	Unsure	Suitable	Suitable	Inconclusive
Unsuitable	Unsure	Suitable	Suitable	Inconclusive
Suitable	Unsuitable	Suitable	Suitable	Suitable
Unsure	Unsuitable	Suitable	Suitable	Inconclusive
Unsuitable	Unsuitable	Suitable	Suitable	Inconclusive
Suitable	Suitable	Unsure	Suitable	Suitable
Unsure	Suitable	Unsure	Suitable	Inconclusive
Unsuitable	Suitable	Unsure	Suitable	Inconclusive
Suitable	Unsure	Unsure	Suitable	Inconclusive
Unsure	Unsure	Unsure	Suitable	Inconclusive
Unsuitable	Unsure	Unsure	Suitable	Inconclusive
Suitable	Unsuitable	Unsure	Suitable	Inconclusive
Unsure	Unsuitable	Unsure	Suitable	Inconclusive
Unsuitable	Unsuitable	Unsure	Suitable	Inconclusive
Suitable	Suitable	Unsuitable	Suitable	Suitable
Unsure	Suitable	Unsuitable	Suitable	Inconclusive
Unsuitable	Suitable	Unsuitable	Suitable	Inconclusive
Suitable	Unsure	Unsuitable	Suitable	Inconclusive
Unsure	Unsure	Unsuitable	Suitable	Inconclusive
Unsuitable	Unsure	Unsuitable	Suitable	Inconclusive
Suitable	Unsuitable	Unsuitable	Suitable	Inconclusive
Unsure	Unsuitable	Unsuitable	Suitable	Inconclusive
Unsuitable	Unsuitable	Unsuitable	Suitable	Unsuitable
Suitable	Suitable	Suitable	Unsure	Suitable
Unsure	Suitable	Suitable	Unsure	Inconclusive
Unsuitable	Suitable	Suitable	Unsure	Inconclusive
Suitable	Unsure	Suitable	Unsure	Inconclusive
Unsure	Unsure	Suitable	Unsure	Inconclusive
Unsuitable	Unsure	Suitable	Unsure	Inconclusive
Suitable	Unsuitable	Suitable	Unsure	Inconclusive
Unsure	Unsuitable	Suitable	Unsure	Inconclusive
Unsuitable	Unsuitable	Suitable	Unsure	Inconclusive
Suitable	Suitable	Unsure	Unsure	Inconclusive
Unsure	Suitable	Unsure	Unsure	Inconclusive
Unsuitable	Suitable	Unsure	Unsure	Inconclusive
Suitable	Unsure	Unsure	Unsure	Inconclusive

Unsure	Unsure	Unsure	Unsure	Inconclusive
Unsuitable	Unsure	Unsure	Unsure	Inconclusive
Suitable	Unsuitable	Unsure	Unsure	Inconclusive
Unsure	Unsuitable	Unsure	Unsure	Inconclusive
Unsuitable	Unsuitable	Unsure	Unsure	Inconclusive
Suitable	Suitable	Unsuitable	Unsure	Inconclusive
Unsure	Suitable	Unsuitable	Unsure	Inconclusive
Unsuitable	Suitable	Unsuitable	Unsure	Inconclusive
Suitable	Unsure	Unsuitable	Unsure	Inconclusive
Unsure	Unsure	Unsuitable	Unsure	Inconclusive
Unsuitable	Unsure	Unsuitable	Unsure	Inconclusive
Suitable	Unsuitable	Unsuitable	Unsure	Inconclusive
Unsure	Unsuitable	Unsuitable	Unsure	Inconclusive
Unsuitable	Unsuitable	Unsuitable	Unsure	Unsuitable
Suitable	Suitable	Suitable	Unsuitable	Suitable
Unsure	Suitable	Suitable	Unsuitable	Inconclusive
Unsuitable	Suitable	Suitable	Unsuitable	Inconclusive
Suitable	Unsure	Suitable	Unsuitable	Inconclusive
Unsure	Unsure	Suitable	Unsuitable	Inconclusive
Unsuitable	Unsure	Suitable	Unsuitable	Inconclusive
Suitable	Unsuitable	Suitable	Unsuitable	Inconclusive
Unsure	Unsuitable	Suitable	Unsuitable	Inconclusive
Unsuitable	Unsuitable	Suitable	Unsuitable	Unsuitable
Suitable	Suitable	Unsure	Unsuitable	Inconclusive
Unsure	Suitable	Unsure	Unsuitable	Inconclusive
Unsuitable	Suitable	Unsure	Unsuitable	Inconclusive
Suitable	Unsure	Unsure	Unsuitable	Inconclusive
Unsure	Unsure	Unsure	Unsuitable	Inconclusive
Unsuitable	Unsure	Unsure	Unsuitable	Inconclusive
Suitable	Unsuitable	Unsure	Unsuitable	Inconclusive
Unsure	Unsuitable	Unsure	Unsuitable	Inconclusive
Unsuitable	Unsuitable	Unsure	Unsuitable	Unsuitable
Suitable	Suitable	Unsuitable	Unsuitable	Inconclusive
Unsure	Suitable	Unsuitable	Unsuitable	Inconclusive
Unsuitable	Suitable	Unsuitable	Unsuitable	Unsuitable
Suitable	Unsure	Unsuitable	Unsuitable	Inconclusive
Unsure	Unsure	Unsuitable	Unsuitable	Inconclusive
Unsuitable	Unsure	Unsuitable	Unsuitable	Unsuitable
Suitable	Unsuitable	Unsuitable	Unsuitable	Unsuitable
Unsure	Unsuitable	Unsuitable	Unsuitable	Unsuitable
Unsuitable	Unsuitable	Unsuitable	Unsuitable	Highly Unsuitable