Linkage analysis in cases of serial burglary: comparing the performance of university students, police professionals, and a logistic regression model

Craig Bennell[a]*, Sarah Bloomfield[a], Brent Snook[b], Paul Taylor[c] and Carolyn Barnes[a]

[a]*Department of Psychology, Carleton University, Ottawa, Canada;* [b]*Department of Psychology, Memorial University of Newfoundland, Psychology, St John's, Canada;* [c]*Department of Psychology, Lancaster University, Lancaster, UK*

University students, police professionals, and a logistic regression model were provided with information on 38 pairs of burglaries, 20% of which were committed by the same offender, in order to examine their ability to accurately identify linked serial burglaries. For each offense pair, the information included: (1) the offense locations as points on a map, (2) the distance (in km) between the two offenses, (3) entry methods, (4) target characteristics, and (5) property stolen. Half of the participants received training informing them that the likelihood of two offenses being committed by the same offender increases as the distance between the offenses decreases. Results showed that students outperformed police professionals, that training increased decision accuracy, and that the logistic regression model achieved the highest rate of success. Potential explanations for these results are presented, focusing primarily on the participants' use of offense information, and their implications are discussed.

Keywords: linkage analysis; comparative case analysis; serial burglary; criminal behavior; decision making

Introduction

Linkage analysis involves using crime scene information to decide if two (or more) offenses were committed by the same offender (Bennell & Canter, 2002). Failure to link offenses accurately can waste police resources, prolong criminal investigations, and prevent the apprehension of the actual offender (Grubin, Kelly, & Brunsdon, 2001). Unfortunately, linkage decisions are often made under conditions of considerable uncertainty because physical evidence (e.g. DNA) is lacking or insufficient (Hazelwood & Warren, 2003). In the absence of physical evidence, investigators will have to base their inferences on behavioral information (Bennell & Canter, 2002; Bennell & Jones, 2005). Little is known, however, about the ability of investigators to use crime scene behavior to make accurate linkage decisions (Woodhams, Hollin, & Bull, 2007).

Although studies have shown that it is possible to link offenses from offense behaviors using statistical techniques (e.g. Bennell & Canter, 2002; Bennell & Jones,

---

2005; Goodwill & Alison, 2006; Green, Booth, & Biderman, 1976; Grubin et al., 2001; Santtila, Fritzon, & Tamelander, 2005; Santtila et al., 2008; Tonkin, Grant, & Bond, 2008; Woodhams, Grant, & Price, 2007; Woodhams & Toye, 2007), only one published study has examined how people perform on this task. Santtila, Korpela, and Häkkänen (2004) examined the ability of four distinct groups of individuals to accurately link vehicle offenses. They presented experienced vehicle offense investigators, experienced general investigators (i.e. investigators with no specialized training in vehicle crime investigation), novice general investigators, and naive participants with offense information relating to 30 offenses committed by 10 known offenders (three offenses each). The participants were asked to review the offense information and determine which of the offenses they believed were linked. The results showed that investigators, as a group, were significantly more accurate than naive participants, but there were no accuracy differences between the types of investigators (with each group correctly identifying about half of the possible links). However, experienced vehicle offense investigators focused their attention on a smaller sub-set of offense behaviors when making their linking decisions compared to all other groups.

Santtila et al.'s (2004) study suggests that there are theoretical and practical reasons to study how people perform linkage analysis. Theoretically, linkage analysis provides a reasonably well-defined task that allows researchers to gather knowledge about human discrimination processes across a range of varying conditions (e.g. levels of decision-maker expertise). Practically, understanding how people perform linkage analysis will determine if there is a need for training in this area (e.g. providing information about the most effective linking cues). If people continue to perform poorly after receiving appropriate training, then there is a case for the development and implementation of a decision aid.

The performance of decision makers on the linking task

Given the inherent ambiguity of offender behavior, and the sheer number of offense behaviors that can occur in any given offense, it is unclear how well people will perform when required to make linkage decisions. On the one hand, judgment and decision-making research has demonstrated that people have a limited ability to process information (e.g. Miller, 1963; Neath, 1998) and that they often rely on heuristics to deal with the complexity of the real world (Kahneman & Tversky, 1973). The majority of this research concludes that, because heuristic processing does not match the performance of the rational choice model, people are prone to exhibit a host of errors and biases, such as overconfidence, confirmation bias, insensitivity to sample size, illusory correlations, and so on (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996; Kahneman & Tversky, 1973). This body of 'heuristics and biases' research therefore suggests that people (particularly novice decision makers) will not be able to cope effectively with the complexities inherent in linkage analysis, and will perform poorly on the task as a result (see Payne, Bettman, & Johnson, 1993, for a discussion of how researchers have defined task complexity by considering the number of decision alternatives, the amount of time to process information, the number of attributes upon which a decision is based, etc.).

On the other hand, a number of more recent studies have shown that simple heuristics can, under certain conditions, lead to accurate decisions (Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & The ABC Research Group, 1999; Todd

& Gigerenzer, 2001). In particular, these studies have demonstrated that heuristics can be successful when the structural properties of a heuristic match the structure of discriminating information within the decision environment (i.e. ecologically rational heuristics; Martignon & Hoffrage, 1999). For example, if a police officer uses an 'error minimization' heuristic to predict the location of an at-large serial offender's home based on where his offenses were committed (e.g. by choosing a point located in the middle of his offenses), the likely result is an accurate prediction because this heuristic matches the empirical regularity that the majority of serial offenders reside within their area of criminal activity (Bennell, Snook, Taylor, Corey, & Keyton, 2007; Snook, Canter, & Bennell, 2002; Snook, Taylor, & Bennell, 2004; Taylor, Snook, & Bennell, 2009). This body of 'bounded rationality' research therefore suggests that people will perform well on the linking task if they use heuristics that match the structure of the decision environment.

Taken together, these studies highlight the importance of studying the cognitive strategies that people use when making linking decisions and determining whether those strategies match what we know about the information that is useful for making these decisions. Recently, Bennell and his colleagues (Bennell, 2002; Bennell & Canter, 2002; Bennell & Jones, 2005) have demonstrated that the distance between burglary locations (i.e. inter-crime distance) is the single most reliable indicator of whether or not burglaries are committed by the same offender, with shorter distances indicating a greater likelihood that the offenses are linked (see Tonkin et al., 2008, and Woodhams & Toye, 2007, for similar findings with other crime types). They also showed that three commonly used linking cues (entry method, target characteristics, and property stolen) were relatively ineffective for linkage analysis when applied to cases of burglary. Thus, we might reasonably predict that people are more likely to make accurate linking decisions if they know that burglars typically commit their crimes close together in space. In contrast, they are more likely to make bad decisions if they assume, for example, that the same burglar always enters their targets in the same way.

The kinds of information that people attend to in this decision-making task, and what role police experience plays in the selection of relevant linking cues and subsequent linking performance, remains unclear. Santtila et al.'s (2004) results suggest that investigative experience (although not necessarily experience related to the exact type of crime under consideration) plays an important role, allowing people to draw on previously acquired schemas to select, process, and implement relatively effective linking strategies, at least compared to naive participants. This accords well with research in other domains, which suggests that experience improves decision making (e.g. Corcoran, 2007; Lamond & Farnell, 1998). However, as highlighted by Santtila et al., there is also a body of research suggesting that experience may not play a central role in successful decision making (e.g. Dawes, 1989; Garb, 1989).

Training decision makers to consider relevant information

It is important to train people to make better decisions if they rely on irrelevant information. In many domains, novices are taught to improve their decisions by focusing on relevant information (e.g. Jacoby et al., 2001). Research suggests that it is also possible to train experts to consider relevant information, and thus increase their decision accuracy, especially in complex decision tasks (e.g. Gaeth & Shanteau, 1984). In the forensic domain, for example, research suggests that many professional

lie detectors rely on inappropriate verbal and non-verbal cues when attempting to detect deception, potentially explaining why these individuals perform so poorly (often no better than chance) on lie detection tasks (Vrij, 2000, 2008). However, providing training on how to use relevant cues can apparently reduce the effect of irrelevant information on lie detection accuracy. For example, when Porter, Wood-worth, and Birt (2000) provided training to Canadian parole officers, their decision accuracy increased significantly. Importantly, Porter et al. also showed that the training had a relatively long-term effect on an individual's ability to accurately detect lies (i.e. performance was maintained over a 5-week period).

In contrast to this research, other research has found that training may not result in substantial improvements to decision-making performance. In some instances, cognitive overload caused by the training may be partly to blame (Clark, Nguyen, & Sweller, 2005). In other cases, 'expert' trainees may exhibit resistance to training, preferring to rely on their past experiences (which can be biased) rather than the empirical evidence. This resistance can be attributed, at least in part, to the type of cognitive conceit discussed by Myers (2002), whereby people think they know more than they actually do. For example, Memon, Holley, Milne, Kohnken, and Bull (1994) examined whether training police officers on the cognitive interview, a technique known to increase the amount of correct information elicited from witnesses, would improve their interviewing performance. Although similarly trained students improve their performance (Milne & Bull, 1999), the police officers in Memon et al.'s study did not. The lack of improvement by the police officers was partially due to their reluctance to give up their old interviewing methods (cf. Dando, Wilcock, & Milne, 2008, 2009).

Developing statistical models to aid decision making

If training in an area is ineffective, it may be necessary to develop and implement decision aids (Dawes & Hastie, 2001; Swets, Dawes, & Monahan, 2000). The development of decision aids can also serve as a normative benchmark against which human performance can be measured. Researchers have developed statistical models in many domains where cognitive limitations have been shown to result in decision errors (Swets et al., 2000), including domains that are relevant to criminal justice professionals (e.g. Jones & Bennell, 2007; Rice & Harris, 1995; Yun, 2007). The primary question with respect to these decision aids is whether or not they provide any advantages over human judgments. Historically, statistical models have been considered the preferred method for a range of consequential decisions where the primary goal is to maximize decision accuracy. Meehl (1954), for example, reviewed a wide array of empirical research that compared the accuracy of human judges to statistical models and concluded that statistical models are superior to clinical judgments (see Grove & Meehl, 1996, for a more recent review that reaches a similar conclusion). This also appears to be the case in many criminal justice contexts (e.g. Hanson & Bussière, 1998; Szucko & Kleinmuntz, 1981; Walters, White, & Greene, 1988).

However, some previous research has also highlighted the possibility that statistical models may not always outperform human decision makers. Indeed, research has shown that individuals drawing on *appropriate* cognitive heuristics can sometimes make decisions that are as accurate as decisions based on statistical models (Bennell et al., 2007; Martignon & Schmitt, 1999). Paulsen (2006), for

example, compared the ability of students and a range of statistical and computational models to predict where serial offenders live on the basis of their crime site locations. He found that students who used appropriate (i.e. ecologically rational) heuristics performed as well as simple statistical models and more complex computationally-based procedures designed for the purpose of making geographic profiling predictions.

The current study

In the current study, we compare university students, police professionals, and a logistic regression model on their ability to accurately link serial burglaries. Based on a review of the literature, we expected that individuals who possess police experience would achieve significantly higher levels of decision accuracy on the linking task compared to individuals with no police experience given that police professionals are more likely to have knowledge of relevant linking cues. We also expected that, regardless of their level of police experience, providing human judges with information about relevant linking cues would significantly increase their decision accuracy on the linking task. However, we expected that the level of decision accuracy achieved by human judges would be significantly lower than that achieved by a statistical model designed for the purpose of linking serial burglaries, even after the judges receive training.

Method

Participants

Participants ($n = 71$) were 40 undergraduate students who took part in the study for course credit at Carleton University and 31 police professionals who were recruited primarily through postings on police-related newsgroups.[1] In this study, a police professional was defined as anyone having experience in crime analysis, police work, or a related field (e.g. police psychology). Approximately half of the participants in each group ($n = 20$ students and $n = 14$ professionals) were given training on how to conduct the linking task (discussed in more detail below), while the remaining participants in each group ($n = 20$ students and $n = 17$ professionals) were not trained.

The mean age of the student group was 21.1 years (SD = 3.2), and there were no significant age differences between the trained ($M = 20.7$ years, SD = 3.9) and untrained ($M = 20.6$ years, SD = 2.3) groups. With regards to gender, the student sample contained a total of 18 men and 22 women, and there were no significant gender differences between the trained (men = 9, women = 11) and untrained (men = 9, women = 11) groups. According to the demographic information provided by each student, none of them had any prior experience in crime analysis, police work, or linkage analysis.

Twenty-nine police professionals provided demographic information. Of these 29 participants, there were 14 crime analysts, nine police officers, and six other professionals (e.g. police psychologists). Fifteen participants were recruited from the USA, eight from the UK, and six from Canada. The mean age of the professional group was 38.1 years (SD = 13.1), and there was a significant age difference between the trained ($M = 45.5$ years, SD = 12. 8) and untrained ($M = 32.7$ years, SD = 10.1)

groups, $t_{(27)} = 3.06$, $p < 0.01$. The professional sample consisted of 13 men and 16 women, and there was a significant gender difference between the trained (men = 10, women = 4) and untrained (men = 3, women = 12) groups, $\chi^2 = 8.57$, d.f. = 28, $p < 0.01$. There was no significant difference between the trained ($M = 2.3$ years, SD = 5.4) and untrained ($M = 3.5$ years, SD = 3.6) groups in terms of their crime analysis experience, nor was there a significant difference between the trained ($M = 8.2$ years, SD = 4.3) and untrained ($M = 9.3$ years, SD = 4.6) groups with respect to their police experience. Finally, there was no significant difference between the trained ($M = 2.2$ years, SD = 5.2) and untrained ($M = 1.7$ years, SD = 3.1) groups with respect to previous linkage analysis experience.

Materials

All participants were required to complete an experimental booklet, which consisted of an informed consent form and a package containing 38 pairs of commercial burglaries. Each burglary was committed in a large urban city in the UK between January 1999 and January 2000, and was randomly extracted from a larger database of burglaries collected from the area as part of a previous research project (Bennell & Canter, 2002). Approximately 20% of these burglary pairs (8 out of 38) represented linked pairs, which is consistent with the estimated proportion of linked pairs that is encountered when working with large numbers of commercial burglaries in the area where these burglaries were committed. Two versions of this package were constructed, with the 38 pairs of burglaries presented in a different random order in each package to control for order effects (no order effect was found and, therefore, order of presentation is not considered in subsequent analyses).

For each of the offense pairs, information relating to the following behaviors was provided: (1) a map (size = 17.8 cm × 24.1 cm) consisting of all 76 offense locations, with the relevant offense pair highlighted, (2) the straight-line distance (km) between the offenses, (3) how the offender entered the buildings, (4) the characteristics of the buildings that were targeted in each offense, and (5) the property that was stolen from the buildings. The information was provided in such a way that all information for a particular offense pair could be viewed at the same time. This information was extracted from a burglary database maintained by the police force where these offenses had occurred as part of a previous study (Bennell & Canter, 2002). The goal of that study was to develop a logistic regression model for identifying relevant linking cues (discussed in more detail below).

Procedure

Human performance

Students were tested in a research laboratory, in groups consisting of one, two, or three participants. The students were asked to work individually through the experimental booklet at their own pace and were unsupervised throughout the duration of the study to reduce any potential demand characteristics. Completion of all tasks in the booklet took approximately 45 minutes. Police professionals received the experimental booklet in the mail and returned it to the first author upon completion.[2]

Before participants examined the 38 offense pairs contained in the booklet, they were each provided with an information sheet. This sheet informed all participants that 20% of the offense pairs that they would be examining were linked offense pairs. Providing base rate information to participants is not uncommon in studies like the current one (e.g. Santtila et al., 2004; Swets et al., 1991). This sort of information is also sometimes available in naturalistic settings, although the exact base rate typically varies across jurisdictions.[3] In addition, participants in the trained group were provided with instructions on how to decide whether or not offense pairs were linked (untrained participants received no additional information). Specifically, these instructions indicated that: 'When determining whether commercial burglars have been committed by the same offender, previous research has indicated that the closer two offenses are to one another geographically, the more likely it is that the same offender committed them'. These instructions were based on past research, which has shown that the most reliable indicator of whether or not burglaries are linked is inter-crime distance (Bennell, 2002; Bennell & Canter, 2002; Bennell & Jones, 2005).

For each offense pair, participants were asked to indicate, on a 10-point scale (1 = 'not at all confident that the offenses are linked', 10 = 'extremely confident that the offenses are linked'), how confident they were that the same offender committed the pair of offenses. They were also asked to indicate, for each offense pair, the extent to which they based their decision on each of the provided cues, again on a 10-point scale (1 = 'not at all' to 10 = 'very much').[4] After completing the package of 38 offense pairs, the participants were asked to fill out a demographic questionnaire and to provide comments about the study if they wished. The participants were then debriefed, with the police professionals sent a debriefing form after completing the study.

Logistic regression model

To examine the relative accuracy of a statistical model, the logistic regression model developed by Bennell and Canter (2002) was applied to each of the 38 offense pairs (no alterations were made to the original regression model). The offense pairs used in the current study were not part of the sample of offenses that was used to construct Bennell and Canter's regression model, but rather they comprised the validation sample. Thus, the results of the regression model used in the current study will not be artificially inflated, as they would have been if offenses from the original construction sample had been used.

The model used in this research, which is discussed more thoroughly by Bennell and Canter (2002), was based solely on inter-crime distances, and took the form:

$$\log\left(\frac{p}{1-p}\right) = -2.82 - 0.88(\text{inter-crime distance})$$

where p is the probability of an offense pair being linked. The probabilities that resulted when this model was applied to each of the offense pairs in the experimental booklet are the equivalent of participant ratings. To establish differences in decision accuracy across the statistical model and our participants, these probabilities were compared to participant ratings using the procedure outlined below.

Measuring decision-making accuracy

Decision accuracy was measured using receiver operating characteristic (ROC) analysis. ROC analysis is a procedure that has been adopted in a variety of medical and industrial settings to evaluate the accuracy of diagnostic decisions (Swets, 1996). Regardless of whether ROC analysis is used to examine decisions made by people or statistical models, the general procedure for conducting the analysis is the same. Hit probabilities (pH) and false alarm probabilities (pFA) are calculated for various decision thresholds set along a rating scale, such as the 10-point scale used by our participants. These values are then plotted on a graph, with pH on the vertical axis and pFA on the horizontal axis. The result of plotting these values and connecting the various points is a concave-downward curve known as a ROC curve.

Each point along a ROC curve represents the pH/pFA ratio that occurs when using a specific decision threshold. The height of a ROC curve represents the overall level of discrimination accuracy achieved by the decision-maker (i.e. higher ROC curves are characterized by greater pH/pFA ratios) (Swets, 1996). Height in this case is measured by calculating the proportion of the graph's area falling under the ROC curve, which is referred to as the area under the curve (AUC).[5] Values for the AUC typically range from 0.50 to 1.00, where 0.50 corresponds to a ROC curve falling along the positive diagonal, indicating chance accuracy, and 1.00 corresponds to a ROC curve falling along the left and upper axes of the graph, indicating perfect accuracy. According to criteria proposed by Swets (1988), AUCs between 0.50 and 0.70 indicate low accuracy, AUCs between 0.70 and 0.90 indicate moderate accuracy, and AUCs between 0.90 and 1.00 indicate high accuracy. The advantage of the AUC over other measures of accuracy (e.g. percentage correct) is that it is not biased by where the decision threshold is placed (Swets, 1996). This is because the AUC corresponds to the position of the entire ROC curve rather than any single ROC point (Swets et al., 2000).

The ROC analysis sub-routine in SPSS (v. 17) was used to conduct all ROC analyses. This program allows users to calculate a variety of standard ROC measures, including the AUC, the standard error associated with the AUC, and 95% confidence intervals (CI$_{95}$) around the AUC. In this study, ROC curves were derived from pooling the results from participants within each group and running the analysis across pooled ROC curves to determine if one group exhibited significantly greater accuracy than the other groups (Swets & Pickett, 1982).

Results

Decision-making accuracy

As can be seen in Figure 1, the AUCs for the ROC curves associated with untrained and trained students are 0.70 (SE $=0.03$; CI$_{95}$ $=0.65$–0.76; d' $=0.75$) and 0.79 (SE $=0.02$; CI$_{95}$ $=0.74$–0.84; d' $=1.15$), respectively. The AUCs for the ROC curves associated with untrained and trained professionals are 0.64 (SE $=0.03$; CI$_{95}$ $=0.59$–0.70; d' $=0.51$) and 0.71 (SE $=0.03$; CI$_{95}$ $=0.66$–0.76; d' $=0.79$), respectively. The ROC curve for the logistic regression model represents an AUC of 0.87 (SE $=0.02$; CI$_{95}$ $=0.82$–0.91; d' $=1.60$). All AUCs in Figure 1 are significantly above chance (all ps <0.001).
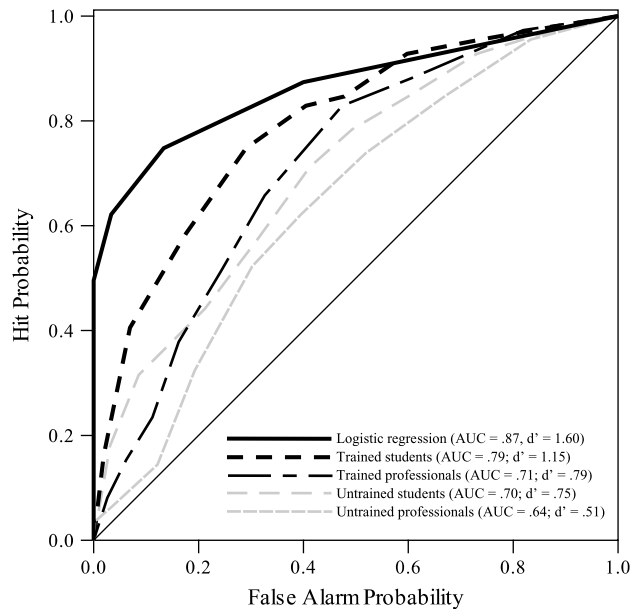
Figure 1. ROC graph of accuracy scores for students, professionals, and the logistic regression model.

To understand the influence of expertise and training on decision accuracy, a 2 (Expertise: student and professional) $\times 2$ (Training: untrained and trained) ANOVA was conducted with linking accuracy scores (i.e. AUCs) as the dependent variable. As was expected, we found a significant main effect for Expertise, $F(1,67) = 9.09$, $p < 0.01$, $\eta^2 = 0.12$, but in contrast to what we hypothesized, students ($M = 0.76$, SD $= 0.14$) outperformed professionals ($M = 0.67$, SD $= 0.13$).[6] In line with our expectations, a significant main effect was also found for Training, $F(1,67) = 9.71$, $p < 0.01$, $\eta^2 = 0.13$, with those who received training ($M = 0.77$, SD $= 0.13$) outperforming those who did not ($M = 0.67$, SD $= 0.13$). The two-way interaction between Expertise and Training was not significant, indicating that training had the desired effect of increasing linking accuracy regardless of the degree to which the judges possessed police experience.

To compare the performance of the human judges to the statistical model, a one-sample t-test was conducted ($\alpha = 0.05$). As expected, the accuracy of the human judges ($M = 0.72$, SD $= 0.14$) was significantly lower than the statistical model ($M = 0.87$), $t_{(70)} = 8.96$, $p < 0.001$, $d = 1.07$. This was also found to be the case for each of the four separate groups of human judges when their accuracy scores were compared to the statistical model (all ts > 3.02).

Reliance on behavioral cues

In order to gain a better understanding of the accuracy results, the reliance ratings provided by the human judges were examined. Figure 2 shows the average reliance scores for each piece of information for the untrained and trained students and professionals.
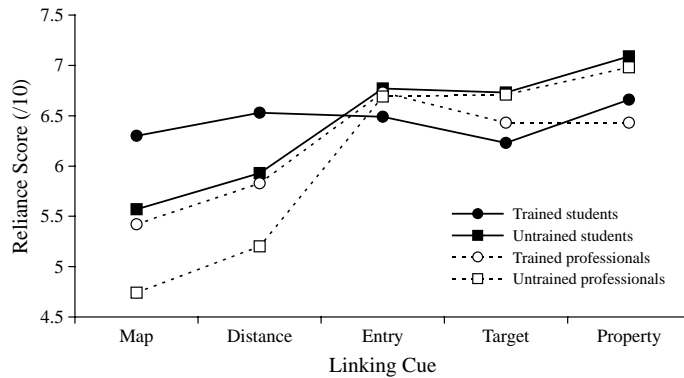
Figure 2.　Student and professional reliance ratings as a function of training and the five different linking cues.

To understand the influence of expertise and training on the reliance scores, a 5 (Cue: map, distance, target, entry, and property) $\times 2$ (Expertise: student and professional) $\times 2$ (Training: untrained and trained) mixed ANOVA was conducted with reliance scores as the dependent variable. We found no significant main effects of Expertise or Training, but a significant main effect of Cue was found, $F(4,268) = 25.78$, $p < 0.001$, $\eta^2 = 0.28$. This main effect was subsumed by a significant two-way interaction between Expertise and Cue, $F(4,268) = 4.16$, $p < 0.01$, $\eta^2 = 0.06$, and between Training and Cue, $F(4,268) = 6.63$, $p < 0.001$, $\eta^2 = 0.09$. There was no two-way interaction between Expertise and Training, and no three-way interaction between Expertise, Training, and Cue.

Use of behavioral cues

To assess whether the participants placed significantly more emphasis on one behavioral cue than another, we performed a series of paired-samples t-tests (Bonferroni corrected $\alpha = 0.005$). These tests indicated that participants relied on inter-crime distance significantly more than all other pieces of information (all $ts >$ 3.02). Our participants also relied on the map information significantly more than all other pieces of information (all $ts > 4.37$), with the exception of inter-crime distance. All other comparisons were not significant.

Expertise and behavioral cues

A series of independent-samples t-tests were then used to compare reliance ratings across students and professionals for each of the five behavioral cues (Bonferroni corrected $\alpha = 0.01$). There was a significant difference in the use of maps, with the students ($M = 5.9$, SD $= 1.4$) relying on them more than the professionals ($M = 5.0$, SD $= 1.6$), $t_{(69)} = 2.52$, $p < 0.01$, $d = 0.61$. There was also a strong, albeit non-significant trend found indicating that students placed more weight on inter-crime distance ($M = 6.2$, SD $= 1.2$) compared to the professionals ($M = 5.5$, SD $= 1.4$), $t_{(69)} = 2.44$, $p < 0.05$, $d = 0.59$. All other comparisons were not significant.

Training and behavioral cues

A series of independent-samples *t*-tests were also used to compare reliance ratings across untrained and trained participants for each of the five behavioral cues (Bonferroni corrected $\alpha = 0.01$). No significant differences were found. However, there were strong trends found indicating that the trained individuals placed more weight on the maps ($M = 5.9$, SD $= 1.2$) compared to untrained individuals ($M = 5.2$, SD $= 1.7$), $t_{(69)} = 2.12$, $p < 0.05$, $d = 0.51$, as well as more weight on inter-crime distances ($M = 6.2$, SD $= 1.2$) compared to untrained individuals ($M = 5.6$, SD $= 1.4$), $t_{(69)} = 2.12$, $p < 0.05$, $d = 0.51$. All other comparisons were not significant.

Discussion

This study examined the ability of university students, police professionals, and a logistic regression model to infer from behavioral information whether or not the same offender committed two burglaries. Overall, the results showed that: (1) students outperformed professionals with respect to decision accuracy, (2) providing information to participants about relevant linking cues resulted in higher levels of decision accuracy, and (3) the logistic regression model significantly outperformed all participants. These findings have implications for our understanding of how people perform linkage analysis and whether or not there is a need for training and/or decision aids to assist with this task.

The role of police experience

The finding that both untrained students (AUC $= 0.70$) and professionals (AUC $= 0.64$) performed significantly above chance levels on the current linking task was surprising. None of the students that took part in this study reported any experience with policing, crime analysis, or linkage analysis, and previous research has demonstrated that police professionals perform relatively poorly on the linking task (Santtila et al., 2004). Therefore, it would not have been unreasonable to assume that participants in this study would have performed at around chance levels under conditions where they were provided with no training.

Given their above chance performance, it is tempting to conclude that the untrained participants were aware of the fact that an offender's spatial behavior could be used for linking purposes. However, while the reliance scores provided by these participants indicate that they focused *some* of their attention on the relative location of the burglaries and the distances between them, they focused significantly more attention on the other less relevant cues (e.g. property stolen). While this strategy is not ideal, it does ensure moderate levels of accuracy since effective linking cues are focused on in addition to the irrelevant cues.

The finding that untrained students generally outperformed the untrained professionals was also somewhat surprising, especially in light of Santtila et al.'s (2004) results, which showed that police investigators significantly outperformed naive participants in a linking task involving vehicle crimes. Of course, it could simply be the case that experience is unrelated to success on the linking task and that our results are correct. This has certainly been found to be the case in other domains (e.g. Dawes, 1989; Garb, 1989). However, there are also potential problems with our

study that provide alternative explanations for why these differences between the two studies emerged. These problems may have negatively impacted participant performance, especially in the case of our police professionals.

First, unlike the data provided to investigators in Santtila et al.'s (2004) study, the participants in the current study were provided with data from an area that was potentially very different from where most of them resided and/or worked.[7] While this issue is unlikely to seriously impact student performance (because students know little of how burglars behave), it could impact the performance of police professionals who may be using *locally derived* analytical knowledge that does not match the criminal patterns found within the study's data. For example, while a police officer serving in a rural Canadian town may know that inter-crime distances in his jurisdiction tend to be about 8 km, this will not serve him well when applying that knowledge to crimes committed in a densely populated urban center in the UK where the average inter-crime distance is 2 km (as it is with the data set being used in this study). A number of the police professionals in the current study noted that knowledge of the area in which the offenses were committed would have increased their performance and it will be important to examine this issue in the future.

Second, providing participants (especially the police professionals) with additional information about the crime pairs could have increased their performance. The absence of temporal information (i.e. when the crimes were committed) was a particular concern, with over 50% of the professionals, but none of the students, indicating that having access to temporal information would have allowed them to make more accurate decisions. For example, one professional commented, 'To me, a crucial omission in your variables is the date and time of each burglary. These details would be vital in determining a particular burglar's mobility – could he rob two places 5 km apart on the same day?'. Unfortunately, temporal information was not available for the current study and often is not in cases of burglary due to the fact that burglaries often occur when homes are unoccupied (Ratcliffe, 2002). However, when it is available, temporal information can be useful for making linking decisions, as indicated by the fact that the use of timing was highly related to increased accuracy in the study by Santtila et al. (2004). Thus, the need to evaluate the importance of temporal information in linkage analysis is a valid concern and something that future research should address.

The role of training

While all participants focused on irrelevant cues when making linking decisions, the current study did find that training had an effect on linking accuracy, with both trained students and trained professionals outperforming untrained students and untrained professionals, respectively. Again, the reliance scores indicate that this general increase in performance was a result of trained participants incorporating, although not to a significant degree, the instructions on how to best go about deciding whether or not two offenses are linked. In this way, the results are similar to those reported by other researchers (e.g. Gaeth & Shanteau, 1984; Porter et al., 2000).

While training improved performance for all groups of participants, it had a particularly strong influence on the students. This finding is, in large part, a result of the trained students placing more reliance on relevant cues than their counterparts

(especially police officers). For example, the students placed more emphasis on both the maps and distance information compared to the professionals once they had received training. Comments from the police professionals indicate that this may have happened because these individuals often believed their experience and knowledge should be given more weight than the training. For example, a number of the trained professionals stated that they disagreed that it was fruitful to use inter-crime distance to link offenses. One participant stated that, '... priority was given to targets followed by the MO for entry ... in my experience criminals committing multiple offenses usually follow this sequence'. Another remarked, 'I discounted distance in the main as burglars are very mobile and I concentrated on MO and what was stolen to detect possible correlations'. Empirical research suggests that these assumptions are incorrect for the majority of serial burglary cases (Snook, 2004; Wiles & Costello, 2000), and they are certainly incorrect for the burglaries in the current data set (Bennell & Canter, 2002). Students in our study did not indicate such resistance, presumably because they lack prior experience with this type of task and do not have preconceived ideas about cues that are (supposedly) effective for making linking decisions.

The value of a statistical model

With an accuracy score of 0.87, the logistic regression model developed by Bennell and Canter (2002) significantly outperformed all groups of human judges, even after training. This suggests that it might be useful for the police to adopt a simple statistical approach for linkage analysis. This result, and conclusion, is consistent with the majority of research that has compared statistical and clinical approaches to decision making (e.g. Meehl, 1954; Grove & Meehl, 1996). There are many possible explanations for the superiority of the regression model in this study (Dawes et al., 1989). Perhaps the most logical one is that the model only contains the significant spatial predictor, while the participants were free to use any linking cue (and the participants clearly did not rely solely on the spatial information).

In addition, it is clear from the results that participants were either not able to, or decided not to, incorporate the base rate information into their decision process. Recall that the initial instructions provided to all participants included a statement that approximately 20% of the offense pairs were linked. However, it was not unusual for participants to indicate that as many as 50% of the offense pairs were linked. This occurred even for participants in the trained conditions, despite the fact that, after training, they had access to a very effective linking strategy (i.e. they could have separated the 20% of offense pairs with the shortest inter-crime distance from the rest of the pairs and declare these to be the linked offenses).

Limitations of the current study

Despite the potential importance of the current findings there are at least three limitations within the current study that need to be considered when interpreting the results (beyond those that have already been highlighted and the obvious problem of sample size).

First, there are potential confounds in the current study between the training received by the police professionals and other demographic variables. Specifically,

the trained group of professionals was significantly older than the untrained group and was made up of significantly more men and less women. While there is no obvious reason we can think of to suspect that age or gender would be related to performance on the linking task, it is difficult to know without further study what influence these variables had in the current study and how they impacted the training that was provided. Importantly, other demographic variables, which are arguably more important (e.g. previous crime analysis and linking experience), did not differ across the trained and untrained professional groups.

Second, it is unclear what effect the provision of base rate information (i.e. that 20% of the crime pairs were linked) had on the performance of the participants in the current study. It seems unlikely that this information favored one group over any other. However, this information could have had: (1) a performance enhancing impact on all of the groups, allowing them to perform better than they would have without this information, or (2) no impact at all (see Tversky & Kahneman, 1982, for a discussion of base rate neglect). While providing such information is not uncommon in research of this type (e.g. Santtila et al., 2004; Swets et al., 1991), base rate information will not always be known in real-world situations where linkage analysis is conducted. Therefore, it will be important in the future to examine how this information impacts linking performance. It will also be important to examine the impact of providing base rate information in different ways (e.g. as a frequency instead of a percentage) given that the format of presentation is potentially important (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995).

Third, one must consider the relatively impoverished nature of the training that was provided to participants in the current study and how this might have influenced their performance. Indeed, by modifying the training, the performance of participants could be enhanced, perhaps to a point where they could make linking decisions that are as accurate as the statistical model. We are currently examining four revisions to the training to determine whether performance can be further improved: (1) offering the training through a more authoritative source (e.g. an experienced police officer), (2) explicitly informing participants that they should ignore irrelevant cues (e.g. property stolen) while focusing on relevant cues, (3) providing participants with a specific decision threshold to use when considering inter-crime distances (e.g. 'link the offense pair if the distance between the two offense locations is less than 2 km'), and (4) presenting feedback following each linking decision to inform participants whether their decision was correct or not.

Conclusion

The results of this study may have implications for the future of linkage analysis. Although this procedure is vital to the success of serial crime investigation, it would seem that the methods currently used by police professionals are open to improvement. Specifically, this study demonstrated a worthwhile improvement in linking accuracy following brief training in the use of relevant linking cues. However, the results also support previous research from other domains in that humans appear unable to perform as well as a statistical model developed specifically for the purpose of linkage analysis, at least under the current testing conditions. This suggests that it might be useful for the police to adopt a statistical approach for linkage analysis.

Notes

1. It is difficult to provide an accurate response rate for the professional group given that it is impossible to know how many professionals were reached through our various recruitment postings. However, we can report that 51 packages were sent to police professionals who indicated an interest in taking part in the study. Given that 31 professionals sent back completed packages, the return rate was 60.8%.
2. The fact that a different procedure was used to collect data from the students and professionals could have introduced a number of confounding variables, most of which we believe would favor the police professionals' performances on the linking task. For example, the professionals could have taken more time to complete their packages than the students, and they could have relied more on external sources for assistance (e.g. crime analysis packages). The results of this study, however, do not support this view. As will be discussed in the results section, the students generally outperformed the professionals with respect to linking accuracy.
3. None of the participants in this study were from the jurisdiction where the burglary data originated and, consequently, these participants would not be aware of the base rate unless told about it explicitly.
4. The potential for problems when using self-report measures to determine cue reliance should be noted. There is a relatively large literature, which suggests that people may not have access to their mental processes and, even if they do, they may not be able to articulate anything about those processes (e.g. Nisbett & Wilson, 1977). Thus, the results related to our reliance scores should be treated with an appropriate level of caution.
5. Given that some readers may be more familiar with d' than the AUC, these values will also be provided. As with the AUC, a higher d' value indicates greater linking accuracy.
6. In an attempt to understand why the professionals did not perform as well as the students, in either the untrained or trained condition, the accuracy scores of the various sub-groups comprising the professional group were scrutinized with respect to their background (policing, crime analysis, or other) and location (USA, UK, or Canada). While the small number of judges in each sub-group prevented us from conducting formal significance tests, the accuracy scores for the professionals appear to be low due largely to the relatively poor performance of police officers, particularly in the trained condition. No obvious differences in accuracy scores were found across professionals from different countries, in either the untrained or trained condition.
7. Note that Santtila et al. (2004) also used data from a location that was different from the location where their participants resided and/or worked, but both locations were major cities in Finland (P. Santtila, personal communication, 11 March 2009). It is likely that these cities are much more similar than the various locations relied on in the current study, which consisted of towns and cities from three different countries, with different population densities, road networks, land-use patterns, etc.

References

Bennell, C. (2002). *Behavioral consistency and discrimination in serial burglary.* Unpublished doctoral thesis, University of Liverpool, Liverpool, UK.

Bennell, C., & Canter, D.V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science and Justice, 42,* 153–164.

Bennell, C., & Jones, N.J. (2005). Between a ROC and a hard place: A method for linking serial burglaries by modus operandi. *Journal of Investigative Psychology and Offender Profiling, 2,* 23–41.

Bennell, C., Snook, B., Taylor, P.J., Corey, S., & Keyton, J. (2007). It's no riddle, choose the middle: The effect of number of crimes and topographical detail on police officer predictions of serial burglars' home locations. *Criminal Justice and Behavior, 34,* 119–132.

Clark, R., Nguyen, F., & Sweller, J. (2005). *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* San Francisco, CA: Pfeiffer.

Corcoran, S.A. (2007). Planning by expert and novice nurses in cases of varying complexity. *Research in Nursing & Health, 9,* 155–162.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition, 58*, 1–73.

Dando, C., Wilcock, R., & Milne, R. (2008). Victims and witnesses of crime: Police officers' perceptions of interviewing practices. *Legal and Criminological Psychology, 13*, 59–70.

Dando, C., Wilcock, R., & Milne, R. (2009). The cognitive interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology, 23*, 138–147.

Dawes, R.M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences and the Law, 7*, 457–467.

Dawes, R.M., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgments. *Science, 243*, 1668–1674.

Dawes, R.M., & Hastie, R. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage.

Gaeth, G.J., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experienced decision-makers. *Organizational Behavior and Human Performance, 33*, 263–282.

Garb, H.N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

Gigerenzer, G., Todd, P., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goodwill, A.M., & Alison, L.J. (2006). The development of a filter model for prioritising suspects in burglary offences. *Psychology, Crime & Law, 12*, 395–416.

Green, E., Booth, C., & Biderman, M. (1976). Cluster analysis of burglary M/Os. *Journal of Police Science and Administration, 4*, 382–388.

Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323.

Grubin, D., Kelly, P., & Brunsdon, C. (2001). *Linking serious sexual assaults through behavior*. London: Home Office.

Hanson, R.K., & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348–362.

Hazelwood, R., & Warren, J. (2003). Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior, 8*, 587–598.

Jacoby, J., Morin, M., Johar, G., Gurhan, Z., Kuss, A., & Mazursky, D. (2001). Training novice investors to become more expert: The role of information accessing strategy. *Journal of Psychology and Financial Markets, 2*, 69–79.

Jones, N.J., & Bennell, C. (2007). The development and validation of statistical prediction rules for discriminating between genuine and simulated suicide notes. *Archives of Suicide Research, 11*, 219–233.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Lamond, D., & Farnell, S. (1998). The treatment of pressure sores: A comparison of novice and expert nurses' knowledge, information use and decision accuracy. *Journal of Advanced Nursing, 27*, 280–286.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.

Martignon, L., & Schmitt, M. (1999). Simplicity and robustness of fast a frugal heuristics. *Minds and Machines, 9*, 565–593.

Meehl, P.E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Memon, A., Holley, A., Milne, R., Kohnken, G., & Bull, R. (1994). Towards understanding the effects of interviewer training in evaluating the cognitive interview. *Applied Cognitive Psychology*, *8*, 641–659.

Miller, G. (1963). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*. Chichester, UK: Wiley.

Myers, D.G. (2002). *Intuition: Its powers and perils*. New Haven, CT: Yale University Press.

Neath, I. (1998). *Human memory: An introduction to research, data, and theory*. Pacific Grove, CA: Brooks/Cole.

Nisbett, R.E., & Wilson, D.T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Paulsen, D. (2006). Human versus machine: A comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling*, *3*, 77–89.

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge Press.

Porter, S., Woodworth, M., & Birt, A.R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, *24*, 643–658.

Ratcliffe, J. (2002). Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology*, *18*, 23–43.

Rice, M.E., & Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*, 737–748.

Santtila, P., Fritzon, K., & Tamelander, A.L. (2005). Linking arson incidents on the basis of crime scene behavior. *Journal of Police and Criminal Psychology*, *19*, 1–16.

Santtila, P., Korpela, S., & Häkkänen, H. (2004). Expertise and decision making in the linking of car crime series. *Psychology, Crime & Law*, *10*, 97–112.

Santtila, P., Pakkanen, T., Zappala, A., Bosco, D., Valkama, M., & Mokros, A. (2008). Behavioural crime linking in serial homicide. *Psychology, Crime & Law*, *14*, 245–265.

Snook, B. (2004). Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling*, *1*, 53–66.

Snook, B., Canter, D.V., & Bennell, C. (2002). Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system. *Behavioral Sciences and the Law*, *20*, 109–118.

Snook, B., Taylor, P.J., & Bennell, C. (2004). Geographic profiling: The fast, frugal, and accurate way. *Applied Cognitive Psychology*, *18*, 105–121.

Szucko, J.J., & Kleinmuntz, B. (1981). Statistical versus clinical lie detection. *American Psychologist*, *36*, 488–496.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.

Swets, J.A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Mahwah, NJ: Lawrence Erlbaum Associates.

Swets, J.A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*, 1–26.

Swets, J.A., & Pickett, R.M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.

Swets, J.A., Getty, D.J., Pickett, R.M., D'Orsi, C.J., Seltzer, S.E., & McNeil, B.J. (1991). Enhancing and evaluating diagnostic accuracy. *Medical Decision Making*, *11*, 9–18.

Taylor, P.J., Snook, B., & Bennell, C. (2009). The bounds of cognitive heuristic performance on the geographic profiling task. *Applied Cognitive Psychology*, *23*, 410–430.

Todd, P.M., & Gigerenzer, G. (2001). Putting natural decision making into the adaptive toolbox. *Journal of Behavioral Decision Making*, *14*, 381–383.

Tonkin, M., Grant, T.D., & Bond, J.W. (2008). To link or not to link: A test of the case linkage principles using serial car theft data. *Journal of Investigative Psychology and Offender Profiling*, *5*, 59–77.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge, UK: Cambridge University Press.

Yun, M. (2007). Hostage taking and kidnapping in terrorism: Predicting the fate of a hostage. *Professional Issues in Criminal Justice: A Professional Journal*, *2*, 23–40.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, UK: Wiley.

Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection: A plea to change police. *Criminal Justice and Behavior*, *35*, 1323–1336.

Walters, G.D., White, T.W., & Greene, R.L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, *56*, 111–117.

Wiles, P., & Costello, A. (2000). *The 'road to nowhere:' Evidence for travelling criminals*. London: Home Office.

Woodhams, J., Grant, T.D., & Price, A.R.G. (2007). From marine ecology to crime analysis: Improving the detection of serial sexual offences using a taxonomic similarity measure. *Journal of Investigative Psychology and Offender Profiling*, *4*, 17–27.

Woodhams, J., Hollin, C., & Bull, R. (2007). The psychology of linking crimes: A review of the evidence. *Legal and Criminological Psychology*, *12*, 233–249.

Woodhams, J., & Toye, K. (2007). An empirical test of the assumptions of case linkage and offender profiling with serial commercial burglars. *Psychology, Public Policy, and Law*, *13*, 59–85.