Psychology, Crime & Law
Vol. 17, No. 4, May 2011, 359–380

Another look at across-crime similarity coefficients for use in behavioural linkage analysis: an attempt to replicate Woodhams, Grant, and Price (2007)

Tamara Melnyk[a]*, Craig Bennell[a], Donna J. Gauthier[a] and Donald Gauthier[b]

[a]*Department of Psychology, Carleton University, Ottawa, Ontario, Canada;* [b]*Mechanical and Aerospace Engineering, Carleton University, Ottawa, Ontario, Canada*

(Received 4 June 2009; final version received 18 August 2009)

In the absence of physical evidence, investigators must often rely on offence behaviours when determining whether several crimes are linked to a common offender. A variety of factors can potentially influence the degree to which accurate linking is possible, including the similarity coefficient used to assess across-crime similarity. The current study examines the performance of two similarity coefficients that have recently been compared to one another, Jaccard's coefficient (J) and the taxonomic similarity index (Δ_s), using samples of two crime types, serial homicide ($N = 237$) and serial burglary ($N = 210$). In contrast to previous research, the results indicate that Δ_s does not significantly outperform J with respect to linking accuracy. In addition, both coefficients lead to higher levels of linking accuracy in cases of serial homicide compared to serial burglary. Potential explanations for these findings are presented and their implications are discussed.

Keywords: linkage analysis; taxonomic similarity index; Jaccard's coefficient; ROC analysis; serial homicide; serial burglary

Introduction

Police investigators must frequently determine whether a series of unsolved crimes has been committed by the same offender (Grubin, Kelly, & Brunsdon, 2001). In the absence of physical evidence, such as DNA, links between crimes must often be established through an analysis of behavioural evidence (Woodhams, Hollin, & Bull, 2007). Using an investigative technique known as behavioural linkage analysis (BLA), an attempt is made to identify behavioural patterns across crime scenes to determine if the same offender is responsible for all the crimes.

Researchers typically focus on two assumptions when considering whether it is possible to use crime scene behaviours to link crimes (Canter, 1995). First, it is assumed that offenders must exhibit relatively high levels of *behavioural stability* across their respective crime series, reflecting the degree to which each individual manifests the same behaviours across his or her crimes. Second, it is assumed that offenders must exhibit relatively high levels of *behavioural distinctiveness* for linking to be possible, whereby the actions that a given serial offender exhibits across his or her crimes differ from those exhibited by other offenders. The extent to which these

*Corresponding author. Email: tmelnyk@connect.carleton.ca

ISSN 1068-316X print/ISSN 1477-2744

DOI: 10.1080/10683160903273188
http://www.informaworld.com

two assumptions are valid is thought to determine the degree to which it is possible to discriminate between crimes committed by different offenders, a process we call *behavioural discrimination*.

Woodhams, Hollin et al. (2007) recently conducted a comprehensive review of the empirical literature dealing with BLA. A substantial amount of empirical evidence for the linking assumptions was provided. A variety of factors were also discussed that have the potential to influence the degree of stability and distinctiveness that exists in the behaviours of serial offenders (and therefore the degree of behavioural discrimination that is possible). Included amongst these factors were the type of crime scene behaviour under examination, the experience level of the criminal, and the period over which crimes have been committed. The current study is concerned with a factor that was not mentioned in this previous review, but has received attention recently – the type of similarity coefficient used to measure across-crime similarity.

The similarity coefficient is the basis for many approaches to BLA, as these coefficients are used to quantify the degree of similarity that exists between crimes. The degree of similarity is usually expressed as a value between 0 (no similarity) and 1 (total similarity), with higher scores expected across crimes committed by the same offender (i.e. indicating behavioural stability) and lower scores expected across crimes committed by different offenders (i.e. indicating behavioural distinctiveness). Research conducted in other contexts demonstrates that the use of different similarity coefficients influences the results on various discrimination tasks (e.g. Baroni-Urbani & Buser, 1976; Gower & Legendre, 1986; Kosman & Leonard, 2005). There is no reason to believe that this will not also be true with respect to BLA.

Recently, several individuals have begun to discuss the possibility that the choice of similarity coefficient will influence discrimination accuracy in BLA (e.g. Bennell, Jones, & Melnyk, 2007; Gauthier, 2008; Woodhams, Grant, & Price, 2007). Specifically, questions have been raised about the suitability of using Jaccard's coefficient, J, for measuring across-crime similarity. This is the most commonly used similarity measure in the context of BLA at the moment (Woodhams, Hollin et al., 2007), despite the wide variety of measures that could be used (e.g. see Liebetrau, 1983). Part of the appeal of J is its simplicity. For a pair of crimes, A and B, J is:

$$J = \frac{a}{a + b + c}$$

where a equals the number of behaviours common to both crimes and b and c equal the number of behaviours unique to Crimes A and B, respectively (Jaccard, 1908).

In addition to this potential advantage, it is commonly argued that J is the most appropriate coefficient for use in BLA because joint non-occurrences of a specific behaviour are not included in its calculation (e.g. Bennell & Canter, 2002; Goodwill & Alison, 2006; Woodhams & Toye, 2007). Within the investigative domain, it is thought that ignoring joint non-occurrences in this way is wise because the absence of a behaviour in any given crime, or crime report, may be due to factors other than its actual non-occurrence (e.g. a behaviour might not have been witnessed by the victim, remembered, or reported to the police, and even if the behaviour is reported it may not be accurately recorded; Alison, Snook, & Stein, 2001).

Beyond debates over the validity of this argument, the obvious drawback of J is that it is a very crude similarity metric (Woodhams, Grant et al., 2007). For example, J only accounts for across-crime similarity at the most discrete behavioural level and, therefore, it is very sensitive to even slight variations in behaviour across crimes. Issues such as these have led researchers to recommend that other similarity coefficients be examined to determine if they might be more suitable for BLA (e.g. Bennell et al., 2007; Gauthier, 2008; Woodhams, Grant et al., 2007). One coefficient that has been recently put forward as a potential candidate is the taxonomic similarity index, Δ_s (Izsak & Price, 2001; Woodhams, Grant et al., 2007).

Developed in marine ecology, Δ_s takes an expanded view of across-crime similarity by utilizing hierarchical information (Izsak & Price, 2001). In other words, just as biological classification is organized into a hierarchy (i.e. ascending from species, genus, family, etc.), this measure assumes that a crime scene can be conceptualized as a hierarchy of behaviours (e.g. ascending from discrete behaviours, sub-types of behaviours, types of behaviours, etc.; Woodhams, Grant et al., 2007). Thus, in contrast to J, Δ_s is not limited to the specific crime scene behaviours that are present in two crimes when calculating across-crime similarity; it also capitalizes on across-crime similarity that may be present at higher levels of the behavioural hierarchy.

For the purpose of illustrating how Δ_s is calculated, consider the hypothetical behavioural hierarchy for sexual assault that was provided by Woodhams, Grant et al. (2007; see Figure 1). The calculation requires two steps. First, the taxonomic distance between Crimes A and B is calculated:

$$TD(A,\ B) = \frac{\sum_i w_{iB} + \sum_j w_{jA}}{n_A + n_B}$$

where w_{iB} is the minimum path length between behaviour i in Crime A and all behaviours in Crime B, w_{jA} is the minimum path length between behaviour j in Crime B and all behaviours in Crime A, and n_A and n_B are the number of behaviours in Crime A and B, respectively.

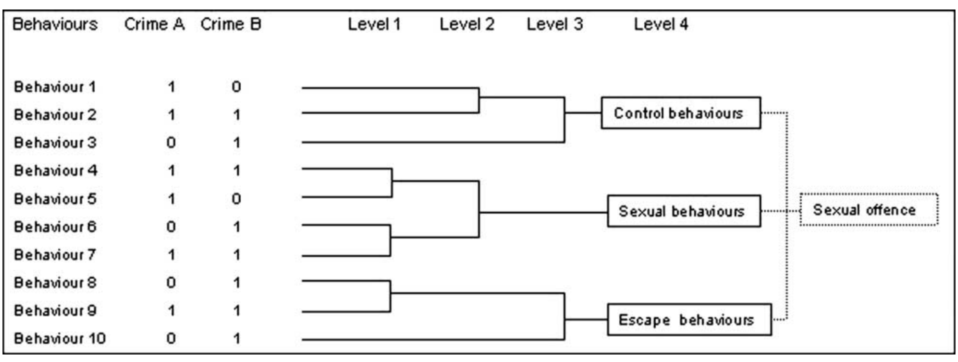

Figure 1. A hypothetical behavioural hierarchy of crime-scene behaviours in sexual assault. The first two columns indicate the presence or absence of 10 specific behaviours across the two crimes, A and B. (Source: Woodhams, Grant et al., 2007.)

Second, Δ_s is calculated by:

$$\Delta_s(A, B) = 1 - \frac{TD}{L - 1}$$

where L is the number of levels in the hierarchy of behaviours and $L - 1$ is the maximum path length between a pair of behaviours. The purpose of the second step is to express Δ_s as a value ranging from 0 to 1, similar to J.[1]

In the first examination of Δ_s within the investigative domain, Woodhams, Grant et al. (2007) compared J to Δ_s using behavioural data from 16 sexual offences committed by seven juvenile offenders. By drawing on a behavioural hierarchy consisting of four levels and 55 offence behaviours, the results of this study demonstrated that both Δ_s and J resulted in significantly higher similarity scores for linked compared to unlinked crimes (with the similarity scores for both linked and unlinked crimes being larger when using Δ_s versus J). However, as predicted, the effect size was greater for Δ_s than it was for J (Cohen's $d = 1.68$ versus 1.43, respectively) indicating that the use of higher-order behavioural information may make Δ_s more effective than J when attempting to discriminate between linked and unlinked offences. Interestingly, the significant differences that existed between the similarity scores for linked and unlinked crimes remained across conditions of data degradation (i.e. with 10%, 20%, and 50% of behaviours randomly removed) when using Δ_s, but not when using J. This last finding was viewed as useful by the study authors, given that crime scene data will often be 'missing'.

The current study

Woodhams, Grant et al. (2007) suggested that their results provide evidence that Δ_s may be more suitable for BLA than J, especially when a substantial amount of crime scene data is missing. While we agree that there are potential problems with using J for the purpose of BLA, and believe that Δ_s may be a suitable replacement for it, there are several reasons to be cautious when interpreting the results of Woodhams, Grant et al.'s study. For example, as these authors themselves pointed out, the results of their study are based on only one particular crime type and on only a very small sample of offences. Therefore, before we accept that Δ_s is a useful coefficient for BLA, it is important to replicate the analysis of Woodhams, Grant et al. using other types of crimes and larger sample sizes.

In addition, it seems to us that there is also a potential danger with using Δ_s in that it increases the degree of across-crime similarity that can be found for both linked *and* unlinked crimes (compared to J). This seems to be an inevitable consequence of using a coefficient that takes into account across-crime similarity at levels beyond discrete behaviours (i.e. there will almost always be some level of across-crime similarity that can be found in a behavioural hierarchy, regardless of who committed the crimes). As has recently been argued, using a coefficient that increases behavioural stability (i.e. higher similarity across crimes committed by the same offender) may have little impact on discrimination accuracy, if the same coefficient also decreases behavioural distinctiveness (i.e. higher similarity across crimes committed by different offenders; Bennell, Jones, & Melnyk, 2009). We accept that this did not appear to be a problem in Woodhams, Grant et al.'s (2007) study, since the effect sizes associated with Δ_s were larger than those for J. However, it is not clear whether this will continue to be the case when larger sample sizes are used, or when different crimes types are examined.

To address these concerns, the current study will attempt to replicate the study conducted by Woodhams, Grant et al. (2007) using samples of serial homicide and serial burglary. To accomplish this, the performance of Δ_s and J will be compared using both of these datasets, while varying the size of the samples and the degree of data degradation that exists. The findings from this study will further our general understanding of the conditions under which BLA is most effective and will help researchers and practitioners decide whether J or Δ_s should be used to study/conduct BLA at the moment.

Method

Samples

Serial homicide data

The serial homicide data used in the present study represents a subset of data originally collected by Godwin (1998). It contains information on 39 crime scene behaviours from 79 male serial killers from the US who committed a total of 237 homicides. The data was restricted to three crimes per offender. This was done to maximize the number of crimes in the sample, while also trying to ensure that the analysis was not biased by undue weight being assigned to highly prolific offenders displaying particularly high (or low) levels of behavioural stability and/or distinctiveness. This is common practice in BLA research (e.g. Bennell & Canter, 2002; Bennell & Jones, 2005; Santtila, Junkkila, & Sandnabba, 2005; Woodhams & Toye, 2007). The data was extracted from various police databases and the crime scene behaviours were coded in dichotomous form as present (1) or absent (0) for each of the crimes. Because the data was originally coded by the police, levels of inter-rater reliability are not available. However, other research suggests that this type of crime scene data can be coded reliably (e.g. Alison & Stein, 2001; Hakkanen, Puolakka, & Santtila, 2004; Santtila et al., 2005).

Serial burglary data

The serial burglary data used in the present study represents a subset of data originally collected by Bennell (2002). It contains information on 28 crime scene behaviours from 42 male serial burglars from the UK who committed a total of 210 residential burglaries. The data was restricted to five crimes per offender for reasons discussed above. The data was extracted from a burglary database managed by a police service in the UK and the crime scene behaviours were coded in dichotomous form as present (1) or absent (0) for each of the crimes. Again, due to the fact that the data was coded by the police, levels of inter-rater reliability are not available for this data.

Procedure

Each of the two datasets was used to examine differences that emerge when using J versus Δ_s to discriminate between crimes committed by the same versus different offenders. The procedure for carrying out this comparison involved several steps.

Step 1: construction of behavioural hierarchies

Hierarchies of crime scene behaviours were developed for each dataset in order to calculate Δ_s. There is no set procedure for deriving these hierarchies within the context of BLA. The hierarchies for each dataset were thus derived both from published literature on each type of serial crime (top-down) and a cursory examination of the data (bottom-up). An attempt was made wherever possible to construct the hierarchies on the basis of empirical results, primarily from studies using multidimensional scaling and cluster analytic techniques to derive the underlying structure of offence behaviour for each respective crime type. However, a degree of subjectivity was also required in the construction of the hierarchies when making certain decisions about variable placement. Inter-rater reliability scores were calculated to ensure that variable placement was being carried out in a reliable manner.

Step 2: calculation of J and Δ_s

In order to calculate J and Δ_s across the linked and unlinked crime pairs, a specially designed computer program, which we refer to as *CrimeSolver*, was used (the program was written using *MathCad* (v. 12) by the fourth author). For each dataset, all of the dichotomously coded crime scene behaviours were submitted to the program as Microsoft *Excel* files. In addition, path length tables illustrating the relationships between all behaviours in the datasets were constructed for each hierarchy, and were also input into *CrimeSolver* as Microsoft *Excel* files. On this basis, the program then calculated J and Δ_s for every possible pair of crimes in each dataset, and provided this information as output.[2]

Step 3: Descriptive and comparative analyses

CrimeSolver produced raw similarity scores for linked and unlinked crime pairs for both similarity coefficients. This output was exported to SPSS (v. 16) where the data was analysed to examine the differences between the linked and unlinked crimes.

Step 4: ROC analysis

CrimeSolver also produced empirical receiver operating characteristic (ROC) graphs, showing the level of discrimination accuracy achieved for the two similarity coefficients. Because the distribution of similarity scores for linked and unlinked crimes (for both datasets) were not normally distributed, we decided not to use Cohen's d as a measure of effect size in this study (as Woodhams, Grant et al., 2007 did). Instead, the area under the ROC curve (AUCs) was used, as it has been in several studies of BLA (e.g. Bennell & Canter, 2002; Bennell & Jones, 2005; Tonkin, Grant, & Bond, 2008; Woodhams & Toye, 2007). When used as a measure of discrimination accuracy, AUCs close to 0.50 are generally considered non-informative, AUCs between 0.50 and 0.70 represent low levels of accuracy, AUCs between 0.70 and 0.90 represent good levels of accuracy, and AUCs between 0.90 and 1.00 represent high levels of accuracy (Swets, 1988).

In an attempt to replicate the findings of Woodhams, Grant et al. (2007) two additional procedures were added to the above steps: a data degradation procedure to examine the impact of missing data and a re-sampling procedure to examine the impact of variations in sample size.

Data degradation procedure

The data degradation procedure utilized by Woodhams, Grant et al. (2007) was replicated to test the robustness of the two similarity coefficients to missing information. This procedure involved progressively removing randomly selected discrete behaviours from each sample (accounting for approximately 10%, 25%, and 50% of each dataset).[3] For each degradation condition, *CrimeSolver* produced raw similarity scores (that could be exported to SPSS) and conducted a ROC analysis. In addition to carrying out the data degradation procedure on the original sets of linked and unlinked crimes, the procedure was also carried out across various sample sizes (see below).

Re-sampling procedure

A re-sampling procedure was utilized to examine the influence of sample size on the relative discrimination accuracy of J and $\Delta_{s.}$ For each crime type, *CrimeSolver* generated different numbers of linked and unlinked crime pairs. Two different re-sampling procedures were used. The first procedure randomly drew equal numbers of linked and unlinked crime pairs of various sizes. For each sample size, 10 draws were made and the average accuracy scores were calculated (e.g. 10 draws of 10 linked and 10 unlinked pairs, 10 draws of 50 linked and 50 unlinked pairs, 10 draws of 100 linked and 100 unlinked pairs, etc.). This procedure attempts to replicate the procedure adopted by Woodhams, Grant et al. (2007) where they examined 11 linked and 11 unlinked crime pairs.

The second procedure randomly drew linked and unlinked crime pairs in numbers that were proportional to the number of actual pairs in each dataset (one linked pair: 118 unlinked pairs for homicide; one linked pair: 52.25 unlinked pairs for burglary). Again, for each sample size, 10 draws were made and average accuracy scores were calculated (e.g. 10 draws of 10 linked and 522 unlinked pairs, etc. for burglary). This was done in an attempt to more accurately reflect the circumstances encountered in naturalistic settings where there will always be many more unlinked crimes than linked crimes (this procedure also replicates the procedure used by Bennell & Canter, 2002, and Bennell & Jones, 2005). For each re-sampling condition, *CrimeSolver* produced raw similarity scores (that could be exported to SPSS) and conducted a ROC analysis.

Results

Study 1: serial homicide

Behavioural hierarchy

The serial homicide hierarchy was constructed around two main branches, consisting of organized and disorganized behaviours, given that this dichotomy is the most common way of categorizing serial homicide behaviour (Beauregard, Goodwill, Taylor, & Bennell, 2007; Bloomfield, 2006; Hazelwood & Douglas, 1980; Holmes & DeBurger, 1988; Prentky & Burgess, 2000; Ressler, Burgess, Douglas, Hartmann, & D'Agostino, 1986; however, see Canter, Alison, Alison, & Wentink, 2004). Essentially, organized crime scene behaviours are those that involve a greater degree of planning and control than their disorganized counterparts (Ressler et al., 1986).

The organized branch was further divided into planning, control, and ritualistic behaviours and the disorganized branch was divided into impulsive, control, and ritualistic behaviours. Planning behaviours include any organized activities, which indicate the offender prepared for the crime before it was committed, and these activities directly contrast impulsive behaviours in the disorganized domain (Bloomfield, 2006). Control behaviours are characterized by actions designed to create and maintain an environment in which the crime can successfully take place, and often involve behaviours that are meant to control the victim (Hodge, in press). Lastly, ritualistic behaviours are those that are more symbolic and excessive, typically going beyond what is necessary to commit the offence (Hazelwood & Warren, 2003). Thus, the organized and disorganized branches in the hierarchy often include the same categories of behaviour, but the specific behaviours included in those categories are distinctly organized or disorganized.

The 39 serial homicide behaviours available in the dataset were incorporated into what was deemed the most appropriate branch. Using a coding scheme whereby each individual branch of the hierarchy was given a code, the researcher and two independent coders assigned each of the behaviours a code indicating where they thought it belonged in the hierarchy. Based on code assignments, the average Kappa across the three raters was 0.83, indicating a satisfactory level of inter-rater reliability (Landis & Koch, 1977). The resulting hierarchy of serial homicide behaviours is presented in Figure 2.

Descriptive and comparative analyses

Results suggested that the distributions of across-crime similarity scores for both coefficients were skewed and therefore non-parametric tests were used to compare the similarity scores from the two distributions (i.e. linked and unlinked crimes). As illustrated in Table 1, the results of these analyses indicated that the mean across-crime similarity scores were higher for linked crimes than for unlinked crimes regardless of whether Δ_s ($p < 0.001$) or J was used ($p < 0.001$).

ROC analysis

The results of the descriptive analyses presented in Table 1 demonstrate that Δ_s is capable of achieving higher across-crime similarity scores for linked homicides compared to *J*. However, Δ_s also generated substantially higher scores for unlinked

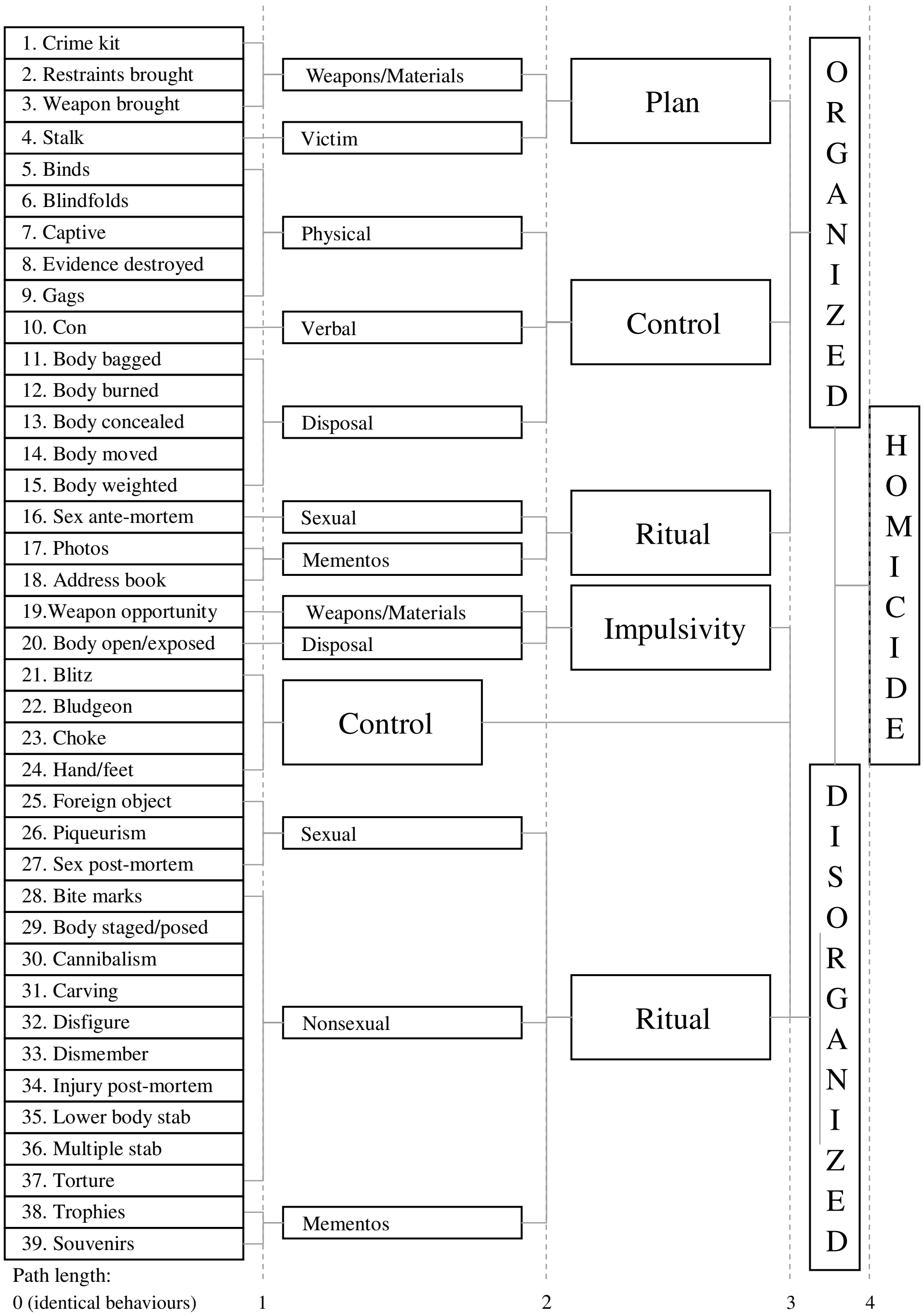

Figure 2. Serial homicide behavioural hierarchy.

Table 1. Descriptive statistics for Δ_s and J for linked/unlinked homicides and burglaries.

	Homicide				Burglary			
	Δ_s		J		Δ_s		J	
Statistic	Linked ($n = 237$)	Unlinked ($n = 27\ 729$)	Linked ($n = 237$)	Unlinked ($n = 27\ 729$)	Linked ($n = 420$)	Unlinked ($n = 21\ 945$)	Linked ($n = 420$)	Unlinked ($n = 21\ 945$)
Min.	0.48	0.00	0.10	0.00	0.00	0.00	0.10	0.00
Max.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Median	0.98	0.75	0.89	0.27	0.71	0.62	0.27	0.19
Mean	0.94	0.72	0.81	0.28	0.64	0.62	0.29	0.22
SD	0.09	0.13	0.22	0.14	0.18	0.18	0.19	0.16

homicides as well. ROC analysis was thus used to evaluate the relative discrimination accuracy of the two similarity coefficients. The ROC curves generated by *CrimeSolver* are presented in Figure 3. As illustrated, both similarity coefficients were capable of differentiating between linked and unlinked homicides at a level significantly greater than chance ($p < 0.001$). However, a comparison of the confidence intervals (CIs) for the two AUCs indicates that Δ_s (AUC = 0.93, CI_{95} = 0.91–0.96) does not significantly outperform J (AUC = 0.96, CI_{95} = 0.94–0.98).

Data degradation

Data was progressively degraded from the dataset to test the robustness of the two coefficients across conditions of data degradation. Testing involved progressively removing 4, 9 and 19 randomly selected behaviours, accounting for approximately 10%, 25% and 50% of the dataset, respectively. Table 2 reports the results of this analysis. The removal of crime scene behaviours from the homicide dataset can be seen to have only a slight impact on both coefficients (all CIs overlap across levels of degradation). Interestingly, across the conditions, the AUC values associated with Δ_s were not significantly, or consistently, higher than J (all CIs overlap). Thus, both coefficients were able to discriminate between linked and unlinked homicides even at high levels of data degradation, but in contrast to previous findings, Δ_s was not found to outperform J.

Re-sampling procedure

Up to this point, the results contradict those of Woodhams, Grant et al. (2007). A re-sampling procedure was thus used to test whether Woodhams, Grant et al.'s small sample size may be an explanation for the discrepancy in findings. Figure 4 illustrates

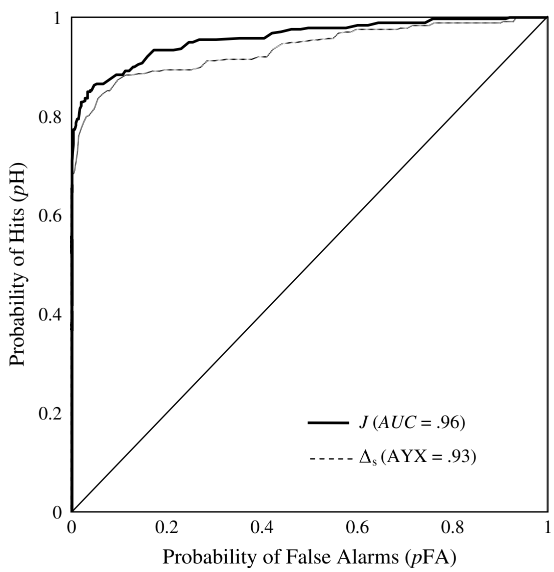

Figure 3. ROC curves illustrating comparative linking accuracy of Δ_s and J for homicides.

Table 2. Influence of data degradation on the discrimination accuracy of Δ_s and J for homicide and burglary.

% of	Homicide				Burglary			
	Δ_s		J		Δ_s		J	
behaviours	AUC	CI_{95}	AUC	CI_{95}	AUC	CI_{95}	AUC	CI_{95}
100%	0.93	0.91–0.96	0.93	0.91–0.96	0.59	0.54–0.65	0.62	0.57–0.68
90%	0.94	0.91–0.96	0.94	0.91–0.96	0.57	0.51–0.62	0.60	0.54–0.65
75%	0.94	0.91–0.96	0.94	0.91–0.96	0.56	0.50–0.62	0.59	0.53–0.64
50%	0.93	0.90–0.95	0.93	0.90–0.95	0.54	0.48–0.59	0.54	0.49–0.60

the effect of the re-sampling procedures using the non-degraded data. The top plot presents the average accuracy rates (the mean AUC across 10 draws per sample size) that were calculated using an equal number of randomly selected linked and unlinked homicides, whereas the bottom plot presents the average accuracy rates for the proportional re-sampling procedure.

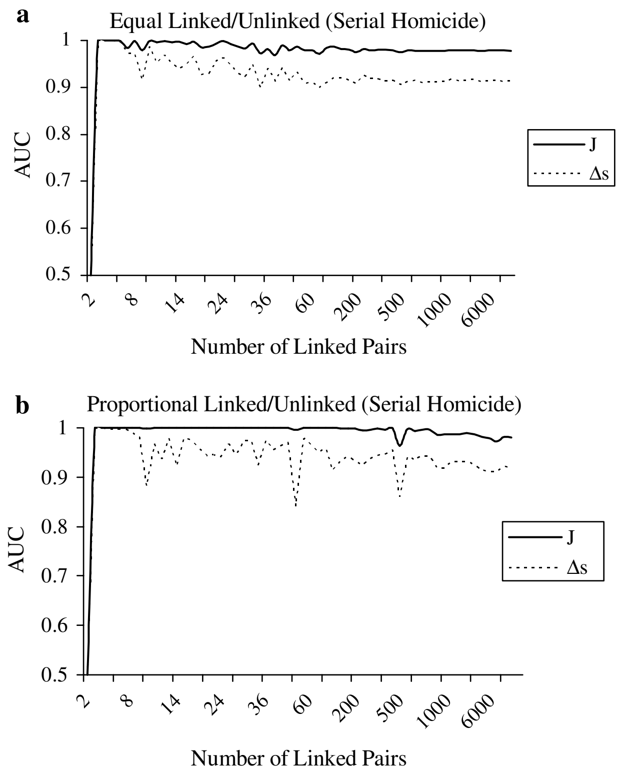

Figure 4. Effect of increasing sample size on linking accuracy for: (a) equal numbers of linked and unlinked homicides and (b) numbers of linked and unlinked homicides proportional to the dataset.

These graphs clearly demonstrate that *J* consistently achieves higher AUC values than Δ_s regardless of sample size. Furthermore, as the number of crime pairs increases, the results become more stabilized in both graphs. These results provide support for the notion that the findings reported by Woodhams, Grant et al. (2007) may be attributed to their use of a very small sample and possibly to their choice of crime type (or perhaps a fortuitous dataset). The exact same pattern of results emerged when running the analysis with degraded data, except that the results were slightly more erratic for the proportional analysis at high levels of data degradation.

Study 2: serial burglary

Behavioural hierarchy

The serial burglary hierarchy was constructed around two main branches, consisting of high and low skill behaviours (Merry & Harsent, 2000). High skill behaviours are those associated with a greater level of expertise, planning, intelligence, and manual dexterity in the commission of the burglary, which is in direct contrast to low skill behaviours (Canter, 1994; Merry & Harsent, 2000). The high skill branch was further sub-divided into high interaction/involvement behaviours, as well as target characteristics, whereas the low skill branch was sub-divided into high interaction/involvement behaviours, low interaction/involvement behaviours, and target characteristics. High interaction/involvement behaviours are those of a more intimate, invasive, and private nature, which directly contrasts with their low interaction/involvement counterparts (Merry & Harsent, 2000). Target characteristics refer to information about the property targeted in the offence.

The 28 serial burglary behaviours available in the dataset were incorporated into what was deemed the most appropriate branch. Inter-rater reliability, calculated as above, was found to be satisfactory (Kappa $=0.93$). The resulting hierarchy of serial burglary behaviours is presented in Figure 5.

Descriptive and comparative analyses

Results suggested that the distributions of across-crime similarity scores for both coefficients were skewed and therefore non-parametric tests were used to compare the similarity scores from the two distributions (i.e. linked and unlinked crimes). As illustrated in Table 1, the results of this analysis indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of whether Δ_s ($p < 0.001$) or *J* ($p < 0.001$) was used.

ROC analysis

The descriptive analyses demonstrate that Δ_s is capable of achieving higher across-crime similarity scores for linked burglaries than *J*. However, Δ_s also generated higher scores for unlinked burglaries as well. ROC analysis was thus conducted to evaluate the relative discrimination accuracy achieved using the two similarity coefficients. The ROC curves generated by *CrimeSolver* are presented in Figure 6. Both similarity coefficients were capable of differentiating between linked and unlinked burglaries at a level significantly greater than chance ($p < 0.001$).

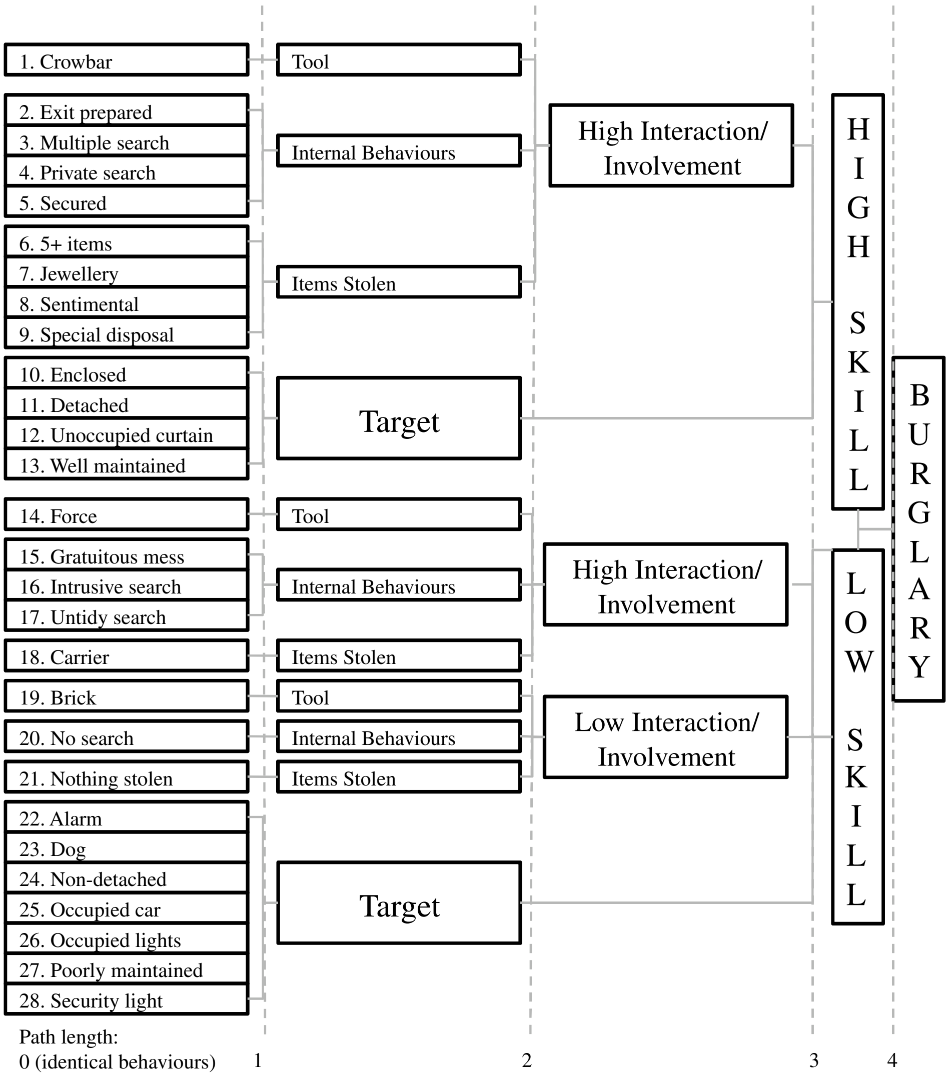

Figure 5. Serial burglary behavioural hierarchy.

A comparison of the CIs for the two AUCs indicates that Δ_s (AUC = 0.59, CI_{95} = 0.54–0.65) does not significantly outperform J (AUC = 0.62, CI_{95} = 0.57–0.68).

Data degradation

Data was progressively degraded from the dataset to test the robustness of the two coefficients. Testing involved progressively removing 3, 7 and 14 randomly selected behaviours, accounting for approximately 10%, 25% and 50% of the dataset, respectively. As is apparent from the results in Table 2, the removal of crime scene behaviours from the burglary dataset had only a small impact on both coefficients (all CIs overlap across levels of degradation). However, across the conditions, the

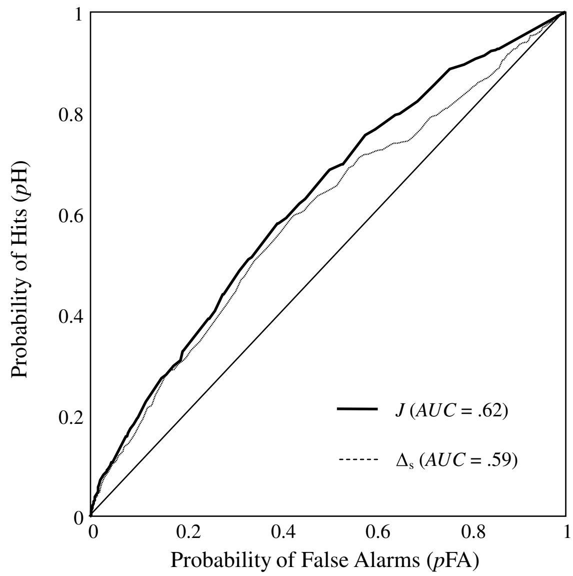

Figure 6. ROC curves illustrating comparative linking accuracy of Δ_s and J for burglaries.

AUC values associated with Δ_s were not significantly, or consistently, higher than J (all CIs overlap). Thus, both coefficients were able to discriminate between linked and unlinked burglaries even at high levels of data degradation, but once again, Δ_s was not found to outperform J.

Re-sampling procedure

The results of the re-sampling procedures are presented in Figure 7 for both Δ_s and J, for the non-degraded burglary dataset. The top plot reflects AUC values that were calculated using an equal number of linked and unlinked burglaries, whereas the bottom plot reflects AUC values that were calculated using the proportional re-sampling procedure. Across both graphs it can clearly be seen that the effect size is much more erratic for the two coefficients when the sample size is smaller (especially in the bottom plot). However, as the number of crime pairs increases, the results become more stabilized in both graphs. Thus, these graphs demonstrate that J consistently achieves higher AUC values than Δ_s when large sample sizes are used, again suggesting that the results of Woodhams, Grant et al. (2007) may be attributed to their use of a small sample size. The same pattern of results emerged when running the same analysis with degraded data, except that the results were more erratic, particularly for the proportional analysis and at high levels of data degradation.

Discussion

The stability and distinctiveness of serial homicide and burglary behaviour

The purpose of this study was to compare the degree to which J and Δ_s could discriminate between linked and unlinked homicides and burglaries across a range of

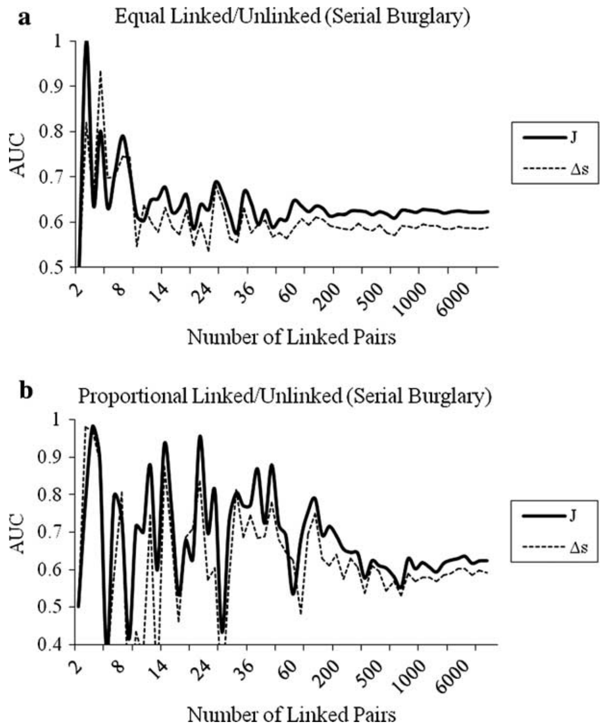

Figure 7. Effect of increasing sample size on linking accuracy for: (a) equal numbers of linked and unlinked burglaries and (b) numbers of linked and unlinked burglaries proportional to the dataset.

conditions. The results demonstrated that, regardless of what coefficient was used, or which crime type was examined, crimes committed by the same offender tended to be associated with significantly higher levels of across-crime similarity than crimes committed by different offenders. Thus, the serial offenders examined in this study do appear to behave in a somewhat stable and distinct fashion across their crimes, which accords well with other research that has examined these issues (Woodhams, Hollin et al., 2007). This suggests that many serial offenders committing homicides and burglaries are predisposed to behave in a particular way when committing their crimes and that these tendencies are, to some extent at least, unaffected by situational variations that exist across crimes. Of course, this result could also be influenced by the fact that only solved crimes were examined in this study, and crimes may be solved, at least in part, because of such high levels of stability and distinctiveness. Unfortunately, until further research is conducted (e.g. prospective BLA research on unsolved crimes), we will never know if this is the case.

Which similarity coefficient is best suited for linkage analysis?

Comparisons of the two coefficients across both crime types revealed that, contrary to previous findings, Δ_s did not outperform *J*. These findings suggest that Δ_s may not be as powerful for linking purposes as was originally thought. This begs the question of why discrepancies emerged between this study and the study by Woodhams, Grant et al. (2007). One obvious possibility is that the sample size in Woodhams, Grant et al.'s study was simply not large enough to produce reliable results. Certainly, the findings presented in this study suggest that sample size does have a large impact on the reliability of linking results and can potentially lead to erroneous conclusions being drawn about the relative accuracy of the two coefficients. Specifically, the re-sampling analyses indicated that when smaller sample sizes are used (of the sort examined by Woodhams, Grant et al.), levels of discrimination accuracy can vary dramatically between analyses making it difficult to determine which coefficient is best. This is not the case when larger samples of crimes are used.

However, this explanation does not help us understand why, when drawing on larger samples of crimes, Δ_s does not significantly outperform *J*. While there are many potential answers to this question, a reconsideration of the fundamental assumptions underlying BLA raises one possibility. Recall that the ultimate task in BLA is to increase our ability to accurately discriminate between crimes committed by different offenders. This can be done in a number of ways: (1) by finding a way to increase the behavioural stability found between crimes committed by the same offender, (2) by finding a way to decrease the behavioural stability found between crimes committed by different offenders, or (3) by finding a way to accomplish both (1) and (2). Based on the results of the current study, Δ_s clearly does a good job of accomplishing (1). Compared to *J*, the across-crime similarity scores for linked crimes are always higher when using Δ_s. However, as we highlighted earlier, the use of Δ_s also appears to adversely influence the across-crime similarity scores calculated across unlinked crimes. That is, the use of Δ_s produces higher similarity scores for *both* linked and unlinked crimes. When using large sample sizes, this appears to make Δ_s slightly inferior to *J* for the purposes of BLA.

Having said this, there are a variety of things that can be explored in future research that might increase the degree of discrimination accuracy that can be achieved when using Δ_s, and until these things are done we are reluctant to draw strong conclusions about the value of Δ_s. For example, alternative hierarchies, which might be based on other psychologically plausible classification systems, may increase the degree of linking accuracy that can be achieved using Δ_s. On a related note, objectively derived hierarchies, which could be accomplished through the use of cluster analytic techniques, may prove much more useful than the somewhat subjective hierarchies relied on in the current study. Indeed, preliminary examination of this issue in our lab has suggested that larger effects can be achieved when objective methods are used to construct the hierarchies. Until this research has been conducted, questions will remain as to whether *J* or Δ_s is best suited for BLA.

What is the influence of data degradation on J and Δ_s?

To test Woodhams, Grant et al.'s (2007) argument that Δ_s will outperform *J* under conditions of data degradation, given that its hierarchical structure allows for more opportunities on which to base similarity, a progressive data degradation procedure was implemented across both crime types. Despite the logic of Woodhams, Grant et al.'s argument, the present results did not find Δ_s to be superior to *J* under conditions of degradation, regardless of crime type. These findings challenge the results of Woodhams, Grant et al. and suggest that their findings may be due to the small sample size that was used to test this hypothesis.

In the future it would be interesting to examine how other forms of missing data impact the performance of Δ_s and J. So far, data degradation has involved reducing the dataset by progressively omitting more and more behaviours. However, it seems more likely to us that the problem plaguing police data is the omission of a particular instance of a behaviour in a particular crime, as opposed to the omission of specific crime scene behaviours across an entire sample of crimes. As a result, the data degradation procedure adopted across these studies may not adequately highlight the strength of Δ_s in dealing with missing data. Unfortunately, *CrimeSolver* is currently unable to handle missing data values (as opposed to missing behaviours) and it was practically not possible to examine this issue by other means. Exploring the impact of different types of data degradation should therefore be a priority in future research.

What is the impact of crime type?

Another point of interest is that clear differences emerged across crime type with respect to the level of linking accuracy that was achieved, although similar patterns of results with respect to J and Δ_s were found for each crime type. Specifically, linking accuracy was noticeably lower in the case of serial burglary compared to serial homicide. Why these differences emerged across the crime types is not entirely clear, but we can speculate.

In fact, the types of results obtained from the burglary analyses in the current study were not all that surprising and accord well with some other studies of burglary in the published literature (e.g. Bennell & Canter, 2002; Bennell & Jones, 2005). While those studies have certainly reported higher linking accuracy scores, these were only found when using very specific subsets of behaviours, most notably the distance between crime locations (with shorter distances indicating an increased likelihood that the crimes were committed by the same offender). For the sorts of behaviours examined in the current study, the scores that were observed are in line with those in the published literature, presumably reflecting the impact that situational factors have on burglary behaviour (e.g. preferred entry points can be blocked, which forces a burglar to adapt; Bennell & Canter, 2002; Bennell & Jones, 2005). Unfortunately, data on crime locations were not available for analysis in the current study, making this a fruitful line for future research.

There are also reasons to expect serial killers to be highly consistent across their crimes, and relatively distinct in their style of offending, as was found in the current study. For example, there has been suggestions that serial killers can possess reasonably high levels of psychopathology (e.g. Pinizzotto & Finkel, 1990), and this might impact the degree of behavioural stability/distinctiveness that an individual exhibits (e.g. Moos, 1968, 1969). In addition, there have been strong suggestions that the behaviour of violent serial offenders is largely guided by scripts that have often been well-rehearsed, are deeply engrained, and are typically rooted in personal fantasies (e.g. Davies, 1992; Hazelwood & Warren, 2003; Keppel, 1997).[4] If this is the case, then perhaps it is not surprising that these offenders maintain their individual differences in offending style across the crimes they commit to a greater extent than do serial burglars.

Another possible explanation for the lower linking accuracy scores for serial burglary relates to the difference between the two datasets in terms of the number of crimes per offender that were included in the analyses (three for homicide versus five for burglary). This difference reflects the fact that burglary series do tend to be longer than homicide series, and this difference could have had an influence on our results.[5] For example, behaviour can evolve over an offender's crimes (e.g. Harbort & Mokros, 2004; Hazelwood, Reboussin, & Warren, 1989; Lussier, Leclerc, Healey, & Proulx, 2008; Warren et al., 1999), and the degree of behavioural change that is observed across any two crimes in a series likely depends on their degree of separation (e.g. the degree of change will likely be less across crimes committed in succession). Given that many more of the across-crime similarity scores for linked serial burglaries were based on crimes that were not committed in sequence (e.g. crime 1 compared to crimes 3, 4 and 5), perhaps it is not surprising that behavioural stability, and thus linking accuracy, was lower for this crime type. That being said, future research is of course needed to support this possibility.

The fact that the level of linking accuracy that was achieved for the serial homicide cases in the current study was so impressive warrants further empirical exploration. Indeed, given the variation in linking accuracy scores reported across studies of serial homicide (e.g. Bateman & Salfati, 2007; Godwin, 1998; Salfati & Bateman, 2005; Santtila et al., 2008), it is crucial to determine if the results reported here generalize across other data sets. Until such research is carried out, it is important to view the serial homicide results with an appropriate degree of caution (the same of course is true for the serial burglary results).

Conclusion

While there are clearly limitations with the current study, the results are potentially important for understanding the conditions under which BLA will be most effective. In contrast to previous research, the results suggest that J is as effective as Δ_s when used as a measure of similarity in BLA, and may slightly outperform Δ_s under certain conditions (e.g. when large samples are examined). Also in contrast to previous research, data degradation of the type examined in this study does not appear to have a detrimental effect on linking accuracy, so long as large sample sizes are used. Finally, both similarity coefficients result in significantly higher levels of linking accuracy in cases of serial homicide compared to serial burglary, suggesting that serial killers may be more stable and distinct than serial burglars. Future research in this area is needed to confirm that these conclusions are valid and to uncover additional factors that may influence the degree to which it is possible to link serial crime.

Acknowledgements

We would like to thank Alyssa Taylor and Rebecca Mugford for their help with data coding. We would also like to thank Jessica Woodhams for her helpful comments on an earlier draft of this paper. This research formed part of an M.A. thesis completed by the first author under the supervision of the second author.

Notes

1. The following example was provided by Woodhams, Grant et al. (2007) as an illustration of how Δs is calculated based on the hierarchy presented in Figure 1. The minimum path length is determined between each behaviour in Crime A and all those in Crime B, and vice

versa. For example, the values for the minimum path lengths for behaviours in Crime A are as follows: 2 for Behaviour 1, as it is absent as a specific behaviour in Crime B, but present in Crime B at level 2, 0 for Behaviour 2 because it is also present in Crime B, 0 for Behaviour 3, etc. This process is also completed for Crime B. Once this is done the path lengths for each crime are added together (i.e., [2+0+0+1+0+0] + [0+3+0+1+0+1+0+3]) and divided by the sum of the total number of behaviours present for each crime (i.e., 6+8) to obtain the taxonomic distance. In this case, *TD (A,B)* = .79. To obtain Δs, *TD* is divided by the number of taxonomic levels minus 1, the total of which is subtracted from 1 (i.e., 1 − [.79/3] = .74).
2. The accuracy of *CrimeSolve*r (for all of the different types of analyses reported in this study) was verified by Gauthier (2008).
3. These levels of degradation varied slightly from those used by Woodhams, Grant et al. (2007). In their study, they removed 10%, 20%, and 50% of behaviours.
4. For example, consider the case of notorious serial killer Edmund Kemper ("The Co-Ed Butcher") who targeted young female co-eds. Kemper "reportedly spent inordinate amounts of time envisaging all the murderous actions he could perform upon the young co-eds. Moreover, it has been estimated that in the year preceding the onset of his crime series, Kemper picked up and safely delivered in excess of 150 female hitchhikers as he rehearsed the preliminary steps towards the physical execution of his fantasy" (Jones, 2005, p. 97).
5. While we know of no published literature to directly support this statement, it is true of the samples that the current data were extracted from, and of other serial burglary and homicide databases that we have access to.

References

Alison, L.J., Snook, B., & Stein, K.L. (2001). Unobtrusive measurement: Using police information for forensic research. *Qualitative Research*, *1*, 241–254.

Alison, L.J., & Stein, K.L. (2001). Vicious circles: Accounts of stranger sexual assault reflect abusive variants of conventional interactions. *Journal of Forensic Psychiatry*, *12*, 515–538.

Baroni-Urbani, C., & Buser, M.W. (1976). Similarity of binary data. *Systematic Zoology*, *25*, 251–259.

Bateman, A.L., & Salfati, C.G. (2007). An examination of behavioral consistency using individual behaviors or groups of behaviors in serial homicide. *Behavioral Sciences and the Law*, *25*, 527–544.

Beauregard, E., Goodwill, A.M., Taylor, P.J., & Bennell, C. (2007). *Typologies of sexual murderers: A test of the organized/disorganized model.* Paper presented at the annual meeting of the Association for the Treatment of Sexual Abusers, San Diego, California, USA.

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

Bennell, C., & Canter, D.V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science & 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., Jones, N.J., & Melnyk, T. (2007). *Linking serial rapes: A test of the behavioural frequency hypothesis.* Poster presented at the annual conference of the Canadian Psychological Association/North American Correctional and Criminal Justice Psychology Conference, Ottawa, Ontario, Canada.

Bennell, C., Jones, N.J., & Melnyk, T. (2009). Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. *Legal and Criminological Psychology*, *14*, 293–310.

Bloomfield, S. (2006). *Classifying serial sexual homicide: Validating Keppel and Walter's (1999) model* (Unpublished masters thesis). Carleton University, Ottawa, Ontario, Canada.

Canter, D.V. (1994). *Criminal shadows.* London, UK: Harper Collins.

Canter, D.V. (1995). Psychology of offender profiling. In R. Bull, & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 343–355). Chichester, UK: Wiley.

Canter, D.V., Alison, L.J., Alison, E., & Wentink, N. (2004). The organized/disorganized typology of serial murder: Myth or model? *Psychology, Public Policy, and Law*, *10*, 293–320.
Davies, A. (1992). Rapists' behaviour: A three part aspect model as a basis for analysis and the identification of serial crime. *Forensic Science International*, *55*, 173–194.
Gauthier, D.J. (2008). *Does a taxonomic similarity index increase the ability to link serial crime?* Unpublished honours thesis, Carleton University, Ottawa, Ontario, Canada.
Godwin, G.M. (1998). *Inner themes – outer behaviours: A multivariate facet model of U.S. serial murderers' crime scene actions* (unpublished doctoral dissertation). University of Liverpool, Liverpool, UK.
Goodwill, A.M., & Alison, L.J. (2006). The development of a filter model for prioritizing suspects in burglary offences. *Psychology, Crime & Law*, *12*, 395–416.
Gower, J.C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, *3*, 5–48.
Grubin, D., Kelly, P., & Brunsdon, C. (2001). *Linking serious sexual assaults through behaviour.* London, UK: Home Office.
Hakkanen, H., Puolakka, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in arsons. *Legal and Criminological Psychology*, *9*, 197–214.
Harbort, S., & Mokros, A. (2004). Serial murderers in Germany from 1945 to 1995: A descriptive study. *Homicide Studies*, *5*, 311–334.
Hazelwood, R., & Douglas, J.E. (1980). The lust murderer. *FBI Law Enforcement Bulletin*, *4*, 18–22.
Hazelwood, R., Reboussin, R., & Warren, J. (1989). Serial rape: Correlates of increased aggression and the relationship of offender pleasure to victim resistance. *Journal of Interpersonal Violence*, *4*, 65–78.
Hazelwood, R., & Warren, J. (2003). Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior*, *8*, 587–598.
Hodge, S. (in press). A multivariate model of serial sexual murder. In D.V. Canter & L.J. Alison (Eds), *Profiling rape and murder.* Aldershot, UK: Ashgate.
Holmes, R.M., & DeBurger, J. (1988). *Serial murder.* Beverly Hills, CA: Sage.
Izsak, C., & Price, A.R.G. (2001). Measuring β-diversity using a taxonomic similarity index, and its relation to spatial scale. *Marine Ecology Progress Series*, *215*, 69–77.
Jaccard, P. (1908). Nouvelle recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *44*, 223–270.
Jones, N.J. (2005). *The temporal stability and cross-situational consistency of behavioural themes in serial homicide* (Unpublished masters thesis). Carleton University, Ottawa, Ontario, Canada.
Keppel, R.D. (1997). *Signature killers.* New York, NY: Pocket Books.
Kosman, E., & Leonard, K.J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology*, *14*, 415–424.
Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
Liebetrau, A.M. (1983). *Measures of association*. Beverly Hills, CA: Sage.
Lussier, P., Leclerc, B., Healey, J., & Proulx, J. (2008). Generality of deviance and predation: Crime-switching and specialization patterns in persistent sexual offenders. In M. DeLisi, & P.J. Conis (Eds.), *Violent offenders: Theory, research, public policy, and practice* (pp. 97–118). Sudbury, MA: Jones and Bartlett.
Merry, S., & Harsent, L. (2000). Intruders, pilferers, raiders, and invaders: The interpersonal dimension of burglary. In D.V. Canter, & L.J. Alison (Eds.), *Profiling property crimes* (pp. 31–56). Aldershot, UK: Dartmouth.
Moos, R.H. (1968). Situational analysis of a therapeutic community milieu. *Journal of Abnormal Psychology*, *73*, 49–61.
Moos, R.H. (1969). Sources of variance in response to questionnaires and in behavior. *Journal of Abnormal Psychology*, *74*, 405–412.
Pinizzotto, A.J., & Finkel, N.J. (1990). Criminal personality profiling: An outcome and process study. *Law and Human Behavior*, *14*, 215–233.

Prentky, R.A., & Burgess, A.W. (2000). *Forensic management of sexual offenders*. New York, NY: Kluwer Academic/Plenum.
Ressler, R.K., Burgess, A.W., Douglas, J.E., Hartmann, C.A., & D'Agostino, R.B. (1986). Sexual killers and their victims – Identifying patterns through crime scene analysis. *Journal of Interpersonal Violence*, *1*, 288–308.
Salfati, C.G., & Bateman, A.L. (2005). Serial homicide: An investigation of behavioral consistency. *Journal of Investigative Psychology and Offender Profiling*, *2*, 121–144.
Santtila, P., Junkkila, J., & Sandnabba, N.K. (2005). Behavioural linking of stranger rapes. *Journal of Investigative Psychology and Offender Profiling*, *2*, 87–103.
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.
Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.
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.
Warren, J., Reboussin, R., Hazelwood, R.R., Gibbs, N.A., Trumbetta, S.L., & Cummings, A. (1999). Crime scene analysis and the escalation of violence in serial rape. *Forensic Science International*, *100*, 37–56.
Woodhams, J., Grant, T., & Price, A. (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.R., & 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 robberies. *Psychology, Public Policy, and Law*, *13*, 59–85.