

Linking Property Crime Using Offender Crime Scene Behaviour: A Comparison of Methods

Authors: Tonkin, M.^{1*}, Lemeire, J.², Santtila, P.^{3,4}, & Winter, J. M.^{5,6}

* Corresponding author

¹ Dr. Matthew Tonkin, Department of Criminology, University of Leicester, 156 Upper New Walk, Leicester, LE1 7QA, United Kingdom. E-Mail: mjt46@le.ac.uk

² Dr. Jan Lemeire, Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

³ Prof. Pekka Santtila, Department of Psychology and Logopedics, Åbo Akademi University, 20500 Turku, Finland.

⁴ NYU Shanghai, Shanghai, China.

⁵ Mr. Jan M. Winter, National Police agency, Zoetermeer, the Netherlands.

⁶ Department of Clinical and Life Span Psychology (KLEP), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

CONFLICT OF INTEREST STATEMENT: The authors have no conflicting interests to declare in relation to this submission.

ACKNOWLEDGEMENTS: The authors would like to acknowledge the assistance of the British Academy who provided a grant that supported the work described within this manuscript.

Linking Property Crime Using Offender Crime Scene Behaviour: A Comparison of Methods

Abstract

This study compared the ability of seven statistical models to distinguish between linked and unlinked crimes. The seven models utilized geographical, temporal, and Modus Operandi information relating to residential burglaries ($n = 180$), commercial robberies, ($n = 118$), and car thefts ($n = 376$). Model performance was assessed using Receiver Operating Characteristic (ROC) analysis and by examining the success with which the seven models could successfully prioritize linked over unlinked crimes. The regression-based and probabilistic models achieved comparable accuracy and were generally more accurate than the tree-based models tested in this study. The Logistic algorithm achieved the highest Area Under the Curve (AUC) for residential burglary (AUC=0.903) and commercial robbery (AUC=0.830) and the SimpleLogistic algorithm achieving the highest for car theft (AUC=0.820). The findings also indicated that discrimination accuracy is maximized (in some situations) if behavioural domains are utilized rather than individual crime scene behaviours, and that the AUC should not be used as the sole measure of accuracy in behavioural crime linkage research.

Keywords: crime linkage; decision-support tools; logistic regression; classification tree analysis; probabilistic models; offender crime scene behaviour

Introduction

Behavioural crime linkage (BCL) seeks to address serial offending by linking crimes based on the fact they share similar (but distinctive) offender crime scene behaviours (Woodhams, Bull, & Hollin, 2007). BCL is utilized by law enforcement agencies around the world (Snook, Luther, House, Bennell, & Taylor, 2012) and can yield a range of benefits (provided it is performed accurately). These benefits include providing the police with a greater quantity and quality of evidence with which to prosecute offenders (Grubin, Kelly, & Brunsdon, 2001) and helping the police to work in a more streamlined and efficient way (Woodhams et al., 2007).

Over the last 15 years academic research has sought to develop statistical methods for linking crimes that could support BCL in practice, with the hope that such methods might (in the future) be incorporated within computerized decision-support tools (e.g., see Tonkin et al., 2017). These tools would be able to process vast quantities of crime scene information in a quick and efficient manner, highlighting to law enforcement practitioners those crimes that are most likely to be linked and providing a summary of behavioural similarities and differences between the various crimes (e.g., Canter & Youngs, 2008; Grubin et al., 2001; Oatley, Ewart, & Zeleznikow, 2006)¹. Not only would this help analysts to process more cases in less time, but it may also help to increase BCL accuracy by reducing the cognitive load on analysts.²

Before such tools can be developed, however, a variety of statistical methods must be tested across a range of crime types and datasets to identify the most appropriate statistical algorithms to underpin these tools. While a growing body of research has developed different statistical approaches to BCL, there are key limitations to this research that must be addressed before researchers can begin developing decision-support tools with confidence. It is the aim of this paper to address some of these key limitations.

The Empirical Research on BCL

¹ Note that such tools would not be a replacement for human analytical expertise; instead, they would assist analysts to make decisions (with the analyst retaining ultimate control over decision-making).

² Cognitive load has been shown to lead to decision-making errors in a range of contexts (see Adcock, 2000).

A growing body of research has sought to investigate the most reliable and accurate methods of using offender crime scene behaviour to identify linked crime series. More than 30 empirical research papers have been published in this area, examining a range of person- and property-oriented crimes, including commercial and personal robbery, residential and commercial burglary, arson, sexual assault, homicide, and car theft; using data from a range of countries, including the US, Canada, the UK, Finland, South Africa, Italy, and Japan (see Woodhams & Bennell, 2014, for an overview).

Within this research, several different statistical methods have been tested to determine whether they have the potential to underpin BCL decision-support tools. These include logistic regression (e.g., Bennell & Jones, 2005; Burrell, Bull, & Bond, 2012; Melnyk, Bennell, Gauthier, & Gauthier, 2011; Tonkin, Grant, & Bond, 2008; Woodhams & Labuschagne, 2012; Woodhams & Toye, 2007), classification tree analysis (e.g., Tonkin, Woodhams, Bull, Bond, & Santtila, 2012), and Bayesian analysis (e.g., Kringen, 2014; Porter, 2014; Salo et al., 2013; Winter et al., 2013). Many of these studies have, however, only tested a single statistical method and have not compared different approaches. This is problematic because discrimination accuracy can vary significantly depending on the choice of statistical model, as demonstrated by findings from a variety of domains, including medical diagnosis (e.g., Ayer et al., 2010), risk prediction in homicide (e.g., Neuilly, Zgoba, Tita, & Lee, 2011), predicting prison inmate misconduct (e.g., Ngo, Govindu, & Agarwal, 2015), vegetation mapping (Prasad, Iverson, & Liaw, 2006), and chemical modelling (Svetnik et al., 2003).

Furthermore, changes to the statistical model used can fundamentally change the way in which predictions are made and how they are displayed to the user. Logistic regression, for example, tends to adopt a 'one-size-fits-all' approach to prediction that applies the same predictor variables in the same way to all cases (Monahan et al., 2001). While this has the advantage of simplicity, it does not allow for any flexibility in predictive methods. Classification tree analysis, on the other hand, allows for different predictive methods to be used for different

sub-groups of cases (Steadman et al., 2000), thereby providing a degree of flexibility in decision-making that logistic regression does not (Tonkin et al., 2017). But, such flexibility can lead to classification tree models becoming very complicated, which not only makes them difficult for practitioners to understand and use but can also lead to over-fitting. Over-fitting happens when the model becomes so closely matched to the data that were used to develop it that it cannot make accurate predictions when applied to new datasets (e.g., Liu, Yang, Ramsay, Li, & Coid, 2011; Thomas et al., 2005). As such, the model loses any practical value. It is, therefore, important to directly compare different statistical methods, which (in the context of BCL) will help to determine the most appropriate statistical methods upon which to base the future development of decision-support tools.

Fortunately, a small number of studies have compared different statistical methods in the context of BCL. Tonkin et al. (2012), for example, compared logistic regression and iterative classification tree analysis in terms of their ability to distinguish between linked and unlinked crimes using residential burglary and car theft data. While this study found comparable accuracy between regression- and tree-based models, over-fitting was observed for the classification tree models (particularly those developed for the burglary sample), leading to significant reductions in predictive accuracy when the tree-based models were applied to new data. The regression-based models did not experience such problems. Other studies have compared regression, classification tree, and Bayesian-based analysis using samples of sexual assault (Tonkin et al., 2017) and breaking and entering offences (Porter, 2014). Tonkin et al. (2017) found that a Bayesian-based approach led to the highest level of predictive accuracy, followed by classification tree analysis and logistic regression; whereas Porter (2014) found comparable levels of accuracy when comparing these three types of statistical approach³.

Limitations of the BCL Research

³ It should be noted that the exact methods used by Tonkin et al. (2017) and Porter (2014) differ (see Tonkin et al., 2017 for more information).

Despite a small number of comparative studies, there is a distinct lack of research comparing different statistical approaches to BCL. For example, no research has yet to compare regression, classification tree, and Bayesian methods with samples of car theft and robbery. Without such research it is difficult for researchers to make recommendations regarding the most useful statistical methods for linking these types of crime, which hinders the development of decision-support tools and limits the value of existing research for police practitioners.

A further limitation is the lack of replication research. For example, there are very few studies that have examined BCL with commercial robbery and car theft offences, and the potential value of Bayesian analysis has yet to be explored with these crime types. Replication is crucial in any applied area of research because reliable practical recommendations cannot be generated until research findings are shown to be robust and generalizable.

Another important limitation is concerned with the way in which much previous research has utilized offender crime scene information when developing statistical approaches to BCL. Rather than utilizing the individual crime scene behaviours as predictors of linkage status, much previous research has chosen to combine several different behaviours into so-called behavioural domains (e.g., combining behaviours associated with controlling the victim in a robbery), partly because of the computational challenges this presents and the need for specialized software to perform these analyses. This approach is problematic because it forces any subsequent analysis to rely on a composite variable rather than allowing the analysis to work with the (larger) number of individual crime scene behaviours that are available. As explained above, allowing greater complexity into statistical models does not always lead to greater accuracy, but it can do in some situations (and almost certainly would lead to a very different approach to linking crimes). Given that statistical approaches such as regression, classification tree analysis, and Bayesian analysis can work equally well with the individual crime scene behaviours, it is surprising that previous research has not compared these two approaches to determine if one approach leads to greater linkage accuracy. Answering this question would have important

implications for both future research on BCL and the development of BCL decision-support tools.

Current Study

It is the aim of the current study to address these limitations by comparing seven different statistical models that derive from three statistical approaches that are commonly used to investigate the behavioural linking of crimes: binary logistic regression analysis (including the SimpleLogistic and Logistic algorithms in Weka 3.7.10; Hall et al., 2009), classification tree analysis (including the J48, Logistic Model Tree and RandomForest algorithms) and Bayesian/probabilistic models (including the BayesNet and Naïve Bayes algorithms in Weka). These approaches will be applied to three datasets: residential burglaries committed in Finland (dataset 1); car thefts committed in the UK (dataset 2); and commercial robberies committed in Finland (dataset 3). The current study, therefore, contributes to the literature in several important ways:

- 1) It provides much-needed research that compares three of the most promising statistical approaches currently available to BCL researchers (never before done for robbery and car theft crimes). This will help to answer important methodological and applied questions regarding which is the most appropriate statistical method/s for linking crimes using offender behaviour.
- 2) It explores the novel application of statistical methods to the task of BCL.
- 3) It seeks to replicate previous BCL research (particularly in relation to commercial robbery, where only one published study currently exists). This will help to build a more robust evidence base from which practical recommendations can be developed to guide the behavioural linking of crimes in practice.

4) It compares a domain-based approach to BCL with an approach that relies on individual crime scene behaviours.

Method

Data

Residential burglary data.

The residential burglary data consisted of 160 residential burglaries committed by 80 serial burglars in the Greater Helsinki region of Finland⁴ between 1990 and 2001. These data were originally collected as part of a previous project (Laukkanen, Santtila, Jern, & Sandnabba, 2008; Santtila, Ritvanen, & Mokros, 2008) and have been utilized in previous studies of BCL (Tonkin, Santtila, & Bull, 2012; Tonkin et al., 2012). Two crimes per offender were randomly selected from the total number of offences that they had committed during this time period. Previous research has considered it necessary to select a constant number of offences per offender so as to prevent highly prolific offenders with unusually consistent or inconsistent offence behaviour having an undue influence on the findings (Bennell, 2002).

For each burglary a range of behavioural data were recorded, including the location of the crime (stored as an x, y coordinate), the time and day of the week the offence occurred, the type of property burgled, the method of entry, the offender's search behaviour once inside the property, and the type and cost of property stolen (see Tonkin et al., 2012a, for further details). Apart from the location information, the data were stored in a binary format (1 = present in the crime; 0 = absent).

Car theft data.

The car theft data consisted of 376 vehicle theft crimes committed by 188 serial car thieves in Northamptonshire, UK, between January 2004 and May 2007. These data have been

⁴ The greater Helsinki region of Finland covers an area of approximately 815 km² that contains the capital of Finland, Helsinki, and the neighbouring cities of Espoo and Vantaa.

used previously to investigate BCL (Tonkin et al., 2012b). Two crimes per offender were randomly selected from the total number of offences that they had committed during this time period (Bennell, 2002).

For each car theft a range of behavioural data were recorded, including the location of the crime (an x, y coordinate), the type of car that was stolen, the age of the vehicle, the time and day of the week the vehicle was stolen, how the vehicle was entered and started, and the physical state in which the vehicle was recovered (see Tonkin et al., 2008, for further details). Apart from the location information, the data were stored in a binary format.

Commercial robbery data.

The commercial robbery data consisted of 118 commercial robberies committed by 59 serial robbers in the Greater Helsinki region of Finland between 1991 and 2002. These data were originally collected as part of a previous project (Laukkanen, 2007), but have never before been used to investigate BCL. Two crimes per offender were randomly selected from the total number of offences that they had committed during this time period (Bennell, 2002).

For each commercial robbery a range of behavioural data were recorded, including the location of the crime, the type of business robbed, the time of day and day of the week the offence occurred, whether a disguise was worn, weapon use, the number of offenders, use of violence, language used, and the type and cost of property stolen (contact the first author for the full list of behaviours).

Procedure

First, a number of behavioural domains were created for each dataset, with the domains containing clusters of individual offender behaviours. For the burglary data, six behavioural domains were created: 1) **target characteristics** (e.g., the type of property burgled, time of day the offence was committed etc.); 2) **entry behaviours** (e.g., the point and method of entry); 3) **internal behaviours** (e.g., search behaviour); 4) **property stolen** (e.g., cash, keys etc.); 5) **inter-crime distance** (the geographical distance in kilometers between two offence locations); and 6)

temporal proximity (the number of days between two offences). These domains were derived from previous BCL studies of burglary and the behaviours were placed into domains according to their placement in previous research (e.g., Bennell & Jones, 2005; Markson, Woodhams, & Bond, 2010; Tonkin et al., 2012a,b).

For the car theft data, four behavioural domains were created: 1) **target selection choices** (e.g., the type and age of the vehicle stolen, time of day of the theft etc.); 2) **target acquisition behaviour** (e.g., the method and point of entry); 3) **disposal behaviour** (e.g., whether property was stolen from the vehicle and the condition of the vehicle when recovered, e.g., burnt out, damaged etc.); and 4) **inter-crime distance** (in kilometers). These domains were similar to those used in previous research (e.g., Tonkin et al., 2008).

For the commercial robbery data, eight behavioural domains were created: 1) **target selection choices** (e.g., type of business premises targeted, time of day the robbery occurred); 2) **planning behaviours** (e.g., whether a disguise or gloves were worn, whether a getaway vehicle was used etc.); 3) **control behaviours** (e.g., whether a weapon was used and what type, number of offenders, type of language and violence used etc.); 4) **property stolen** (e.g., cash, jewelry etc.); 5) **inter-crime distance** (in kilometers); 6) **temporal proximity** (in days); 7) **difference in the value of items stolen** (between the two crimes in each pair, in Euro); and 8) **approach behaviours** (e.g., number of clients/employees present, the manner in which the offender approached the victim). These domains and the placement of behaviours therein were based on previous BCL research (Woodhams & Toye, 2007).

Next, these data were used to create linked and unlinked crime pairs. The linked pairs contained two crimes committed by the same offender and the unlinked pairs contained two crimes committed by different offenders. From the 160 residential burglary crimes, 80 linked and 12,640 unlinked burglary pairs were created. From the 376 car thefts, 188 linked and 70,312 unlinked car theft pairs were created. From the 118 commercial robberies, 59 linked and 6,844

unlinked commercial robbery pairs were created. This represented every possible linked and unlinked pair that could be created from the three datasets.

For every crime pair an inter-crime distance, temporal proximity value, and/or a Jaccard's coefficient were calculated for each behavioural domain. Jaccard's coefficient can range from 0.00 (indicating no behavioural similarity) to 1.00 (indicating complete behavioural similarity). This coefficient has been used by numerous previous researchers of BCL (e.g., Bennell & Jones, 2005; Markson et al., 2010; Melnyk et al., 2011; Woodhams & Labuschagne, 2012). The Jaccard's coefficients, inter-crime distance, and temporal proximity values were input as the predictor variables in the subsequent analyses.

Data Analysis

Three types of statistical model were used in the current study: logistic regression models, classification tree models, and probabilistic models. All models were built using Weka 3.7.10 (Hall et al., 2009). The accuracy of the models was evaluated based on a 10-fold cross-validation method.

Logistic regression models.

In the current study, the two types of logistic regression model available in Weka were tested: SimpleLogistic and Logistic. The SimpleLogistic function in Weka used LogitBoost with simple regression functions as base learners to fit the logistic models (Landwehr, Hall, & Frank, 2005; Sumner, Frank, & Hall, 2005). LogitBoost is a commonly-used algorithm that iteratively fits a linear function to the data while reducing bias and variance. The second logistic method, called Logistic in Weka (le Cessie & van Houwelingen, 1992), built a multinomial logistic regression model using ridge estimators, which help to ensure more reliable estimates of the regression model parameters (e.g., logit/Beta coefficients).

Classification tree models.

We tested three different types of classification tree model on the data. The first (and most simple) algorithm, called J48 in Weka (Quinlan, 1993), builds a simple decision tree with

the C4.5 algorithm. The second tree model utilized was a Logistic Model Tree, which is a classification tree but with logistic regression functions at the leaves (Landwehr et al., 2005). The inclusion of such an algorithm is particularly useful in this context, as it allows us to examine the classification performance of a hybrid model that combines two of the three most commonly used statistical approaches in the BCL literature (logistic regression and classification tree models). This provides a useful methodological comparison with the J48 algorithm, which does not utilize logistic regression functions and only uses tree-based methods to classify linked and unlinked crime pairs. There are, however, potential problems with the two previously described approaches, as they develop just one decision tree from the data, which is then used to make classification decisions. As explained by Breiman (1994), this can lead to unstable models that are over-fitted to the data. Given that over-fitting has been discussed as a potential problem when applying classification tree models to the BCL task (Tonkin et al., 2012b), it was useful to consider a third method, the RandomForest model (Breiman, 2001), which utilized the procedure known as bagging (Breiman, 1994). Bagging involves splitting the data into multiple sub-samples and developing a separate decision tree for each sub-sample. The final classification decision is based on the classification that is reached by the majority of trees. By combining multiple decision tree models, bagging can provide a more stable and accurate model for generating classifications (Breiman, 1994). As far as the authors are aware, this particular tree-based approach has not previously been explored in the context of BCL.

Probabilistic models.

We tested two different types of probabilistic model available in Weka: BayesNet and Naïve Bayes. BayesNet models the probabilistic relations among the variables with a graph, while the Naïve Bayes Classifier assumes that the evidence variables are probabilistically independent. Although this assumption is rarely valid in practice, it results in a simple and very practical approach, which yields good results (Domingos & Pazzani, 1997).

Measuring discrimination accuracy.

The above analyses were each used to produce a predicted probability value for every crime pair in the three datasets, indicating the predicted likelihood that the two crimes in each pair were committed by the same person (0.00 = a low predicted likelihood; 1.00 = a high predicted likelihood). These values were subsequently used to construct 21 ROC curves that gave an insight into the accuracy with which the various statistical models could distinguish between linked and unlinked crime pairs (one ROC curve for each of the seven statistical methods across all three datasets, $3 \times 7 = 21$).

The Area Under the Curve (AUC) produced by ROC analysis can range from 0 (indicating perfect negative prediction) to 1.00 (indicating perfect positive prediction), with a value of 0.50 indicating a chance level of accuracy. Typically, AUC values of 0.50 indicate no discrimination accuracy, values of 0.70 – 0.80 are considered acceptable, values of 0.80 - 0.90 are excellent, and values of 0.90 - 1.00 are outstanding (Hosmer & Lemeshow, 2000). The ROC analyses were based on 25 iterations of analysis, whereby different subsets of the data were selected and used to calculate a ROC curve, with the final AUC value representing the mean AUC across all 25 iterations.

Results

As indicated by the ROC analyses reported in Table 1, the regression-based models were the most successful across all three datasets, with the Logistic algorithm achieving the highest AUC for residential burglary (AUC = 0.903) and commercial robbery (AUC = 0.830) and the SimpleLogistic algorithm achieving the highest AUC for the car theft data (AUC = 0.820). When the AUCs in this table were statistically compared, the regression-based models (Logistic and SimpleLogistic) and probabilistic models (BayesNet and Naïve Bayes) were comparable in terms of the AUC, but the tree-based models (J48, Logistic Model Tree and RandomForest) achieved statistically lower AUCs ($p < .05$). The only exception was the Logistic Model Tree with the

commercial robbery data, which was statistically comparable to regression-based and probabilistic models.

[INSERT TABLE 1 ABOUT HERE]

Within the BCL literature, it has been suggested that statistical methods such as those tested in the current study could be applied to a police crime dataset and used to rank order crime pairs that are behaviourally similar, thereby providing the analyst with an evidence-based way of focusing their attention on those crimes that are most likely to be linked (e.g., Tonkin et al., 2017). To test whether this was possible using the current data, all crime pairs were sorted from highest to lowest probability value (based on the values given by the most successful statistical method for each dataset). The number of linked crime pairs (i.e., true positives), the percentage of linked pairs, and the cumulative percentage of linked pairs was then calculated for different ranges within that list (e.g., the number of linked pairs within the top 10 ranks was calculated). These findings are presented in Table 2 and summarized visually in Figure 1. If the findings have the potential to assist BCL practice, we would expect the percentage of linked crime pairs to be higher at low ranks in the list and to decrease as one progresses further down this sorted list. From Table 2 and Figure 1, we can see that this general trend is supported for the commercial robbery and residential burglary datasets, but not for the car theft dataset. For example, 60% of the linked burglary pairs were within the top 3.931% of that prioritized list and 65.7% of the linked robbery pairs were within the top 7.243%. However, only 5.8% of the linked car thefts were within the top 500 ranks (0.709% of the prioritized list) and only 57.2% were within the top 10,000 ranks (14.184% of the prioritized list).

[INSERT TABLE 2 AND FIGURE 1 ABOUT HERE]

The above analyses indicate the predictive value of all behavioural domains when used in combination to distinguish between linked and unlinked crime pairs. It is also useful to summarize the predictive value of each domain separately, as this can help to identify those types of offender crime scene behaviour that are most and least suitable for the purposes of BCL

(see Table 3). These findings were produced by entering each domain's coefficient of similarity (i.e., Jaccard's coefficient, inter-crime distance, and temporal proximity) directly into a ROC analysis as the predictor/independent variable, while linkage status (linked/unlinked) served as the outcome/dependent variable. The behavioural domains with the highest discrimination accuracy across all three datasets were temporal proximity and/or inter-crime distance, with these two domains achieving statistically larger AUCs than all other domains tested ($p < .05$). The exceptions to this were in the burglary data where internal behaviours (AUC = 0.757) were statistically comparable to temporal proximity (AUC = 0.818) and in the robbery data where planning (AUC = 0.680) and control behaviours (AUC = 0.702) were statistically comparable to inter-crime distance (AUC = 0.745).

[INSERT TABLE 3 ABOUT HERE]

As explained in the introduction, few previous studies have compared discrimination accuracy when using individual crime scene behaviours compared to using a domain-based approach that combines multiple behaviours into a single predictor. Table 4 presents the findings when using individual crime scene behaviours as predictors, which can be compared to the findings in Table 1 (where a domain-based approach was used). In general, there were few statistical differences in discrimination accuracy between the two approaches. There was, however, a non-significant trend towards greater accuracy when using the domains, with a higher AUC observed in 15 out of 21 comparisons (three of which were statistically significant, $p < .001$ - the Logistic algorithm with the burglary and robbery data and the Logistic Model Tree with the robbery data).

Discussion

BCL can potentially play an important role in the investigation of serial offenders by providing the police with a greater quantity and quality of evidence with which to prosecute offenders, as well as helping the police to work in a more efficient and streamlined manner (Grubin et al.,

2001; Woodhams et al., 2007). These potential benefits have led to specialist units being established around the world to support BCL and to a growing body of empirical research that has sought to identify statistical methods that might be able to support these units. There are, however, important limitations to this research that limit the theoretical and practical value of this work. The current study aimed to address these issues by comparing seven statistical methods (across three datasets) in terms of their ability to distinguish between linked and unlinked crime pairs.

Comparing Different Statistical Methods in terms of Discrimination Accuracy

The regression-based and probabilistic models tested in this study were found to yield the highest accuracy when distinguishing between linked and unlinked crime pairs (based on the AUC). In general, these models achieved statistically larger AUCs than the tree-based models, which is perhaps unsurprising given the problems of over-fitting that are sometimes associated with classification tree models (e.g., Tonkin et al., 2012b; Liu et al, 2011; Thomas et al., 2005). Although, it is important to note that regression and probabilistic models do not always outperform tree-based models (Porter, 2014).

There are a multitude of reasons to explain the lack of consistency across studies in terms of which statistical models achieve the highest level of discrimination accuracy. These reasons include the use of different crime types, the use of data from different countries (where different behavioural information might be collected, using different data collection protocols etc.), and the use of different statistical methods (e.g., the methods of classification tree analysis, logistic regression, and Bayesian analysis differ across studies). In the absence of more focused research to explore these issues, it should be concluded that there is not robust, generalizable evidence to support the superiority of any statistical method over others. It is likely that the answer to this question will depend on the particular linkage context in which the algorithms are applied, with the best approach depending on a number of factors, including (but not limited to) the crime type under investigation, the base rate frequencies of linked versus unlinked crimes, and the types of

offender behaviour available for use when linking. Furthermore, it is important to recognise that the current study has compared the statistical methods purely on the basis of discrimination accuracy; while important there are a range of other important factors that should be considered when determining which is the optimal approach to use in practice (e.g., which methods provide the most easy-to-understand and easy-to-implement analytical output). The answers to questions like this should be the focus of future research. While such questions currently remain unanswered, the choice of linking algorithm is clearly an important issue that may have significant real-world consequences in terms of how (in)accurately crimes are linked during live investigations, as demonstrated by the wide-ranging AUC values achieved in this study (burglary AUCs ranged from 0.474 – 0.903; car theft AUCs ranged from 0.442 – 0.820; and robbery AUCs ranged from 0.553 – 0.830).

Comparing a Domain-Based Approach to BCL with the Use of Individual Crime Scene Behaviours

For the first time this study examined the impact on discrimination accuracy of adopting a domain-based approach to BCL (Table 1) compared to utilizing individual offender behaviours to make linkage predictions (Table 4). In general, there was a trend towards greater accuracy when using domains (although, it should be noted that only three of these comparisons were statistically significant). When developing BCL decision-support tools and conducting BCL research in the future, researchers should, therefore, carefully consider whether they use behavioural domains or individual offender behaviours, as this decision may in some situations (but not all) impact on discrimination accuracy. It's possible that these findings are a reflection of the fact that situational circumstances can change from one crime to the next. As such, an offender may not be afforded with the opportunity to repeat the exact same behaviour, but they may be able to demonstrate a different behaviour that represents the same underlying behavioural theme (e.g., Canter, 2000). The use of behavioural domains captures consistency at this more thematic level whereas individual offender behaviours do not.

Comparing Approaches to Measuring BCL ‘Performance’

In this study two different approaches were used to examine BCL ‘performance’. The first approach relied on the AUC produced by ROC analysis (which is traditionally thought of as the ‘gold standard’ approach to measuring BCL accuracy; Bennell et al., 2014). The second approach involved using the statistical methods to create prioritized/ranked lists of crime pairs and then examining to what extent these lists had successfully prioritized linked over unlinked pairs. These two methods provide very different approaches to evaluating BCL ‘performance’, and also highlight why researchers should not rely solely on the AUC to examine accuracy. Consider the car theft findings, for example, where the highest AUC achieved was 0.820 (based on the SimpleLogistic algorithm). These findings imply that BCL is theoretically possible, but when examining how successfully this algorithm was able to prioritize linked crimes it is clear that considerable difficulties were experienced (see Table 2 and Figure 1). Had we relied solely on the AUC, this would have led to a misleading impression of BCL accuracy with car theft crimes.

The reason for this apparent contradiction in the findings is the class imbalance problem (see Longadge, Dongre, & Malik, 2013, for a general discussion of this issue). In this context, when the number of unlinked crime pairs vastly outweighs the number of linked pairs, a statistical model will tend to give a very low predicted likelihood of linkage for all pairs in the sample. Consequently, one is likely to make very accurate predictions when identifying unlinked crime pairs (true negative predictions/correct rejections), and by virtue of the fact there are so many unlinked pairs this will often lead to a high AUC regardless of how accurately the model can identify linked pairs (true positive predictions/hits). It is, therefore, vital that future researchers do not rely solely on the AUC when examining BCL accuracy and that they employ additional measures of ‘performance’. The approach taken in the current study is one example of an additional measure that can be used, but several others exist (e.g., see Tonkin et al., 2017, for another example).

Comparing Different Behavioural Domains in terms of Discrimination Accuracy

This study also compared different types of offender crime scene behaviour in terms of their ability to support BCL (Table 3). The inter-crime distance and/or temporal proximity achieved the highest level of discrimination accuracy across all three datasets, which is a finding consistent with many previous studies (see Bennell et al., 2014). Various explanations have previously been given for these findings, all of which apply equally to the current findings. For example, it has been suggested that the decision over where and when to commit crimes is more under the control of the offender than other behaviours (such as what to steal, whether violence is used etc.), which depend to some extent on situational circumstances. Offender-driven behaviours that are relatively less influenced by the offending situation are, therefore, predicted to be more consistent from one crime to the next; thereby making them more suited to BCL (e.g., Bennell & Jones, 2005). Alternatively, it might be that geographical and temporal information are more easily/objectively recorded on police databases than behaviours such as whether a property was searched in a tidy or untidy manner. The superior performance of inter-crime distance and temporal proximity may, therefore, be a result of differences in data quality rather than due to inherent differences in the consistency and/or distinctiveness of offender behaviour (Bennell & Jones, 2005).

The superiority of inter-crime distance and temporal proximity lends support to several seminal theories of offender behaviour (such as rational choice theory, routine activities theory, and crime pattern theory; Brantingham & Brantingham, 1984; Clarke & Felson, 1993), which suggest that offenders seek to minimize the efforts and risks involved in offending (e.g., by returning to geographical locations that are familiar to them to commit multiple offences in a relatively short period of time). These findings also lend support to the notion that the near-repeat phenomenon can be explained by the same offender returning to that geographical area to commit further crimes (Bernasco, 2008).

Despite the superior performance of inter-crime distance and temporal proximity, there were other domains that reached an acceptable level of accuracy (e.g. commercial robbery control behaviours, $AUC = 0.702$). This is perhaps surprising because robbery involves an interpersonal interaction between victim and offender and the level of control required (e.g., in terms of aggression) will be influenced to some extent by the behaviour of the victim and will therefore be driven by situational factors that are out of the offender's control. Nevertheless, these findings are consistent with previous work (Woodhams & Toye, 2007), which has found control behaviours to be a useful feature for distinguishing between linked and unlinked commercial robberies in the UK ($AUC = 0.90$). It is possible that there are a sufficient number of offender-driven behaviours within this domain that enable a relatively high level of discrimination accuracy to be achieved (e.g., decisions about how many co-offenders there are, whether to bring a weapon etc.).

Limitations and Future Directions

There are a number of limitations to the current study that must be considered. The primary limitation is in terms of ecological validity, as the current study comprised solved, serial offences only. When conducting BCL in practice, police crime analysts would typically have to identify linked crimes from databases that contain both solved and unsolved crimes, as well as a mixture of serial and one-off offences. Unsolved and one-off offences were not used in the current study, which means caution must be exercised when generalizing the current findings to real-world datasets. Future work must, therefore, seek to replicate these findings using datasets that contain both unsolved and one-off offences.

Future research must also begin to explore how statistical approaches (such as those tested in the current study) might be implemented in practice to support BCL. One obvious possibility is that computerized decision-support tools are developed, which can extract relevant behavioural information from police databases, subject them to statistical analysis, and then provide the analyst with a prioritized list of crime pairs for further investigation. As mentioned

previously, though, such tools would not be a replacement for human analysts; instead, they would assist analysts to make decisions (with the analyst retaining ultimate control over decision-making). Such an approach might have a number of advantages. First, it can perform many thousands of behavioural comparisons between crimes in a quick and efficient manner that is simply not possible for a human to do unaided. This might reduce the time it takes to perform BCL, which is advantageous at a time when police resources are limited. Second, an automated tool can help an analyst to deal with the vast quantity of information that must be considered when conducting BCL (including potentially hundreds of crimes and multiple offender behaviours). This reduces the cognitive load on the analyst, potentially increasing the accuracy and efficiency of decision-making. Third, given the focus on evidence-based policing around the world, it is important that BCL units base their working practices on reliable methods. Using decision-support tools that have been developed from empirical research would be an effective way for the police to adhere to these principles of evidence-based policing. However, it is important not to assume that statistical decision-support tools will have the expected benefits; this is a question that must be subjected to empirical testing. Indeed, it is only by developing, trialing, and evaluating such tools in realistic environments (ideally ongoing police investigations) that the true practical value of BCL research will be determined. This represents an important direction for future research.

References

- Adcock, A. B. (2000). *Effects of cognitive load on processing and performance*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.387&rep=rep1&type=pdf>
- Ayer, T., Chhatwal, J., Alagoz, O., Kahn, C. E., Woods, R. W., & Burnside, E. S. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Informatics in Radiology*, 30. doi: 10.1148/rg.301095057
- Bennell, C. (2002). *Behavioural consistency and discrimination in serial burglary* (Unpublished doctoral dissertation). University of Liverpool, Liverpool, UK.
- 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. doi: 10.1002/jip.21
- Bennell, C., Mugford, R., Ellingwood, H., & Woodhams, J. (2014). Linking crimes using behavioural clues: Current levels of linking accuracy and strategies for moving forward. *Journal of Investigative Psychology and Offender Profiling*, 11, 29-56. doi: 10.1002/jip.1395
- Bernasco, W. (2008). Them again?: Same-offender involvement in repeat and near-repeat burglaries. *European Journal of Criminology*, 5, 411-431. doi: 10.1177/1477370808095124
- Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. New York, NY: Macmillan.
- Breiman, L. (1994). *Bagging predictors* (Technical Report No. 421). Retrieved from <http://digitalassets.lib.berkeley.edu/sdtr/ucb/text/421.pdf>

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. doi: 10.1023/A:1010933404324

Burrell, A., Bull, R., & Bond, J. W. (2012). Linking personal robbery offences using offender behaviour. *Journal of Investigative Psychology and Offender Profiling*, 9, 201-222. doi: 10.1002/jip.1365

Canter, D. V. (2000). Offender profiling and criminal differentiation. *Legal and Criminological Psychology*, 5, 23-46. doi: 10.1348/135532500167958

Canter, D., & Youngs, D. (2008). Interactive Offender Profiling System (IOPS). In S. Chainey & L. Tompson (Eds.), *Crime mapping case studies: Practice and research* (pp. 153-160). Chichester, UK: Wiley.

Clarke, R. V., & Felson, M. (1993). *Routine activity and rational choice*. New Brunswick, NJ: Transaction.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29, 103-130. doi: 10.1023/A:1007413511361

Grubin, D., Kelly, P., & Brunson, C. (2001). *Linking serious sexual assaults through behaviour* (Home Office Research Study 215). London, UK: Home Office Research, Development and Statistics Directorate.

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1). Retrieved from <http://www.sigkdd.org/newsletter/explorations-july-2009-11-1>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.
- Kringen, J. A. (2014). *Validating a Bayesian model for linking serial crimes through simulation* (Unpublished doctoral dissertation). Texas State University, Texas, US.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *59*, 161-205. doi: 10.1007/s10994-005-0466-3
- Laukkanen, M. (2007). *Geographic profiling: Using home to crime distances and crime features to predict offender home location* (Unpublished doctoral dissertation). Åbo Akademi University, Turku, Finland.
- Laukkanen, M., Santtila, P., Jern, P., & Sandnabba, K. (2008). Predicting offender home location in urban burglary series. *Forensic Science International*, *176*, 224-235. doi: 10.1016/j.forsciint.2007.09.011
- le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *41*, 191-201. doi: 10.2307/2347628
- Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., & Coid, J. W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent

re-offending. *Journal of Quantitative Criminology*, 27, 547-573. doi: 10.1007/s10940-011-9137-7

Longadge, R., Dongre, S. S., & Malik, L. (2013). Class imbalance problem in data mining: A review. *International Journal of Computer Science and Network*, 2. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1305/1305.1707.pdf>

Markson, L., Woodhams, J., & Bond, J. W. (2010). Linking serial residential burglary: Comparing the utility of *modus operandi* behaviours, geographical proximity, and temporal proximity. *Journal of Investigative Psychology and Offender Profiling*, 7, 91-107. doi: 10.1002/jip.120

Melnyk, T., Bennell, C., Gauthier, D. J., & 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. doi: 10.1080/10683160903273188

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., ... Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford, UK: Oxford University Press.

Neuilly, M.-A., Zgoba, K. M., Tita, G. E., & Lee, S. S. (2011). Predicting recidivism in homicide offenders using classification tree analysis. *Homicide Studies*, 15, 154-176. doi: 10.1177/1088767911406867

Ngo, F. T., Govindu, R., & Agarwal, A. (2015). Assessing the predictive utility of logistic regression, classification and regression tree, Chi-squared automatic interaction detection, and neural

network models in predicting inmate misconduct. *American Journal of Criminal Justice*, 40, 47-74. doi: 10.1007/s12103-014-9246-6

Oatley, G. C., Ewart, B. W., & Zeleznikow, J. (2006). Decision support systems for police: Lessons from the application of data mining techniques to “soft” forensic evidence. *Artificial Intelligence and Law*, 14, 35–100. doi: 10.1007/s10506-006-9023-z

Porter, M. D. (2014). *A statistical approach to crime linkage*. Retrieved from <http://arxiv.org/pdf/1410.2285v1.pdf>

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9, 181-199. doi: 10.1007/s10021-005-0054-1

Quinlan, R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann Publishers.

Salo, B., Sirén, J., Corander, J., Zappalà, A., Bosco, D., Mokros, A., & Santtila, P. (2013). Using Bayes' theorem in behavioural crime linking of serial homicide. *Legal and Criminological Psychology*, 18, 356-370. doi: 10.1111/j.2044-8333.2011.02043.x

Santtila, P., Ritvanen, A., & Mokros, A. (2004). Predicting burglar characteristics from crime scene behaviour. *International Journal of Police Science & Management*, 6, 136-154. Retrieved from: <http://www.vathek.com/ijpsm/contents.php?vi=6.3>

- Snook, B., Luther, K., House, J. C., Bennell, C., & Taylor, P. J. (2012). The violent crime linkage analysis system: A test of interrater reliability. *Criminal Justice and Behavior*, *39*, 607-619. doi: 10.1177/0093854811435208
- Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., ... Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, *24*, 83-100. doi: 10.1023/A:1005478820425
- Sumner, M., Frank, E., & Hall, M. (2005). Speeding up logistic model tree induction. *Lecture Notes in Computer Science*, *3721*, 675-683. doi: 10.1007/11564126_72
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Modeling*, *43*, 1947-1958. doi: 10.1021/ci034160g
- Thomas, S., Leese, M., Walsh, E., McCrone, P., Moran, P., Burns, T., ... Fahy, T. (2005). A comparison of statistical models in predicting violence in psychotic illness. *Comprehensive Psychiatry*, *46*, 296-303. doi: 10.1016/j.comppsy.2004.10.001
- Tonkin, M., Grant, T., & 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. doi: 10.1002/jip.74
- Tonkin, M., Pakkanen, T., Sirén, J., Bennell, C., Woodhams, J., Burrell, A. ... & Santtila, P. (2017). Using offender crime scene behavior to link stranger sexual assaults: A comparison of three

statistical approaches. *Journal of Criminal Justice*, 50, 19-28. doi:
10.1016/j.jcrimjus.2017.04.002

Tonkin, M., Santtila, P., & Bull, R. (2012a). The linking of burglary crimes using offender behavior: Testing research cross-nationally and in more realistic settings. *Legal and Criminological Psychology*, 17, 276-293. doi: 10.1111/j.2044-8333.2010.02007.x

Tonkin, M., Woodhams, J., Bull, R., Bond, J. W., & Santtila, P. (2012b). A comparison of logistic regression and classification tree analysis for behavioural case linkage. *Journal of Investigative Psychology and Offender Profiling*, 9, 235-258. doi: 10.1002/jip.1367

Woodhams, J., & Bennell, C. (2014). *Crime linkage: Theory, research, and practice*. Boca Raton, FL: CRC Press.

Woodhams, J., Bull, R., & Hollin, C. R. (2007). Case linkage: Identifying crimes committed by the same offender. In R. N. Kocsis (Ed.), *Criminal profiling: International theory, research, and practice* (pp. 117-133). Totowa, NJ: Humana Press Inc.

Woodhams, J., & Labuschagne, G. (2012). A test of case linkage principles with solved and unsolved serial rapes. *Journal of Police and Criminal Psychology*, 27, 85-98. doi: 10.1007/s11896-011-9091-1

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. doi: 10.1037/1076-8971.13.1.59

Tables

Table 1

ROC Analyses for the Seven Statistical Methods Across All Three Datasets

	Residential Burglary	Car Theft	Commercial Robbery
	AUC (SE) [95% CI]	AUC (SE) [95% CI]	AUC (SE) [95% CI]
Logistic	0.903 (0.0013) [0.891 – 0.915]	0.818 (0.0007) [0.811 – 0.825]	0.830 (0.0012) [0.818 – 0.842]
SimpleLogistic	0.898 (0.0010) [0.888 – 0.908]	0.820 (0.0005) [0.815 – 0.825]	0.825 (0.0044) [0.781 – 0.869]
BayesNet	0.866 (0.0014) [0.852 – 0.880]	0.776 (0.0008) [0.769 – 0.784]	0.752 (0.0009) [0.743 – 0.761]
Naïve Bayes	0.892 (0.0010) [0.882 – 0.902]	0.796 (0.0008) [0.788 – 0.804]	0.753 (0.0037) [0.717 – 0.789]
J48 Decision Tree	0.474 (0.0012) [0.462 – 0.486]	0.442 (0.0008) [0.434 – 0.450]	0.553 (0.0019) [0.535 – 0.571]
Logistic Model Tree	0.497 (0.0015) [0.482 – 0.512]	0.500 (0.0000) [0.500 - 0.500]	0.802 (0.0032) [0.771 – 0.833]
RandomForest	0.741 (0.0009) [0.732 – 0.750]	0.528 (0.0005) [0.523 – 0.533]	0.732 (0.0009) [0.723 – 0.741]

Note. The highest Area Under the Curve (AUC) for each dataset is highlighted in **bold**. CI = Confidence Interval.

Table 2

True Positives for Different Ranges of the Rank-Ordered Crime Pair List

<i>Residential Burglary</i>			
Rank Ranges	Number of linked crime pairs within the range (i.e., number of true positives)	% of pairs within the range that were linked	Cumulative percentage of linked pairs
[1 - 10]	1.8	18.4%	2.3%
[11 - 25]	3.6	23.7%	6.7%
[26 - 50]	10.3	41.1%	19.5%
[51 - 100]	11.4	22.8%	33.7%
[101 - 250]	14.4	9.6%	51.6%
[251 - 500]	6.7	2.7%	60.0%
[501 - 1000]	6.9	1.4%	68.5%
[1001 - 2500]	6.0	0.4%	76.0%
[2501 - 5000]	11.4	0.5%	90.2%
[5001 - 12720]	7.7	0.2%	100.0%
<i>Car Theft</i>			
Rank Ranges	Number of linked crime pairs (i.e., true positives) within the range	% of pairs within the range that were linked	Cumulative percentage of linked pairs
[1 - 10]	0	0.0%	0.0%
[11 - 25]	0.1	0.5%	0.0%
[26 - 50]	0.6	2.6%	0.4%
[51 - 100]	1.3	2.6%	1.1%
[101 - 250]	3.3	2.2%	2.8%
[251 - 500]	5.5	2.2%	5.8%

[501 - 1000]	7.8	1.6%	9.9%
[1001 - 2500]	19.2	1.3%	20.1%
[2501 - 5000]	27.0	1.1%	34.4%
[5001 - 10000]	43.1	0.9%	57.2%
[10001 - 25000]	54.4	0.4%	86.0%
[25001 - 50000]	19.6	0.1%	96.4%
[50001 - 70500]	6.9	0.0%	100.0%
<i>Commercial Robbery</i>			
Rank Ranges	Number of linked crime pairs (i.e., true positives) within the range	% of pairs within the range that were linked	Cumulative percentage of linked pairs
[1 - 10]	6.5	64.8%	10.7%
[11 - 25]	5.6	37.1%	19.9%
[26 - 50]	3.6	14.2%	25.8%
[51 - 100]	7.1	14.2%	37.5%
[101 - 250]	6.2	4.2%	47.8%
[251 - 500]	10.8	4.3%	65.7%
[501 - 1000]	5.3	1.1%	74.4%
[1001 - 2500]	5.6	0.4%	83.8%
[2501 - 5000]	3.0	0.1%	88.8%
[5001 - 6903]	6.8	0.1%	100.0%

Note. The burglary and robbery pairs in this table are rank-ordered according to the predicted probability values produced by the Logistic algorithm and the car theft pairs according to the Simple Logistic algorithm. All figures are based on the average findings across 25 iterations.

Table 3

ROC Analyses by Behavioural Domain for the Three Datasets

Behavioural Domains	AUC (SE) [95% CI]
<i>Residential Burglary Domains</i>	
Inter-crime Distance (km)	0.854 (0.0018) [0.836 – 0.872]
Temporal Proximity (days)	0.818 (0.0016) [0.802 – 0.834]
Target Characteristics	0.674 (0.0020) [0.654 – 0.694]
Entry Behaviours	0.719 (0.0028) [0.692 – 0.746]
Internal Behaviours	0.757 (0.0014) [0.743 – 0.771]
Property Stolen	0.649 (0.0016) [0.633 – 0.665]
<i>Car Theft Domains</i>	
Inter-crime Distance (km)	0.820 (0.0010) [0.810 – 0.830]
Target Selection Choices	0.555 (0.0014) [0.541 – 0.569]
Target Acquisition Behaviour	0.512 (0.0028) [0.485 – 0.539]
Disposal Behaviour	0.535 (0.0026) [0.510 – 0.560]
<i>Commercial Robbery Domains</i>	
Inter-crime Distance (km)	0.745 (0.0022) [0.723 – 0.767]
Temporal Proximity (days)	0.816 (0.0012) [0.804 – 0.828]
Target Selection Choices	0.645 (0.0026) [0.620 – 0.670]
Planning Behaviours	0.680 (0.0028) [0.653 – 0.707]
Control Behaviours	0.702 (0.0022) [0.680 – 0.724]
Property Stolen	0.623 (0.0026) [0.598 – 0.648]
Value Stolen (Euro)	0.571 (0.0040) [0.532 – 0.610]
Approach Behaviours	0.562 (0.0018) [0.544 – 0.580]

Note. The highest AUC per dataset is highlighted in **bold**. CI = Confidence Interval.

Table 4

ROC Analyses Using Individual Crime Scene Behaviours as Predictor Variables

	Residential Burglary AUC (SE) [95% CI]	Car Theft AUC (SE) [95% CI]	Commercial Robbery AUC (SE) [95% CI]
Logistic	0.770 (0.0011) [0.760 – 0.781]	0.811 (0.0004) [0.807 – 0.815]	0.665 (0.0014) [0.651 – 0.679]
SimpleLogistic	0.878 (0.0010) [0.868 – 0.888]	0.813 (0.0006) [0.807 – 0.819]	0.790 (0.0041) [0.750 – 0.831]
BayesNet	0.845 (0.0013) [0.833 – 0.857]	0.776 (0.0008) [0.769 – 0.784]	0.762 (0.0014) [0.749 – 0.775]
Naïve Bayes	0.877 (0.0010) [0.867 – 0.887]	0.792 (0.0007) [0.786 – 0.799]	0.745 (0.0039) [0.707 – 0.783]
J48 Decision Tree	0.454 (0.0063) [0.392 – 0.516]	0.442 (0.0008) [0.434 – 0.450]	0.592 (0.0021) [0.571 – 0.613]
Logistic Model Tree	0.514 (0.0018) [0.497 – 0.531]	0.500 (0.0000) [0.500 - 0.500]	0.555 (0.0036) [0.520 – 0.590]
RandomForest	0.694 (0.0012) [0.682 – 0.706]	0.526 (0.0005) [0.521 – 0.531]	0.678 (0.0012) [0.666 – 0.690]

Note. The highest Area Under the Curve (AUC) for each dataset is highlighted in **bold**. CI = Confidence Interval.

Figures

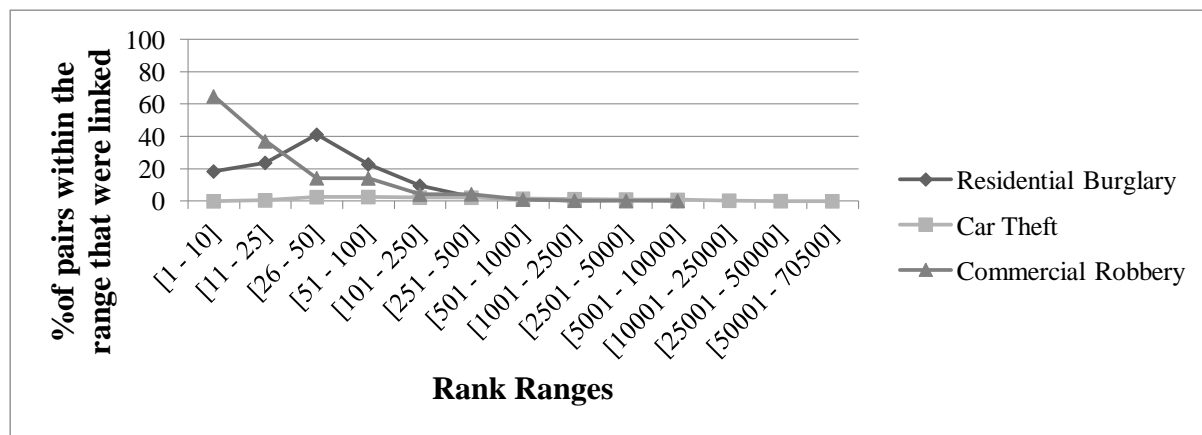


Figure 1

The Percentage of Linked Crime Pairs as a Function of Predicted Probability Rank