Journal of Investigative Psychology and Offender Profiling
J. Investig. Psych. Offender Profil. **10**: 1–27 (2013)
Published online 19 June 2012 in Wiley Online Library (wileyonlinelibrary.com). **DOI**: 10.1002/jip.1364

Examining the Role of Similarity Coefficients and the Value of Behavioural Themes in Attempts to Link Serial Arson Offences

HOLLY ELLINGWOOD[1,*], REBECCA MUGFORD[1], CRAIG BENNELL[1], TAMARA MELNYK[1] and KATARINA FRITZON[2]

[1]*Department of Psychology, Carleton University, Canada*
[2]*Faculty of Humanities and Social Sciences, Bond University, Australia*

Abstract

When relying on crime scene behaviours to link serial crimes, linking accuracy may be influenced by the measure used to assess across-crime similarity and the types of behaviours included in the analysis. To examine these issues, the present study compared the level of linking accuracy achieved by using the simple matching index (S) to that of the commonly used Jaccard's coefficient (J) across themes of arson behaviour. The data consisted of 42 crime scene behaviours, separated into three behavioural themes, which were exhibited by 37 offenders across 114 solved arsons. The results of logistic regression and receiver operating characteristic analysis indicate that, with the exception of one theme where S was more effective than J at discriminating between linked and unlinked crimes, no significant differences emerged between the two similarity measures. In addition, our results suggest that thematically unrelated behaviours can be used to link crimes with the same degree of accuracy as thematically related behaviours, potentially calling into the question the importance of theme-based approaches to behavioural linkage analysis. Copyright © 2012 John Wiley & Sons, Ltd.

Key words: linkage analysis; simple matching index; Jaccard's coefficient; ROC analysis; logistic regression; serial arson

One of the challenges police investigators sometimes face is the task of correctly linking unsolved crimes to the same offender (Grubin, Kelly, & Brunsdon, 2001). This is a particularly difficult task when physical evidence is not available for analysis. Under these circumstances, investigators often rely on behavioural information obtained from crime scenes to establish any crime linkages by using a technique known as behavioural linkage analysis (BLA) (Woodhams, Hollin, & Bull, 2007). The need to rely on BLA may be particularly common in cases of serial arsons because physical evidence will often be destroyed by the fire.

For it to be possible to successfully link serial arsons by using crime scene behaviours, one must consider two key assumptions (Canter, 1995). First, arsonists must exhibit

*Correspondence to: Holly Ellingwood, Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada.
E-mail: hellingw@connect.carleton.ca.

relatively high levels of *behavioural stability* across their respective crime series by exhibiting the same or similar behaviours each time an arson is committed. Second, arsonists must exhibit relatively high levels of *behavioural distinctiveness* across their crimes by exhibiting behaviours that are not exhibited by other arsonists. When serial arsonists do exhibit both stability and distinctiveness, it should be possible to discriminate between arsons committed by different offenders (i.e., behavioural discrimination).

A variety of factors can impact the degree of behavioural stability and distinctiveness that is observed across a series of crimes and thus the degree to which it is possible to accurately link those crimes. For example, in a comprehensive review of empirical studies of BLA, Woodhams, Hollin, and Bull (2007) indicated that the type of crime scene behaviour under examination, the experience level of the criminal, and the period over which crimes have been committed can all impact stability/distinctiveness. The current study focuses on two other factors that might have an impact on our ability to link arsons (or any crimes for that matter): the type of similarity coefficient used to assess across-crime similarity and whether thematically related or unrelated behaviours are focused on in the analysis.

The potential importance of across-crime similarity coefficients in BLA

In order to determine the degree to which behavioural stability and distinctiveness exists across crimes, researchers often use similarity coefficients. Although there are many similarity coefficients to choose from for this purpose (e.g., Liebetrau, 1983), all of the coefficients allow the researcher to quantify how similar two crimes are to one another with respect to the behaviours that are present (or absent) at a crime scene. The degree of across-crime similarity that is found is typically expressed as a value between 0 (no similarity) and 1 (total similarity). It has been argued previously that behavioural stability is demonstrated by relatively high scores being observed across crimes committed by the same offender and that behavioural distinctiveness is demonstrated by relatively low scores being found across crimes committed by different offenders (Bennell, Jones, & Melnyk, 2009). It is this pattern of high scores (across linked crimes) and low scores (across unlinked crimes) that allow us to discriminate between crimes committed by different offenders.

Previous research conducted in non-forensic domains has clearly demonstrated that the results on a variety of discrimination tasks are influenced by which type of coefficient is used (e.g., Baroni-Urbani & Buser, 1976; Gower & Legendre, 1986; Kosman & Leonard, 2005). Recent studies in the area of BLA suggest that this may also be true in this context (e.g., Bennell, Gauthier, Gauthier, Melnyk, & Musolino, 2010; Emeno, Bennell, Melnyk, & Jones, 2008; Melnyk, Bennell, Gauthier, & Gauthier, 2011; Woodhams, Grant, & Price, 2007). That is, the type of coefficient used to assess across-crime similarity may impact the degree to which it is possible to accurately link two or more crimes together to form a crime series.

A number of these studies have focused on the use of Jaccard's coefficient (J; Jaccard, 1908), which is arguably the coefficient of choice in studies of BLA (Woodhams, Hollin, & Bull, 2007). One of the reasons for its popularity in BLA research is that J is very easy to calculate (Melnyk *et al.*, 2011). Quite simply, 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. Another reason for its

DOI: 10.1002/jip

popularity is that joint non-occurrences of behaviour (across crimes) are not *typically* included in the calculation of *J* (e.g., Bennell & Canter, 2002; Goodwill & Alison, 2006; Woodhams & Toye, 2007).[1] This has been seen by some as an advantage because the absence of a behaviour in any given crime may be due to factors other than its actual non-occurrence (e.g., a behaviour might not have been reported to the police; Alison, Snook, & Stein, 2001).

Despite these potential advantages, researchers have started to recommend that other similarity coefficients be examined to determine if they might be more suitable for the purpose of BLA. For example, Woodhams, Grant, and Price (2007) have recently argued that *J* may be problematic because it only accounts for across-crime similarity at the most discrete behavioural level. Indeed, an obvious drawback of *J* is that it is very sensitive to even slight variations in behaviours across crimes. These researchers suggested the taxonomic similarity index (Δ_s) as a more appropriate measure of across-crime similarity. This is a coefficient that can make use of higher-order behavioural information when attempting to link crimes, rather than relying solely on discrete behaviours. Thus, even if stability and distinctiveness cannot be observed at the level of discrete behaviours (e.g., punching a victim), useful levels of stability and distinctiveness may still be captured at higher levels of the behavioural hierarchy (e.g., the expression of physical aggression).

The first study to directly compare Δ_s and *J* in the context of BLA suggested that the use of higher-order behavioural information may make Δ_s more effective than *J* when attempting to discriminate between linked and unlinked crimes, at least in cases of juvenile serial sex offences (Woodhams, Grant, & Price, 2007). However, more recent investigations have found less support for Δ_s. For example, drawing on a larger sample than in the original investigation, Bennell *et al.* (2010) found that Δ_s did not significantly outperform *J* with respect to linking accuracy when analysing adult serial sexual assaults. This finding was replicated by Melnyk *et al.* (2011) who used samples of serial homicide and burglary. Both of these more recent studies suggest that one of the potential problems with Δ_s is that its emphasis on higher-order behavioural categories results in reasonably high levels of behavioural similarity. However, this is the case for crimes committed by the same offender *and* for crimes committed by different offenders. Thus, whilst the use of Δ_s appears to increase the degree of behavioural stability that can be observed in a sample of crimes, it simultaneously decreases the degree of behavioural distinctiveness that can be observed, which negatively impacts one's ability to effectively discriminate between linked and unlinked crimes (Bennell *et al.*, 2010).

In addition to its reliance on discrete behaviours, the fact that joint non-occurrences of behaviour tend to be ignored when calculating *J* may also be potentially problematic. Indeed, it is possible that stable patterns of behavioural non-occurrence may be as much a part of an offender's 'behavioural fingerprint' as the actions that they do exhibit, and thus these patterns of non-occurrences may prove useful for linking purposes. In other words, by ignoring these patterns, *J* may be excluding behavioural information that is important for BLA. A more appropriate measure of across-crime similarity may be a coefficient that *does* include joint non-occurrences in its calculation.

[1]This is not to say that methods cannot be devised to include the joint absence of behaviours (i.e., non-occurrences) in the calculation of *J*. For example, it is possible to code variables in such a way that their presence in a data set actually reflects the absence of a behaviour at a crime scene (e.g., where a positive coding of 1 indicates that cash was not stolen from a property despite it being available to steal). However, not only does such a procedure have the potential to cause confusion but it also makes assumptions that cannot be tested (e.g., that the offender actually did see the cash at the crime scene). Furthermore, this approach requires decisions to be made on the part of the data coder about which non-occurrences should be coded (a wide variety of behaviours are *not* exhibited by offenders when committing crimes, all of which could potentially be coded to reflect their absence).

DOI: 10.1002/jip

One such measure that may serve as a potential candidate is the simple matching index (S; Baroni-Urbani & Buser, 1976; Gower & Legendre, 1986; Kosman & Leonard, 2005). Similar to J, the simplicity of S is appealing. For a pair of crimes, A and B, S is

$$S = \frac{a + d}{a + b + c + d}$$

where a equals the number of behaviours common to both crimes, b and c equal the number of behaviours unique to crimes A and B, respectively, and d equals the number of behaviours absent from both crimes. The obvious difference when S is compared with J is thus the inclusion of d, joint non-occurrences of behaviour, potentially rendering S more capable of capturing patterns of behavioural stability and distinctiveness that are harder to capture using J. This may ultimately prove important for linking purposes.

The potential importance of behavioural themes in BLA

There are many reasons to suspect that high levels of behavioural stability and distinctiveness will not always be found across crimes when analysing crime scene behaviours. Indeed, a wide variety of factors can influence the expression of behaviours at a crime, including learning, maturation, and situational interruptions (Douglas & Munn, 1992). To the extent that such factors do influence the expression of crime scene behaviours, one's ability to accurately link crimes by using behavioural information will obviously be negatively affected, perhaps to the point where it is impossible to establish links.

It is particularly probable that not all crime scene behaviours will be useful for linking purposes to the same degree. Indeed, why, from a psychological perspective, should we expect that all crime scene behaviours will be exhibited in a stable and distinct fashion, even in the absence of learning, maturation, or situational interruptions? Instead, it may be more reasonable to assume that certain sets of behaviours will be exhibited in a more stable and distinct fashion across crimes compared with other sets of behaviours (see Funder & Colvin, 1991, for a discussion of similar issues in non-forensic contexts). For example, we may be more likely to observe higher levels of stability and distinctiveness when examining behaviours that reflect 'styles' of offending that are psychologically meaningful to offenders committing specific types of crimes. These 'offending styles' may relate to underlying predispositions within offenders to behave in a particular way when interacting with their victim/target (Canter, 1994).

In the case of arson, numerous studies have attempted to classify the behaviours exhibited by arsonists to determine the dominant offending styles that characterise these offenders. Much of this research has been guided by the work of Canter and Fritzon (1998). They attempted to develop a classification system of arsonists by applying the Action Systems Framework (ASF) to 175 solved arson cases committed in the UK. The ASF was first proposed by Shye (1985) to explain an individual's actions in terms of the source of an action (the emergence or birth of the event) and a target (the actualisation or manifestation of the event), both of which may be determined either internally or externally.

As a result of combining the source and target of an action, the ASF identifies four different modes of functioning: adaptive (external source, external target), conservative (external source, internal target), integrative (internal source, internal target), and expressive (internal source, external target). Canter and Fritzon (1998) hypothesised a model of arson

DOI: 10.1002/jip

corresponding to each of these modes of interaction and tested their hypothesised model by examining behaviours contained within police records of solved arsons by using multidimensional scaling (MDS).

The resulting MDS plot was supportive of the proposed model, with behaviours forming four distinct clusters or themes:

1. Instrumental person (conservative): behaviours reflecting that the fire was provoked by an emotional response associated with the breakdown of personal relationships (e.g., retaliation or revenge to perceived wrong-doing).
2. Instrumental object (adaptive): behaviours reflecting that the fire was opportunistic and served to benefit the offender (e.g., burning a car to destroy evidence of a crime).
3. Expressive person (integrative): behaviours reflecting that the fire was the result of internal distress, often with suicidal purpose (e.g., setting fire to oneself or surrounding objects to deliberately endanger life).
4. Expressive object (expressive): behaviours reflecting that the fire was a means of emotional acting out to derive attention (e.g., setting fire to buildings of symbolic significance such as churches or hospitals).

The specific behaviours included within each theme are presented in Table 1 (Canter & Fritzon, 1998).

These four themes have been found in subsequent studies of arson (e.g., Almond, Duggan, Shine, & Canter, 2005; Fritzon, Canter, & Wilton, 2001; Häkkänen, Puolakka, & Santtila, 2004; Santtila, Fritzon, & Tamelander, 2004; Wachi *et al*., 2007). This suggests that these themes may be capturing the major underlying forces that drive arsonists to behave in the way they do. If this is the case, it may be that compared to thematically unrelated arson behaviours, the behavioural indicators of these themes will be exhibited in a particularly stable and distinct fashion across crimes committed by serial arsonists. It may also be that behaviours representing some of these themes are more useful than others for the purpose of BLA.

In the only study to attempt BLA on the basis of this model, Santtila *et al*. (2004) applied principal components analysis (PCA) and discriminant function analysis (DFA) to a sample of 248 arson cases from Finland committed by 42 offenders. Content analysis of

Table 1. Arson themes and their associated behavioural features as proposed by Canter and Fritzon (1998)

Instrumental person	Instrumental object	Expressive person	Expressive object
Car	School	Residence	Business
Targeted property	Misc./uninhabited property	Self	Public building
Planned	Set fire directly	Own home	Hospital/institution
Victim known	Did not alert anyone	Multiple seats of fire	Prior arson
Victim (ex-)partner	Spree	Lives endangered deliberately	Multiple items set on fire
Prior argument	Weekday	Lives endangered by location	Drug use
Prior threats	Travelled <1 mile	Suicide note	Serial
Prior arson threats	Illegal entry		Daytime
Accelerant used	Theft		Non-specific trigger
Material brought	Multiple offenders		Remains at/returned to scene
Alcohol use	Outside		
Witness	Public view		
Trigger specific to victim			

DOI: 10.1002/jip

crime scene behaviours was followed by PCA to identify underlying themes of arson behaviour, the results of which supported Canter and Fritzon's (1998) classification model. Summary scores reflecting the resulting themes were then calculated for each case, serving as the predictor variables for DFA (with the series each case belonged to serving as the grouping variable). Discrimination accuracy was then determined by examining the DFA probabilities. Using the discriminant functions, they classified 32% of the sample as belonging to the correct series, well beyond that expected by chance (3%). Moreover, for 52% of the cases, the correct crime series (i.e., the series to which the case actually belonged to) was listed amongst the top 10 most probable series.

Although the results of this study provide some initial support for conducting BLA on the basis of the arson themes originally derived by Canter and Fritzon (1998), there is no evidence provided by Santtila *et al*. (2004) that a reliance on such themes is necessary for accurate links to be established. Indeed, similar (or even higher) levels of accuracy might have been achieved with the use of arson behaviours that are not thematically related. If this was found to be the case, it might lead one to question the value or necessity of relying on behavioural themes when conducting BLA.

In addition, we believe that there are certain methodological limitations associated with the study of Santtila *et al*. (2004), which limit our ability to assess the validity of their results. For example, the results reported by Santtila *et al*. (2004) are dependent on the threshold they selected for deciding what series a particular crime is likely to belong to (i.e., the top 10 most probable series). The degree of linking effectiveness associated with the approach to BLA by Santtila *et al*. would vary as a function of this threshold, making it difficult to assess how effective their general approach actually is (Bennell *et al*., 2009).[2] The current study will not only examine in a more direct way the value of relying on arson themes for the purpose of BLA, using both Jaccard's coefficient (*J*) and the simple matching index (*S*), it will also adopt an analytical approach that will address some of the methodological issues that we believe exist in the study reported by Santtila *et al*.

THE CURRENT STUDY

As discussed, researchers have traditionally argued that *J* is the most appropriate similarity coefficient for use in BLA because it typically omits joint non-occurrences of behaviour. Although we agree that the omission of joint non-occurrences may be advantageous in some cases, there are reasons to believe that *J* may not always be the most suitable coefficient for linking purposes. In comparison with similarity coefficients that do not ignore joint non-occurrences, *J* may potentially exclude behavioural information important for BLA. This might limit its ability to capture across-crime similarity and distinctiveness, which will in turn hinder one's ability to discriminate between linked and unlinked crimes. Therefore, before we accept *J* as the best coefficient for use in BLA, it is important to compare its performance with other coefficients not subject to these criticisms, using data from various crime types. Examining this issue was the primary goal of the current study.

[2]Santtila *et al*. (2004) manually tested a variety of thresholds, which supports our point. For example, when they lowered the threshold to include instances where the correct series is among the five most probable series, the percentage of correct classifications decreased to just over 30%. In contrast, when they raised the threshold to include instances where the correct series is among the 25 most probable series, classification accuracy increased to just over 70%.

DOI: 10.1002/jip

A secondary goal of the current study was to examine the value of relying on behavioural themes for the purpose of conducting BLA. To accomplish this goal, we will examine how linking accuracy varies across the sorts of behavioural themes proposed by Canter and Fritzon (1998), and we will compare the accuracy achieved when relying on behavioural indicators of these themes with the accuracy associated with crime scene behaviours that are not thematically related. This analysis should demonstrate whether there is added value in analysing behaviours that are reflective of underlying styles of offending, as compared with the use of crime scene behaviours more generally.

Given that we will examine how linking accuracy varies across behavioural themes for both types of similarity coefficients, our analyses will also allow us to determine whether there are potential interaction effects between the types of behavioural themes being examined and the similarity coefficients being used. Given the nature of the behaviours representing certain themes, it may be more or less important to ignore non-occurrences of certain crime scene behaviours. For example, certain behavioural themes in cases of arson may be characterised by behaviours that are more difficult to verify (e.g., because the behaviours are not able to be observed directly at the crime scene). When analysing the behaviours that represent these particular themes, it may be more beneficial to ignore joint non-occurrences of behaviour across offences (i.e., to use *J* instead of *S*).

The degree to which the different similarity measures and behavioural themes can be used to discriminate between arsons committed by different offenders will be examined with the use of a common method for studying BLA: logistic regression analysis followed by receiver operating characteristic (ROC) analysis for validation purposes (Bennell & Jones, 2005; Tonkin, Grant, & Bond, 2008; Woodhams & Toye, 2007). As discussed in more detail below, this procedure will allow us to address some of the limitations that we perceive in previous linking research on arson offences, particularly the use of linking accuracy metrics that are threshold specific.

METHOD

Sample

The data used in the current study represent a subset of data originally collected for previous arson research (i.e., Canter & Fritzon, 1998; Fritzon *et al.*, 2001). Unlike these previous studies, the current data set was restricted to cases of serial arson. Specifically, the data used in this study contain information on 42 crime scene behaviours from 114 solved arson offences committed by 37 offenders in the UK. The offence series range in length from two to nine crimes. In an attempt to be consistent with the only prior study conducted on linking serial arsons (Santtila *et al.*, 2004), series length was not restricted to a specific number of crimes per offender in this study.

In order to collect the data, Canter and Fritzon (1998) content-analysed records from various police forces across the UK to identify crime scene characteristics that could potentially be used to distinguish between arson offences (see the Appendix for a detailed explanation of each variable). All of the offences were coded for the presence (1) or absence (0) of these characteristics with each characteristic serving as a feature of one of the four arson themes as outlined previously in Table 1. Because of the nature of the data, inter-rater reliability could not be determined. However, previous research suggests that this type of crime scene data can be coded reliably (Alison & Stein, 2001; Häkkänen, Lindlöf, & Santtila, 2004).

DOI: 10.1002/jip

Procedure

Identifying behavioural themes

Given that the data used in the current study represent only a subset of the data that were originally used to establish the four-theme structure of arson, it was important to confirm that the same structure exists within this data set. To do this, the serial arson data were subjected to an MDS procedure known as Proximity Scaling (PROXSCAL), which is a subroutine in SPSS (v. 20) (Commandeur & Heiser, 1993; SPSS Inc.). As with other MDS procedures, PROXSCAL allows the user to plot a set of variables as points in space, with the distance between points indicating their level of association (variables that appear closer to one another co-occur more frequently). The degree of fit between the variable plot and the actual associations between variables is given by the measure of normalised raw stress, which ranges from 0 (perfect fit) to 1 (complete lack of fit) (Kruskal & Wish, 1978). Customarily, a stress measure under 0.10 indicates a good degree of fit.

The identification of the structure inherent in the MDS plot is based on the principle of contiguity, which states that variables tied to a common theme will be more highly associated than variables associated with different themes (Canter & Heritage, 1990). Therefore, the former variables will be closer in proximity within the MDS plot. On this basis, it was possible to delineate regions within the plot, which reflected different themes. Kuder–Richardson 20 (K-R 20) coefficients were calculated for each of the themes that could be identified in the plot. This is an index of internal reliability, which is essentially the equivalent of Cronbach's alpha, but for dichotomous data (Kuder & Richardson, 1937).

Calculating J and S

The dependent variable in the present study is the dichotomous classification of whether the same offender or different offenders committed a pair of arsons (i.e., whether a crime pair is actually linked or unlinked). The independent variables are the continuous across-crime similarity scores (using *J* and *S*) calculated for each pair of crimes in the sample with the use of different sets of variables: all 42 crime scene behaviours and subsets of these characteristics representing one of the themes of arson identified by the PROXSCAL analysis.

It is expected that a higher degree of similarity will be exhibited across crimes committed by the same offender. Thus, we predict that, compared with unlinked crimes, linked crimes will be characterised by higher across-crime similarity scores (given the way in which *S* is calculated, values associated with this similarity coefficient should be higher than the values associated with *J* for both linked and unlinked crimes). With respect to the values of the similarity scores found for the linked and unlinked crimes, it is currently unclear how the analysis of all 42 crime scene characteristics will compare with the analysis of thematically related behaviours.

A specially designed computer program known as B-LINK (Bennell, 2002) was used to calculate *J* and *S* for all possible crime pairs. The program accepts as input a Microsoft *Excel* file that contains a series of dichotomously coded variables pertaining to the presence or absence of each of the relevant behaviours in each of the arson offences. On the basis of this information, *J* and *S* were first calculated with the use of all 42 behaviours. This process was then repeated several times, each time using a different subset of behaviours representing one of the themes of arson identified in the MDS plot. Each time the program is run, B-LINK provides as output a Microsoft *Excel* file containing the *J* and *S* values for

DOI: 10.1002/jip

all crime pairs (the program also indicates which crime pairs are actually linked). These files were then imported into SPSS (v. 20) to provide the data for the subsequent regression and ROC analyses, as described in the succeeding discussions.

Developing the regression models

Once the various *J* and *S* values were calculated, logistic regression analyses were performed with the use of each similarity coefficient separately to examine the extent to which the various linking features could be used to accurately link arsons committed by the same offender. The use of logistic regression analysis in this case is appropriate given the dichotomous nature of the dependent variable (Tabachnick & Fidell, 2007). Specifically, for both *J* and *S*, individual logistic regression analyses were conducted for all of the 42 behaviours combined and behaviours from each of the identified themes. This was carried out to examine the extent to which these simple regression models could be used to accurately predict linked arson pairs. Forward stepwise logistic regression was then conducted in order to determine the optimal combination of themes for predicting whether crime pairs are linked. This optimal model was also developed separately for *J* and *S*.

In order to reduce the potential for bias that stems from developing and testing the regression models on the same sample of data, split-half validation was used (Efron, 1982). This procedure involves splitting the sample randomly in half to form development samples (upon which the regression models for *J* and *S* were developed) and validation samples (upon which the regression models for *J* and *S* were validated). Results from the validation samples should be indicative of how the logistic regression models might perform on arsons that have yet to be observed. The results from the validation samples are what will be focused on in this study.

Evaluating the regression models

Receiver operating characteristic analysis was then carried out on the validation samples to evaluate the ability of the various logistic regression models to accurately classify arson pairs as linked or unlinked. Briefly, this procedure allows one to plot the probability of hits on the linking task (determining that a crime pair is linked when it is) against the probability of false alarms (determining that a crime pair is linked when it is not) across each of the possible thresholds that can be used to make these decisions (i.e., across various levels of across-crime similarity). When these points are plotted on a graph (hits on the *y*-axis and false alarms on the *x*-axis) and the points are connected, the result is a concave curve. The area under the curve (*AUC*) can be used as a measure of linking accuracy (Bennell, 2005).

The *AUC* can range from 0 (total inaccuracy) to 1 (total accuracy), although most ROC curves fall above the positive diagonal on the graph, which represents an *AUC* of 0.50 (chance accuracy). Importantly, given that the *AUC* represents the location of the entire ROC curve in the ROC graph, this measure provides an index of linking accuracy (for both *J* and *S*) that is not specific to any single decision threshold. In this way, the *AUC* provides a more valid measure of linking accuracy (for other advantages associated with ROC analysis, see Bennell, 2005; Bennell *et al.*, 2009).

All ROC analyses were performed with the use of the ROC analysis subroutine in SPSS (v. 20). The precise data entered into the ROC analyses were the estimated linkage probabilities for every arson pair comprising the validation samples, along with the data representing whether the arson pairs were actually linked or unlinked. As indicated, the *AUC* derived from ROC analysis provides a numerical index of discrimination accuracy. According to commonly accepted guidelines, *AUC*s between 0.50 and 0.70 represent

DOI: 10.1002/jip

low levels of accuracy, *AUC*s between 0.70 and 0.90 represent good levels of accuracy, and *AUC*s between 0.90 and 1.00 represent high levels of accuracy (Swets, 1988). An examination of the *AUC*s corresponding to the models developed with the use of *J* versus *S* allow for a direct comparison of the ability of these coefficients to accurately link arson offences across the various subsets of behaviours being tested.

The use of thematically unrelated behaviours for comparison purposes

To determine if there is value associated with the use of behavioural themes for the purpose of conducting BLA, it was deemed important to compare the results of the theme-based analyses with analyses of behaviours that are thematically unrelated (i.e., not indicative of any single behavioural theme). To accomplish this, the same steps outlined earlier were followed for randomly selected subsets of behaviours so that we could compare the results from analyses of these behaviours to the results that emerged when each of the behavioural themes were examined.

For each comparison, the number of behaviours selected was determined by the number of behaviours included in the theme that was being assessed, thus allowing us to control for this potentially important variable. For example, the instrumental person theme that was identified in the PROXSCAL analysis consists of 13 behaviours. Thus, the sample of behaviours used for comparison purposes in this case included 13 behaviours. To increase reliability, 10 different draws of random behaviours were made for each comparison (e.g., 10 draws of 13 behaviours for the comparison with behaviours comprising the instrumental person theme).

Using the same split-half validation procedure described earlier, we constructed separate logistic regression models for each random draw and we subjected the predicted probabilities resulting from the validation samples to ROC analysis. We compared the combined *AUC*s resulting from these analyses with the *AUC*s resulting from the analyses of the behavioural themes to determine if statistically significant differences emerged. These comparisons should tell us something about the potential value of using behavioural themes for the purpose of BLA.

To calculate the combined *AUC*s for each set of random samples, and the standard errors (SE) and 95% confidence intervals (CI_{95}) associated with these *AUC*s, we used the procedure outlined by Borenstein, Hedges, and Rothstein (2007). Specifically, we calculated the combined *AUC* for each set of 10 random samples by multiplying the *AUC* for each sample by the inverse variance of the sample, summing these weighted *AUC*s, and then dividing that sum by the sum of the weights (inverse variances). We calculated the SE for this combined *AUC* by taking the square root of the variance of the combined *AUC* (1 divided by the sum of the variances). Finally, we calculated the CI_{95} for the combined *AUC* by adding (for the upper limit) or subtracting (for the lower limit) the SE of the combined *AUC*, multiplied by 1.96, to the combined *AUC*.

RESULTS

PROXSCAL analysis

Figure 1 presents the PROXSCAL plot of arson behaviours in two dimensions (along with the K-R 20 values for each theme). The normalised raw stress score associated with this plot is 0.08, indicating a good degree of fit between the plot and the actual associations that exist between the variables. What the PROXSCAL plot and K-R 20 values make clear is that the original themes identified by Canter and Fritzon (1998) hold up reasonably well

DOI: 10.1002/jip

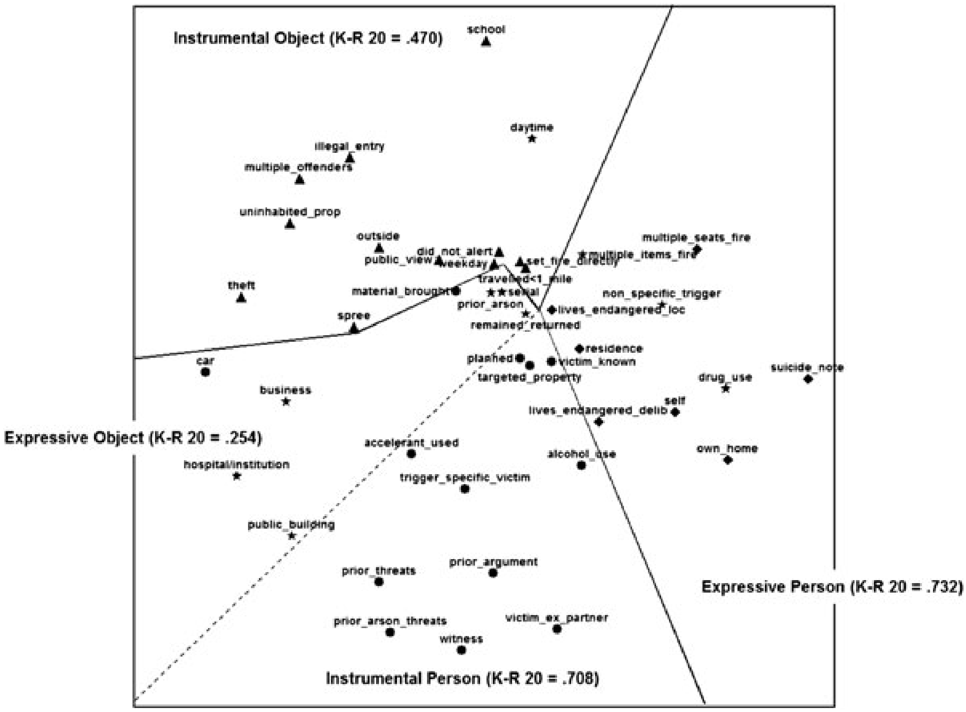

Figure 1. PROXSCAL plot of serial arson behaviours. Behaviours belonging to Canter and Fritzon's (1998) original themes are identified with the use of different symbols (● = instrumental person; ♦ = expressive person; ▲ = instrumental object; ★ = expressive object).

in this sample of serial arsons. In particular, the behaviours in the original instrumental object, instrumental person, and expressive person themes cluster together in the PROXSCAL plot and are associated with either moderately low (0.470 in the case of the instrumental object theme) or relatively high K-R 20 values (0.708 and 0.732 for the instrumental person and expressive person themes, respectively). Although the majority of behaviours in the original expressive object theme also cluster together, the K-R 20 value associated with this theme is low (0.254). This K-R 20 value suggests that this theme is not represented well in the current sample of serial arsons, which we reflect in the PROXSCAL plot by the use of a dashed partition line. On the basis of this analysis, a decision was made to remove the expressive object behaviours from further analysis. Despite the moderately low K-R 20 value associated with the instrumental object theme, it was retained for further analysis given the results of the PROXSCAL analysis.

Descriptive analysis

Prior to conducting the main analyses, a descriptive analysis of the similarity scores was conducted (Tables 2 and 3). Separately for *J* and *S*, descriptive statistics were calculated across the distributions of linked crime pairs and unlinked crime pairs for all behaviours combined (with the exception of expressive object behaviours) and for the three arson themes. Significance tests of differences in these distributions were also conducted to determine

Table 2. Descriptive statistics for the linked and unlinked distributions of *J* and *S* scores for the sampled behaviours

	Range		Median		Mean (SD)	
Variables	L	UL	L	UL	L	UL
Jaccard's						
All behaviours	0.23–1.00	0.05–0.92	0.83	0.38	0.75 (0.24)	0.40 (0.13)
Instrumental person	0.00–1.00	0.00–1.00	1.00	0.29	0.69 (0.41)	0.32 (0.27)
Instrumental object	0.00–1.00	0.00–1.00	0.83	0.43	0.76 (0.26)	0.45 (0.20)
Expressive person	0.00–1.00	0.00–1.00	1.00	0.17	0.61 (0.46)	0.27 (0.31)
Simple matching						
All behaviours	0.56–1.00	0.44–0.98	0.95	0.71	0.90 (0.11)	0.71 (0.08)
Instrumental person	0.31–1.00	0.15–1.00	1.00	0.69	0.94 (0.11)	0.73 (0.15)
Instrumental object	0.33–1.00	0.25–1.00	0.92	0.67	0.87 (0.15)	0.68 (0.14)
Expressive person	0.14–1.00	0.00–1.00	1.00	0.71	0.91 (0.17)	0.69 (0.21)

Note. L, linked crime pairs ($n = 176$); UL, unlinked crime pairs ($n = 6265$); SD, standard deviation.

whether the crimes committed by the same offender were characterised by higher levels of behavioural stability (i.e., had higher *J* and *S* scores) than the crimes committed by different offenders. As tests of normality indicated that all the *J* and *S* distributions for linked and unlinked crime pairs were significantly different from a normal distribution (all p's < 0.001), non-parametric tests were used to compare the similarity scores across these distributions.

As shown in Table 2, greater behavioural similarity is evident in the higher mean *J* and *S* coefficients for the linked crime pairs than the unlinked crime pairs for all behaviours and across each arson theme. As expected, for all the linked and unlinked distributions, the mean *S* is consistently higher than the mean *J*. As illustrated in Table 3, significance tests comparing the linked versus unlinked distributions revealed that the across-crime similarity scores were higher for linked crime pairs than for unlinked crime pairs, regardless of whether *J* (all p's < 0.001) or *S* (all p's < 0.001) was used. The effect sizes reported in Table 3 also support this, particularly in the case of *S* where large effects were consistently found. Nevertheless, as is evident in the range of similarity scores across the distributions for both *J* and *S* in Table 2, the substantial overlap in the linked versus unlinked distributions does suggest that it is probably not possible to achieve perfect linking accuracy when using any of the behaviours extracted from the current sample of offences.

Table 3. Significance tests of the differences in *J* and *S* coefficients for the linked versus unlinked distributions

	J			*S*		
Variables	Wilcoxon	*p*-value	Effect size	Wilcoxon	*p*-value	Effect size
All behaviours	−10.87	<0.001	0.82	−10.84	<0.001	0.82
Instrumental person	−5.76	<0.001	0.43	−9.75	<0.001	0.73
Instrumental object	−11.14	<0.001	0.84	−9.84	<0.001	0.74
Expressive person	−4.97	<0.001	0.37	−9.00	<0.001	0.68

Note. Effect size $= r = z/\sqrt{N}$ (0.00–0.30 = small effect; 0.30–0.50 = moderate effect; 0.50– = large effect).

DOI: 10.1002/jip

Logistic regression analysis

Separately for *J* and *S*, a series of simple logistic regression models were initially developed to determine the predictive accuracy of (1) all arson behaviours combined (with the exception of expressive object behaviours), (2) instrumental person behaviours, (3) instrumental object behaviours, and (4) expressive person behaviours. Forward stepwise logistic regression analysis was then conducted in an attempt to identify the optimal combination of arson themes for achieving the highest degree of linking accuracy for both *J* and *S*.

Results of the initial simple regressions are provided in Table 4. Model coefficients confirm that, relative to unlinked arsons, linked arsons are consistently characterised by higher levels of across-crime similarity when all arson behaviours are combined and in each of the arson themes separately. These findings hold regardless of whether *J* or *S* is employed as the similarity coefficient. Similarly, the Wald's and chi-square tests indicate that all regression models accurately predict whether arsons are linked or unlinked (all *p*'s 0.001). With that said, the corresponding R^2 values suggest that some models fit the data better than others, for both *J* and *S*. The model containing all arson behaviours fits the data the best for both *J* ($R^2 = 0.46$) and *S* ($R^2 = 0.48$). In contrast, the model containing behaviours from the expressive person theme is the poorest fitting model when using *J* ($R^2 = 0.11$) or *S* ($R^2 = 0.19$).

Results of the forward stepwise logistic regression analyses are provided in Tables 5 (for *J*) and 6 (for *S*). As indicated by the chi-square tests, the optimal model for both *J* and *S* accurately predicts whether arsons are linked or unlinked (all p's < 0.001). Furthermore, as expected, the optimal model for both *J* and *S* outperformed the single-predictor models displayed in Table 4, with the exception of the models containing all behavioural information. As further outlined in Tables 5 and 6, the optimal models differ depending on whether *J* or *S* is employed. Specifically, although all themes are retained in the optimal model for *S*, the expressive person theme is omitted in the optimal model for *J*, indicating that the behaviours from the expressive person theme do not increase the predictive accuracy of this model above and beyond that achieved when behaviours from the other two themes are taken into account. The contribution of each individual theme to the overall predictive accuracy of the optimal models for *J* and *S* also differed. For *J*, the best predictor of linkage status was behaviours from the instrumental object theme, followed by behaviours from the instrumental person theme. Conversely, for *S*, the best predictor of linkage status was behaviours from the instrumental person theme, followed by behaviours from the instrumental object and expressive person themes, respectively.

ROC analysis

As displayed in Table 2, *S* is capable of achieving higher across-crime similarity scores for linked arsons compared to *J*. However, *S* also generated higher across-crime similarity scores for unlinked arsons as well. Similarly, although the logistic regression models constructed from the development samples indicate that arson behaviours can be used to accurately predict which arsons are linked, the predictive accuracy of these models varies as a function of the behaviours included in the model, as well as whether *J* or *S* is used as the similarity coefficient.

ROC analysis was thus used to evaluate the relative discrimination accuracy of the two similarity coefficients across the different regression models developed on the development samples. To perform these analyses, we used the regression models presented in Tables 4–6

DOI: 10.1002/jip

Table 4. Summary of the simple logistic regression results for each similarity coefficient

	J				*S*			
Model	*B* (SE)	Wald (*df*)	χ^2 (*df*)	R^2	*B* (SE)	Wald (*df*)	χ^2 (*df*)	R^2
All behaviours	11.05 (0.71)	240.66 (1)	374.51 (1)	0.46	23.43 (1.54)	232.51 (1)	394.41 (1)	0.48
Instrumental person	4.42 (0.39)	130.57 (1)	151.51 (1)	0.19	16.87 (1.37)	151.01 (1)	260.43 (1)	0.33
Instrumental object	6.67 (0.51)	168.84 (1)	196.69 (1)	0.25	10.56 (0.91)	133.82 (1)	177.23 (1)	0.22
Expressive person	2.67 (0.30)	81.84 (1)	85.00 (1)	0.11	8.81 (0.92)	91.09 (1)	148.15 (1)	0.19

Note. Linked crime pairs ($n = 176$); unlinked crime pairs ($n = 6265$); SE, standard error; χ^2, model chi-square; *df*, degrees of freedom; R^2, Nagelkerke index; all p's < 0.001.

DOI: 10.1002/jip

Table 5. Results of the forward stepwise logistic regression displaying the optimal model of arson themes for predicting linkage classification using *J*

Theme	*B* (SE)	Wald (*df*)	χ^2 (*df*)	R^2
Instrumental object	5.03 (0.52)	94.09 (1)	263.49 (2)	0.33
Instrumental person	2.96 (0.38)	59.81 (1)		

Note. Linked crime pairs ($n = 176$); unlinked crime pairs ($n = 6265$); SE, standard error; χ^2, model chi-square; *df*, degrees of freedom; R^2, Nagelkerke index; all p's < 0.001.

Table 6. Results of the forward stepwise logistic regression displaying the optimal model of arson themes for predicting linkage classification using *S*

Theme	*B* (SE)	Wald (*df*)	χ^2 (*df*)	R^2
Instrumental person	10.25 (1.38)	55.32 (1)	366.76 (3)	0.45
Instrumental object	6.32 (0.95)	44.03 (1)		
Expressive person	5.20 (0.95)	248.80 (1)		

Note. Linked crime pairs ($n = 176$); unlinked crime pairs ($n = 6265$); SE, standard error; χ^2, model chi-square; *df*, degrees of freedom; R^2, Nagelkerke index; all p's < 0.001.

to calculate estimated probabilities for every arson pair in the validation samples. These probabilities were then used to construct separate, cross-validated ROC curves for each logistic regression model, including the optimal models for *J* and *S*. Because of space constraints, only the ROC graphs associated with the optimal theme-based model for *J* and *S* were provided for illustrative purposes (Figures 2 and 3). The results of all the ROC analyses are presented in Table 7, including the results from the analyses of thematically unrelated behaviours (K-R 20 values for each set of behaviours are provided).

Consistent with the logistic regression analyses, the ROC analyses indicated that all models resulted in overall levels of accuracy beyond that expected by chance (all p's < 0.001). Using Swets' (1988) guidelines, we achieved good to high levels of discrimination accuracy across all the models for *J* and *S*, with the highest levels of discrimination accuracy (although not necessarily to a significant degree) achieved for the models containing more behavioural information (e.g., models based on all behaviours) as opposed to those containing less behavioural information (e.g., models based on individual themes).

In terms of the predictive accuracy of the models based on individual arson themes, the instrumental object theme ($AUC = 0.82$) is the most accurate for predicting linkage status when using *J*, followed by the instrumental person theme ($AUC = 0.77$), and the expressive person theme ($AUC = 0.72$), respectively. However, all of the CIs associated with these *AUC*s overlap, suggesting that none of the *AUC*s are significantly different from one another. In contrast, for *S*, the instrumental person theme is associated with the highest level of predictive accuracy ($AUC = 0.90$), followed by the instrumental object theme ($AUC = 0.83$) and the expressive person theme ($AUC = 0.82$), respectively. As indicated by the CIs, the instrumental person theme achieves *significantly* higher levels of predictive accuracy when compared with the expressive person theme. No other differences are significant.

In terms of the relative discrimination accuracy of the two similarity coefficients, the *AUC*s associated with the models developed with the use of *S* are consistently higher than the *AUC*s associated with the models developed with the use of *J*, regardless of what

DOI: 10.1002/jip

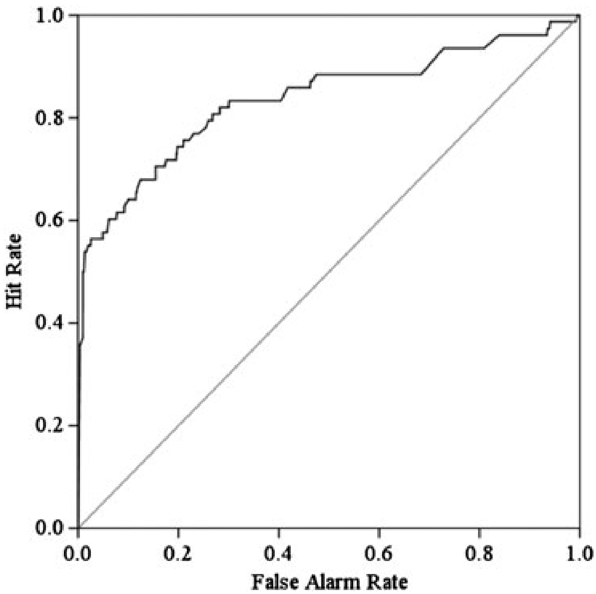

Figure 2. ROC graph for the optimal theme-based model using *J* as the similarity coefficient ($AUC = 0.84$).

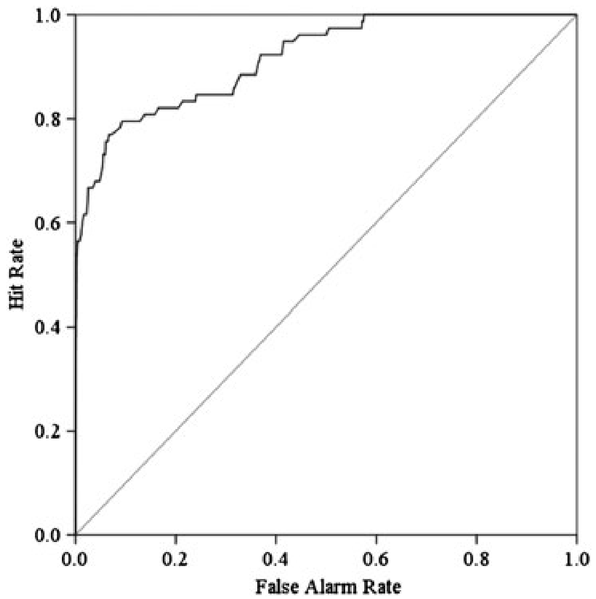

Figure 3. ROC graph for the optimal theme-based model using *S* as the similarity coefficient ($AUC = 0.92$).

regression model is considered. However, a comparison of the CIs for the two coefficients reveals that S ($CI_{95} = 0.87–0.94$) outperforms J ($CI_{95} = 0.69–0.84$) to a significant degree for the instrumental person theme only ($AUC = 0.90$, $AUC = 0.77$, respectively). Although the CIs associated with the optimal models for J and S overlap only slightly, the CIs associated with all remaining models overlap to a substantial degree.

DOI: 10.1002/jip

Table 7. Summary of ROC analyses for each similarity coefficient across all regression models

		J			*S*		
Model	K-R 20	*AUC*	SE	CI_{95}	*AUC*	SE	CI_{95}
Optimal	—	0.84	0.03	0.78–0.90	0.92	0.02	0.88–0.95
All behaviours	0.465	0.89	0.02	0.85–0.94	0.93	0.02	0.89–0.96
Instrumental person	0.708	0.77	0.04	0.69–0.84	0.90	0.02	0.87–0.94
Unrelated (IP)	0.204	0.84	0.02	0.80–0.88	0.88	0.02	0.84–0.92
Instrumental object	0.470	0.82	0.03	0.76–0.88	0.83	0.03	0.78–0.89
Unrelated (IO)	0.162	0.83	0.02	0.79–0.87	0.87	0.02	0.83–0.90
Expressive person	0.732	0.72	0.04	0.64–0.79	0.82	0.03	0.76–0.87
Unrelated (EP)	0.144	0.72	0.01	0.70–0.74	0.80	0.01	0.78–0.83

Note. Linked crime pairs ($n = 176$); unlinked crime pairs ($n = 6265$); K-R 20, Kuder–Richardson 20; *AUC*, area under the curve; SE, standard error; CI_{95}, 95% confidence intervals.

Finally, with respect to the analysis of the thematically unrelated behaviours, the results in Table 7 indicate that there are no significant differences between the average *AUC*s calculated for these behaviours and the *AUC*s associated with the behavioural themes these behaviours were compared with. Indeed, reasonably high levels of accuracy were found for each set of unrelated behaviours, suggesting that relying on behavioural themes, at least in the present study, is not necessary to achieve a high level of linking success. In line with the analysis of behavioural themes, the analyses of the thematically unrelated behaviours also indicates that regression models based on *S* perform as well as models that use *J* (in one case, EP, the model based on *S* significantly outperforms the model based on *J*).

DISCUSSION

Although substantial overlap existed between the distributions of similarity scores for linked and unlinked crimes, the results of this study demonstrated that crimes committed by the same offender tended to be associated with higher levels of across-crime similarity compared with crimes committed by different offenders. This was true for every analysis, regardless of what combination of behaviours was used to assess across-crime similarity, or whether *J* or *S* was used as the similarity coefficient. Given this set of results, it was unsurprising that the logistic regression models we developed fit the data well and were highly predictive of linkage status for the crime pairs under examination. Indeed, on the basis of current guidelines for interpreting *AUC*s, each of the regression models was associated with good to high levels of predictive accuracy (Swets, 1988).

These general results provide support for the core assumptions underlying BLA; that is, that serial arsonists will exhibit crime scene behaviours in a relatively stable, but distinct fashion across the crimes they commit, thus allowing crimes committed by the same offender to be distinguished from crimes committed by different offenders. This finding is consistent with previous studies that have examined other crime types (e.g., Bennell & Jones, 2005; Melnyk *et al*., 2011; Woodhams & Toye, 2007) and suggests that serial arsonists are predisposed to behave in a particular way when committing their crimes. To some extent at least, these behavioural tendencies appear to be unaffected by situational variations that might exist across arsons. However, it should be noted that these results may be

DOI: 10.1002/jip

partly because only solved arsons were examined in this study. As Bennell and Jones (2005) have argued previously, crimes may be solved, at least in part, because the offenders committing the crimes exhibit high levels of stability and distinctiveness. In addition, the fact that the current data set excludes all of the one-off arsons that real crime analysts would have to sift through is a limitation of the study. Not only does this limit the ecological validity of the research but the levels of linking accuracy that are achieved with such a sample are also likely to be higher than what can actually be achieved in naturalistic settings (Woodhams & Labuschagne, 2012).

Jaccard's coefficient versus the simple matching index

One of our primary goals in this study was to examine whether the use of *J* as a measure of across-crime similarity would prove to be more effective than the use of *S*. The majority of previous studies that have examined BLA have relied on *J*, and many researchers have argued that, because of its simplicity and its treatment of joint non-occurrences, it should be the similarity coefficient of choice. Analyses presented in the current study lead us to question this recommendation. Although both *J* and *S* were able to accurately distinguish linked from unlinked crimes to a significant degree, the descriptive statistics indicated that *S* consistently produced higher levels of discrimination accuracy than *J*. Although the results from the ROC analyses indicate that these differences between *J* and *S* were not statistically significant in the majority of cases, *S* did significantly outperform *J* when the instrumental person theme was examined (the CIs associated with the *AUC*s did not overlap in this case) and almost did when the expressive person theme was examined (the CIs associated with the *AUC*s barely overlap in this case). Although it is obviously important to replicate these results before drawing strong conclusions, these findings regarding *J* and *S* are nonetheless very interesting.

Although there are a number of possible explanations for the trends favouring *S* over *J*, the most obvious difference between *S* and *J* is that *S* includes in its calculation joint non-occurrences of behaviour. Thus, the most probable explanation for the generally superior performance of *S* lies in the fact that additional information is included in the regression models that were based on this coefficient; information that was valuable for establishing accurate links between the crimes included in this sample. This result, combined with the comparative analyses involving behavioural themes, suggests that linking accuracy might be maximised by including as much behavioural information as possible in the analyses of serial crimes.

In the current study at least, the result of including more behavioural information in the analysis of the crimes (by using *S* versus *J*) was that the similarity scores emerging from the analyses were more fine grained. Specifically, a wider variety of similarity scores were produced when using *S* versus *J*, and these scores were more evenly distributed across the possible range of similarity scores (from 0 to 1). From a purely mathematical perspective, this presumably allows for a greater degree of differentiation between crimes committed by different offenders, especially when many crimes pairs are included in the sample under examination (i.e., if only 1 similarity score resulted from the analysis, no differentiation would be possible, and as the number of different similarity scores increases, a higher degree of differentiation becomes possible). As mentioned, further support for the idea that 'more behavioural information is better' comes from our analysis of behavioural themes in arson offences.

DOI: 10.1002/jip

Thematically related versus unrelated behaviours

Beyond examining the potential role of similarity coefficients in BLA, a secondary goal of the present study was to investigate the predictive accuracy of regression models based on arson themes and to understand their possible utility in BLA. Specifically, we sought to determine whether there was any value (in terms of improving discrimination accuracy) in basing prediction models on the themes of arson found by Canter and Fritzon (1998), which have been subsequently replicated. It makes sense that if these themes reflect important ways that offenders interact with targets in cases of arson, then the behaviours representing these themes might be exhibited in a particularly stable and distinct fashion across crimes. Indeed, if these thematically based behaviours reflect underlying predispositions within arsonists to behave in a certain way, then presumably the level of linking accuracy that could be accomplished by drawing on these thematically based behaviours would be greater than the accuracy that could be achieved by drawing on combinations of thematically unrelated behaviours.

To ensure that the themes identified by Canter and Fritzon (1998) generalised to this study, a PROXSCAL analysis was conducted and K-R 20 values were calculated. This analysis revealed a model of serial arson that was conceptually consistent with Canter and Fritzon's original action systems model. Indeed, reasonably strong evidence for the previous model was found, with the exception perhaps of the expressive object theme. A number of behaviours predicted to fall within that theme were not highly associated with other expressive object behaviours and the K-R 20 value associated with these behaviours was low (0.254). Although it is not entirely clear why the expressive object theme was not represented well in the current sample of arsons, this finding presumably relates to the fact that, unlike the larger sample of arsons from which our data were drawn, the current sample is composed solely of serial arsons. Specifically, the movement of some of the expressive object items into other themes, particularly the expressive person theme, might speak to the degree to which these variables are associated with serial versus single offenders. Relative to the one-off offenders that made up a large portion of Canter and Fritzon's sample, it appears that serial arsonists may be more likely to have a non-specific trigger, take drugs, and set fire to multiple items and that for serial arsonists, these behaviours may be associated with communicative behaviours targeted at themselves (expressive person) rather than external objects (expressive object).

As a result of the PROXSCAL analysis, and the extremely low K-R 20 value for behaviours that were originally associated with the expressive object theme, the behaviours related to this theme were excluded from the linkage analysis. In general, the results from that analysis suggest that the individual arson themes can be used to link serial arsons, in that good to high levels of discrimination accuracy were found across the single-theme models for both J (AUCs ranging from 0.72 to 0.82) and S (AUCs ranging from 0.82 to 0.90). Similarly, good to high levels of discrimination accuracy were found when the themes were combined, for both J ($AUC = 0.89$) and S ($AUC = 0.93$). However, when using J, forward stepwise regression revealed that only two of the three themes held significant predictive value, with the expressive person theme not uniquely adding to the predictive accuracy of the optimal theme-based model. For S, all three themes were found to uniquely contribute to the predictive accuracy of the model. It is not clear why the expressive person theme lacks predictive power when relying on J, and a scan of the behaviours included in that theme does not reveal any obvious clues. The fact that the expressive person theme is associated with the highest degree of internal consistency of all three themes (K-R 20 = 0.732) suggests that the issue is not one of reliability.

DOI: 10.1002/jip

On the basis these results, our study appears to be consistent with previous examinations of BLA in arson cases (Santtila *et al*., 2004), and it provides general support for the themes of arson originally proposed by Canter and Fritzon (1998) and found by researchers since then (e.g., Almond *et al*., 2005). With that said, the models using all of the behavioural information (not separated by theme) performed as well as the optimal theme-based models for both J and S, suggesting that separating the behaviours into themes may not be a necessary prerequisite for successful linking to occur. The analysis of the randomly selected, thematically unrelated behaviours directly confirms this. The K-R 20 values associated with these randomly selected behaviours indicated that these behaviours exhibit lower levels of internal consistency than the thematically related behaviours. Despite this, the *AUC*s generated from these unrelated behaviours were not significantly different from the *AUC*s based on thematically based behaviours, even after controlling for the number of behaviours entered into the analysis. This represents reasonably convincing evidence that little might be gained in BLA by considering the degree to which the behaviours being analysed represent important offending styles or behavioural themes. With that said, there are many different ways of examining the role of behavioural themes in the context of BLA (e.g., Salfati & Bateman, 2005), and it would be unwise to dismiss the value of behavioural themes for the purpose of BLA until more thorough research has been conducted.

CONCLUSION

Although there are limitations with the current study, the results are potentially important for our understanding of BLA, particularly with respect to serial arson. Although any definite conclusions concerning the relative effectiveness of S at this time are premature, results of this study suggest that S is a suitable similarity measure for use in BLA. Indeed, this index appears to be as suitable (if not more suitable) than the more commonly used J. In addition, whereas our results suggest that behaviours representing arson themes can be used to accurately link arson offences, the results that emerged from our analysis of thematically unrelated behaviours suggests that it is not necessary to rely on thematically based behaviours for linking purposes. Given the difficulties in solving arson crimes, it is crucial to determine if the results reported here can be replicated with the use of other arson data sets. Of course, future research should also continue to examine other similarity measures, beyond S and J, and to investigate the value of other categorisation schemes, beyond the themes that we tested here. Such research would help to determine not only which similarity measure should be used in BLA but also which crime scene behaviours to use in order to maximise linking accuracy.

REFERENCES

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.

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

Almond, L., Duggan, L., Shine, J., & Canter, D. (2005). Test of the arson action system model in an incarcerated population. *Psychology, Crime & Law*, *11*, 1–15.

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

DOI: 10.1002/jip

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

Bennell, C. (2005). Improving police decision making: General principles and practical applications of receiver operating characteristic analysis. *Applied Cognitive Psychology*, *19* (9), 1157–1175.

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. (2009). Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. *Legal and Criminological Psychology*, *14*, 293–310.

Bennell, C., Gauthier, D., Gauthier, D., Melnyk, T., & Musolino, E. (2010). The impact of data degradation and sample size on the performance of two similarity coefficients used in behavioural linkage analysis. *Forensic Science International*, *199*, 85–92.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Introduction to meta-analysis. Retrieved from: http://www.meta-analysis.com/downloads/Meta%20Analysis%20Fixed%20vs%20Random%20effects.pdf

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., & Fritzon, K. (1998). Differentiating arsonists: A model of firesetting actions and characteristics. *Legal and Criminological Psychology*, *3*, 73–96.

Canter, D. V., & Heritage, R. (1990). A multivariate model of sexual offence behaviour: Developments in 'offender profiling'. *Journal of Forensic Psychiatry*, *1*, 185–212.

Commandeur, J. J. F., & Heiser, W. J. (1993). Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices. Leiden, NL: Department of Data Theory, University of Leiden.

Douglas, J. E., & Munn, C. (1992). Violent crime scene analysis: Modus operandi, signature, and staging. *FBI Law Enforcement Bulletin*, *61*, 1–10.

Efron, B. (1982). The jackknife, the bootstrap, and other re-sampling plans. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Emeno, K., Bennell, C., Melnyk, T., & Jones, N. J. (2008). Linking serial rapes: A comparison of similarity coefficients. Poster presented at the annual conference of the American Psychology-Law Society, Jacksonville, Florida, USA.

Fritzon, K., Canter, D. V., & Wilton, Z. (2001). The application of an action systems model to destructive behaviour: The examples of arson and terrorism. *Behavioral Sciences & the Law*, *19*, 657–690.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, *60*, 773–794.

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.

Häkkänen, H., Lindlöf, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in a sample of Finnish stranger rapes. *Journal of Investigative Psychology and Offender Profiling*, *1*, 17–32.

Häkkänen, H., Puolakka, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in arsons. *Legal and Criminological Psychology*, *9*, 197–214.

Jaccard, P. (1908). Nouvelle recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *44*, 223–270.

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.

DOI: 10.1002/jip

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills, CA: Sage Publications.
Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*, 151–60.
Liebetrau, A. M. (1983). Measures of association. Beverly Hills, CA: Sage Publications.
Melnyk, T., Bennell, C., Gauthier, D., & Gauthier, D. (2011). Another look at across-crime similarity coefficients for use in behavioural linkage analysis: An attempt to replicate Woodhams, Grant, and Price (2007). *Psychology, Crime & Law*, *17*, 359–380.
Salfati, C. G., & Bateman, A. L. (2005). Serial homicide: An investigation of behavioural consistency. *Journal of Investigative Psychology and Offender Profiling*, *2*(2), 121–144.
Santtila, P., Fritzon, K., & Tamelander, A. L. (2004). Linking arson incidents on the basis of crime scene behavior. *Journal of Police & Criminal Psychology*, *19*, 1–16.
Shye, S. (1985). Nonmetric multivariate models for behavioural action systems. In D. Canter (Ed.), Facet theory approaches to social research. New York, NY: Springer Verlag.
Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.
Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics. New York, NY: Pearson Education.
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.
Wachi, T., Watanabe, K., Yokota, K., Suzuki, M., Hoshino, M., Sato, A., & Fujita, G. (2007). Offender and crime characteristics of female serial arsonists in Japan. *Journal of Investigative Psychology and Offender Profiling*, *4*, 29–52.
Woodhams, J., & Labuschagne, G. (2012). A test of case linkage principles with solved and unsolved serial rapes. *Journal of Police and Criminal Psychology*, *27*(1), 85–98.
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.
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.

APPENDIX
CONTENT DICTIONARY

All crime-scene variables listed below are dichotomous. That is, they have values based on the presence (1) or absence (0) of each category of behaviour. A description of the categorisation scheme is given below.

Instrumental person

1. Car/vehicle

Any type of vehicle that is used for transportation of goods or people is coded as car/vehicle, including bicycles and boats.

2. Targeted property

If there is any evidence to suggest that a specific property was fired for a particular reason, then this is coded as targeted. In other words, it must be apparent or readily inferred that the offender(s) would not have set fire to anything other than that object. For example, if the offender travelled any great distance to the target, bypassing other buildings with similar properties. Also, if the victim was known, and the fire followed a dispute, then it can be inferred that the victim was targeted.

DOI: 10.1002/jip

3. Planned

For example, if materials were brought to the scene, such as petrol or matches, then this would suggest planning. Also, if the individual made an effort to avoid detection, for example, wearing gloves when handling petrol containers.

4. Victim known

This would generally go along with targeting and includes institutions or governing bodies that the offender has been involved with, for example, a school he or she has attended, or council owned property if he or she is a council tenant.

5. Victim (ex-)partner

This variable would also be coded as present if the offender fires property belonging to someone close to his or her (ex-)partner, for example, a family member or new partner. The rationale for this is that that person would not have been targeted were it not for their association with the (ex-)partner.

6. Prior violence/argument with victim

This refers to any dispute, preferably heated, occurring within a reasonable time-frame (usually not more than a month) of the arson.

7. Prior threats towards victim

This includes verbal or physical threats of an overt or implicit nature.

8. Prior threat of arson

If the offender has made any threatening remarks with reference to fires, even in an abstract sense such as, ‘I once knew someone who’s house burned down’, or ‘be careful you don’t leave matches lying around; someone might get hold of them’, then these count as threats of arson.

9. Accelerant used

Again, there is usually mention of an accelerant in the fire investigator’s report.

10. Material brought

Anything that the offender brought for the specific purpose of starting or accelerating the fire would be coded as this. It is important that the material is something that he would not normally be carrying, for example, matches or a cigarette lighter is ambiguous particularly if the individual is a smoker.

11. Alcohol use

The offender may not state that he or she has consumed alcohol, but if a police officer or witness mentions that the offender appeared to be drunk or smelled of alcohol, then this is coded.

12. Witness

If the fire setting takes place in front of another person who is not a willing participant, that is, explicitly or implicitly does not condone the act, then he or she is coded as a witness. It is important that the offender knows that the other person is present; therefore, a passer-by who happens to see the fire setting would not be coded as a witness.

DOI: 10.1002/jip

13. Trigger specific to victim

If the fire setting occurs immediately following, or within a reasonable period of an argument or other, usually emotional trigger, and is targeted at a specific person or property, then that is a victim-specific trigger.

Instrumental object

1. School

A fire that occurs in any area of an educational establishment would be coded as school. For example, if a fire is set to waste bins outside the school, this would be coded as both miscellaneous and school.

2. Misc./uninhabited/derelict property

Misc. applies to items fired that were not inside a property, for example, a rubbish bin or park bench. However, anything that is fired inside a property will be coded as that property, for example, a rubbish bin inside a school is coded as school. Uninhabited or derelict properties can be both commercial and residential properties that are currently not in use.

3. Set fire

If the offender has actually placed a burning object (e.g., match or lighted piece of paper) to the property he or she wants to fire, then this is a set fire. If the burning object has been thrown, for example, a petrol bomb, or burning pieces of paper have been dropped onto an object from above, then this is not coded as a set fire.

4. Did not alert anyone

If the offender left the scene of the fire without subsequently alerting either the fire brigade or any other person, then this variable is coded.

5. Spree

If the offender sets more than one fire with a gap of no more than 24 hours, then this is coded as spree fire setting.

6. Weekday

A weekday is classified as being between 00:01 on a Monday and 16:59 on a Friday.

7. Distance travelled less than 1 mile

This is coded if the offence occurs less than a mile from where the offender either lives or was based immediately before the fire setting. In other words, if the offender was at school all day, and then set a fire on the way from school to home, then the important measurement would be from the school to the offence rather than from the offence to the home.

8. Forced/illegal entry

If the offender were required to make some effort to obtain entry to the fired property, then this would be coded as forced/illegal entry. Also, if the offender could be said to be trespassing, for example, in a hay barn which has open access, this variable would be coded as present.

DOI: 10.1002/jip

9. Theft from premises

This variable would be coded if any property were taken either before or after the fire setting.

10. More than one offender

The other individual need not be instrumental in the actual setting of the fire, for example, they could be acting as a lookout. If another person is present during the fire setting and they do not actually try to stop the offender, then they are counted as a co-offender.

11. Outside

If the fired object is itself outside, or the individual sets fire to a house by throwing a fire bomb or inserting lighted material through the letter box, then this is coded as being outside.

12. Public view

If the fire setting occurs in a place and time where the offender could potentially be seen by passers-by, then this is coded as being in public view. If the fire setting occurs at a time where there are unlikely to be other people around, but in a place which usually has CCTV, for example, a car park, then this would also be coded as public view.

Expressive person

1. Residential

This refers to a property that at the time of the fire was being used for residential purposes. If the property was derelict or uninhabited (as opposed to simply unoccupied) at the time, then it would not be coded as residential. An exception to this would be an uninhabited flat contained within a block of flats some of which were inhabited. Also, a property that was known to contain 'squatters' would be classified as residential.

2. Self

If an individual starts a fire in his or her own home, and then makes no attempt to leave or alert anyone, then this is coded as self.

3. Own home

This is coded in addition to residential and/or self.

4. Multiple seats of fire

This refers to initial ignition points of the item(s) fired. For example, if a house is fired by pouring petrol in one room and holding a match to a curtain in another room, then the fire would be coded as having multiple seats. The numbers of seats of a fire are usually stated in the investigating fire officer's report.

5. Lives endangered deliberately

If the offender knew that the property was occupied at the time of the fire and made no attempt to alert the occupants, then this is coded.

DOI: 10.1002/jip

6. Lives endangered by location

A fire in any residential property, or building attached to a residence which is not completely detached, has the potential to endanger lives.

7. Suicide note

This is coded not only in the presence of an actual suicide note but also if the offender has alerted anyone prior to the fire of their intention or wishes to commit suicide.

Expressive object

1. Business

Again, the property would have to currently be in use as business premises. A disused unit on an industrial estate would not be coded as business. Other exceptions include allotments and pigeon lofts, which would be coded as uninhabited.

2. Public building

This includes any type of building to which the public have access, for example, library, church, town hall, law courts, and police station.

3. Hospital/institution

Again, if the fire is set on any part of the institution's grounds, then it is coded as institution.

4. Prior arson

This is coded if the offender has set any fires prior to the current offence. Although this variable is duplicated in the Offender Variable list, it is included here in order to identify which other actions are associated with prior arson.

5. Multiple items fired

This refers to the objects that have actually ended up on fire, rather than secondary objects used to start that fire. In other words, if multiple waste bins or skips are fired, then this variable would be coded as present, but if multiple bits of newspaper are used to set fire to one waste bin, then this variable would not be coded.

6. Drug use

This refers to any recreational, that is, non-prescription drug, including solvents during the commission of the offence.

7. Serial

If the offender sets more than one fire with a gap of more than 24 hours, then this is coded as serial fire setting. However, if the gap is a matter of years rather than weeks or months, then this would not be serial, but the offender would be coded as having prior arson in his history.

8. Daytime

If the offence occurs during daylight hours, this is classified as daytime. Note that this will depend on the time of year; for example, 21:00 in July would be daytime whereas in November it would not.

DOI: 10.1002/jip

9. Non-specific trigger

If the fire setting occurs immediately following, or within a reasonable period of an argument or other, usually emotional trigger, and there is no obvious targeting of a specific person or property, then that is a non-specific trigger.

10. Remains at/returned to scene

This refers to any cases where the offender remains at the scene of their crime, returns to the scene whilst the fire is still burning, or returns to the scene of a previous crime to set another fire.

DOI: 10.1002/jip