FACTORS THAT INFLUENCE THE ACCURACY OF BEHAVIOURAL LINKAGE ANALYSIS IN CASES OF SERIAL SEXUAL ASSAULT, HOMICIDE, AND BURGLARY

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree

Masters of Arts

by

Tamara Melnyk

Department of Psychology
Carleton University

September 2008

Your file Votre référence
ISBN: 978-0-494-43479-6
Our file Notre référence
ISBN: 978-0-494-43479-6

Abstract

In the absence of physical evidence, investigators must rely on offense behaviours when determining whether several crimes are linked to a common offender. This is known as behavioural linkage analysis (BLA). A variety of factors may influence the degree to which it is possible to link serial crime, including the crime type under consideration, the similarity coefficient used to assess across-crime similarity, and the frequencies of the behaviours used to conduct the analysis. The current thesis consists of two studies designed to examine these factors. Study 1 compares the linking accuracy of two similarity coefficients – Jaccard's coefficient (J) and the taxonomic similarity index (Δ_s) – across three crime types – serial sexual assault (n = 126), serial homicide (n = 237), and serial burglary (n = 210). Study 2 compares the linking accuracy that can be achieved across the three crime types when using various subsets of behaviours defined by their frequency of occurrence. The degree to which these factors influence linking accuracy is determined through the use of receiver operating characteristic (ROC) analysis. The results of Study 1 indicate that no significant differences exist between the linking ability of J and Δ_s across a variety of conditions, although J slightly outperforms Δ_s when relying on larger sample sizes. Both coefficients lead to much higher levels of linking accuracy in cases of serial homicide and sexual assault compared to serial burglary. The results of Study 2 indicate that no significant differences emerge when four behavioural frequency recommendations are compared to one another, indicating that it might be most productive to base linkage analyses on all crime scene behaviours included in a crime sample. These various findings are discussed, along with their implications. The thesis ends with a presentation of some limitations with the research and some suggestions for how these limitations can be overcome in the future.

Acknowledgements

Dr. Craig Bennell, you are a true advisor, teacher, and mentor. I am so fortunate to have had the opportunity to learn from you. My achievements over the course of this MA are in large part due to who you are as an advisor, and your exceptional dedication to your work and students. It has not gone unrecognized. You have provided me with the unwavering support, encouragement, and understanding that made it possible for me to complete this degree while also facing tremendous loss in my life. I have learned so much from you, not only about research, and could never express in words the true extent of my appreciation for your countless hours of guidance. Few people will have influenced the course of the rest of my life as much as you have.

To my committee, Dr. Shelley Brown, Dr. Adelle Forth, Dr. Ron Saunders, and Dr. Ralph Serin – thank you so very much for your time, support, and constructive criticism in helping me strive to produce the very best work that I possibly can. Dr. Forth, I have learned so much from you over the course of my degree about what it means to be a teacher. You have provided me with experience and incredible support in so many ways, and I thank you.

Dr. Donald and Donna Gauthier, my sincere thanks to you both for your time and assistance during the analysis phase of this project. I appreciate so very much your interest, enthusiasm, incredible kindness and words of encouragement. I could never thank you enough for all that you have done.

To the administrative staff of the Carleton University Psychology Department – thank you for always being so kind, helpful, and patient, and for being so supportive of your students.

To my fellow graduate students, Shevaun Corey, Alyssa Taylor, Rebecca Mugford, and especially Karla Emeno – with whom I shared the experience of completing a Masters degree – there is no group of people I would rather study with, and I am so fortunate to have you as my labmates. You have provided me with friendship, inspiration, and support during what proved to be, for many reasons, the most challenging years of my life. Also to Melissa Martineau, thank you for your friendship, understanding, and the opportunities you have provided me.

To my Mom, Sue MacDonald, the Godick, MacDonald, Melnyk, and Maxwell families, and to my friends – I will be forever grateful for you, and extend my love and appreciation to you all. Your support during this time of my life will never be forgotten, and made it possible for me to achieve this goal. Thank you for believing in me when I felt all was lost.

Most importantly, to my husband, Ryan Melnyk – I could not have done this without you. Your love, support, patience, and understanding made this possible for me. Thank you for being who you are, encouraging me, and helping me to achieve this goal while we faced some of life's greatest challenges together. You are everything to me. I love you.

This thesis is dedicated in loving memory of my father, Jim Godick, who passed away unexpectedly the first days of this degree. Thank you, Dad, for the love and support that you provided me throughout my life. You encouraged me to take this program, further my education, and always to strive for my very best. I hope that I can make you proud. Life is not the same without you. I miss you. I love you. I did this for you.

Table of Contents

List of Tables

List of Figures

List of Appendices

Factors that Influence the Accuracy of Behavioural Linkage Analysis in Cases of Serial Sexual Assault, Homicide, and Burglary

Introduction

Police investigators must frequently determine whether a series of crimes has been committed by the same offender (Grubin, Kelly, & Brunsdon, 2001). The analysis of physical evidence (e.g., DNA) provides the most reliable means for accomplishing this task. However, such evidence is frequently not available at a crime scene, and when it is, it is often compromised in some way (Hazelwood & Warren, 2003). Even in cases where physical evidence is available, its processing can be costly and time-consuming (Davies, 1991; Woodhams, Hollin, & Bull, 2007). In the absence of physical evidence, links between crimes are typically established through an analysis of behavioural evidence (Woodhams, Hollin, et al., 2007). Using an approach known as behavioural linkage analysis (BLA), an attempt is made to identify behavioural patterns across crime scenes in an attempt to determine if the crimes have been committed by the same offender.

Accurately identifying a crime series through the analysis of crime scene behaviours can be beneficial to criminal investigators. For example, not only does it allow investigators to pool information from all relevant crime scenes, thus enabling a more efficient use of investigative resources (Grubin et al., 2001), it is also a prerequisite for the successful implementation of other investigative techniques, such as criminal and geographic profiling (Rossmo, 2000). Given the practical importance of BLA, it is surprising that little empirical research has been conducted on the topic. Indeed, as Bennell, Jones, and Melnyk (in press) have argued, "it has only been in the last decade that any notable effort has been made to understand the processes underlying

[BLA]...and to systematically determine the degree to which it is possible to successfully link a series of crimes [using crime scene behaviours]" (p. 3).

Recently, Woodhams, Hollin, et al. (2007) conducted a comprehensive review of the empirical literature on BLA. This review generally supported the idea that BLA is possible, and it concluded by recommending the use of a particular approach for tackling the linking problem. The recommended approach was originally proposed by Bennell (2002), and subsequently developed by Bennell and his colleagues (Bennell, 2005; Bennell, Bloomfield, Snook, Taylor, & Barnes, 2008; Bennell & Canter, 2002; Bennell & Jones, 2005, Bennell et al., in press). In a general sense, the approach is to treat the behavioural linking task as a signal detection problem and to rely on analytical methods from signal detection theory to study the task, particularly *receiver operating characteristic* (ROC) *analysis* (Bennell, 2005; Swets, 1996).

The goal of the present thesis is to contribute further to research on BLA from a signal detection perspective. The specific objectives of the thesis are to: (1) identify the fundamental constructs that underlie the behavioural linking task, (2) present arguments supporting the re-conceptualization of BLA as a signal detection problem, (3) demonstrate how ROC analysis can be used to clarify important linking issues, and (4) explore several factors that potentially influence linking accuracy.

Underlying Constructs: Behavioural Stability, Distinctiveness, and Discrimination

Developing Operational Definitions

Researchers investigating BLA focus on two assumptions when considering whether or not it is possible to accurately link crimes (Canter, 1995; Grubin et al., 2001; Woodhams, Hollin, et al., 2007). First, it is assumed that offenders must exhibit high

levels of *behavioural stability* across their respective crime series, reflecting the degree to which each individual manifests the same behaviours across his/her own crimes. Second, it is assumed that offenders must exhibit high levels of *behavioural distinctiveness* for linking to be possible, whereby the actions that a given serial offender exhibits across his/her crimes differ from those exhibited by other offenders committing similar types of crimes. The extent to which these two constructs exist is thought to determine the degree to which it is possible to discriminate between crimes committed by different offenders, a process referred to as *behavioural discrimination.*

To illustrate, consider a study conducted by Canter et al. (1991). The study examined BLA by utilizing a sample of solved serial sex offences (four different crime series consisting of three crimes each). Each crime was coded for the presence/absence of a set of 74 crime scene behaviours, which served as the basis for calculating across-crime similarity scores (ranging from 0 to 1) for every pair of crimes within the sample. The scores for each crime pair are presented in Table 1 (see the note below the table for a legend of its abbreviations).

Table 1. Across-crime similarity coefficients reported by Canter et al. (1991).

	Same offender		Different offenders								
	2	3	B1	B2	B3	C1	C2	C3	D1	D2	D3
A1	.11	.42*	.27	.32*	.27	.15	.17	.06	.43*	.17	.26
A2		.14	.29	.29	.11	.15	.07	.12	.29	.26	.18
A3			.27	.27	.23	.09	.11	.05	.27	.14	.33*
B1	.45*	.26				.06	.08	.05	.21	.08	.18
B2		.41*				.07	.07	.02	.27	.07	.16
B3						.27	.31*	.27	.12	.14	.06
C1	.38*	.48*							.22	.33*	.16
C2		.36*							.10	.11	.02
C3									.07	.20	.11
D1	.21	.46*									
D2		.17									
D3											

Note. Within this table, the letters A, B, C, and D refer to different serial offenders, and the numbers 1, 2, and 3 refer to different crimes. Thus, A1 refers to the first crime committed by offender A, A2 refers to the second crime committed by offender A, and so on. As an example of how the table should be read, the cell in the upper-left corner of the table (A1-2) refers to the degree of behavioural stability (.11) exhibited across the first and second crimes of offender A (high similarity scores across crimes committed by the same offender equates to high levels of behavioural stability). In contrast, the cell corresponding to A1-B1 refers to the degree of behavioural stability (.27) exhibited across the first crime of offender A and the first crime of offender B (low similarity scores across crimes committed by different offenders equates to high levels of behavioural distinctiveness). The * in this table indicates those instances where the similarity score exceeds an imposed threshold of ≥.30 (discussed further below).

Similarity scores presented on the left-hand side of Table 1 correspond to crime pairs committed by the same offender. These values represent the level of behavioural stability exhibited by the sample (i.e., higher similarity scores reflect higher levels of stability). In contrast, scores presented on the right-hand side of Table 1 correspond to crimes committed by different offenders. These values represent the level of behavioural

distinctiveness exhibited by the sample (i.e., lower similarity scores reflect higher levels of distinctiveness). A comparison of the scores in Table 1 makes it clear that the left-hand ("same offender") scores are generally larger than the right-hand ("different offenders") scores and, therefore, a degree of behavioural stability and distinctiveness exists within the sample. According to common thinking, this suggests that it should be possible to discriminate between the crimes committed by different offenders (to some extent at least).

The degree to which it is possible to do this is determined by: (1) selecting a decision threshold (i.e., a specific similarity score) for deciding when two crimes are similar enough to be considered linked and (2) applying that threshold to the data and calculating the proportion of correct and incorrect linking decisions that result. In Canter et al.'s (1991) study, they selected a threshold of ≥ .30. This particular threshold permitted the correct determination of 7 out of 12 (58.3%) linked crime pairs and 49 out of 54 (90.7%) unlinked crime pairs (discrimination accuracy = 84.84%). Thus, the relatively high levels of behavioural stability and distinctiveness exhibited by the offenders in this sample permitted a reasonable degree of discrimination accuracy to be achieved.

Evidence for the Constructs Underlying BLA

Individuals who study BLA frequently cite research from the field of personality psychology to provide support for these constructs (e.g., Bennell & Canter, 2002; Salfati & Bateman, 2005; Santtila, Korpela, & Hakkanen, 2004; Woodhams, Hollin, et al., 2007; Woodhams, Hollin, & Bull, 2008; Woodhams & Toye, 2007). However, in contrast to this approach, the present review is restricted to research conducted within the domain of

investigative psychology. The reason for this is that studies of behavioural stability and distinctiveness in the domain of personality psychology likely have little bearing on BLA or BLA research, as there are important differences across these two domains with respect to the way these constructs are defined and measured (Bennell, in press; Canter, 2004; Woodhams, Hollin, et al., 2008).

For example, behavioural stability is typically defined in *relative* terms by personality psychologists, such that individuals are said to be stable if they maintain their rank ordering (compared to one another) on some measure of a behavioural trait across multiple situations that each individual participates in (Bem & Allen, 1974; Funder & Colvin, 1991; Furr & Funder, 2004; Roberts & DelVecchio, 2000; Shiner & Caspi, 2003; Shoda, 1999). It is possible for personality psychologists to define and measure stability in this way because they can control the situations that people are exposed to and directly observe their behaviour (Bennell, in press). This contrasts with the *absolute* stability that investigative psychologists typically search for (i.e., the extent to which offenders exhibit consistent behaviour across their crimes). Not only is absolute stability more practically useful in investigative contexts, it is not possible for an investigative psychologist to measure relative stability in the same way that a personality psychologist does. Not only do offenders not participate in the exact same situations as one another, offending behaviour cannot be directly observed, and even if it could, it is unclear what should be observed as the sorts of behavioural traits studied by personality psychologists are arguably irrelevant in the investigative context.

Having said this, other personality psychologists do define stability in more absolute terms, such as when they examine the stability of some measure of a behavioural

trait *within* an individual over time (e.g., Caspi & Herbener, 1990; Furr & Funder, 2004; Mroczek & Spiro, 2003; Shiner & Caspi, 2003) and, on occasion, trait-like dispositions that are potentially relevant to BLA are examined (e.g., psychopathy; Blonigen, Hicks, Krueger, Patrick, & Iacono, 2006; Frick, Kimonis, Dandreaux, & Farell, 2003; Lynam, Loeber, & Stouthamer-Loeber, 2008). However, here too there are important differences between these types of studies and those conducted by investigative psychologists. For example, unlike investigative psychology, where the search for absolute stability relates to behaviours exhibited across discrete situations (i.e., specific crimes), the search for absolute stability in personality psychology is usually concerned with the expression of behaviour across an individual's life span or part of their life span (i.e., not across discrete situations) (Bennell, in press).

In any case, Woodhams, Hollin et al. (2007) have recently highlighted that a substantial amount of empirical evidence has been found for behavioural discrimination from within the investigative domain (thus implying the existence of behavioural stability and distinctiveness). These findings have generally been interpreted as supporting the notion that the behaviours exhibited by serial offenders at their crime scenes are at least partially determined by relatively stable, and somewhat unique, predispositions to commit crimes in a particular way (Woodhams, Hollin, et al., 2007).

For example, in one of the only studies on BLA to appear before the 1990s, Green, Booth, and Biderman (1976) coded a range of burglary behaviours exhibited across 15 burglaries committed by three serial offenders, and subjected the across-crime similarity scores to cluster analysis. By examining the resulting visual plot, contours were

subjectively drawn around clusters of crimes to indicate predicted linkages and an impressive level of discrimination accuracy was achieved (93%).

In another study that used cluster analysis to examine BLA, though in a rather different way to how Green et al. (1976) used the procedure, Grubin et al. (2001) examined whether it was possible to link serial sexual assaults using crime scene behaviours. They used cluster analysis to identify underlying domains of offence behaviour and to further sub-type those domains. The domains they identified consisted of control, sex, escape, and style behaviours and each could be further categorized into four sub-types (e.g., control type 1, 2, 3, and 4). Grubin et al. then examined the degree to which these groupings of behaviour could be used to link the crimes of the serial sex offenders. After assigning each crime in the sample a profile of scores across the domains (e.g., control type 1, sex type 3, escape type 1, style type 4), they treated each crime as a target offence and searched the remaining database for the 10% of crimes that were most similar to the target. When examining series in their sample consisting of three crimes each (totalling 57 offences), 22 target offences had at least one other offence from its series included in this 10% (where only 10 would be expected by chance). Eleven target offences were found to have both other offences from their series included in this 10% (where only 0.5 would be expected by chance).

In a similar study, Santtila, Fritzon, and Tamelander (2004) attempted to link arson cases based on the level of behavioural stability and distinctiveness exhibited by arsonists. Principle components analysis (PCA) and discriminant function analysis (DFA) were applied to a sample of 248 arson cases committed by 42 offenders. Content analysis was first conducted using 45 dichotomous variables relating to various crime scene

behaviours. This was followed by PCA in order to identify underlying themes of arson behaviour. Summary scores reflecting the resulting themes of behaviour were then calculated for each case on the basis of these results. These summary scores served as the predictor variables in the DFA, while the dependent variable, or the grouping variable, was the series that each particular arson case belonged to. As part of its output, DFA produces probabilities of each case (i.e., target offences) belonging to other groups (i.e., other series) in the analysis. To examine the degree of discrimination accuracy that could be achieved through the use of DFA, these probabilities were examined to determine "the proportion of cases for which the actual series was among a specific number of series to which the cases had the highest likelihood of belonging" (p. 9). When this was done, the right series was among the 10 most probable series for 52% of cases, which is well beyond what would be expected by chance.

Most recently, Woodhams and Toye (2007) provided further support for BLA in a study of 160 serial robberies committed by 80 robbers (two crimes per offender). They calculated across-crime similarity scores (for linked and unlinked crime pairs) for a range of behavioural domains, including target selection, planning, and control variables, along with inter-offence distances. They then subjected these scores to logistic regression analysis in order to discriminate between crimes committed by different offenders. Based on their results, each of the predictor variables under examination could be used to link crimes beyond the chance level, with R^2 values ranging from .20 for planning variables (discrimination accuracy = 65.63%) to .59 for control variables (discrimination accuracy = 79.38%). Using the predictor variables in combination resulted in the most accurate linking decisions (R^2 = .69; discrimination accuracy = 89.74%).

Similar results to those reported in the studies discussed above have been presented in a number of other studies, regardless of the analytical method used or the type of crime examined (e.g., Bennell, 2002; Bennell & Canter, 2002; Bennell & Jones, 2005; Ewart, Oatley, & Burn, 2005; Goodwill & Alison, 2006; Salfati & Bateman, 2005; Santtila, Junkkila, & Sandnabba, 2005; Woodhams, Grant, & Price, 2007).

Reconceptualizing BLA as a Signal Detection Problem

Problems with Traditional Approaches to BLA

Although the results from studies like those discussed above appear promising on the surface, it is important to consider what they actually convey. Again, for illustrative purposes, consider Canter et al.'s (1991) study. To what extent are the serial offenders in Canter et al.'s study actually exhibiting behavioural stability and distinctiveness? To what extent can one use the linking approach proposed in that study to discriminate between crimes committed by different offenders? Practically speaking, what does this study indicate with respect to the degree of similarity that must exist between two crimes before investigators should consider them part of the same series?

The primary problem with studies like Canter et al.'s (1991) is that they are unable to address such concerns. In large part, this is because the results of these studies are biased by where researchers decide to place their decision thresholds. Indeed, the answers to each of the questions posed in the previous paragraph depend entirely on the specific threshold that is adopted. For example, while a threshold of ≥ .30 yielded a discrimination accuracy score of 84.84% in Canter et al.'s study, a threshold of ≥ .10 would yield an accuracy score of 40.90%. Using the analytical method adopted by Canter et al., it is simply not possible to answer even the most basic of questions that arise in the

linking task, and this problem is not specific to Canter et al.'s study. The same problem plagues nearly all linking research. Consider the study conducted by Grubin et al. (2001), for instance, where they used the pre-determined 10% cut-off to select crimes that were "similar" to the target offences, or the study by Woodhams and Toye (2007), where their classification results were based on a threshold set by SPSS ($p>.5$) for determining when a pair of crimes should be considered linked. In each case, the results presented in these studies would be different if a different threshold had been used.

By raising this issue, it is not being implied that one should avoid setting decision thresholds when carrying out BLA. On the contrary, that process is an inherent part of the linking task. What is needed is an analytical approach that does not suffer from the same sorts of problems that are encountered when using existing methods. Specifically, what is required is an analytical method that can be used to "quantify the degree of [discrimination] accuracy achieved under any given set of conditions, unbiased by threshold placement…the method of analysis would ideally also guide decision threshold placement such as to maximize [discrimination] accuracy" (Bennell et al., 2008, p. 8). The argument being presented in this thesis is that ROC analysis represents such a method.

Signal Detection Theory and ROC Analysis

ROC analysis originated within the field of signal detection theory, where attempts were being made to understand how people distinguish signals (e.g., blips on a radar screen) from background noise (Green & Swets, 1974). Over time, the term "signal detection task" started to take on a more generic meaning to include any two-alternative (i.e., yes/no type) diagnostic decision making task where some sort of significant event

(e.g., malignant tumours, tornadoes, high risk offenders, etc.) had to be distinguished from more frequent, but less important events (e.g., benign tumours, no storm, low risk offenders, etc.) (Swets, 1996). ROC analysis was originally proposed as a method for modeling performance in signal detection tasks (Green & Swets, 1974). It has since grown to become the method of choice for assisting with diagnostic decision making in fields ranging from radiology to engineering (Swets, Dawes, & Monahan, 2000).

How is BLA similar to these other diagnostic tasks? First, like each of the tasks alluded to above, the task in BLA is to distinguish a relatively rare "signal" (i.e., a linked pair of crimes) from a background of noise (i.e., the majority of crimes being committed that are unlinked). Second, BLA is a probabilistic task in that definitive answers to the question of whether a "signal" is present will rarely, if ever, be possible (due largely to the fact that the evidence upon which the decision is based is often ambiguous). Third, although being a probabilistic task, a dichotomous decision must be made in BLA (i.e., if the pair of crimes is linked or unlinked) and this results in the same four decision outcomes that can occur in any diagnostic task, specifically hits (predict linked/actually linked), correct rejections (predict unlinked/actually unlinked), misses (predict unlinked/actually linked), and false alarms (predict linked/actually unlinked). Finally, the goal of BLA is the same as for all of the other tasks, which is to maximize the probability of rendering a correct decision while minimizing the probability of making an incorrect decision.

Another way in which BLA is similar to other diagnostic tasks is in the way the tasks can be conceptualized. From a signal detection perspective, diagnostic tasks are often conceptualized using a pair of probability distributions (Swets et al., 2000) and this

has proven to be a useful way of thinking about BLA (Bennell, 2005). For example, consider Canter et al.'s (1991) study once again and imagine a graph where the *x*-axis represents the degree of across-crime similarity between crime pairs (ranging from 0 to 1) and the *y*-axis represents the probability that any particular crime pair will possess a certain degree of across-crime similarity (ranging from 0 to 1). If the similarity scores from the left-hand side ("same offender") and right-hand side ("different offenders") of Table 1 were plotted on this graph, two probability distributions that are similar to those presented in Figure 1 would emerge.

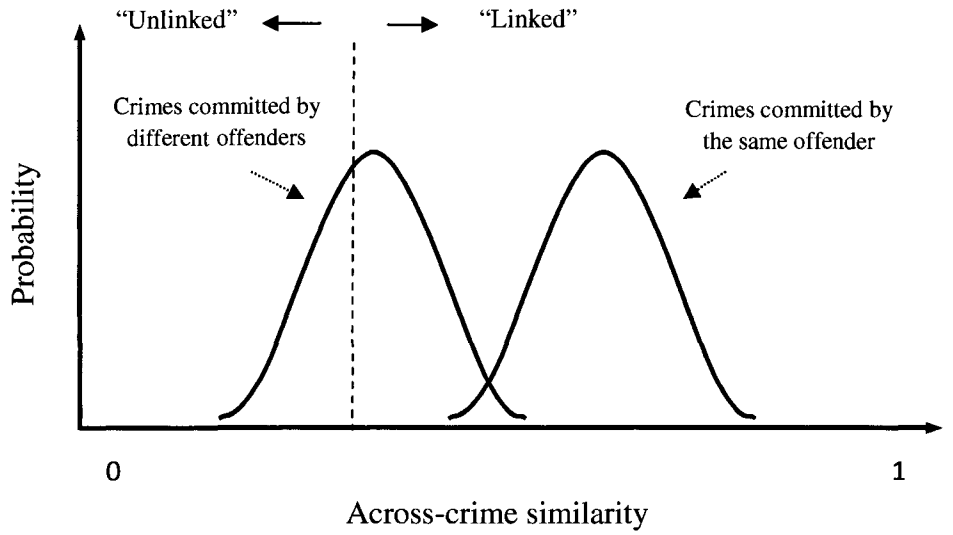


Figure 1. Hypothetical distributions of across-crime similarity scores for crimes committed by the same offender versus different offenders. The *x*-axis represents the degree of similarity (from 0 to 1) between crime pairs and the *y*-axis represents the probability (from 0 to 1) that a crime pair possesses any given degree of similarity. The decision threshold is represented by the dashed line. Any crime pair with a level of across-crime similarity above this threshold would be considered "linked".

As can be seen in Figure 1, the right-hand ("same offender") distribution represents the degree of behavioural stability exhibited by a sample of offenders (the larger the scores, the more stability, and the further up the *x*-axis this distribution lies). The left-hand ("different offenders") distribution, on the other hand, represents the degree of behavioural distinctiveness exhibited by a sample of offenders (the smaller the scores, the more distinctiveness, and the lower down the *x*-axis this distribution lies). Conceptualizing BLA in this way makes it even more obvious that behavioural

discrimination depends on the extent to which offenders in a sample exhibit stability and distinctiveness. Essentially, the amount of overlap present between the two distributions will determine the degree to which it is possible to discriminate between crimes committed by different offenders (the smaller the degree of overlap, the higher the level of discrimination accuracy that will be achieved).

The degree of discrimination accuracy possible when using a particular decision threshold (i.e., similarity score) is also captured in Figure 1. When a threshold (represented by the dashed line) is set for determining when a pair of crimes should be considered linked, the probability of making a hit (*p*H) is represented by the area under the right-hand distribution to the right of the threshold and the probability of making a false alarm (*p*FA) is represented by the area under the left-hand distribution to the right of the threshold. The probabilities of misses (*p*M) and correction rejections (*p*CR) are simply the complements of *p*H and *p*FA and are represented by the areas under the right-hand and left-hand distributions, respectively, to the left of the threshold. In order to calculate *p*H and *p*FA, one simply counts up the frequency of hits (H), correct rejections (CR), misses (M) and false alarms (FA) made when using a particular threshold on a sample of cases and applies the following formulae: $pH = H/(H+M)$ and $pFA = FA/(FA+CR)$. Given that *p*M and *p*CR are the complements of *p*H and *p*FA, respectively, the following formulae can be used to calculate these values: $pM=1-pH$ and $pCR=1-pFA$.

The ratio of *p*H to *p*FA provides the information necessary to assess discrimination accuracy. However, as argued previously, this ratio is specific to the threshold adopted and says nothing about the linking decisions that might be made when

using an alternative threshold. Indeed, as is made clear in Figure 1, this ratio will vary drastically across thresholds even when the position of the underlying probability distributions remains fixed (e.g., making the threshold stricter will decrease *p*FA, but it will also decrease *p*H). ROC analysis is an analytical method that can be used to overcome this problem (Swets, 1996). Specifically, ROC analysis accounts for how the probabilities of making correct and incorrect linking decisions change as decision thresholds are varied. This is done by calculating values of *p*H and *p*FA across multiple thresholds and plotting these coordinates on an ROC graph (*p*H on the *y*-axis and *p*FA on the *x*-axis). As illustrated in Figure 2, when the plotted points are connected, the result is typically a concave downward curve known as an ROC curve, which begins in the lower-left corner of the graph and rises to the upper-right corner. Strict decision thresholds (set high along the *x*-axis in Figure 1) result in ROC points falling on the lower portion of the curve, whereas lenient thresholds (set low along the *x*-axis in Figure 1) result in ROC points falling on the upper portion of the curve.

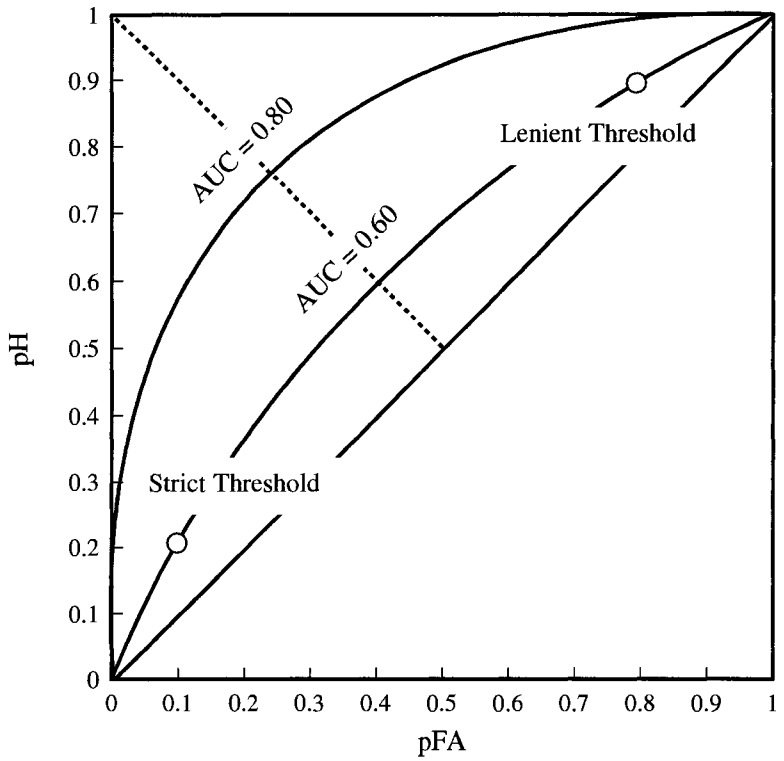


Figure 2. An ROC graph consisting of two ROC curves illustrating different levels of discrimination accuracy. Each ROC curve is made up of ROC points that result from using different decision thresholds.

The *area under the curve* (*AUC*) serves as a measure of discrimination accuracy for the specific linking approach, or linking evidence, that was used to create that particular curve. Less overlap between the two distributions in Figure 1 (i.e., higher levels of behavioural stability and distinctiveness) results in a higher ROC curve, which indicates a greater degree of discrimination accuracy. The *AUC* typically varies from .50 to 1.00. Specifically, an *AUC* of .50 indicates chance discrimination and is represented by

an ROC curve that falls on the positive diagonal of the graph. On the other hand, an *AUC* of 1.00 indicates perfect discrimination and is represented by a curve that runs along the left and upper axes.

Advantages Associated with ROC Analysis When Applied to BLA

There are several advantages associated with ROC analysis (and the *AUC*) when it is applied to BLA (Bennell, 2002). First, and perhaps most importantly, is the fact that the *AUC* is a measure of discrimination accuracy that is independent of the specific decision threshold selected for linking purposes. This is because the *AUC* reflects the position of the entire ROC curve in its graph rather than any single point (i.e., threshold) along the curve. Second, the *AUC* represents a simple measure of discrimination accuracy that can be easily interpreted and applied to naturalistic settings. For example, in the context of BLA, if one were to use an across-crime similarity score to distinguish between crimes committed by the same offender versus different offenders, an *AUC* of .80 indicates that there is an 80% chance that a randomly selected pair of crimes committed by the same offender will have a larger similarity score than a randomly selected pair of crimes committed by different offenders (Bennell et al., in press). Third, the *AUC* is a flexible measure of discrimination accuracy that can be used regardless of the linking approach adopted, the measure of similarity used, or the types of behaviour examined. Fourth, the *AUC* is a measure of accuracy that is unbiased by base rates and, thus, it can be used to compare the level of discrimination accuracy achieved across data sets that vary drastically in their rates of serial crime activity (e.g., a sample of serial homicides vs. serial burglaries). Finally, while the *AUC* is threshold-independent, information is still maintained in the analysis that allows one to identify optimal decision

thresholds for linking purposes (i.e., thresholds that result in the desired balance of decision outcomes). For instance, a common procedure is to identify the threshold corresponding to the ROC point falling closest to the upper-left corner of the ROC graph where *p*H is maximized and *p*FA is minimized. Alternatively, a predetermined rate of *p*FA can be set, which is deemed acceptable for the situation at hand, and a threshold can be adopted that maximizes *p*H without exceeding this limit. For a more detailed discussion of these and other strategies for selecting thresholds, see Bennell (2005).

Factors that May Influence Behavioural Discrimination

Beyond the already stated advantages of using the *AUC* as a measure of discrimination accuracy, conceptualizing BLA as a signal detection task is advantageous because it clarifies what must be done in order to increase linking accuracy. Specifically, from a signal detection perspective, the goal must be to identify factors that, when manipulated, decrease the degree of overlap between the distributions in Figure 1. For the purpose of this thesis, three factors that may influence discrimination accuracy will be selected for analysis: (1) the type of crime under consideration, (2) the type of similarity coefficient used to measure across-crime similarity, (3) and the frequency with which crime scene behaviours occur in a given sample. These factors were selected because they have repeatedly been presented as potentially important variables in the research literature, but they have yet to be examined empirically.

Crime Type

As mentioned previously, BLA research has examined a range of crime types, including burglary, arson, robbery, and sexual assault. Although yet to be empirically tested through direct comparison, researchers have argued that behavioural stability and

distinctiveness will be present to greater degrees in certain types of crime and, as a result, the degree of behavioural discrimination will vary across crime type. Interestingly, two competing hypotheses exist with respect to this factor.

One hypothesis is based on the idea that behavioural discrimination will vary based on the degree of psychopathology possessed by offenders committing a particular type of crime. For example, in their now classic study of criminal profiling, Pinizzotto and Finkel (1990) countered arguments that serial offenders do not exhibit enough stability and distinctiveness across their crimes for profiling to be possible by drawing on research from the field of personality psychology, which suggests that pathological populations exhibit more stability and distinctiveness than "normal" populations (e.g., Alker, 1972; Endler, 1973; Endler & Okada, 1975; Moos, 1968, 1969). It is currently unclear whether the serial rapists and killers that Pinizzotto and Finkel were concerned with in their study actually possess a significant degree of psychopathology, though some research suggests that they probably do (e.g., Blackburn, 1993; Monahan, 1992). Further, the mechanism that leads people suffering from psychopathology to exhibit more stability and distinctiveness than normal individuals has never been articulated. However, the "psychopathology hypothesis" originally proposed by Pinizzotto and Finkel generally leads to the expectation that serial offenders committing interpersonal violent crimes (specifically serial sexual assault and homicide) will exhibit more stability and distinctiveness compared to serial offenders committing higher volume property crimes (such as serial burglary). By extension, this would lead one to expect higher levels of discrimination accuracy in cases of serial sexual assault and homicide compared to serial burglary.

The contrasting hypothesis is based on the idea that behavioural discrimination will vary as a function of how much control the offender has to make decisions within the context of their crimes (versus the decisions being influenced by situational factors). For example, Bennell and Canter (2002), who originally proposed this hypothesis, argued that the stability and distinctiveness of criminal behaviour is determined primarily by how situation-dependent, or context-dependent, the behaviours are. In the case of burglary, for example, the decision about where to commit a crime is clearly less situation-dependent than the decision about what property to steal (given that the items stolen depend largely on what is available to steal). Across crime types, crimes involving face-to-face encounters with victims would presumably be classified as being less under the control of offenders (e.g., due to victim resistance, etc.) than crimes where victims are absent. Thus, in contrast to the psychopathology hypothesis described above, the "context-dependency" hypothesis leads one to expect higher levels of discrimination accuracy in cases of serial burglary compared to serial sexual assault and homicide.

Unlike the psychopathology hypothesis, the context-dependency hypothesis is supported (to some extent at least) by existing research. For example, at a theoretical level, the hypothesis is consistent with studies of non-criminal stability in the field of personality psychology where operant behaviours (i.e., behaviours emitted by the person across a range of situations) are typically exhibited in a more stable fashion than respondent behaviours (i.e., behaviours that require specific, eliciting stimuli within situations) (Funder & Colvin, 1991; Furr & Funder, 2004). At an empirical level, this hypothesis has also received preliminary support. For example, one of the most consistent findings in BLA research is that behaviours that allow offenders to control their

environment and reduce situational interference tend to be exhibited in a more stable and distinct fashion than any other type of offending behaviour (e.g., Bennell, 2002; Bennell & Canter, 2002; Bennell & Jones, 2005; Goodwill & Alison, 2006; Grubin et al., 2001; Woodhams & Toye, 2007). Having said this, all but one of the existing studies in this area has focused on property crime. Therefore, to date, the context-dependency hypothesis has only been explored by comparing the levels of stability found across behaviours that are exhibited by property offenders, with some behaviours expected to be more context-dependent (and therefore less stable) than others.

In sum, despite some initial support for the context-dependency hypothesis, there has yet to be a study that directly compares the degree of discrimination accuracy that can be achieved across crime types using the same analytical method. Thus, this thesis will examine how crime type influences the positioning of the probability distributions associated with ROC analysis and, in turn, how discrimination accuracy varies across crime types. This will be done by incorporating the three crime types that are examined in this thesis – serial sexual assault, homicide, and burglary – into the analysis of two other potentially influential factors – similarity coefficient and behavioural frequency.

Similarity Coefficient

The similarity coefficient can be considered the basis for most approaches to linkage analysis, as it is used to quantify the degree of similarity that exists between crimes. Research conducted in other contexts has demonstrated that the use of different similarity coefficients can drastically influence results in tasks that are similar in many ways to BLA (e.g., Baroni-Urbani & Buser, 1976; Cheetham & Hazel, 1969; Gower & Legendre, 1986; Hubalek, 1982; Kosman & Leonard, 2005; Lamboy, 1994; Simberloff &

Connor, 1979). There is no reason to believe that this will not also be true with respect to BLA, and several individuals conducting research in this domain have begun to discuss the possibility that the choice of similarity coefficient will influence discrimination accuracy (e.g., Bennell, Jones, & Melnyk, 2007; Gauthier, 2008; Woodhams, Grant, et al., 2007).

Specifically, questions have been raised about the suitability of using Jaccard's coefficient, *J*, for measuring across-crime similarity (Jaccard, 1908). By far, this is the most commonly used similarity measure in the context of BLA (e.g., Bennell, 2002; Bennell & Canter, 2002; Bennell & Jones, 2005; Canter et al., 1991; Goodwill & Alison, 2006; Salfati, 2000; Salfati & Bateman, 2005; Woodhams & Toye, 2007), despite the wide variety of measures that are available for use (Liebetrau, 1983). Part of its appeal is probably its simplicity. Indeed, for a pair of crimes (Crime A and B), *J* is simply equal to:

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

where *a* equals the number of behaviours common to both crimes and *b* and *c* equal the number of behaviours unique to Crime A and B, respectively.

In addition, it is commonly argued that *J* is the most appropriate coefficient for use in BLA because joint non-occurrences of behaviour are not included in its calculation (Woodhams, Hollin, et al., 2007). That is, if a specific behaviour is absent from the two crimes in question, *J* will not increase. Within the investigative domain, it is thought that ignoring joint non-occurrences in this way is wise because the absence of a behaviour in any given crime, or crime report, may be due to factors other than its actual non-

occurrence (e.g., a specific behaviour might not have been witnessed by the victim, remembered, or reported to the police, and even if the behaviour was reported it may not have been accurately recorded; Alison, Snook, & Stein, 2001).

Beyond debates over the validity of this argument, the other major drawback of *J* is that it is a very crude similarity metric (Woodhams, Grant, et al., 2007). For example, *J* only accounts for across-crime similarity at the most discrete behavioural level and, therefore, it is likely going to be sensitive to missing information (which is an extremely common occurrence in police data). Issues such as these have led researchers to suggest that other similarity coefficients be examined to determine if they might be more suitable for BLA (e.g., Bennell et al., 2007). One coefficient in particular seems to be particularly promising in this regard – the taxonomic similarity index, Δ_s (Izsak & Price, 2001; Woodhams, Grant, et al., 2007).

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

Consider the following example provided by Gauthier (2008). During Crime A, the offender uses manual gagging to physically control the victim, while during Crime B

the offender uses manual restraints to physically control the victim. Using *J*, these behaviours would contribute nothing to the across-crime similarity score. However, Δ_s would identify a degree of similarity across the crimes because there was evidence of physically controlling tactics used by the offenders in both crimes (albeit less similarity than if there was an exact match at the level of the discrete behaviours).

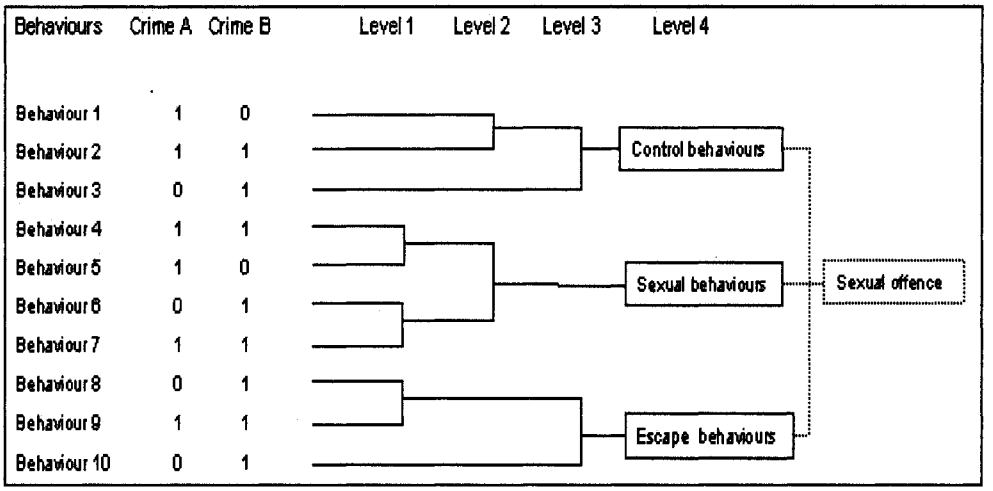


Figure 3. A hypothetical behavioural hierarchy of crime scene behaviours in sexual assault. The first two columns indicate the presence/absence of 10 specific behaviours across the two crimes, A and B. (Source: Woodhams, Grant, et al., 2007).

For the purpose of illustrating how Δ_s is calculated, consider the hypothetical behavioural hierarchy for sexual assault that is presented in Figure 3. The calculation requires two steps. First, the taxonomic distance between Crime A and Crime B is calculated:

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

where w_{iB} is the minimum path length between behaviour i at Crime A and all behaviours at Crime B, w_{jA} is the minimum path length between behaviour j at Crime B and all behaviours at Crime A, and n_A and n_B are the number of behaviours at Crime A and B, respectively. The path length is an integer that represents the difference in levels of the hierarchy. Second, Δ_s is calculated by:

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

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

In the first examination of Δ_s within the investigative domain, Woodhams, Grant, et al. (2007) compared J to Δ_s using behavioural data from 16 sexual offences committed by seven offenders. By drawing on a behavioural hierarchy consisting of 4 levels and 55 offence behaviours the results of this study demonstrated, as predicted, that the similarity scores for both linked and unlinked crimes were significantly higher when using Δ_s

[1] As an example of how Δ_s is calculated, consider the scenario presented in Figure 3. The minimum path length is determined between each behaviour in Crime A and all those in Crime B, and vice versa. For example, the values for the minimum path lengths for behaviours in Crime A are as follows: 2 for Behaviour 1, as it is absent as a specific behaviour in Crime B, but present in Crime B at level 2, 0 for Behaviour 2 because it is also present in Crime B, 0 for Behaviour 3, etc. This process is also completed for Crime B,and the path lengths for each crime are added together (i.e., [2+0+0+1+0+0] + [0+3+0+1+0+1+0+3]) and divided by the sum of the total number of behaviours present for each crime (i.e., 6+8) to obtain the taxonomic distance. In this case, *TD (A,B)* = .79. To obtain Δ_s, *TD* is divided by the number of taxonomic levels minus 1, the total of which is subtracted from 1 (i.e., 1-[.79 /3] = .74).

compared to *J*. In addition, and as predicted, both *J* and Δ_s resulted in significantly higher similarity scores for linked compared to unlinked crimes. Perhaps most importantly, the significant differences that existed between linked and unlinked crimes remained across conditions of data degradation (i.e., with 10%, 20%, and 50% of behaviours randomly selected and removed) when using Δ_s, but not when using *J*.

Woodhams, Grant, et al. (2007) suggest that these results offer clear evidence that Δ_s may be more suitable for BLA than *J*, especially when a substantial amount of data is missing. However, from a signal detection perspective, it is unclear whether this is the case. For example, it is difficult to determine from the results of this study whether the use of Δ_s decreases the degree of overlap between the distributions in Figure 1, or whether it simply shifts both distributions up the *x*-axis. From the perspective adopted in this thesis, using a coefficient that increases behavioural stability will have little impact on discrimination accuracy, if the same coefficient also decreases behavioural distinctiveness.

This is precisely what Gauthier (2008) found in an attempt to replicate Woodhams, Grant, et al.'s (2007) study. In contrast to what was reported in that previous study, Gauthier reported no significant difference in linking performance when comparing the two coefficients, regardless of sample size or degree of data degradation. To a large extent, these results appeared to be due to the much larger sample size that was utilized by Gauthier (i.e., the small sample size utilized in the previous study likely results in unreliable findings). This thesis will explore further the impact that a similarity coefficient can have on discrimination accuracy. This will be done by using ROC analysis to compare the performance of Δ_s and *J* across datasets consisting of serial

sexual assault, homicide, and burglary behaviours, and by examining the underlying probability distributions for these two coefficients. Issues related to sample size and data degradation will also be examined.

Behavioural Frequency

The last factor to be discussed is behavioural frequency. A common assumption held by many researchers who investigate BLA is that behavioural frequency will also influence discrimination accuracy (e.g., Canter, Alison, Alison, & Wentink, 2004; Canter, Bennell, Alison, & Reddy, 2003; Canter & Fritzon, 1998; Salfati & Bateman, 2005; Salfati & Dupont, 2006; Santtila, Junkkila, et al., 2005; Woodhams, Hollin, et al. 2008). In other words, it is thought that basing across-crime similarity measures on behaviours that vary with respect to their frequency of occurrence in a sample will influence the results of BLA.

For example, many researchers assume that a higher degree of discrimination accuracy in BLA will be achieved when omitting very high frequency behaviours. For example, Canter et al. (2003) stated that "when attempting to link a potential series of unsolved [sexual assaults] to the offender responsible, care should be taken to avoid spurious links that may result if too much weight is given to the sexual behaviours exhibited by the offender(s)" (p. 171). These researchers made this claim because their multidimensional scaling analysis of sexual assaults indicated that sexual behaviours (e.g., vaginal penetration) represented the highest frequency behaviours in sexual assault, which should come as no surprise given that many of these behaviours are the centrally defining features of this crime. Although such hypotheses have not been tested, it makes sense that basing BLA on a set of crime scene behaviours that are exhibited by nearly

every offender in a sample will be of little use because behavioural distinctiveness under such circumstances will be virtually non-existent.

Other researchers also caution against the use of high frequency behaviours for the purpose of BLA, but are more lenient than Canter et al. (2003) with respect to their definition of what constitutes a "high frequency" behaviour. For example, Salfati and Bateman (2005) recommend that behaviours with frequencies of 50% or greater should be removed from any linkage analysis given that these behaviours will be unable to discriminate between crimes committed by different offenders (due to a lack of behavioural distinctiveness). However, these researchers have yet to establish that this is in fact the case, as they have not conducted any formal comparative analysis. In addition, it is difficult to visualize what the impact will be (on the underlying probability distributions in Figure 1) of removing all behaviours from a sample of crimes that exceed a frequency cut-off of 50%.

At the other extreme to the recommendations made by Canter et al. (2003) and Salfati and Bateman (2005), are recommendations to remove very low frequency behaviours when linking serial crimes. For example, Santtila, Junkkila, et al. (2005) argue that very low frequency behaviours (<1% occurrence in a sample) are of little use in the context of BLA. As above, others make similar arguments, but use less stringent criteria, such as removing any behaviours that occur in less than <10% of cases (e.g., Woodhams, Hollin, et al., 2008). While these researchers have never articulated the rationale behind this recommendation, from a signal detection perspective it appears logical. In contrast to high frequency behaviours, low frequency behaviours may have no negative impact on behavioural distinctiveness, but it seems clear that they could

decrease the degree of behavioural stability present in a sample (especially if the behaviours are so infrequent that they cannot occur across the entire crime series of even a single offender).

The only research to date that has empirically examined the behavioural frequency hypothesis (i.e., that behavioural frequencies influence discrimination accuracy) is a study conducted by Melnyk, Bennell, and Jones (in press). They analyzed a sample of serial sex offences and examined how discrimination accuracy varied when using low (<20%), moderate (20-50%), and high frequency (>50%) behaviours. Across-crime similarity scores (*J*) were calculated for every pair of crimes in the sample using the behaviours from each frequency band. These scores were then subjected to logistic regression analysis and ROC analysis. The results indicated that it was possible to discriminate between crimes committed by different offenders (beyond chance levels) using behaviours from all frequency bands, however the degree of discrimination accuracy varied across the bands. Specifically, moderate frequency behaviours were the most effective for linking purposes ($AUC = .70$), followed by high ($AUC = .62$) and low ($AUC = .62$) frequency behaviours. Interestingly, it was when all of the behaviours were combined that the highest level of discrimination accuracy were achieved ($AUC = .75$).

Given the limited nature of the study conducted by Melnyk et al. (in press), this thesis will explore the frequency issue further by using ROC analysis, and by examining the underlying probability distributions that result when linking crimes using behaviours that vary with respect to their frequencies. Specifically, the similarity coefficient (J or Δ_s) found to be optimal in the first set of analyses will be used to examine the various

recommendations that are being made in this area. These issues will be examined using serial sexual assault, homicide, and burglary data.

The Current Study

The goal of this study is to contribute further to research on BLA from a signal detection perspective. The thesis accomplishes this goal by relying on ROC analysis to examine a range of factors that may influence the degree of discrimination accuracy that is found in the investigative context. Due to the preliminary nature of research on the topic of BLA, directional hypotheses will not be tested. However, three factors will be examined – crime type, similarity coefficient, and behavioural frequency – to determine empirically their impact on discrimination accuracy. The findings from this research will contribute to our understanding of the conditions under which serial offenders exhibit behavioural stability and distinctiveness, and this increased understanding has the potential to contribute to linking efforts in naturalistic settings (e.g., by providing the parameters under which the greatest degree of discrimination accuracy is likely to be achieved).

Method

Data

Three datasets will be used to examine the research questions in this thesis. The datasets relate to the crimes of serial sexual assault, homicide, and burglary.

Serial Sexual Assault Data

The serial sexual assault data used for the present analyses was originally collected by Canter, Wilson, Jack, and Butterworth (1996). The dataset contains information on 36 crime scene behaviours for 42 convicted male serial rapists from the

UK who committed a total of 126 sexual assaults (three crimes per offender) during the early 1990s. The original sampling procedure followed the common practice in BLA research of limiting the data to three crimes per offender in order to ensure that each offender was equally represented in the analyses (e.g., Bennell & Canter, 2002; Bennell & Jones, 2005; Santtila, Junkkila, et al., 2005; Woodhams & Toye, 2007). Capping crimes in this way prevents analyses from being biased by undue weight being assigned to highly prolific offenders displaying particularly high or low levels of behavioural stability and/or distinctiveness.

In this case, the first three detected crimes in each offender's series comprised the data. Given that an offender's crimes may not be detected in the order they are committed, and that an offender's crime series likely contains non-detected crimes interspersed with those that are known to the police, the choice of the first three detected crimes for use in this analysis is likely as representative of the offender's series as any other selection method (e.g., using the first, middle, and last crimes in a series). Of more importance than this sampling issue is the fact that three crimes were sampled for each offender. Three crimes per offender was considered the minimum number of crimes that could be used in this thesis because, currently, serial offenders are defined as offenders who commit three or more crimes on three separate occasions (e.g., Holmes & DeBurger, 1988).

Unlike the other datasets used in this thesis, all of the data in the serial sexual assault dataset was extracted directly from victim statements that were collected by police officers as a part of criminal investigations. These victim statements were provided to the previous researchers by various police forces from around the UK and were content

analyzed by trained researchers (Canter et al., 1996). In typical fashion, the content categories were derived from the published literature on sexual assault and from a thorough qualitative analysis of the victim statements. For each crime, behaviours were coded as being present (1) or absent (0). Levels of inter-rater reliability were not available for the sexual assault data and such levels could not be calculated retroactively, since it was not possible to access the actual victim statements (however, see pp. 33-34 for an estimate of reliability). See Appendix A for a complete content dictionary of the serial sexual assault variables.

Serial Homicide Data

The serial homicide data used in the present analyses was originally collected by Godwin (1998). In addition to other types of data, this dataset contains information on 39 crime scene behaviours for 79 convicted male serial killers from the US who committed a total of 237 homicides between 1950 and 1998. The data is restricted to three crimes per offender (for reasons discussed above) and these offences represent the first three detected crimes in each offender's series.

The majority of the data (approximately 75%) in Godwin's (1998) dataset was extracted directly from the Homicide Investigation and Tracking System (HITS), which is housed in the Attorney General's Office in Seattle, Washington. The HITS database was developed in 1987 by Keppel with the original goal of standardizing the collection of information in interpersonal crime investigations occurring in Washington (since that time, the data capture area of HITS has been expanded to include Oregon, California, Idaho, Kansas, and parts of Canada) (Keppel & Weiss, 1992). The remainder of Godwin's sample (approximately 25%) was extracted from various other police

databases, such as the FBI's Violent Criminal Apprehension Program (VICAP) (Ressler, Burgess, & Douglas, 1988).

The information included within these police databases comes directly from reports prepared by police officers as a part of criminal investigations. The information relates to a range of variables, including the behaviour of the offender at their crime, the existence of physical (and any other sort of) evidence, information about the victim, environmental factors associated with the crime (e.g., whether the offence occurred indoors or outdoors), and offender information obtained post-arrest (e.g., background characteristics). While much of this information is included in Godwin's (1998) original dataset, for the purpose of the current thesis only information pertaining to crime scene behaviours was of interest. This information was coded in the serial homicide dataset in dichotomous form with a 1 indicating the presence of a behaviour in a crime, and a 0 indicating its absence. Again, levels of inter-rater reliability were not available for this data (however, see pp. 33-34). See Appendix B for a complete content dictionary of the serial homicide variables.

Serial Burglary Data

The serial burglary data used in the present analyses was originally collected by Bennell (2002). The dataset contains information on 28 crime scene behaviours for 42 convicted male serial burglars from the UK who committed a total of 210 residential burglaries (five crimes per offender) from 1997 to 1998. These offences represent the first five detected crimes in each offender's series (for reasons discussed above). The data was extracted directly from a burglary database managed by the Dorset Police Service in Dorset, UK.

The information in this database relates to a range of variables, including the behaviour of the offender at their crime(s), the existence of physical (and any other sort of) evidence, and information about the targeted property. For the purpose of the current thesis, only information pertaining to crime scene behaviours and target characteristics was of interest. For each crime, behaviours and characteristics were coded as being present (1) or absent (0). Levels of inter-rater reliability were not available for this data (however, see below). See Appendix C for a complete content dictionary of the serial burglary variables.

A Note on the Datasets

There are two important limitations inherent in the use of police data for research purposes that warrant discussion. First, rarely is information available about the degree of inter-rater reliability associated with police data and this is the case for the current datasets. While there are several factors that are likely to decrease the reliability of police data (e.g., failure on the part of the police to accurately record details from victims; Alison, et al. 2001), there are also a variety of factors that will likely increase its reliability (e.g., the data is collected with the expectation that it will likely be used, and potentially challenged, in a court of law; Bennell, Alison, Stein, Alison, & Canter, 2001). It is also important to point out that crime scene behaviours across each of the three datasets were coded using data collection proformas that were specifically constructed to be as objective as possible. The few studies that do report levels of inter-rater reliability for police data indicate that the type of data used in the current study is likely to have been coded in a relatively reliable fashion. For example, Alison and Stein (2001) report disagreement levels in the range of 3-4% in their study of serial sexual assault where the

data was extracted from victim statements. Santtila, Junkkila, et al. (2005) report similar levels of high reliability, as do Hakkanen, Puolakka, and Santilla (2004) (where in both studies the data was extracted from case files containing both victim statements and police reports).

Second, using investigative information as a data source means that the data is obviously only representative of serial crimes that the police are aware of (i.e., detected and solved offences). Such data may tell us relatively little about unreported or unsolved serial crimes, which could detract from the practicality of the results reported in this thesis (Bennell & Canter, 2002). This problem may be particularly important in the current study because one of the reasons why serial crimes may be solved in the first place is because they are associated with high levels of behavioural stability and distinctiveness. Unfortunately, this problem exists in all research on the topic of BLA, in addition to many other areas of research in the domain of investigative psychology (e.g., Mokros & Alison, 2002; Snook, Taylor, & Bennell, 2005), and there is no way of avoiding this limitation.

Having raised these two limitations, it is also important to point out the obvious advantage of using police data for the type of research being conducted here. Perhaps the most important advantage is that the data that is being used in the current study takes the exact same form as the data currently being used by the police to perform their daily tasks, including BLA. Indeed, the primary purpose of databases like HITS and VICAP is to link serial crimes (Keppel & Weiss, 1992; Ressler et al., 1988), and this is also the case with the database being used by the Dorset Police Service (Merry, 2000). Data extracted from victim statements is also regularly used by the police for the purpose of BLA

(Alison et al., 2001). Thus, given the relatively high level of external validity associated with police data, we can be fairly confident that the results found in the current study will have the potential to support BLA efforts in naturalistic settings.

Procedure

Study 1 – Similarity Coefficient and Crime Type

Each of the three datasets were used to examine differences that potentially emerge when using Jaccard's coefficient (*J*) versus the taxonomic similarity index (Δ_s) to discriminate between crimes committed by the same offender versus different offenders. The procedure for carrying out this comparison replicated the approach adopted by Gauthier (2008) and as such involved several steps:

1. Behavioural hierarchies for each of the three crime types were constructed.
2. The two similiarity coefficients under investigation, *J* and Δ_s, were calculated across each of the following: (1) 126 linked and 7,749 unlinked sexual assault crime pairs, (2) 237 linked and 27,729 unlinked homicide crime pairs, and (3) 420 linked and 21,945 unlinked burglary crime pairs.
3. Distributions of the across-crime similarity scores were constructed for linked and unlinked crime pairs for *J* and Δ_s, across each crime type. For each dataset, the distributions of scores were plotted on two separate graphs (one for *J* and one for Δ_s), in a similar way to what was done in Figure 1.
4. Descriptive and comparative analyses were conducted for these distributions, across each crime type. Statistics included the range, median, mean and standard deviations, from which Cohen's *d* values were also calculated. In addition, non-

parametric *t*-tests assessed mean differences between linked and unlinked distributions for statistical significance.

5. ROC curves were created on the basis of the constructed distributions and *AUC*s were calculated.
6. A data degradation procedure was used on each dataset. This involved progressively removing approximately 10%, 25%, and 50% of randomly selected behaviours from the datasets, and proceeding with Steps 2-5.
7. A re-sampling procedure was used with respect to the original datasets of each crime type in order to examine the influence of sample size on the relative linking accuracy of *J* and Δ_s. This procedure involved generating new sample sizes ranging from 10 to 10,000 crime pairs, and proceeding with Steps 2-5. For each of the three crime types, the re-sampling procedure was conducted using equal numbers of linked and unlinked crime pairs and numbers of linked and unlinked crime pairs that were in the right proportions for each dataset (i.e., more unlinked crimes than linked crimes). The re-sampling procedure was also conducted across each level of data degradation.
8. Finally, ROC analysis was used to explore the impact of setting different decision thresholds.

The above eight steps are described in more detail in the following sections.

Construction of behavioural hierarchies. First, for each of the separate datasets, hierarchies of crime scene behaviours were developed in order to calculate Δ_s. There is no set procedure for deriving these hierarchies within the context of BLA. In the current study, the hierarchies for each dataset were derived both from published literature on

each type of serial crime (top-down) and a cursory examination of the data (bottom-up). An attempt was made wherever possible to construct the hierarchies on the basis of empirical results, primarily from studies using multidimensional scaling and cluster analytic techniques to derive the underlying structure of offence behaviour for each respective crime type. However, at times, such results seriously lacked face validity. Because of this, a degree of subjectivity was also required in the construction of the hierarchies when making decision about variable placement.[2]

Calculation of J and Δ_s. In order to calculate J and Δ_s for each of the separate datasets, a specially designed computer program called CrimeSolver (written using Mathcad) was used for analysis. A flow chart summarizing its procedure, as outlined by Gauthier (2008), is provided in Appendix D. For each of the separate datasets, all of the dichotomously coded crime scene behaviours were submitted as input to CrimeSolver as Microsoft Excel files for the purpose of calculating both similarity coefficients. Using Microsoft Excel, path lengths tables illustrating the relationships between all behaviours in the datasets were constructed for each hierarchy and were also input into CrimeSolver in order to calculate Δ_s (see Appendices E-G). On this basis, CrimeSolver then calculated J and Δ_s for every possible pair of crimes in each dataset, and provided this information as output. CrimeSolver's accuracy was previously verified by Gauthier (2008).

---

[2] It should be noted that, although it is important to construct a hierarchy that is as sound as possible, it is impossible to construct the "perfect" hierarchy for any crime type. The particular branches to be included, or the placement of specific behaviours in the hierarchy, could always be debated and, as such, this can be regarded as one of the inherent limitations associated with using Δ_s. Every attempt was made to construct the hierarchies in this thesis on the basis of published literature, however it should be emphasized that the purpose of the study is to examine the logic of Δ_s (versus J) as opposed to establishing what constitutes "true" behavioural domains of each crime type. This sort of work goes well beyond the scope of the current thesis.

Descriptive and comparative analyses. CrimeSolver provided descriptive statistics for each dataset on the basis of its generated distributions of linked and unlinked crime pairs. The distributions associated with both J and Δ_s were plotted for each crime type. Descriptive statistics included the range, median, mean, and standard deviation for each of the distributions. CrimeSolver also provided Cohen's d values for both similarity coefficients as a measure of their ability to discriminate between linked and unlinked crimes. The output from CrimeSolver was exported to SPSS (v. 16). In order to determine whether mean differences between the linked and unlinked crime distributions for each crime type were statistically significant, non-parametric t-tests were used.

ROC analysis. CrimeSolver was also capable of conducting ROC analysis for both similarity coefficients (and providing AUCs and CIs_{95}). The ROC subroutine of CrimeSolver was previously validated by Gauthier (2008) using worked-out examples provided by Swets (1996) and an ROC analysis applet (Eng, 2008).

Data degradation. The data degradation procedure utilized by Woodhams, Grant, et al. (2007) and Gauthier (2008), was replicated in order to test the robustness of the two similarity coefficients to missing information. This procedure involved progressively removing randomly selected behaviours from each file (accounting for approximately 10%, 25%, and 50% of each dataset). The same procedure as outlined above was then followed for each degraded dataset across crime type, in order to evaluate the discrimination accuracy of the two similarity coefficients under such conditions.

Re-sampling procedure. The re-sampling procedure utilized by Gauthier (2008) was replicated in order to test the influence of sample size on the relative discrimination accuracy of J and Δ_s. For each crime type, CrimeSolver generated both equal numbers of

linked and unlinked crime pairs, as well as numbers that were proportional to the linked and unlinked crime pairs in each dataset, where the total number of pairs varied from 10 to 10,000. Equal numbers of linked and unlinked crime pairs were selected in an attempt to replicate Gauthier (2008), who used this procedure because Woodhams, Grant et al. (2007) restricted their sample of crime pairs in this way for their analysis. Numbers of linked and unlinked crime pairs that were proportional to the number of actual pairs in each dataset were selected in an attempt to more appropriately reflect the circumstances encountered in naturalistic settings where there will always be many more unlinked crimes than linked crimes.

This procedure was followed for each original dataset as well as at each level of degradation. Each of the previous steps in the procedure was then followed for each of the newly generated samples in order to evaluate the discrimination accuracy of the two similarity coefficients under conditions where sample size varied. CrimeSolver plotted *d* values on re-sampling graphs as a function of the number of random pairs in the samples. This procedure was repeated 10 times for each sample size.

Impact of setting different decision thresholds. Finally, ROC analysis was used to explore the impact of setting different decision thresholds. Specifically, an examination was carried out looking at how ratios of *p*H to *p*FA changed across various thresholds. The two strategies discussed above were examined to see what *p*H to *p*FA ratios emerged. To reiterate, this involved: (1) examining the ROC point falling closest to the upper-left corner of the ROC graph in order to maximize *p*H while minimizing *p*FA and (2) setting a pre-determined rate for pFA (.50) and maximizing *p*H without exceeding this value.

Study 2 – Behavioural Frequency and Crime Type

As *J* was the similarity coefficient resulting in the highest degree of discrimination accuracy for each crime type in Study 1, it was used in Study 2 to carry out all analyses. Once again, each of the three datasets was used to examine the potential impact that behavioural frequency has on discrimination accuracy. As in Study 1, the procedure for carrying out the analyses involved several steps:

1. Each of the three datasets was separated into subsets representing the four recommendations that have been made by researchers with respect to behavioural frequency. Specifically, the following subsets were created for each crime type: (1) behaviours with frequencies > 90% were removed from the dataset in order to test Canter et al.'s (2003) recommendation of omitting very high frequency behaviours, (2) behaviours with frequencies > 50% were removed from the dataset in order to test Salfati and Bateman's (2005) recommendation, (3) behaviours with frequencies < 10% were removed from the dataset in order to test Woodhams, Hollin, et al.'s (2008) recommendation, and (4) behaviours with frequencies < 1% were removed from the dataset in order to test Santtila, Junkkila, et al.'s (2005) recommendation.
2. For each recommendation, *J* was calculated between every possible linked and unlinked crime pair for each crime type using a program called *B-LINK* (for more information on this program, see Bennell, 2002). The relevant data from Study 1 (i.e., *J* values for all behaviours, irrespective of their

frequencies) was also examined. *B-LINK* output was then exported to SPSS (v. 16) for analysis.

3. Distributions of the *J* values were constructed for linked and unlinked crime pairs for each recommendation, across each crime type. For each dataset, the distributions of scores were plotted on separate graphs for each recommedation, in a similar way to what was done in Figure 1.
4. Descriptive and comparative analyses were calculated for these distributions, across each crime type. Statistics included the range, median, mean and standard deviations. In order to assess mean differences between linked and unlinked distributions for statistical significance, non-parametric *t*-tests were also conducted.
5. ROC curves were created for each recommendation, across each crime type, on the basis of the constructed distributions. *AUC*s were calculated.

Results

Study 1 – Similarity Coefficient and Crime Type

Serial Sexual Assault Data

Serial sexual assault behavioural hierarchy. The hierarchy for sexual assault that was developed by Woodhams, Grant, et al. (2007) was used as the starting point in the development of the current hierarchy. For example, in line with their hierarchy, the current hierarchy included control behaviours and sexual behaviours as major branches. This is in accordance with existing classification research (e.g., Canter & Heritage, 1990; Grubin et al., 2001; Knight & Prentky, 1990). In brief, control behaviours include those verbal and physical behaviours that are necessary "to create and maintain an environment

in which the crime can [successfully] take place" (Grubin et al., 2001, p. 14), whereas sexual behaviours include any verbal or physical sexual acts exhibited by the offender or requested from the victim (Woodhams, Grant, et al., 2007). In contrast to the previous hierarchy, no branches for escape or style behaviours were included in the current hierarchy. Not only is there little research to support the existence of these branches, many of the behaviours included in these domains were not included in the current data set, and those that did exist could arguably be categorized more appropriately in other domains.

Also in contrast to the previous hierarchy, three new branches were proposed, consisting of pseudo-intimacy, aggression, and theft categories. These three branches are well established in existing sexual assault classification systems (e.g., Canter et al., 2003; Hakkanen, Lindloff, & Santtila, 2004; Marshall, 1989) and have been studied in the context of BLA research (e.g., Gauthier, 2008; Melnyk et al., in press; Santtila, Junkkila, et al., 2005). The category of pseudo-intimacy includes behaviours that are of a more "intimate" nature, potentially reflecting the offender's desire for social contact with the victim, the aggression category includes verbal and physical behaviours that are explicitly destructive, forceful, or violent, and the theft category includes behaviours that reflect attempts to steal something from the victim for instrumental reasons.

Based on existing classification research, and some degree of logical inference, the 36 serial sexual assault behaviours available in the dataset were incorporated into what was deemed the most appropriate branch. To reduce the degree of subjectivity with this assignment process, inter-rater reliability agreement was assessed. This was done to ensure that independent raters could assign the crime scene behaviours to the same

branch when provided with definitions of the crime scene behaviours and the branches of the hierarchy (for an example of these definitions see Appendix A and H, respectively). Using a coding scheme whereby each individual branch of the hierarchy was given a code (e.g., aggression – physical = 1), the researcher and two independent coders who were not involved in the development of the hierarchy assigned each of the 36 behaviours a code. Based on code assignments, the average Kappa across the three raters was .81, indicating a satisfactory level of reliability (Landis & Koch, 1977). Discussion over the disagreements took place and the resulting hierarchy of sexual assault behaviour was established (see Figure 4). A path length table for this hierarchy, which illustrates the relationship between every behaviour in the dataset, is presented in Appendix E.

Note that, as was done by Woodhams, Grant, et al. (2007), and Gauthier (2008), the objective in the construction of the hierarchy was to reduce the path length between specific behaviours and hierarchial node wherever possible (e.g., see the theft and pseudo-intimacy branches). According to Woodhams and her colleagues (2007),

> This follows the original development of Δ_s in allowing the calculation of the shortest possible path length . . . The mathematical effect of this approach is to emphasize the similarity between offences at the lower subordinate levels where the alternative decision of increasing path lengths would emphasize dissimilarity at these levels. The overall effect, therefore, is to increase similarity between any two offences (whether linked or unlinked). (p. 21)

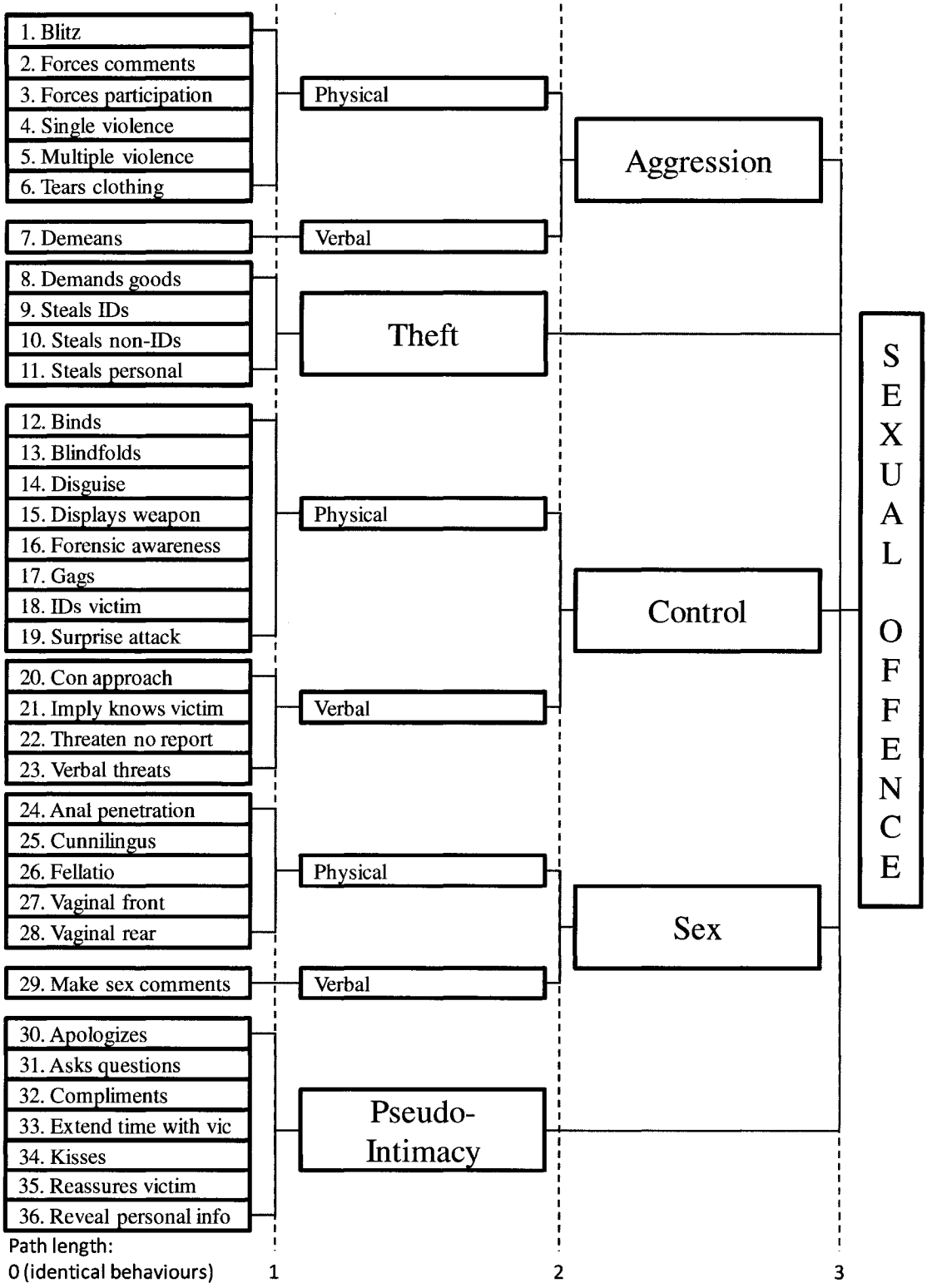


Figure 4. Serial sexual assault behavioural hierarchy.

Descriptive and comparative analyses. Distribution normality was examined by calculating CIs_{95} around the skewness value for each distribution (if the CIs_{95} include 0 the distributions are normally distributed) (Dunn, 2001). Results suggested negatively skewed linked and unlinked across-crime similarity score distributions for Δ_s (skew linked = -.90, SE = .22, CI_{95} = -1.33 to -.47; skew unlinked = -.91, SE = .03, CI_{95} = -.97 to -.85), and for J, normally distributed linked across-crime similarity scores (skew linked = .08, SE = .22; CI_{95} = -.35 to 0.51) and positively skewed distributions for unlinked across-crime similarity scores (skew unlinked = .55, SE = .03, CI_{95} = .49 to .61) (see Figure 5). This suggests that non-parametric tests should be used to compare the similarity scores from the two distributions. However, note that ROC analysis is not dependent on standard statistical assumptions and is largely unaffected by violations of normality (Swets, 1996).

In order to determine whether linked sexual assaults had higher across-crime similarity scores than unlinked sexual assaults, descriptive statistics were calculated (see Table 2). As normality was violated, non-parametric t-tests were used when comparing linked and unlinked distributions for Δ_s and J. Results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of whether Δ_s (mean rank linked = 5,724.56, mean rank unlinked = 3,908.95, U = 263,080.50, p <.001) or J (mean rank linked = 6,138.32, mean rank unlinked = 3,902.22, U = 210,947.00, p < .001) was used. However, it is important to note the substantial overlap that is apparent in the distributions of the linked and unlinked crimes, indicating that some linked sexual assaults were characterized by lower across-crime similarity

scores than some unlinked sexual assaults. As discussed, overlap of the distributions in such a way decreases the ability to link serial crimes.

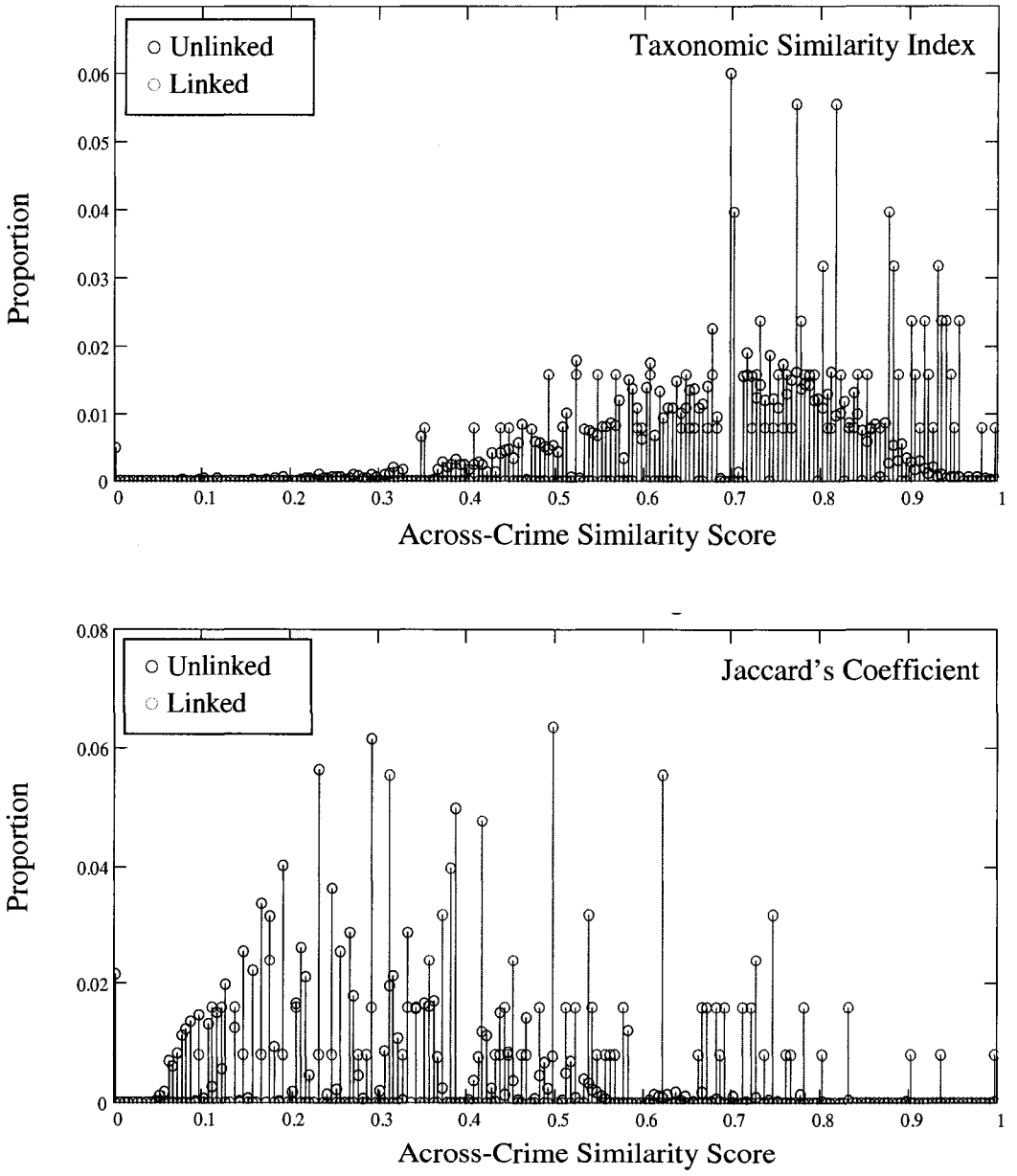


Figure 5. Distributions of across-crime similarity scores for linked and unlinked sexual assaults using Δ_s (top) and J (bottom).

Table 2. Descriptive statistics for Δ_s and *J* comparing linked and unlinked sexual assaults.

Statistic	Δ_s		*J*	
	Linked (n = 126)	Unlinked (n = 7,749)	Linked (n = 126)	Unlinked (n = 7,749)
Minimum	.33	.00	.08	.00
Maximum	.95	.96	.80	.86
Median	.76	.65	.39	.22
Mean	.75	.63	.39	.23
Standard Deviation	.13	.14	.16	.11

ROC analysis. The descriptive analysis demonstrates that Δ_s is capable of achieving higher across-crime similarity scores than *J* (M = 0.75 for Δ_s vs. M = 0.39 for *J*). However, it must be noted that, although this is the case, Δ_s also generated substantially higher scores for unlinked sexual assaults as well (M = 0.63 for Δ_s vs. M = 0.23 for *J*). As mentioned, the ability to discriminate between linked and unlinked crimes depends largely on the degree of overlap between the distributions. ROC analysis was conducted on the serial sexual assault data in order to evaluate the relative linking ability of the two similarity coefficients.

ROC curves produced by CrimeSolver, which illustrate the discrimination accuracy of Δ_s and *J*, are presented in Figure 6. As indicated in this figure, both similarity coefficients were capable of differentiating between linked and unlinked sexual assaults at a level significantly greater than chance. Comparison of the *AUC*s indicate that *J* (*AUC* = .81, SE = .02, CI_{95} = .77 to .85) outperforms Δ_s (*AUC* = .74, SE = .03, CI_{95} = .69 to .79) in terms of relative linking accuracy, although not to a significant degree (the *CIs*

associated with the two indices overlap). Furthermore, linking accuracy was greater for J than for Δ_s across the vast majority (if not all) decision thresholds, as is apparent in Figure 6.

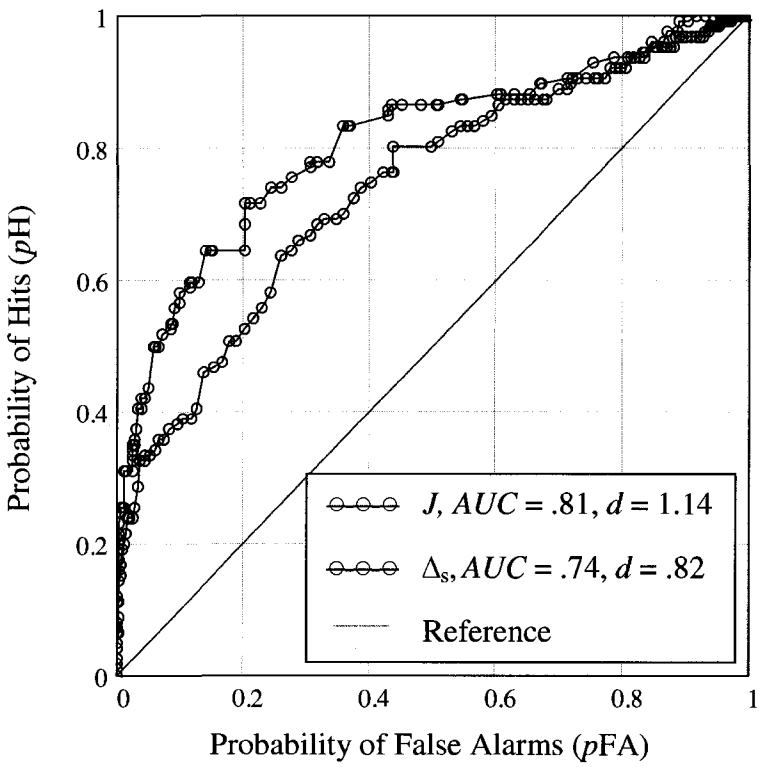


Figure 6. ROC curves illustrating comparative linking accuracy of Δ_s and J for sexual assaults.

Data degradation. In line with the Woodhams, Grant et al. (2007) study, data was progressively degraded from the dataset in order to test the robustness of the two coefficients across conditions of data degradation. Testing involved progressively

removing 4, 9, and 18 randomly selected behaviours, accounting for approximately 10%, 25%, and 50% of the dataset, respectively.[3] Table 3 reports the result of this analysis.

Table 3. The influence of data degradation on the discrimination accuracy of Δ_s and *J* for sexual assault

Number of Bhvs.	Δ_s				*J*			
	AUC	*SE*	CI_{95}	*d*	*AUC*	*SE*	CI_{95}	*d*
36 (100%)	.74	.02	.69 ≤ AUC ≤ .79	.82	.81	.02	.69 ≤ AUC ≤ .79	1.14
32 (90%)	.72	.02	.68 ≤ AUC ≤ .77	.72	.78	.02	.74 ≤ AUC ≤ .83	1.10
27 (75%)	.72	.03	.68 ≤ AUC ≤ .77	.78	.78	.02	.74 ≤ AUC ≤ .83	1.08
18 (50%)	.72	.03	.67 ≤ AUC ≤ .77	.76	.76	.02	.72 ≤ AUC ≤ .81	1.01

As is apparent from the results in Table 3, the removal of crime scene behaviours from the sexual assault dataset had only a slight impact on both coefficients (all *CIs* overlap). Interestingly, across the conditions, the *AUC* and *d* values associated with *J* remained consistently higher than those of Δ_s. Thus, both coefficients were able to discriminate between linked and unlinked sexual assaults even at high levels of data degradation (i.e., 50% of behaviours removed), particularly *J*.

Re-sampling procedure. All of the results up to this point contradict those of Woodhams, Grant, et al. (2007), who found the effect size associated with Δ_s to be higher

[3] It should be noted that an attempt was also made to remove corresponding amounts of values from the datasets, as opposed to behaviours, as this scenario is more likely to resemble problems that emerge when handling police data. However, CrimeSolver is currently incapable of handling missing values. The program is being adapted in order to manage missing values in its calculations of similarity coefficients, which will present an interesting avenue for future research.

than that of *J*. As suggested by Gauthier (2008), the very small sample size of 11 linked and 11 unlinked sexual assaults that was used in this initial study may be one explanation for the discrepancy in findings. The re-sampling procedure used by Gauthier was utilized as a way of examining this issue in the current study.

Figure 7 illustrates the effect of the re-sampling procedure. Specifically, this figure presents a pair of graphs illustrating the effect of sample size on the discrimination accuracy of both Δ_s and *J* as measured by *d,* for the complete (i.e., non-degraded) sexual assault dataset. The top plot reflects *d* values that were calculated using an equal number of randomly selected linked and unlinked sexual assaults in each sample (as was done by Woodhams, Grant, et al., 2007). The bottom plot reflects *d* values that were calculated using the number of linked and unlinked sexual assaults proportional to the dataset (1: 62.5) (as would occur in naturalistic settings). Each plot is based on an average of 10 repeats of the random trials.

Across both graphs it can clearly be seen that the effect size is more erratic for the two coefficients when the sample size is smaller. This is especially true in the bottom plot. That is, when analyses are based on relatively small sample sizes (e.g., <100 pairs) the effect sizes for both coefficients vary dramatically across the trials and, on occasion, Δ_s outperforms *J*. However, as the number of crime pairs increases, the results become more stabilized in both graphs, though slightly less so in the bottom plot. These graphs clearly demonstrate that *J* consistently achieves higher *d* values than Δ_s when a relatively large number of crime pairs are examined, thus confirming that the results of Woodhams, Grant et al. (2007) could be attributed to their use of a very small sample. The exact same pattern of results emerged when running the same analysis with degraded data, except

that the results were slightly more erratic for the proportional analysis at high levels of data degradation.[4]

[4] CrimeSolver cannot currently conduct the re-sampling analysis using *AUC* values. However, several *AUC*s were manually calculated by Gauthier (2008) using equal numbers of linked and unlinked sexual assaults. In line with the results presented here, the findings indicated that effect sizes were unreliable with small sample sizes. As the sample size was increased, *J* was found to consistently outperform Δ_S.

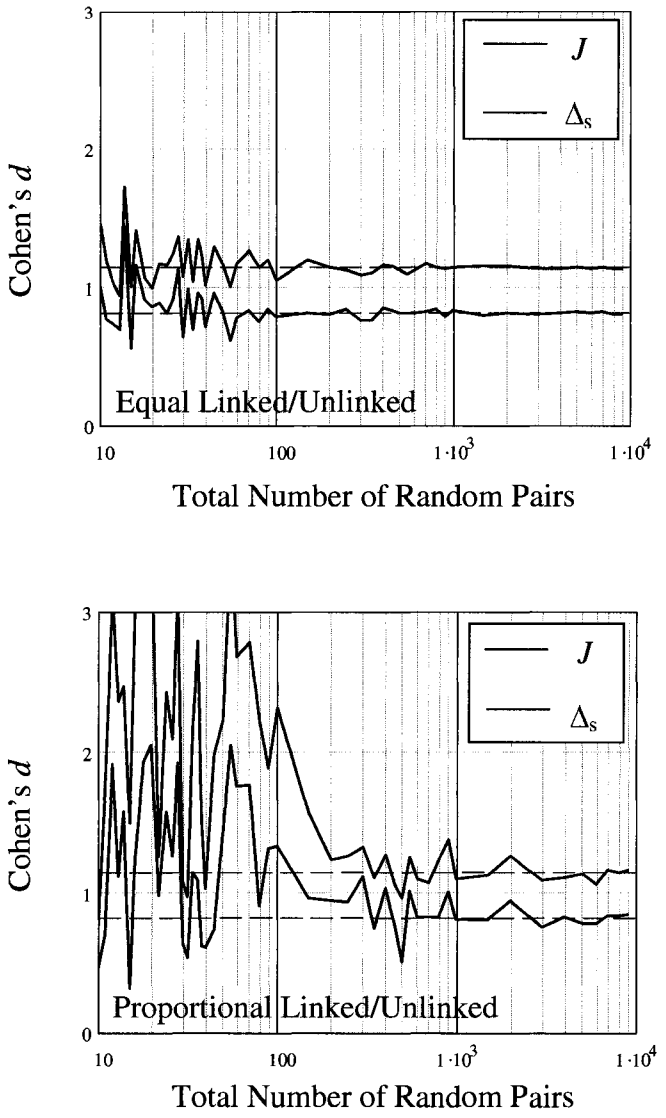


Figure 7. Effect of increasing sample size on linking accuracy for equal numbers of linked and unlinked sexual assaults (top) and numbers of linked and unlinked sexual assaults proportional to the dataset (bottom; 1:62.5).

Impact of setting different decision threholds. Finally, the ROC graph in Figure 6 was used to explore the impact of setting different decision thresholds on the ability to link serial sexual assaults. Again, two specific strategies were tested. Using the non-

degraded sexual assault dataset, Figure 8 illustrates the findings of the first strategy across the two different similarity coefficients. Recall that the strategy was to adopt the threshold (i.e., the specific across-crime similarity score) that resulted in the ROC point falling closest to the upper-left corner of the ROC graph. The *p*H to *p*FA ratios associated with this threshold are presented in Table 4 for each coefficient (it is desirable to have as high a ratio as possible). As Jones and Bennell (2007) explain, the usefulness of this strategy becomes clear when the ratios in Table 4 are compared to ratios that occur at other, more extreme thresholds. For example, relying on a much more lenient threshold when using *J*, say .05 instead of .30, results in a ratio of 1.03. In contrast, a very strict threshold, such as .70, results in a ratio of 0. Importantly, Figure 8 can also be used to confirm previously presented results, which attest to the slight superiority of *J* over Δ_s. For example, when adopting a threshold that maximizes *p*H and minimizes *p*FA, as is the case in Figure 8, it is possible to make three additional hits (.71-.68) for every 100 linked crime pairs that are encountered by using *J* instead of Δ_s (or seven less false alarms for every 100 unlinked crime pairs encountered; .33-.26).

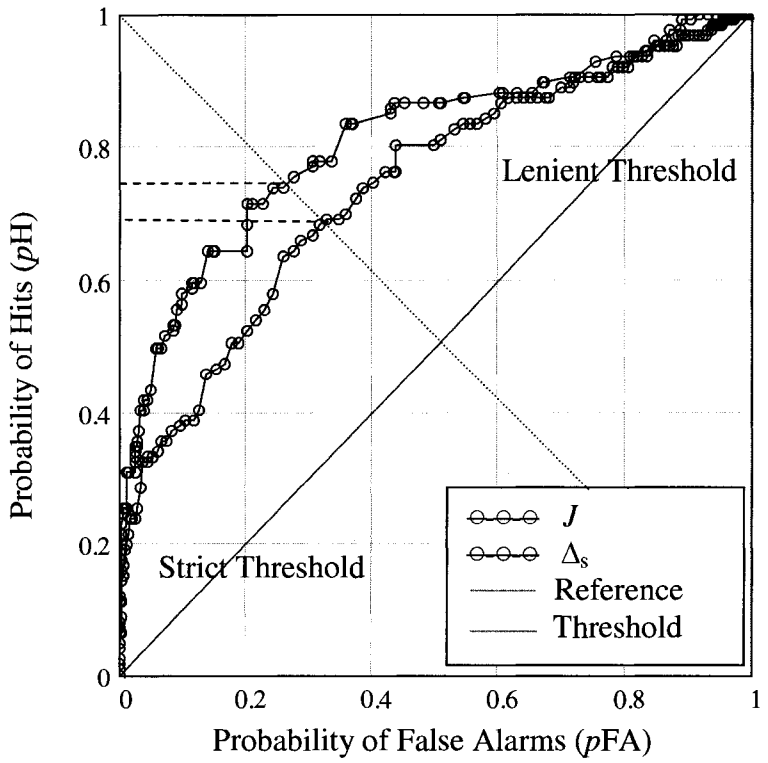


Figure 8. Illustration of decision thresholds that maximize *p*H while minimizing *p*FA, using *J* and Δ_s, for sexual assault behaviours.

Table 4. Ratios of hits to false alarms resulting from thresholds associated with *J* and Δ_s sexual assault ROC curves.

ROC curve	Across-crime similarity score (optimal threshold)	*p*H/*p*FA
J	.30	.71/.26 = 2.73
Δ_s	.71	.68/.33 = 2.06

Figure 9 illustrates the findings of the second strategy using the same data. In this case, a pre-determined value of *p*FA (.50) was selected, and then a threshold was adopted which maximized *p*H without exceeding this pre-specified value. This figure shows the ROC curves associated with both *J* and Δ_s, as well as the imposed limit of *p*FA = .50. The ROC curve associated with *J* illustrates that a hit probability of *p*H = .87 would be achieved when using this threshold (*p*H/pFA = 1.74). In contrast, applying this same decision threshold to the ROC curve associated with Δ_s would result in a hit probability of approximately *p*H = .80 (*p*H/*p*FA = 1.60). This indicates that for every 100 linked crime pairs considered, an additional seven linked crimes can be identified by using *J* instead of Δ_s (when restricting *p*FA to .50).

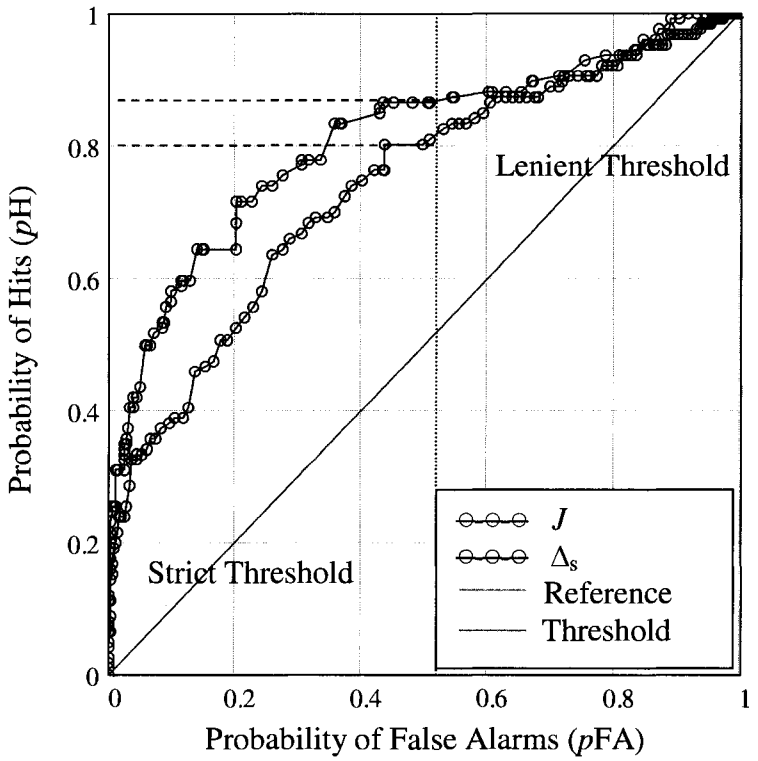


Figure 9. Illustration of decision thresholds reflecting a pre-specified rate of pFA = .50, using J and Δ_s for sexual assault behaviours.

Serial Homicide Data

Serial homicide behavioural hierarchy. No previous attempts have been made to construct a behavioural hierarchy of serial homicide and therefore the starting point in the construction of this hierarchy was the published literature on the classification of serial homicide. Based on this literature, the hierarchy was constructed around two main branches, consisting of organized and disorganized behaviours, given that this dichotomy is the most common way of categorizing serial homicide behaviour (Beauregard, Goodwill, Taylor, Bennell, 2007; Bloomfield, 2006; Hazelwood & Douglas, 1980;

Holmes & DeBurger, 1988; Prentky & Burgess, 2000; Ressler, Burgess, Douglas, Hartmann, & D'Agostino, 1986; however, see Canter et al., 2004). Essentially, organized crime scene behaviours are those that involve a greater degree of planning and control than their disorganized counterparts (Ressler et al., 1986).

The organized behaviours were further divided into planning, control, and ritualistic behaviours, with the categories of impulsive, control, and ritualistic behaviours included within the disorganized domain. Planning behaviours include any organized activities, which indicate the offender prepared for the crime before it was actually committed, and these activities directly contrast impulsive behaviours in the disorganized domain (Bloomfield, 2006). Control behaviours are characterized by actions designed to create and maintain an environment in which the crime can successfully take place, and often involve behaviours that are meant to control the victim (Groth, 1979; Jones, 2005; Canter et al., 2003; Canter & Heritage, 1990; Hodge, 2001). Lastly, ritualistic behaviours are those that are more symbolic and excessive (i.e., signature behaviours), typically going beyond what is necessary to commit the offence (Hazelwood & Warren, 2003). Thus, the organized and disorganized branches in the hierarchy often include the same categories of behaviour, but the specific behaviours included in those categories are distinctly organized or disorganized.

The 39 serial homicide behaviours available in the dataset were incorporated into what was deemed the most appropriate branch and inter-rater reliability, calculated as above, was found to be satisfactory (Kappa = .83). The resulting hierarchy of serial homicide behaviours is presented in Figure 10. A path lengths table for this hierarchy,

which illustrates the relationship between every behaviour in the dataset, is presented in Appendix F.

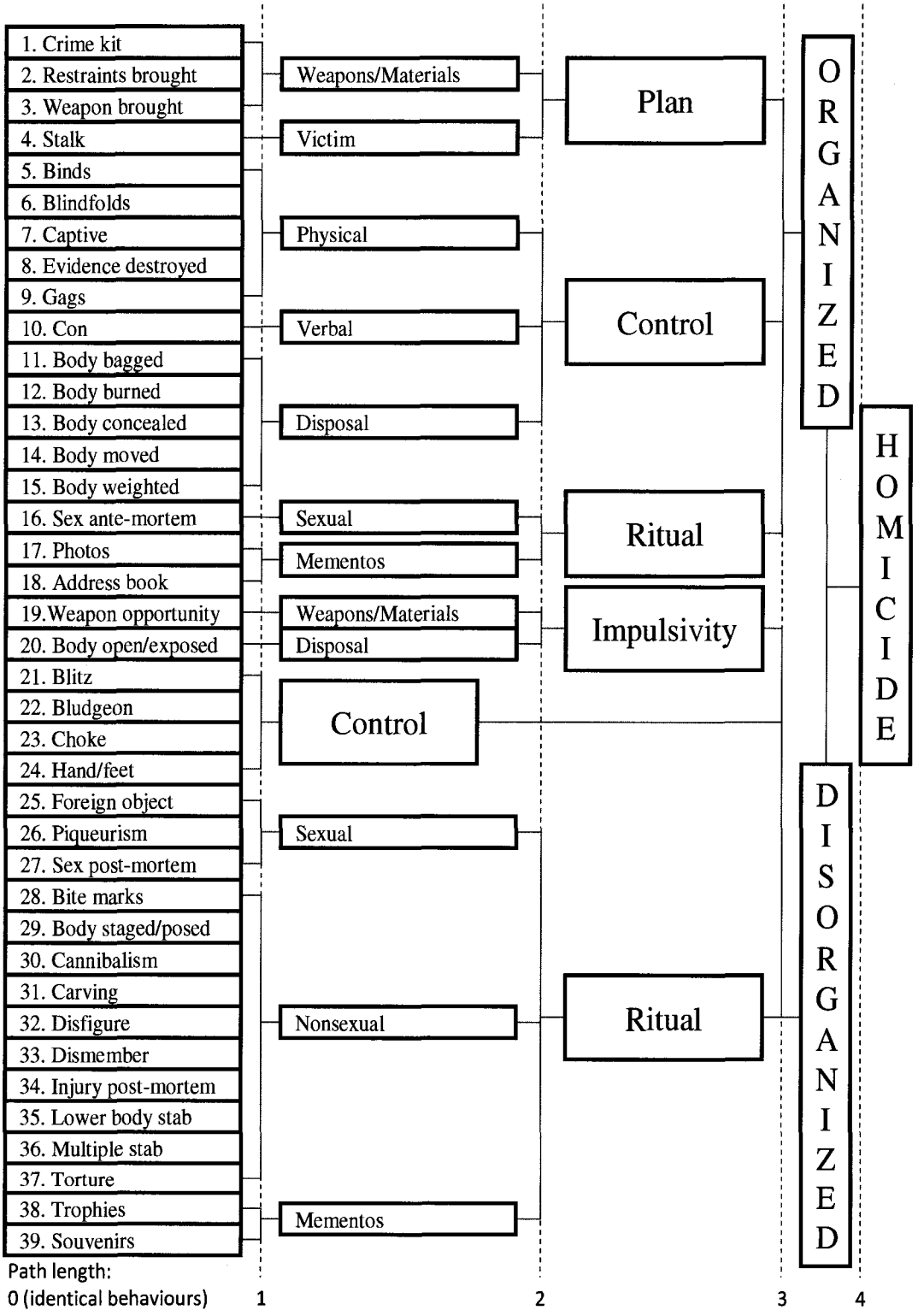


Figure 10. Serial homicide behavioural hierarchy.

Descriptive and comparative analyses. Distribution normality was examined as above. Results suggested negatively skewed distributions for linked and unlinked across-crime similarity scores for Δ_s (skew linked = -2.52, SE = .16, CI_{95} = -2.83 to -2.21; skew unlinked = -.85, SE = .02, CI_{95} = -.89 to -.81), and for J, a negatively skewed linked distribution (skew = -1.15, SE = .16, CI_{95} = -1.46 to -.84) and a positively skewed unlinked distribution (skew = .55, SE = .02, CI_{95} = .52 to .59) (see Figure 11). This suggests that non-parametric tests should be used to compare the similarity scores from the two distributions.

In order to determine whether linked homicides had higher across-crime similarity scores than unlinked homicides, descriptive statistics were calculated (see Table 5). Due to the violation of normality, non-parametric t-tests were used. Results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of whether Δ_s (mean rank linked = 26,035.38, mean rank unlinked = 13,880.49, U = 429,590.00 p <.001), or J (mean rank linked = 26,674.53, mean rank unlinked = 13,875.03, U = 278,113.50, p <.001), was used. However, it is important to note the overlap that is apparent in the distributions of the linked and unlinked crimes, indicating that some linked homicides were characterized by lower across-crime similarity scores than some unlinked homicides. As discussed, overlap of the distributions in such a way decreases the ability to link serial crimes.

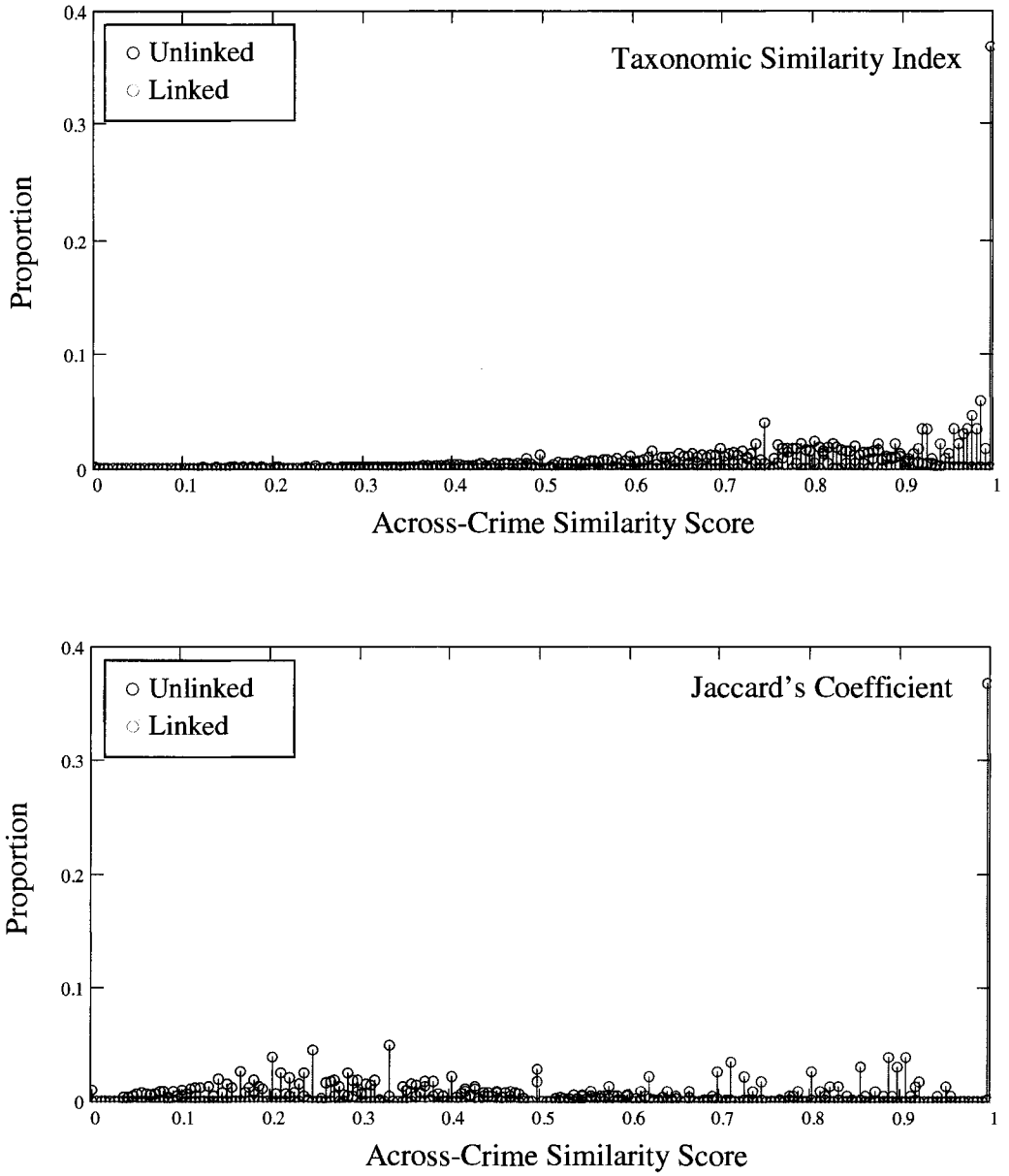


Figure 11. Distributions of across-crime similarity scores for linked and unlinked homicides using Δ_s (top) and J (bottom).

Table 5. Descriptive statistics for Δ_s and J comparing linked and unlinked homicides.

Statistic	Δ_s		J	
	Linked (n = 237)	Unlinked (n = 27,729)	Linked (n = 237)	Unlinked (n = 27,729)
Minimum	.48	.00	.10	.00
Maximum	1.00	1.00	1.00	1.00
Median	.98	.75	.89	.27
Mean	.94	.72	.81	.28
Standard Deviation	.09	.13	.22	.14

ROC analysis. The results of the descriptive analyses presented in Table 5 demonstrate that Δ_s is capable of achieving higher across-crime similarity scores than J (M = 0.94 for Δ_s vs. M = 0.81 for J). Although this is the case, Δ_s once again generated substantially higher scores for unlinked homicides as well (M = 0.72 for Δ_s vs. M = 0.28 for J). As the ability to discriminate between linked and unlinked crimes depends largely on the degree of overlap between the probability distributions, ROC analysis was also conducted on the serial homicide data in order to evaluate the relative linking ability of the two similarity coefficients.

ROC curves generated by CrimeSolver illustrating the comparative linking accuracy of Δ_s and J are presented in Figure 12. As is illustrated, both similarity coefficients were capable of differentiating between linked and unlinked homicides at a level significantly greater than chance. Comparison of the *AUC*s indicate that J (AUC = .96, SE = .01, CI_{95} = .94 to .98) outperforms Δ_s (AUC = .93, SE = .01, CI_{95} = .91 to .96) in terms of relative linking accuracy, although not to a significant degree (the *CIs* associated

with the two indices overlap). Furthermore, linking accuracy was greater for J than Δ_s across the vast majority (if not all) decision thresholds, as is apparent in Figure 12.

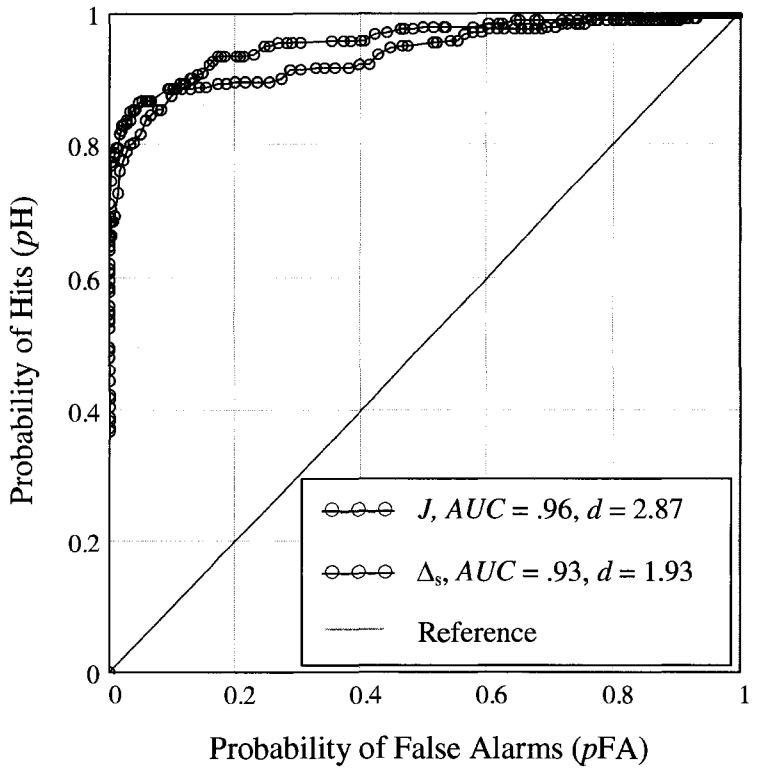


Figure 12. ROC curves illustrating comparative linking accuracy of Δ_s and J for homicides.

Data degradation. Data was progressively degraded from the dataset in order to test the robustness of the two coefficients across conditions of data degradation. Testing involved progressively removing 4, 9, and 19 randomly selected behaviours, accounting for approximately 10%, 25%, and 50% of the dataset, respectively. Table 6 reports the results of this analysis.

Table 6. The influence of data degradation on the discrimination accuracy of Δ_s and *J* for homicide.

Number of Bhvs.	Δ_s				*J*			
	AUC	*SE*	CI_{95}	*d*	*AUC*	*SE*	CI_{95}	*d*
39 (100%)	.93	.01	.91 ≤ AUC ≤ .96	1.93	.96	.01	.94 ≤ AUC ≤ .98	2.87
35 (90%)	.94	.01	.91 ≤ AUC ≤ .96	2.00	.96	.01	.94 ≤ AUC ≤ .98	2.98
26 (75%)	.94	.01	.91 ≤ AUC ≤ .96	2.03	.96	.01	.94 ≤ AUC ≤ .98	2.95
20 (50%)	.93	.01	.90 ≤ AUC ≤ .95	1.96	.95	.01	.93 ≤ AUC ≤ .97	2.76

As is apparent from the results in Table 6, the removal of crime scene behaviours from the homicide dataset had only a slight impact on both coefficients (all *CIs* overlap). Interestingly, across the conditions, the *AUC* and *d* values associated with *J* remained consistently higher than those of Δ_s. Thus, both coefficients were able to discriminate between linked and unlinked homicides even at high levels of data degradation, particularly *J*.

Re-sampling procedure. As was noted in regards to the sexual assault analyses, all of the results up to this point contradict those of Woodhams, Grant, et al. (2007). The re-sampling procedure was once again used in order to test whether Woodhams, Grant, et al.'s small sample size may be an explanation for the discrepancy in findings.

Figure 13 illustrates the effect of the re-sampling procedure using the non-degraded data. The top plot reflects *d* values that were calculated using an equal number of randomly selected linked and unlinked homicides, whereas the bottom plot reflects *d*

values that were calculated using the number of randomly selected linked and unlinked homicides proportional to the dataset (1: 118) (for reasons outlined above). Each plot is based on an average of 10 repeats per sample size. These graphs clearly demonstrate that *J* consistently achieves higher *d* values than Δ_s regardless of sample size. This is especially true in the bottom plot. As the number of crime pairs increases, the results become more stabilized in both graphs, though slightly less so in the bottom plot. This may provide further support for the notion that the results reported by Woodhams, Grant, et al. (2007) can be attributed to their use of a very small sample, and possibly to their choice of crime type as well. The exact same pattern of results emerged when running the analysis with degraded data, except that the results were slightly more erratic for the proportional analysis at high levels of data degradation.

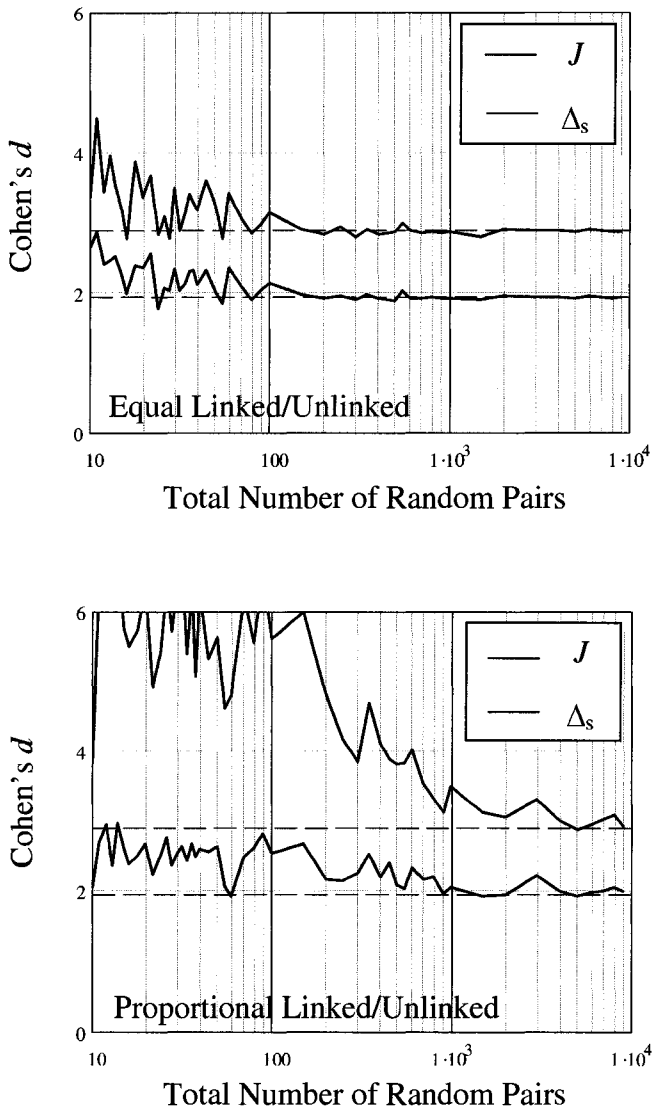


Figure 13. Effect of increasing sample size on linking accuracy for equal numbers of linked and unlinked homicides (top) and numbers of linked and unlinked homicides proportional to the dataset (bottom; 1: 118).

Impact of setting different decision thresholds. The ROC graph in Figure 12 was used to explore the impact of setting different decision thresholds on the ability to link serial homicides. Again, the two strategies outlined above were tested. The *p*H to *p*FA

ratios resulting from the first strategy (maximizing *p*H while minimizing *p*FA) are presented in Table 7 for each coefficient.

Table 7. Ratios of hits to false alarms resulting from thresholds associated with *J* and Δ_s homicide ROC curves.

ROC curve	Across-crime similarity score (optimal threshold)	*p*H/*p*FA
J	.46	.89/.12 = 7.42
Δ_s	.86	.88/.12 = 7.33

The second strategy (limiting *p*FA to .50) resulted in a hit probability of *p*H = .98 when using *J* and a ratio of *p*H/*p*FA = 1.96. In contrast, applying this same decision threshold to the ROC curve associated with Δ_s resulted in a hit probability of *p*H = .95 and a ratio of *p*H/*p*FA = 1.90. This indicates that for every 100 linked homicides considered, an additional three linked crimes could be identified by using *J* instead of Δ_s (when restricting *p*FA to .50).

Serial Burglary Data

Serial burglary behavioural hierarchy. No previous attempts have been made to construct a behavioural hierarchy of serial burglary and therefore the starting point in the construction of this hierarchy was the published literature on the classification of serial burglary. Based on this literature, the hierarchy was constructed around two main branches, consisting of high and low skill behaviours, given that this is the most common way of categorizing serial burglary behaviour (e.g., Merry & Harsent, 2000). High skill

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

The 28 serial burglary behaviours available in the dataset were incorporated into what was deemed the most appropriate branch and inter-rater reliability, calculated as above, was found to be satisfactory (Kappa = .93). The resulting hierarchy of serial burglary behaviours is presented in Figure 14. A path lengths table for this hierarchy, which illustrates the relationship between every behaviour in the dataset, is presented in Appendix G.

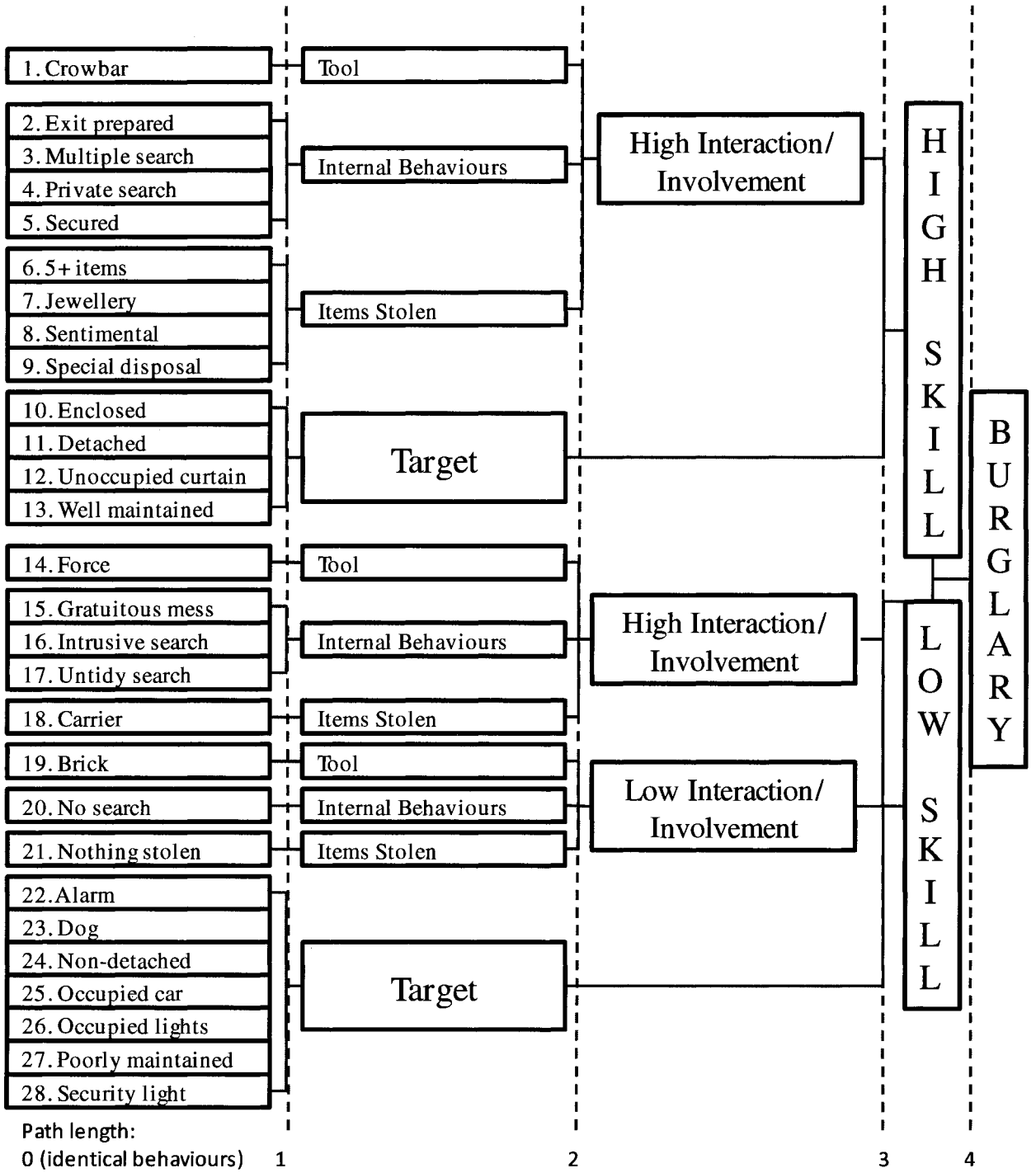


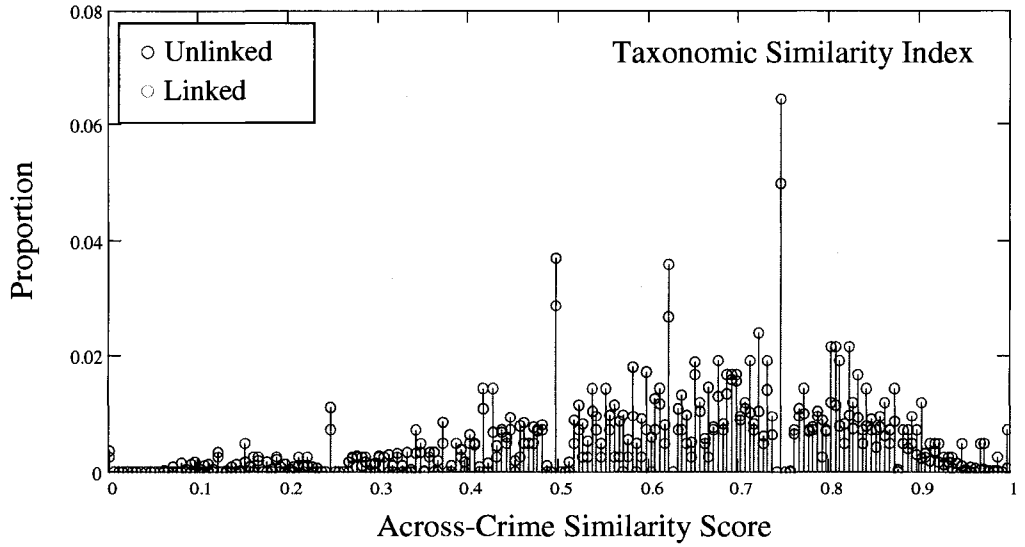
Figure 14. Serial burglary behavioural hierarchy.

Descriptive and comparative analyses. Distribution normality was examined as above. Results suggested negatively skewed distributions of linked and unlinked across-crime similarity scores for Δ_s (skew linked = -.81, SE = .12, CI_{95} = -1.05 to -.57; skew unlinked = -.69, SE = .02, CI_{95} = -.73 to -.65) and positively skewed distributions of linked and unlinked across-crime similarity scores for J (skew linked = .82, SE = .12, CI_{95} = .58 to 1.10; skew unlinked = .81, SE = .02, CI_{95} = .77 to .85) (see Figure 15). This suggests that non-parametric tests should be used to compare the similarity scores from the two distributions.

In order to determine whether linked burglaries had higher across-crime similarity scores than unlinked burglaries, descriptive statistics were calculated (see Table 8). Due to the violation of normality, non-parametric t-tests were used. Results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of whether Δ_s (mean rank linked = 13,011.06, mean rank unlinked = 10,933.22, U = 3,664,266.00 p <.001), or J (mean rank linked = 13,613.10, mean rank unlinked = 10,921.49, U = 3,411,407.50, p <.001), was used. However, it is important to note the substantial overlap (almost entirely) that is apparent in the distributions of the linked and unlinked crimes (as demonstrated in Figure 15). This overlap suggests that many linked burglaries were associated with very low across-crime similarity scores, while many unlinked burglaries were associated with high across-crime similarity scores. As discussed, overlap between the distributions decreases the ability to link serial crimes.

ROC analysis. The descriptive analyses results demonstrate that Δ_s is capable of achieving higher across-crime similarity scores for linked burglaries than J (M = .64 for Δ_s vs. M = .29 for J). However, Δ_s also generated higher scores for unlinked burglaries as

well (M = .62 for Δ_s vs. M = .22 for J). As mentioned, the ability to discriminate between linked and unlinked crimes depends largely on the degree of overlap between the distributions. ROC analysis was conducted on the serial burglary data in order to evaluate the relative linking ability of the two similarity coefficients.



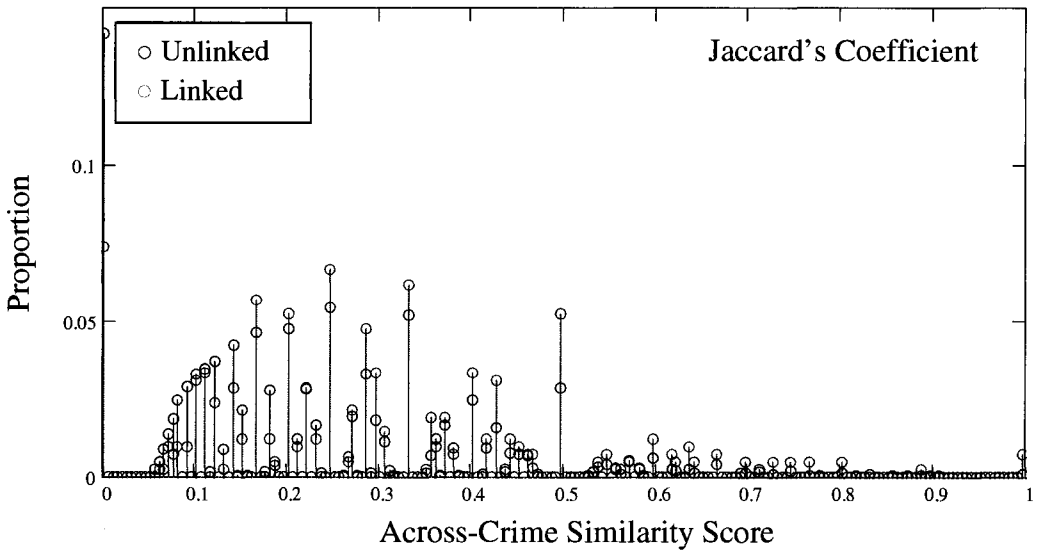


Figure 15. Distributions of across-crime similarity scores for linked and unlinked burglaries using Δ_s (top) and J (bottom).

Table 8. Descriptive statistics for Δ_s and J comparing linked and unlinked burglaries.

Statistic	Δ_s		J	
	Linked (n = 420)	Unlinked (n = 21,945)	Linked (n = 420)	Unlinked (n = 21,945)
Minimum	.00	.00	.00	.00
Maximum	1.00	1.00	1.00	1.00
Median	.71	.62	.27	.19
Mean	.64	.62	.29	.22
Standard Deviation	.18	.18	.19	.16

ROC curves generated by CrimeSolver are presented in Figure 16. As is illustrated, both similarity coefficients were capable of differentiating between linked and unlinked burglaries at a level significantly greater than chance. Comparison of the *AUC*s indicate that J (AUC = .62, SE = .03, CI_{95} = .57 to .67) outperforms Δ_s (AUC = .60, SE = .03, CI_{95} = .54 to .65) in terms of relative linking accuracy, although not to a significant degree (the *CIs* associated with the two indices overlap). Furthermore, linking accuracy was greater for J than for Δ_s across the vast majority (if not all) decision thresholds, as is also apparent in Figure 16.

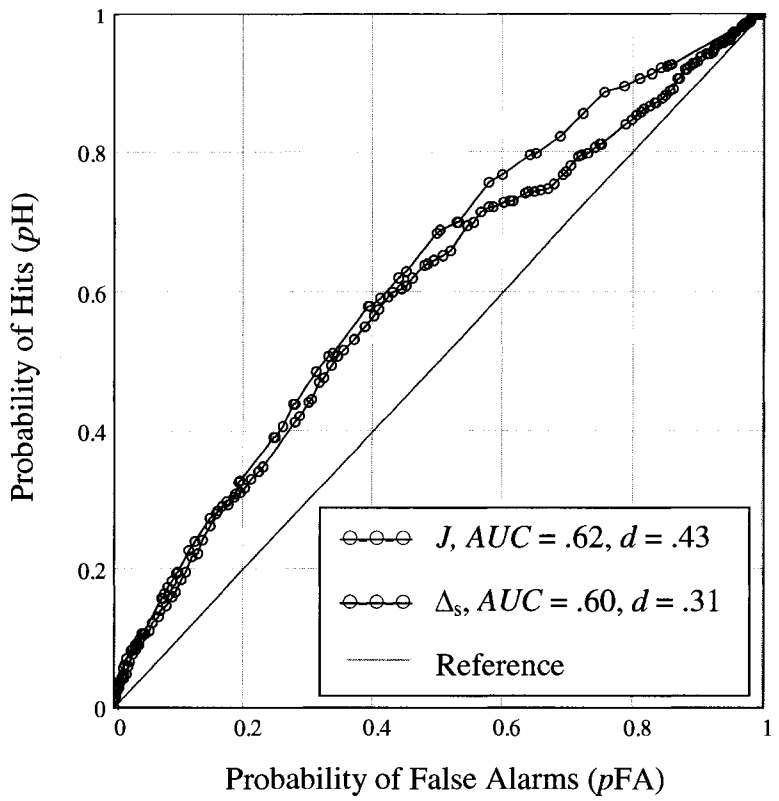


Figure 16. ROC curves illustrating comparative linking accuracy of Δ_s and J for burglaries.

Data degradation. In line with the Woodhams, Grant, et al. (2007) study, data was progressively degraded from the dataset in order to test the robustness of the two coefficients across conditions of data degradation. Testing involved progressively removing 3, 7, and 14 randomly selected behaviours, accounting for approximately 10%, 25%, and 50% of the dataset, respectively. Table 9 reports the results of this analysis.

Table 9. The influence of data degradation on the discrimination accuracy of Δ_s and *J* for burglary.

Number of Bhvs.	Δ_s				*J*			
	AUC	*SE*	CI_{95}	*d*	*AUC*	*SE*	CI_{95}	*d*
28 (100%)	.59	.03	.54 ≤ AUC ≤ .65	.31	.62	.03	.57 ≤ AUC ≤ .68	.43
25 (90%)	.57	.03	.51 ≤ AUC ≤ .62	.22	.60	.03	.54 ≤ AUC ≤ .65	.34
21 (75%)	.56	.03	.50 ≤ AUC ≤ .62	.20	.59	.03	.53 ≤ AUC ≤ .64	.33
14 (50%)	.54	.03	.48 ≤ AUC ≤ .59	.13	.54	.03	.49 ≤ AUC ≤ .60	.18

As is apparent from the results in Table 9, the removal of crime scene behaviours from the burglary dataset had only a slight impact on both coefficients (all *CIs* overlap). Interestingly, across the conditions, the *AUC* and *d* values associated with *J* remained consistently higher than those of Δ_s (with the exception of the 50% degradation level where the *AUC*s were the same). Thus, both coefficients were able to discriminate between linked and unlinked burglaries even at high levels of data degradation, if only slightly, particularly *J*.

Re-sampling procedure. Again, the results of the burglary analyses up to this point contradict those of Woodhams, Grant, et al. (2007). The re-sampling procedure was therefore used again to examine the impact of sample size on effect size. The results are presented in Figure 17 for both Δ_s and *J*, for the non-degraded burglary dataset. The top plot reflects *d* values that were calculated using an equal numbers of linked and unlinked burglaries, whereas the bottom plot reflects *d* values that were calculated using the

number of linked and unlinked burglaries proportional to the dataset (1: 52.25) (for reasons outlined above). Each plot is based on an average of 10 repeats per sample size.

Across both graphs it can clearly be seen that the effect size is much more erratic for the two coefficients when the sample size is smaller (even more so than was seen in the other datasets). This is especially true in the bottom plot. That is, when analyses are based on relatively small sample sizes (e.g., <100 pairs) the effect sizes for both coefficients vary dramatically across the trials, and in many instances Δ_s appears to outperform *J*. However, as the number of crime pairs increases, the results become more stabilized in both graphs, though this requires larger samples in the bottom plot. Thus, these graphs clearly demonstrate that *J* consistently achieves higher *d* values than Δ_s when a relatively large number of crime pairs are examined, which again indicates that the results of Woodhams, Grant, et al. (2007) could be attributed to their use of a small sample size. The exact same pattern of results emerged when running the same analysis with degraded data, except that the results were more erratic, particularly for the proportional analysis and at high levels of data degradation.

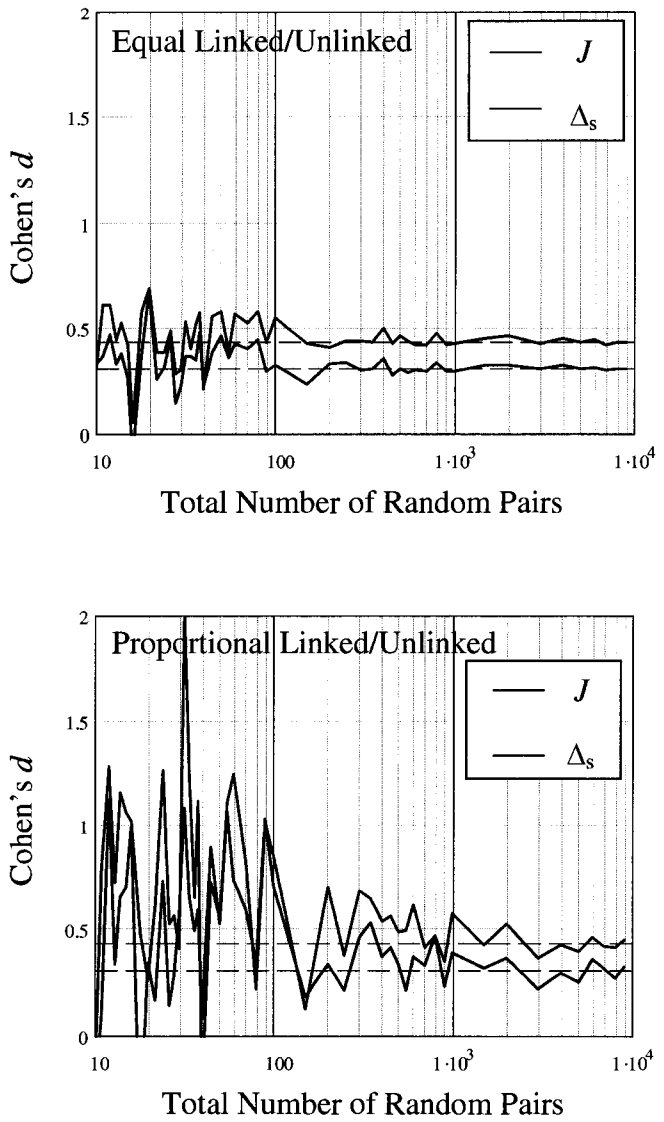


Figure 17. Effect of increasing sample size on linking accuracy for equal numbers of linked and unlinked burglaries (top) and numbers of linked and unlinked burglaries proportional to the dataset (bottom; 1: 52.25).

Impact of setting different decision thresholds. The ROC graph in Figure 16 was used to explore the impact of setting different decision thresholds on the ability to link serial burglaries. Again, two specific strategies were tested. The *p*H to *p*FA ratios

resulting from the first strategy (maximizing *p*H while minimizing *p*FA) are presented in Table 10 for each coefficient.

Table 10. Ratios of hits to false alarms resulting from thresholds associated with *J* and Δ_s burglary ROC curves.

ROC curve	Across-crime similarity score (optimal threshold)	*p*H/*p*FA
J	.23	.58/.41 = 1.43
Δ_s	.69	.57/.40 = 1.41

The second strategy (limiting *p*FA to .50) resulted in a hit probability of *p*H = .68 when using *J* and a ratio of *p*H/*p*FA = 1.36. In contrast, applying this same decision threshold to the ROC curve associated with Δ_s resulted in a hit probability of *p*H = .64 and a ratio of *p*H/*p*FA = 1.28. This indicates that for every 100 linked burglaries considered, an additional four linked crimes could be identified by using *J* instead of Δ_s (when restricting *p*FA to .50).

Study 2 – Behavioural Frequency and Crime Type

Serial Sexual Assault Data

Serial sexual assault frequency subsets. On the basis of a frequency analysis, the serial sexual assault dataset was separated into subsets representing the frequency recommendations of Canter et al. (2003) (>90% removed), Salfati and Bateman (2005) (>50% removed), and Woodhams, Hollin, et al. (2008) (<10% removed). Results

indicated that there were no behaviours present in the dataset that had a frequency <1%, and therefore the recommendation put forward by Santtila, Junkkila, et al. (2005) could not be tested with respect to serial sexual assault. The frequency of occurrence for each behaviour in the sexual assault dataset is reported in Appendix A.

Descriptive and comparative analyses. Distribution normality was examined using the procedure outlined above. This analysis indicated that the linked across-crime similarity scores are normally distributed for all recommendations, with the exception of removing behaviours >50%, which was positively skewed. The distributions for unlinked across-crime similarity scores were positively skewed for all recommendations (see Table 11). This suggests that non-parametric tests should be used to compare the similarity scores from the distributions.

Table 11. Skewness data for linked and unlinked sexual assaults across frequency recommendations

Frequency	Linked (n = 126)			Unlinked (n = 7,749)		
	Skew	*SE*	CI_{95}	Skew	*SE*	CI_{95}
>90% Removed	.17	.22	-.26 to .60	.58	.03	.52 to .64
>50% Removed	.54	.22	.10 to 1.00	.98	.03	.92 to 1.04
<10% Removed	.09	.22	-.34 to .52	.54	.03	.48 to .60
All behaviours	.08	.22	-.35 to .51	.55	.03	.49 to .61

In order to determine whether the linked sexual assaults (within each frequency subset) had higher across-crime similarity scores than unlinked sexual assaults, descriptive statistics were calculated (see Table 12). On the basis of non-parametric *t*-tests, results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of which recommendation was considered (see Table 12). For the sake of brevity, graphic representations of each frequency distribution are not provided. However, it is clear from the results presented in Table 12 that a degree of overlap exists between the distributions of similarity scores associated with linked and unlinked crimes. This is the case for each frequency recommendation. As discussed, overlap of these distributions decreases the ability to link serial crime.

Table 12. Descriptive statistics for linked and unlinked sexual assaults across frequency recommendations.

Statistic	>90% Removed		>50% Removed		<10% Removed		All Bhvs.	
	L	UL	L	UL	L	UL	L	UL
Minimum	.00	.00	.00	.00	.08	.00	.08	.00
Maximum	.75	.83	1.0	.75	.80	.86	.80	.86
Median	.33	.17	.27	.10	.40	.23	.39	.22
Mean	.34*	.18	.29*	.12	.40*	.24	.39*	.23
SD	.17	.12	.19	.12	.17	.12	.16	.11

Note. L = Linked (n = 126), UL = Unlinked (n = 7,749); *: $p < .001$ (for the difference between the average across-crime similarity score for linked versus unlinked crime pairs).

ROC analysis. The results of the descriptive analysis presented in Table 12 indicate that the subsets of behaviour that lead to the highest across-crime similarity scores for linked crimes also result in the highest across-crime similarity scores for unlinked crimes. Since the ability to discriminate between linked and unlinked crimes depends on the degree of overlap between these distributions, ROC analysis was used to evaluate the relative linking accuracy that could be achieved when following through with each of the recommendations.

As above, ROC curves were generated for each recommendation using SPSS. However, for the sake of brevity and clarity, the ROC graphs are not presented here (while it is possible to present a ROC graph with numerous ROC curves, when the curves intersect each other, as they do in the current case, the graph becomes very difficult to interpret). Instead, the summary data from each ROC analysis is presented in Table 13. Based on this data, it appears that each recommendation results in a degree of linking accuracy that significantly exceeds chance.

A comparison of the *AUC*s indicates that the use of all behaviours, as well as the omission of behaviours with a frequency >90% and <10%, results in the same level of linking accuracy (*AUC* = .78). The *AUC*s associated with these recommendations are slightly higher, but not significantly higher, than the *AUC* that results from the removal of behaviours with a frequency >50% (*AUC* = .77). Furthermore, a review of the ROC curves revealed that *p*H/*p*FA ratios were comparable for all frequency subsets, including when all behaviours were used, across the vast majority of decision thresholds. The only exception to this was when behaviours with a frequency >90% were removed from the

analysis. In this case, the ROC curve was slightly higher than the other ROC curves at more lenient thresholds.

Table 13. Summary data from ROC analyses testing frequency recommendations with sexual assault behaviours.

Frequency	*AUC*	*SE*	CI_{95}
>90% Removed	.78	.02	.74 to .83
>50% Removed	.77	.02	.72 to .82
<10% Removed	.78	.02	.74 to .83
All Behaviours	.78	.02	.74 to .83

Serial Homicide Data

Serial homicide frequency subsets. On the basis of a frequency analysis, the serial homicide dataset was separated into subsets representing the frequency recommendations of Salfati and Bateman (2005) (>50% removed) and Woodhams, Hollin, et al. (2008) (<10% removed). Results indicated that there were no behaviours present in the dataset that had a frequency >90% or <1%, and therefore the recommendations put forward by Canter et al. (2003) and Santtila, Junkkila, et al. (2005) could not be tested with respect to serial homicide. The frequency of occurrence for each behaviour in the homicide dataset is reported in Appendix B.

Descriptive and comparative analyses. Distribution normality was examined as above. This analysis indicated that the linked distributions of across-crime similarity scores were negative skewed for each recommendation, while the unlinked distributions

of across-crime similarity scores were positively skewed for each recommendation (see Table 14). This suggests that non-parametric tests should be used to compare the similarity scores from the distributions.

Table 14. Skewness data for linked and unlinked homicides across frequency recommendations.

Frequency	Linked (n = 237)			Unlinked (n = 27,729)		
	Skew	*SE*	CI_{95}	Skew	*SE*	CI_{95}
>50% Removed	-1.10	.16	-1.41 to -.79	1.17	.02	1.13 to 1.21
<10% Removed	-1.18	.16	-1.49 to -.87	.54	.02	.50 to .58
All Behaviours	-1.15	.16	-1.46 to -.84	.55	.02	.51 to .59

In order to determine whether the linked homicides (within each frequency subset) had higher across-crime similarity scores than unlinked homicides, descriptive statistics were calculated (see Table 15). On the basis of non-parametric *t*-tests, results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of which recommendation was considered (see Table 15). As was the case with sexual assault, it is clear from the results presented in Table 15 that a degree of overlap exists between the distributions of similarity scores associated with linked and unlinked crimes for each frequency recommendation. As discussed, overlap of the distributions decreases the ability to link serial crime.

Table 15. Descriptive statistics for linked and unlinked homicides across frequency recommendations.

Statistic	>50% Removed		<10% Removed		All Bhvs.	
	L	UL	L	UL	L	UL
Minimum	.00	.00	.10	.00	.10	.00
Maximum	1.00	1.00	1.00	1.00	1.00	1.00
Median	.83	.13	.89	.28	.89	.27
Mean	.76*	.15	.82*	.29	.81*	.28
SD	.29	.15	.22	.15	.22	.14

Note. L = Linked (n = 237), UL = Unlinked (n = 27,729); *: $p < .001$ (for the difference between the average across-crime similarity score for linked versus unlinked crime pairs).

ROC analysis. As was found in the analysis of sexual assaults, the results of the descriptive analysis presented in Table 15 indicate that the subsets of behaviour which lead to the highest across-crime similarity scores for linked crimes also result in the highest across-crime similarity scores for unlinked crimes. Since the ability to discriminate between linked and unlinked crimes depends on the degree of overlap between these distributions, ROC analysis was used to evaluate the relative linking accuracy that could be achieved when following through with each of the recommendations. The summary data from each ROC analysis is presented in Table 16.

Based on the data that are illustrated in Table 16, it appears that each recommendation results in a degree of linking accuracy that far exceeds chance. A comparison of the *AUC*s indicates that the use of all behaviours, as well as the omission of behaviours with frequencies <10%, results in the same level of linking accuracy (*AUC*

= .96). The *AUC*s associated with these recommendations are slightly higher, but not significantly higher, than the *AUC* that results when behaviours with a frequency >50% are removed from the dataset (*AUC* = .94). Furthermore, a review of the ROC curves revealed that *p*H/*p*FA ratios were comparable for all frequency subsets, including when all behaviours were used, across the vast majority of decision thresholds. The only exception to this was when behaviours with a frequency >50% were removed from the analysis. In this case, the ROC curve was slightly lower than the other ROC curves at more lenient thresholds.

Table 16. Summary data from ROC analyses testing frequency recommendations with homicide behaviours.

Frequency	*AUC*	*SE*	CI_{95}
>50% Removed	.94	.01	.92 to .96
<10% Removed	.96	.01	.94 to .94
All Behaviours	.96	.01	.94 to .97

Serial Burglary Data

Serial burglary frequency subsets. On the basis of a frequency analysis, the serial burglary dataset was separated into subsets representing the frequency recommendations of Salfati and Bateman (2005) (>50% removed), Woodhams, Hollin, et al. (2008) (<10% removed), and Santtila, Junkkila et al. (2005) (<1% removed). Results indicated that there were no behaviours present in the dataset that had a frequency >90%, and therefore the recommendation put forward by Canter et al. (2003) could not be tested with respect to

serial burglary. The frequency of occurrence for each behaviour in the burglary dataset is reported in Appendix C.

Descriptive and comparative analyses. Distribution normality was examined as above. This analysis indicated that the distributions of the linked and unlinked across-crime similarity scores were positively skewed for all recommendations (see Table 17). This suggests that non-parametric tests should be used to compare the similarity scores.

Table 17. Skewness data for linked and unlinked homicides across the frequency recommendations.

Frequency	Linked (n = 420)			Unlinked (n = 21,525)		
	Skew	*SE*	CI_{95}	Skew	*SE*	CI_{95}
>50% Removed	.82	.12	.58 to 1.10	.91	.02	.87 to .95
<10% Removed	.63	.12	.39 to .87	.80	.02	.76 to .84
<1% Removed	.81	.12	.57 to 1.05	.81	.02	.77 to .85
All behaviours	.82	.12	.58 to 1.10	.81	.02	.77 to .85

In order to determine whether the linked burglaries (within each frequency subset) had higher across-crime similarity scores than unlinked burglaries, descriptive statistics were calculated (see Table 18). On the basis of non-parametric *t*-tests, results indicate that the mean across-crime similarity score was higher for linked crimes than for unlinked crimes regardless of which recommendation was implemented (see Table 18). Again, it is clear from the results presented in Table 18 that a degree of overlap exists between the distributions of similarity scores associated with linked and unlinked crimes

for each frequency recommendation, and to a much greater extent than was the case for sexual assault and homicide. As discussed, overlap of these distributions decreases the ability to link serial crime.

Table 18. Descriptive statistics for linked and unlinked burglaries across the frequency recommendations.

Statistic	>50% Removed		<10% Removed		<1% Removed		All Bhvs.	
	L	UL	L	UL	L	UL	L	UL
Minimum	.00	.00	.00	.00	.00	.00	.00	.00
Maximum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Median	.25	.17	.30	.20	.27	.19	.27	.19
Mean	.28*	.20	.32*	.24	.29*	.22	.29*	.22
SD	.19	.17	.20	.18	.19	.16	.19	.16

Note. L = Linked (n = 420), UL = Unlinked (n = 21,525); *: $p < .001$ (for the difference between the average across-crime similarity score for linked versus unlinked crime pairs).

ROC analysis. As was found in both of the previous analyses, the results of the descriptive analysis presented in Table 18 indicates that the subsets of behaviour which lead to the highest across-crime similarity scores for linked crimes also result in the highest across-crime similarity scores for unlinked crimes. Again, ROC analysis was used to evaluate the relative linking accuracy that could be achieved when following through with each of the recommendations. The summary data from each ROC analysis is presented in Table 19.

Based on the data that are illustrated in Table 19, it appears that each recommendation results in a degree of linking accuracy that significantly exceeds chance, but certainly not to the same extent as was found for sexual assaults or homicides. Comparison of the *AUC*s indicates that the removal of behaviours with a frequency >50% resulted in the highest linking accuracy (*AUC* = .63). All of the other subsets of behaviours, including the use of all behaviours, resulted in a slightly lower (but not significantly lower) *AUC* of .62. Furthermore, a review of the ROC curves revealed that linking accuracy was comparable for all frequency subsets across all decision thresholds.

Table 19. Summary data from ROC analyses testing the frequency recommendations with serial burglary behaviours.

Frequency	*AUC*	*SE*	CI_{95}
>50% Removed	.63	.01	.61 to .66
<10% Removed	.62	.01	.60 to .65
<1% Removed	.62	.01	.60 to .65
All Behaviours	.62	.01	.60 to .65

Discussion

The purpose of this thesis was to contribute further to research on BLA from a signal detection perspective. This was to be accomplished by relying on ROC analysis to examine a number of factors that had the potential to influence the degree of discrimination accuracy that could be achieved when linking crimes. Specifically, three factors were examined – crime type, similarity coefficient, and behavioural frequency –

to determine their impact on discrimination accuracy. The results from the two studies, which will be discussed in detail below, increase our understanding of criminal behaviour and of the assumptions inherent in the linking task. The results also have the potential to contribute to linking efforts in naturalistic settings by providing insight into the conditions under which the greatest degree of discrimination accuracy can be achieved.

Study 1 – Similarity Coefficient and Crime Type

As indicated previously, the similarity coefficient can be considered the basis for most approaches to linkage analysis as it is used to quantify the degree of similarity that exists between crimes. Since research in other disciplines has demonstrated that the choice of coefficient can drastically influence results in tasks that are similar to the crime linking task, it was important to test whether or not this was the case in regards to BLA. The most commonly used similarity coefficient in BLA is currently *J*. However, Woodhams, Grant, et al. (2007) have recently advocated for the use of Δ_s, arguing that it is a potentially more appropriate coefficient for quantifying across-crime similarity. The purpose of Study 1 was to compare the degree to which Δ_s and *J* could discriminate between linked and unlinked sexual assaults, homicides, and burglaries across a range of conditions.

Is it possible to link crimes using J and Δ_s? The results of this thesis demonstrated that, regardless of what coefficient was used or which crime type was examined, crimes committed by the same offender tended to be associated with significantly higher levels of across-crime similarity than crimes committed by different offenders. Thus, the serial offenders examined in this study do appear to behave in a somewhat stable and distinct fashion across their crimes. This indicates that the offenders

are likely predisposed to behave in a particular way when committing their crimes. Furthermore, these tendencies are, to some extent at least, unaffected by situational variations that no doubt exist across crimes (although situational factors, such as victim resistance in the case of sexual assault, likely contribute to the fact that absolute stability and distinctiveness were not found).

Given this result, it was not overly surprising to find that both coefficients could be used to link the crimes in each dataset beyond levels that would be expected by chance, but not with perfect accuracy. This finding accords well with previous BLA research, which has shown that linking crimes is definitely possible, but that there are limits as to how accurate one can be when faced with the linking task (e.g., Bennell & Canter, 2002; Goodwill & Alison, 2006; Grubin et al., 2001; Woodhams & Toye, 2007). The fact that this finding emerged across all three datasets, and in particular across the sexual assault and homicide datasets, is particularly important because much of the research that has been conducted on BLA in the past has focused exclusively on property crimes as opposed to interpersonal crimes (e.g., Bennell & Canter, 2002; Bennell & Jones, 2005; Ewart et al., 2005; Goodwill & Alison, 2006; Green et al., 1976; Woodhams & Toye, 2007).

Despite these generally positive findings, clear differences did emerge across the various datasets with respect to the level of linking accuracy that was achieved. More specifically, linking accuracy was noticeably lower in the case of serial burglary compared to sexual assault and homicide. This does not contradict previous BLA research that has focused on property crimes. While much higher *AUC*s have certainly been reported in existing studies of property offences (e.g., Bennell & Canter, 2002;

Bennell & Jones, 2005; Woodhams & Toye, 2007), these *AUC*s have only ever been found for very specific subsets of behaviours, most notably certain aspects of spatial behaviour, such as inter-crime distance (i.e., the shorter the distance between two crimes, the more likely the crimes have been committed by the same offender). For the sorts of behaviours examined in this thesis, the *AUC*s that were found in the current study are in line with those in the published literature. Unfortunately, data on spatial behaviours were not available for analysis in the current study, making this a fruitful line for future research.

With respect to the context-dependency and psychopathology hypotheses that were discussed previously, the fact that linking accuracy was so high in the case of serial homicide (and to a slightly lesser extent sexual assault), but so low in the case of serial burglary, tends to favour the psychopathology hypothesis. Of course, this is if one accepts, as Pinnizzotto and Finkel (1990) do, that offenders committing serial homicides and sexual assaults possess higher levels of psychopathology than do serial burglars. The existing literature is unclear on this point. However, there have been strong suggestions that, in contrast to serial burglary behaviour, the behaviour of both serial killers and serial rapists are largely guided by scripts that have often been well-rehearsed, are deeply engrained, and are typically rooted in personal fantasies (Davies, 1992; Hazelwood & Warren, 2003; Keppel, 1997).[5] If this is the case, then perhaps it is not surprising that these offenders maintain their individual differences in offending style across the crimes

[5] For example, consider the case of notorious serial killer Edmund Kemper ("The Co-Ed Butcher") who targeted young female co-eds. Kemper "reportedly spent inordinate amounts of time envisaging all the murderous actions he could perform upon the young co-eds. Moreover, it has been estimated that in the year preceding the onset of his crime series, Kemper picked up and safely delivered in excess of 150 female hitchhikers as he rehearsed the preliminary steps towards the physical execution of his fantasy" (Jones, 2005, p. 97).

they commit to a greater extent than do serial burglars, despite the fact that these offenders might potentially have to deal with more situational variation in their crimes.

What then of the context-dependency hypothesis? What do the results of this thesis say about the notion that less situationally driven behaviours will be exhibited in a more stable and distinct fashion? It was initially proposed that serial burglary behaviours would need to be expressed in a more stable and distinct fashion than either sexual assault or homicide behaviours for this hypothesis to be confirmed, given that situational variation across crimes was viewed as less of an issue in cases of serial burglary (i.e., in general, serial burglary behaviours would be less situationally driven than either sexual assault or homicide behaviours). This was assumed to be the case due to the fact that victims are often not present during burglaries. At the outset of this study, this seemed like a sensible rational. Indeed, in past studies of serial burglary behaviour, very high *AUCs* have been found, which appeared to confirm the context-dependency hypothesis. In hindsight, however, it is difficult to reject or accept this hypothesis given the nature of the data that was available to be examined in this thesis. Many of the behaviours in serial burglary that are arguably more "offender driven" (e.g., where to go to commit the crime) were simply not able to be analyzed in this study. Given this, it is likely that serial burglary behaviours, as they were examined here, are just as likely to be influenced by situational variations as the sexual assault or homicide behaviours (e.g., items to be stolen vary across crimes, available entry methods vary across crimes, rooms to be searched vary across crimes, etc.).

Which similarity coefficient is best suited for BLA? As indicated, the results of this study demonstrate that either J or Δ_s could be used to discriminate between linked

and unlinked crimes at a level greater than chance. Indeed, comparisons of the two coefficients across the various crimes types revealed only non-significant differences, although *J* consistently outperformed Δ_s. These findings are in line with a recent study conducted by Gauthier (2008), although her analysis was restricted to sexual assault. Together, these findings contradict the results presented by Woodhams, Grant, et al. (2007) and suggest that Δ_s may not be as powerful for linking purposes as was originally thought. This begs the question of why discrepancies emerged between these studies.

One distinct possibility was that the sample size in Woodhams, Grant, et al.'s (2007) study was not large enough to produce reliable results, implying that if they would have tested a larger sample their results would have corresponded with the results of the current study. Certainly, the results presented in this thesis, and those of Gauthier (2008), suggest that the sample size does have a large effect on the reliability of the results and can potentially lead to erroneous conclusions being drawn about the relative accuracy of the two coefficients. Specifically, the re-sampling analyses conducted in the current study indicate that when smaller sample sizes are used (of the sort examined by Woodhams, Grant et al.) levels of linking accuracy are unreliable and can vary dramatically between analyses. Indeed, when drawing on small sample sizes, it is difficult to make a clear determination as to which coefficient is best – sometimes *J* outperforms Δ_s and sometimes the reverse is true. This is not the case when larger samples of crimes are used. Under these conditions, *J* consistently outperforms Δ_s.

Thus, one of the potential reasons for the discrepancy between Woodhams, Grant, et al.'s (2007) study and the current one is that the sample size in that previous study was simply too small to draw any valid conclusions. However, this explanation

doesn't help us understand why, when drawing on larger samples of crimes, *J* outperforms Δ_s. The logic underlying the use of Δ_s makes some intuitive sense, so why does it not outperform *J*? While there are many possible answers to this question, reconceptualizing BLA as a signal detection task raises a likely possibility.

Recall that from a signal detection perspective, BLA can be conceptualized as two probability distributions. One distribution is representative of across-crime similarity scores associated with linked crimes, while the other represents unlinked crimes. The ultimate task in BLA is to identify ways of decreasing the degree of overlap between these two distributions since this will increase our ability to accurately link crimes. This can be done in a number of ways: (1) by finding a way to shift the distribution for linked crimes up the x-axis (i.e., increasing behavioural stability), (2) by finding a way to shift the distribution for unlinked crimes down the x-axis (i.e., increasing behavioural distinctiveness), or (3) by finding a way to accomplish both (1) and (2).

Based on the results of the current study, Δ_s clearly does a good job of accomplishing (1). Compared to *J*, the across-crime similarity scores for linked crimes are always higher when using Δ_s (i.e., behavioural stability increases with the use of Δ_s). However, the use of Δ_s also appears to adversely influence the distribution of across-crime similarity scores calculated across unlinked crimes (i.e., behavioural distinctiveness is decreased with the use of Δ_s). That is, the use of Δ_s produces higher similarity scores for *both* linked and unlinked crimes. The analysis of the serial homicide data is perhaps the best example of this (see Table 5). Thus, the problem with using Δ_s for the purposes of BLA can be summarized as follows:

> Put simply, the degree of distribution overlap is greater when using Δ_s than when using *J*. [As a result] the taxonomic similarity index may be a more suitable measure if one wishes to increase the degree of consistency found between crimes committed by the same offender, but when one wants to balance the degree of consistency that is found between linked crimes with the degree of distinctiveness found between unlinked crimes [e.g., for linking purposes], *J* would appear to be the more suitable choice [especially given its simplicity compared to Δ_s]. (Gauthier, 2008, p. 36)

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

What is the influence of data degradation on J and Δ_s? Despite the fact that *J* seems to outperform Δ_s when large sample sizes are utilized, there may be conditions under which the reverse is true. One potential set of circumstances where this might occur is under conditions of data degradation (e.g., high levels of missing or incomplete data). According to Woodhams, Grant, et al. (2007), this is an important consideration in

the context of BLA given that missing data is a common occurrence with police reports. In fact, they suggest that its robustness to missing data is the greatest strength of Δ_s. The argument for this is that its hierarchical structure allows for more opportunities on which to base similarity than *J*, as it has the ability to generate higher across-crime similarity scores even when little similarity exists at the level of discrete behaviours.

In order to test this argument, Study 1 consisted of a progressive data degradation procedure across all three crime types, replicating that of Woodhams, Grant, et al. (2007), and in turn Gauthier (2008). Despite the logic of Woodhams, Grant et al.'s (2007) argument, the present results did not find Δ_s to be superior to *J* under conditions of data degradation. That is, *J* outperformed Δ_s at 10%, 25%, and 50% degradation levels across all three crime types, achieving either equivalent to or typically higher *AUC*s at each level (with the possible exception of serial burglary at 50% degradation levels). These findings support those of Gauthier (2008), and challenge the results of Woodhams, Grant, et al. As Gauthier suggests, this finding is likely due to the inadequate sample size that was initially used to test these hypotheses, which did not allow "a reliable indication of the relative performance of Δ_s vs *J*" to be established" (p. 37).

However, it is important to emphasize that the degradation procedure used in this thesis, while identical to that used by Woodhams, Grant, et al. (2007) and Gauthier (2008), may not actually address the primary issue of importance. In these studies, data degradation involved reducing the dataset by progressively omitting more and more behaviours, as opposed to slowly increasing the amount of missing values in a dataset. In hindsight, it seems more likely that it is this problem that typically plagues police data, rather than the omission of specific crime scene behaviours across a sample of crimes. As

a result, the procedure adopted across these studies may not adequately highlight the strength of Δ_s (vs. J) in dealing with missing data. Unfortunately, CrimeSolver is currently unable to handle missing data values as opposed to missing behaviours, and it was practically not possible to examine this issue by other means. Exploring the impact of different types of data degradation should therefore be a priority in future research.

What is the impact of using different decision thresholds? ROC analysis was found to be an appropriate tool for the evaluation of discrimination accuracy in the present thesis, as it was able to quantify the degree of discrimination accuracy generated by the two similarity coefficients across crime types, unbiased by threshold placement. That is, ROC analysis allowed the direct comparison of the two measures along a common scale across all possible thresholds. However, in addition to this, the use of ROC analysis allowed for the exploration of how different decision thresholds impact linking performance. This is very important. As argued by Jones and Bennell (2007):

> In a practical context, it is futile to recognize the discriminatory power of a specific piece of evidence without further establishing a threshold at which that evidence is deemed present in sufficient quantity to render a given decision...One of the primary advantages of the ROC approach to such discrimination tasks is that accuracy and utility may be considered simultaneously, without one dimension biasing the other. (p. 229)

In the current study, two strategies were used with respect to selecting thresholds: (1) examining the ROC point falling closest to the upper-left corner of the ROC graph in order to maximize pH while minimizing pFA and (2) setting a pre-determined rate for pFA (.50) and maximizing pH without exceeding this value. As was seen, the ratios of

*p*H to *p*FA fluctuate greatly depending on the particular threshold selected, even when the level of linking accuracy remains constant. Equally important is the fact that the identified thresholds varied, sometimes
quite dramatically, across the difference crimes types. For example, when using the first s trategy, the optimal threshold was found to be .46 for serial homicide, but only .23 for serial burglary (when using *J*). All of these findings emphasize the fact that it is very important to select an appropriate threshold, especially when considering different types of crimes.

However, despite the fact that the analysis conducted in this study demonstrates the importance of selecting an appropriate threshold, it would be untrue to say that either strategy that was tested leads to an "optimal" threshold. The strategies that were tested are perhaps best viewed as being *rational*, but not *optimal*. The only procedure that can identify an optimal threshold requires that one take into account the probabilities of encountering linked and unlinked crimes in the jurisdiction of interest, and the various costs and benefits associated with all types of decision outcomes (hits, correct rejections, false alarms, and misses). Procedures exist for combining these values in order to derive truly optimal thresholds (see Swets, 1992). However, these values are difficult to estimate (e.g., what is the cost of a false alarm in the linking context?). In the future, attempts should be made in this area to identify these values (e.g., by conducting formal cost-benefit analyses). This will likely increase our ability to accurately link serial crimes.

Study 2 – Behavioural Frequency and Crime Type

BLA researchers have offered several suggestions with respect to the level of behavioural frequency most apt to distinguish between linked and unlinked crimes. As

discussed previously, the following recommendations appear to be amongst the most popular: (1) Canter et al.'s (2003) suggestion to omit very high frequency behaviours (i.e., >90%), (2) Salfati and Bateman's (2005) suggestion to omit behaviours that occur in more than 50% of crimes, (3) Woodhams, Hollin, et al.'s (2008) suggestion to omit behaviours with frequencies less than 10%, and (4) Santtila, Junkkila, et al.'s (2005) suggestion to omit very low frequency behaviours (i.e., <1%). Despite how frequently such recommendations are made (and followed) there is to date no empirical basis for making these recommendations. Thus, the purpose of Study 2 was to test each of the recommendations by comparing the degree of linking accuracy that could be achieved if the recommendations were followed, to the degree of linking accuracy that could be achieved if they weren't (i.e., if all behaviours were used). As *J* was found to be the optimal similarity coefficient in Study 1, it was used as the measure of across-crime similarity for all analyses in Study 2.

Which recommendation received the most support? In a sense, the results of this thesis demonstrate that all of the recommendations are supported to a degree, while at the same time none of them are supported. That is, all of the recommendations that could be tested across the three crime types resulted in levels of linking accuracy that exceeded chance, and the specific levels of accuracy that were achieved across the crime types were, unsurprisingly, predicted by the analyses in Study 1 (e.g., very high levels of accuracy were found for serial homicide, but relatively low levels were found for serial burglary). However, in no case did the recommendations lead to *AUCs* that were significantly larger than the *AUC* obtained from using all of the crime scene behaviours. Nor, for that matter, did the *AUCs* obtained by following the various recommendations

differ significantly from one another. In short, based on the results of this study, it did not really seem to matter which recommendation was followed, the linking results were essentially identical across all of the analyses.

However, this is not to say that the various recommendations did not influence any of the results. Indeed, to some extent, the recommendations did have their intended effect on the across-crime similarity scores (but not on the linking results). For example, the recommendation to remove high frequency behaviours before carrying out BLA is based on the idea that the inclusion of such behaviours will decrease linking accuracy due to behavioural distinctiveness being compromised (i.e., if many offenders exhibit these behaviours in their crimes, how will they be useful in discriminating between crimes committed by different offenders?). On examination of the results, these recommendations did in fact have their intended effect of increasing behavioural discrimination. Compared to the analyses where all crime scene behaviours were used, the removal of high frequency behaviours (>90%, or even >50%) resulted in lower across-crime similarity scores for unlinked crimes (i.e., behavioural distinctiveness increased). Unfortunately, the by-product of following these specific recommendations was that the across-crime similarity scores for linked crimes were adversely affected. In other words, compared to when all crime scene behaviours were used, the removal of high frequency behaviours decreased the amount of behavioural stability that could be found. The end result of course is that linking accuracy itself doesn't increase under such conditions.[5]

[5] A similar argument could have been made for the recommendations to remove behaviours with frequencies <1% or <10%. When implementing these recommendations, across-crime similarity scores for linked crimes often increased, but so to did across-crime similarity scores for unlinked crimes.

Are there other reasons for the lack of a behavioural frequency effect? Beyond these conceptual reasons for the lack of a behavioural frequency effect, there are other reasons for why an effect might not have been found in the current study. However, before discussing some of these reasons, it must be reiterated that not all of the recommendations could be examined in this thesis, at least not all within a single crime type. Indeed, there were several recommendations that could not be adequately tested, including: (1) omitting serial sexual assault behaviours with frequencies <1%, (2) omitting serial homicide behaviours with frequencies >90% and <1%, and (3) omitting serial burglary behaviours with frequencies >90%. Thus, at this stage, it is not possible to speak to these specific recommendations as they relate to these specific crime types, and this is an issue that can be addressed in future research. However, a few points can be raised about some of the other recommendations.

One likely reason why some of the recommendations did not receive strong support in this thesis likely has to do with the number of behaviours falling within each of the targeted frequency bands. For example, there were several cases where behaviours within the >50% frequency band were extremely rare and therefore their omission from the dataset would likely have a minimal impact on the overall level of linking accuracy that could be achieved. In the case of serial burglary, for instance, only one behaviour occurred in >50% of cases (well-maintained target). Not only does this probably explain the low level of linking accuracy that was found for the serial burglary sample in general, it also explains why following through with this recommendation in this case had a negligible impact on the *AUC* (compared to all behaviours). A similar problem occurred in other crime types (e.g., only one behaviour had a frequency >90% in the case of serial

sexual assault). As was seen in Study 1, removing large numbers of behaviours from a dataset can seriously affect the *AUC*. Thus, it stands to reason that, if there existed a larger number of behaviours in some of the frequency bands for certain crime types, the *AUCs* appearing throughout Study 2 might have changed more than they did. How they would change is a question that will only be answered through additional research.

Another potentially important issue has to do with the actual frequencies of the behaviours that were omitted from the dataset. This may be particularly important for the recommendations put forward by Canter et al. (2003) and Salfati and Bateman (2005). When one goes back and carefully reviews the recommendations that these researchers have made, it becomes clear that the basis of their recommendations (especially Canter et al.'s) has to do with the removal of what are commonly referred to as "core" behaviours. In brief, core behaviours are defined as those behaviours that essentially define the crime in questions. For example, in cases of stranger rape (as opposed to sexual assault), core behaviours are often thought to include variables such as vaginal penetration, surprise attack, and the use of a weapon, given that these behaviours are almost always exhibited by rapists in this context. In the current study, however, core behaviours are arguably absent from the samples and therefore it is difficult to know what one might find if they removed these behaviours. It would be surprising if their removal did not substantially increase the level of behavioural distinctiveness that was found for a given sample, but the impact that decreased behavioural stability would have on the overall linking results is less clear. A closer examination of core behaviours would be a useful direction for future research.

What are the implications of these results? As is the case for the results of Study 1, the results from the behavioural frequency analyses have several important implications for research and practice if the findings presented here can be replicated. With respect to research, the obvious implication is that researchers should think carefully about implementing any of the recommendations tested here until empirical evidence exists that clearly shows that the recommendations make a difference. This is an extremely important point when one considers the frequency with which researchers do implement these recommendations at present (indeed, it is hard to find a study of BLA where one of the recommendations tested here is not implemented). At this point in time, it looks like the best course of action for researchers is to use all of the crime scene behaviours that are at their disposal.

With respect to carrying out BLA in an investigative context, the results of this analysis also have important implications. Indeed, for some time people have been discussing the possibility that the sort of analyses conducted here might provide insight into the behaviours that should be used by crime analysts and investigators to link serial crime. Specifically, researchers such as Bennell and Canter (2002), have suggested that these results might provide the police with a way of enhancing the degree of efficiency when collecting data for linking purposes given that, currently, the approach is to collect every possible behaviour from a crime scene (e.g., ViCLAS, the current crime linkage system in Canada, is based on over 200 behaviours; Martineau & Corey, in press). If behaviours belonging to a particular frequency band were found to be more conducive to BLA, then the police could save valuable time and money by focusing on this subset of behaviour. However, as discussed above, the results of this thesis suggest that for the

time being it may be best for the police to continue examining all crime scene behaviours when carrying out BLA. This strategy appears to result in an appropriate balance between behavioural stability and distinctiveness, thus ensuring that a reasonable level of linking accuracy is achieved.

Study Limitations and Directions for Future Research

A number of limitations with this thesis have been raised throughout the discussion, as have a number of directions for future research. In this section, the major limitations and directions for the future are summarized.

First, there are potential problems with the use of police data in research of this type. In general, one must be cautious when using police data for research purposes given that it was not collected with this end goal in mind. Without doubt, people's agendas (e.g., victims, police officers, lawyers) likely influence this data, and much of it was likely collected and coded by multiple people. As discussed previously, the fact that inter-rater reliability could not be calculated for any of the datasets is also cause for concern, as is the fact that the data is only representative of solved crimes. Having said this, procedures were put in place to ensure the validity and reliability of each dataset (e.g., standardized proformas were used by the original coders in each case; conviction data was relied on; dichotomous coding was used, etc.), and readers should be confident that the quality of the data analyzed in the current study is very much similar to data used in other published studies. Ideally in the future, the analyses reported here should be replicated using data from police databases where inter-rater reliability has been assessed (e.g., ViCLAS) and some thought should go into how BLA research might be conducted

in a prospective fashion in order to deal with the problems of relying solely on solved offences.

Second, an exploration of how hierarchical structure impacts linking accuracy was simply beyond the scope of the current thesis. Yet, it is likely that the structure of behavioural hierarchies will seriously impact the performance of Δ_s. Thus, there are limits to what can currently be said about the relative performance of J vs. Δ_s. In addition, early in the process of developing the behavioural hierarchies, it became obvious that it was a challenging endeavour requiring a substantial number of subjective decisions. As mentioned previously, this may be an inherent weakness of Δ_s compared to J. Although it can be resolved to some extent by the inter-rater reliability checks that were conducted, there are others ways that could be used to eliminate many of these problems. Future research may wish to thoroughly explore the idea of using cluster analytic techniques to establish the structure of the hierarchies. While this will likely decrease the face validity of the hierarchies, there are reasons to believe that it would increase the degree of linking accuracy that would be achieved. The development of CrimeSolver makes it relatively easy to construct these sorts of hierarchies, and ROC analysis provides an appropriate tool for comparing the linking accuracy that results from subjective vs. objective approaches to establishing hierarchical structures.

Third, while the data degradation procedure explored in this thesis replicated the procedure used by Woodhams, Grant et al. (2007) and Gauthier (2008), other procedures could have been used that would have allowed the examination of missing data in a more productive way than the procedure that was adopted here. Specifically, instead of progressively removing more behaviours from the analysis, it might have been useful to

progressively remove more and more data. It seems that this is the key problem with investigative data, and that Δ_s would be better able to handle situations like this in a much more effective way than *J* (in hindsight, it is not totally clear why Δ_s would perform better than *J* when fewer behaviours are included in the analysis). Unfortunately, Crime Solver does not currently have the capability to examine this particular issue and it would have simply been too time consuming to manually run analyses that examined the issue of missing data. Therefore, in the future, this might be a productive line of research given that it might demonstrate the value of using Δ_s instead of *J*.

Fourth, there was obviously a limit on the extent that the behavioural frequency hypothesis could be examined in the current study given that some of the samples did not include crime scene behaviours that fit into each of the various frequency bands tested. Thus, it is not possible at the present time to speak to some of the recommendations being made by researchers who study BLA, at least not for certain crime types. In the future, the four recommendations that were examined here (which are by far the most common recommendations being discussed in the literature) should be more thoroughly examined. This will no doubt require an examination of datasets that are much larger than the ones that were examined in this thesis. Not only would this increase the chances that all of the recommendations could be examined, it would also increase the number of behaviours that fall within each of the frequency bands, including the very high frequency, core behaviours. As discussed, the removal of these core behaviours from a sample might have a significant impact on linking accuracy, and thus should be examined in future research.

Fifth, and finally, it must be appreciated that the current research is severely limited in the number of factors that were explored. Due to time and space constraints, only three factors – crime type, similarity coefficient, and behavioural frequency – were examined. These factors were selected because they are frequently mentioned in the BLA literature as variables that might influence the degree to which is possible to link crimes. However, there are many other factors that also could have been examined and certainly should be in future research (e.g., the nature of the behaviour, such as whether the crime scene action is a viewed by offenders as "offender driven" or "situation driven", etc.). The factors that were examined can also be expanded in future research. Indeed, there are many more crime types that could be examined, many more similarity coefficients that could be explored, and a number of other frequency recommendations that could be investigated.

Conclusion

The purpose of this thesis was to approach BLA from a signal detection perspective. This was done by using ROC analysis to: (1) compare the linking ability of two similarity coefficients, J and Δ_s, and (2) compare the linking accuracy that results when following four different recommendations regarding behaviours frequencies. In both Study 1 and Study 2, three crime types were examined – serial sexual assault, serial homicide, and serial burglary. The results of this thesis lead to several important conclusions which, if supported in future studies, could have implications for both BLA research and BLA practice.

The major findings of this thesis include the following:

1. Relying solely on crime scene behaviours, crimes committed by the same offender can be accurately differentiated from those committed by different offenders by using either *J* or Δ_s, although *J* slightly outperforms Δ_s when larger sample sizes are used. This seems to be the case because, while Δ_s increases the degree of behavioural stability that can be found, it also decreases the degree of behavioural distinctiveness that can be found.
2. While both *J* and Δ_s are able to differentiate between linked and unlinked crimes, linking accuracy varies substantially across crime types. Linking accuracy was very high for serial homicides, slightly lower for serial sexual assaults, and reasonably low for serial burglaries. This may be due to the fact that offenders who commit very violent interpersonal crimes, such as serial homicide, are guided by unique, very engrained, behavioural scripts.
3. Generally speaking, the ability to discriminate between linked and unlinked crimes decreased (with a few exceptions) as behaviours were randomly removed from the analysis. However, across conditions of data degradation, both similarity coefficients (especially *J*) were still able to distinguish between linked and unlinked crimes, even when 50% of the behaviours were removed from the analysis.
4. The decision of where to place a decision threshold when carrying out BLA is an important one and a variety of procedures exist for the purpose of selecting appropriate thresholds. The ratio of hits to false alarms fluctuates greatly across different thresholds that can be used when examining the same type of

crime. In addition, appropriate thresholds they also vary substantially across crime types. Thus, it is important to go through the process of selecting an appropriate threshold for the situation at hand.

5. The behavioural frequency recommendations that are currently being made by BLA researchers do not seem to have a significant impact on the degree of linking accuracy that can be achieved. Thus, at the present time, relying on all of the crime scene behaviours included in a dataset seems to be a sensible approach to take if one wishes to maximize linking accuracy.

While there are many limitations with the current thesis, the conclusions that are discussed above are reasonable and important. Each of the conclusions allows researchers to better understand the conditions under which it is possible to accurately link serial crime. Having said that, future research in this area is certainly needed to confirm that these conclusions are valid, and to uncover other factors that many influence the degree to which it is possible to link serial crime.

References

Alison, L. J., Snook, B., & Stein, K. L. (2001). Unobtrusive measurement: Using police information for forensic research. *Qualitative Research, 1*(2), 241-254.

Alison, L. J., & Stein, K. L. (2001). Vicious circles: Accounts of stranger sexual assault reflect abusive variants of conventional interactions. *Journal of Forensic Psychiatry, 12*(3), 515-538.

Alker, H. (1972). Is personality situationally specific or intrapsychically consistent? *Journal of Personality, 40,* 1-16.

Baroni-Urbani, C., & Buser, M. W. (1976). Similarity of binary data. *Systematic Zoology, 25*(3), 251-259.

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

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behaviour. *Psychological Review, 81*(6), 506-520.

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. (in press). Linkage analysis. In C. Laney (Ed.), *Course handbook for the M.Sc. in Forensic Psychology.* Leicester, UK: Leicester University.

Bennell, C., Alison, L. J., Stein, K. L., Alison, E. K., & Canter, D. V. (2001). Sexual offences against children as the abusive exploitation of conventional adult-child relationships. *Journal of Social and Personal Relationships, 18*(2), 155-171.

Bennell, C., Bloomfield, S., Snook, B., Taylor, P. J., & Barnes, C. (2008). Discriminating linked and unlinked burglaries: Comparing the performance of university students, forensic professionals, and a logistic regression model. *Manuscript submitted for publication.*

Bennell, C., & Canter, D. V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science & Justice, 42*(3), 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*(1), 23-41.

Bennell, C., Jones, N. J., & Melnyk, T. (2007). *Linking serial rapes: A test of the behavioural frequency hypothesis.* Poster presented at the annual conference of the Canadian Psychological Association/North American Correctional & Criminal Justice Psychology Conference, Ottawa, Ontario.

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

Blackburn, R. (1993). *The psychology of criminal conduct: Theory, research and practice.* Chichester, UK: Wiley.

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2006). Continuity and change in psychopathic traits as measured via normal-range personality: A longitudinal biometric study. *Journal of Abnormal Psychology, 115*(1), 85-95.

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

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

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

Canter, D. V., Alison, L. J., Alison, E., & Wentink, N. (2004). The organized/disorganized typology of serial murder: Myth or model? *Psychology, Public Policy, and Law, 10,* 293-320.

Canter, D. V., Bennell, C., Alison, L. J., & Reddy, S. (2003). Differentiating sex offences: A behaviorally based thematic classification of stranger rapes. *Behavioural Sciences & the Law, 21*(2), 157-174.

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.

Canter, D. V., Heritage, R., Wilson, M., Davies, A., Kirby, S., Holden, R., et al. (1991). *A facet approach to offender profiling.* London, UK: Home Office.

Canter, D. V., Wilson, M., Jack, K., & Butterworth, D. (1996). *The psychology of rape investigations: A study in police decision making.* Liverpool, UK: University of Liverpool.

Caspi, A. & Herbener, E. S. (1990). Continuity and change: Assortative marriage and the consistency of personality in adulthood. *Journal of Personality and Social Psychology, 58*(2), 250-258.

Cheetham, A. H., & Hazel, J. E. (1969). Binary (presence-absence) similarity coefficients. *Journal of Paleontology, 43*(5), 1130-1136.

Davies, A. (1991). The use of DNA profiling and behavioural science in the investigation of sexual offences. *Medicine, Science and Law, 31,* 95-101.

Davies, A. (1992). Rapists' behaviour: A three aspect model as a basis for analysis and the identification of serial crime. *Forensic Science International, 55,* 173-194.

Dunn, D. (2001). *Statistics and data analysis for the behavioral sciences.* New York: McGraw-Hill.

Endler, N. S. (1973). The person versus the situation – A pseudo issue? A response to Alker. *Journal of Personality, 4,* 286-303.

Endler, N. S., & Okada, M. (1975). A multidimensional measure of trait anxiety: The S-R Inventory of General Trait Anxiousness. *Journal of Consulting & Clinical Psychology, 43,* 319-329.

Eng, J. (2008). ROC analysis: Web-based calculator for ROC curves. Retrieved April, 2008, from http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html (by Gauthier, 2008).

Ewart, B. W., Oatley, G. C., & Burn, K. (2005). Matching crimes using burglars' modus operandi: A test of three models. *International Journal of Police Science and Management, 7*(3), 160-174.

Frick, P. J., Kimonis, E. R., Dandreaux, D. M., & Farell, J. M. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Sciences and the Law, 21*(6), 713-736.

Funder, D. C., & Colvin, C. R. (1991). Explorations of behavioural consistency: Properties of persons, situations, and behaviours. *Journal of Personality and Social Psychology, 60,* 773-704.

Gauthier, D. (2008). *Does a taxonomic similarity index increase the ability to link serial crime?* Unpublished honours thesis, Carleton University, Ottawa, Canada.

Godwin, G. M. (1998). *Inner themes – outer behaviours: A multivariate facet model of U.S. serial murderers' crime scene actions.* Unpublished doctoral dissertation, University of Liverpool, Liverpool, UK.

Goodwill, A. M., & Alison, L. J. (2006). The development of a filter model for prioritizing suspects in burglary offences. *Psychology, Crime & Law, 12*(4), 395-416.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 3*(1), 5-48.

Green, E. J., Booth, C. E., & Biderman, M. D. (1976). Cluster analysis of burglary M/O's. *Journal of Police Science and Administration, 4*(4), 382-287.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Oxford, UK: Robert E. Krieger.

Groth, A. N. (1979). *Men who rape: The psychology of the offender*. New York: Plenum.

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

Hakkanen, H., Lindlof, 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.

Hakkanen, H., Puolakka, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in arsons. *Legal and Criminological Psychology*, 9(2), 197-214.

Hazelwood, R., & Douglas, J. E. (1980). The lust murderer. *FBI Law Enforcement Bulletin, 4*, 18-22.

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

Hodge, S. (2001). A multivariate model of serial sexual murder. In D.V. Canter, & L. J. Alison (Eds.), *Profiling, rape, & murder.* Aldershot, England: Ashgate.

Holmes, R. M., & DeBurger, J. (1988). *Serial murder.* Beverly Hills, CA: Sage Publications.

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biological Reviews of the Cambridge Philosophical Society, 57*(4), 669-689.

Izsak, C., & A. R. G., Price (2001). Measuring β-diversity using a taxonomic similarity index, and its relation to spatial scale. *Marine Ecology Progress Series, 215,* 69-77.

Jaccard, P. (1908). Nouvelle recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat., 44*, 223-270.

Jones, N. J. (2005). *The temporal stability and cross-situational consistency of behavioural themes in serial homicide*. Unpublished masters thesis, Carleton University, Ottawa, Canada.

Jones, N. J., & Bennell, C. (2007). The development and validation of statistical prediction rules for discriminating between genuine and simulated suicide notes. *Archives of Suicide Research, 11*(2), 219-233.

Keppel, R. D. (1997). *Signature killers*. New York: Pocket Books.

Keppel, R. D., & Weiss, J. G. (1992). *Improving the investigation of violent crime: The homicide investigation and tracking system (HITS)*. Washington, DC: U.S. Department of Justice, National Institute of Justice.

Knight, R. A., & Prentky, R. A. (1990). Classifying sexual offenders: The development and corroboration of taxonomic models. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp.23-520). New York: Plenum Press.

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.

Lamboy, W. F. (1994). Computing genetic similarity coefficients from RAPD data: The effects of PCR artefacts. *PCR Methods and Applications, 4*(1), 31-37.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Liebetrau, A. M. (1983). Measure of association. Beverly Hills, CA: Sage Publications.

Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2008). The stability of psychopathy from adolescence into adulthood: The search for moderators. *Criminal Justice and Behavior, 35*(2), 228-243.

Marshall, W. L. (1989). Intimacy, loneliness, and sexual offenders. *Behavioural Research in Therapy*, *27*(5), 491-503.

Martineau, M., & Corey, S. (in press). Investigating the reliability of the Violent Crime Linkage Analysis System Crime Report. *Journal of Police and Criminal Psychology.*

Melnyk, T., Bennell, C., & Jones, N. J. (in press). Linking serial rapes: A test of the behavioural frequency hypothesis. *Proceedings of the Canadian Psychological Association*. Ottawa: Canadian Psychological Association.

Merry, S. (2000). Crime analysis: Principles for analyzing everyday serial crime. In D. V. Canter and L. J. Alison (Eds.), *Profiling property crimes* (pp.297-318). Aldershot, UK: Dartmouth.

Merry, S., & Harsent, L. (2000). Intruders, pilferers, rasider, and invaders: The interpersonal dimension of burglary. In D.V. Canter and L. J. Alison (Eds.), Profiling property crimes (pp. 31-56). Aldershot, UK: Dartmouth.

Mokros, A., & Alison, L. J. (2002). Is offender profiling possible? Testing the predicted homology of crime scene sctions and background characteristics in a sample of rapists. *Legal & Criminological Psychology, 7,* 25-43.

Monahan, J. (1992). Mental disorder and violent behaviour: Perceptions and evidence. *American Psychologist, 47*, 511-521.

Moos, R. H. (1968). Situational analysis of a therapeutic community milieu. *Journal of Abnormal Psychology, 73,* 49-61.

Moos, R. H. (1969). Sources of variance in response to questionnaires and in behaviour. *Journal of Abnormal Psychology, 74,* 405-412.

Mroczek, D. K., & Spiro, A. (2003). Personality structure and process, variation between and within: Integration by means of a developmental framework. *Journal of Gerontology, 58*(6), 153-165.

Pinizzotto, A. J., & Finkel, N. J. (1990). Criminal personality profiling: An outcome and process study. *Law and Human Behavior, 14*(3), 215-233.

Prentky, R. A., & Burgess, A. W. (2000). *Forensic management of sexual offenders.* New York, NY: Kluwer Academic/Plenum Publishers.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3-25.

Rossmo, D. K. (2000). *Geographic profiling.* Boca Raton, FL: Chemical Rubber Company.

Ressler, R. K., Burgess, A. W., & Douglas, J. E., (1988). *Sexual homicide: Patterns and motives.* New York, NY: The Free Press.

Ressler, R. K., Burgess, A. W., Douglas, J. E., Hartmann, C. A., & D'Agostino, R. B. (1986). Sexual killers and their victims – Identifying patterns through crime scene analysis. *Journal of Interpersonal Violence, 1*, 288-308.

Salfati, C. G. (2000). Profiling homicide: A multidimensional approach. *Homicide Studies, 4*(3), 265-293.

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.

Salfati, C. G., & Dupont, F. (2006). Canadian homicide: An investigation of crime-scene actions. *Homicide Studies, 10*(2), 118-139.

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), 1-16.

Santtila, P., Junkkila, J., & Sandnabba, N. K. (2005). Behavioural linking of stranger rapes. *Journal of Investigative Psychology and Offender Profiling, 2*(2), 87-103.

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

Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry*, *44*(1), 2-32.

Shoda, Y. (1999). Behavioral expressions of a personality system: Generation and perception of behavioural signatures. In D. Cervone & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 155-191). New York: Guildford Press.

Simberloff, D., & Connor, E. F. (1979). Q-mode and R-mode analyses of biogeographic distributions: Null hypotheses based on random colonization. In G. P. Patil and M. L. Rosenzweig (Eds.), *Contemporary quantitative ecology and related ecometrics, Series 12* (pp. 123-138). Fairland, MD: Statistical Ecology International Cooperative Publishing House.

Snook, B., Taylor, P. J., & Bennell, C. (2005). Shortcuts to geographic profiling success: A reply to Rossmo. *Applied Cognitive Psychology, 19,* 655-661.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stake diagnostics. *American Psychologist, 47*, 522-532.

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

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

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*(2), 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*(2), 233-249.

Woodhams, J., Hollin, C. R., & Bull, R. (2008). Incorporating context in linking crimes: An exploratory study of situational similarity and if-then contingencies. *Manuscript submitted for publication.*

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*(1), 59-85.

Appendix A

Content Dictionary of Serial Sexual Assault Behaviours

Content Dictionary

Thirty-six variables were created from a content analysis of victim statements in order to provide a list of elements common to offences. All variables are dichotomous with values based on the presence (1) or absence (0) of each category of behaviour. A description of the categorization scheme in alphabetical order is given below. Occurrence percentages are provided in brackets after the variable name.

1. Anal (14.3%). This variable refers to the offender penetrating or attempting to penetrate the victim's anus.

2. Apologize (7.1%). This variable refers to the offender using apologetic language directed at the victim at some point during the offence.

3. Ask (24.6%). This variable refers to the offender asking questions or being inquisitive about the victim.

4. Bind (20.6%). This variable refers to the use, at any time during the attack, of any article to bind the victim (excluding restraint by the offender's hands).

5. Blindfold (25.4%). This variable refers to the use, at any time during the attack, of any physical interference with the victim's ability to see (excluding verbal threats to the victim to close eyes or the use of the offender's hands).

6. Blitz (5.6%). This variable refers to the offender using a sudden and violent attack to overpower the victim; distinct from surprise attack because of the use of violence.

7. Compliment (10.3%). This variable refers to the offender complimenting the victim (e.g., on appearance).

8. Con (20.6%). This variable refers to the offender approaching the victim by giving a false impression of legitimacy (i.e., a false story, asking questions, etc.).

9. Cunnilingus (19.8%). This variable refers to the offender performing a sexual act on the victim's genitalia or attempting to perform such a sex act using his mouth.

10. Demand goods (25.4%). This variable refers to the offender approaching the victim with a demand for goods or money. This variable specifically relates to initial demands.

11. Demean (17.5%). This variable refers to the offender demeaning or insulting the victim (e.g., using profanities directed at the victim or women in general).

12. Disguise (15.9%). This variable refers to the offender wearing any form of disguise.

13. Display weapon (51.6%). This variable refers to the offender displaying a weapon in order to control the victim.

14. Extend time (14.3%). This variable refers to the offender extending the time spent with the victim after the actual attack.

15. Fellatio (24.6%). This variable refers to the offender forcing the victim to perform oral sex.

16. Force (victim participation) (38.9%). This variable refers to the offender forcing the victim to physically participate in the sexual aspects of the offence.

17. Force (victim sexual comment) (5.6%). This variable refers to the offender forcing the victim to make sexual comments.

18. Forensic awareness (23.0%). This variable refers to the offender showing knowledge of forensic procedures (e.g., fingerprints, DNA).

19. Gag (15.1%). This variable refers to the use, at any time during the attack, of any article to prevent the victim from making noise (excluding the temporary use of the offender's hand).

20. Identify (26.2%). This variable refers to the offender taking steps to obtain from the victim details that would identify her (e.g., by examining the victim's belongings).

21. Imply (9.5%). This variable refers to the offender implying that he knows the victim.

22. Kiss (37.3%). This variable refers to the offender kissing or attempting to kiss the victim.

23. Reassure (18.3%). This variable refers to the offender using reassuring or comforting language.

24. Reveal (26.2%). This variable refers to the offender revealing information about himself.

25. Sexual comment (52.4%). This variable refers to the offender making sexual comments during the attack.

26. Steal (identifiable) (7.9%). This variable refers to the offender stealing items from the victim that are recognizable as belonging to the victim (e.g., the victim's wallet).

27. Steal (personal) (10.3%). This variable refers to the offender stealing items from the victim that are personal to the victim, but not necessarily of any great value in terms of re-saleable goods (e.g., photographs or letters).

28. Steal (unidentifiable) (34.9%). This variable refers to the offender stealing items from the victim that are not recognizable as belonging to the victim (e.g., cash).

29. Surprise (92.1%). This variable refers to the offender using a method of approach consisting of an immediate attack on the victim.

30. Tear clothing (15.9%). This variable refers to the offender forcibly tearing the victim's clothing.

31. Threat (no report) (29.4%). This variable refers to the offender threatening the victim that she should not report the incident to the police or to any other person.

32. Threat (verbal) (11.9%). This variable refers to the offender threatening the victim using insults and profanity at some time during the attack (excluding threats not to report the incident).

33. Vaginal (front) (74.6%). This variable refers to the offender penetrating or attempting to penetrate the victim's vagina from the front.

34. Vaginal (rear) (17.5%). This variable refers to the offender penetrating or attempting to penetrate the victim's vagina from the rear.

35. Violence (multiple) (12.7%). This variable refers to the offender perpetrating multiple acts of violence against the victim (e.g., multiple punches).

36. Violence (single) (30.2%). This variable refers to the offender perpetrating a single act of violence against the victim (e.g., a single slap).

Appendix B

Content Dictionary of Serial Homicide Behaviours

Content Dictionary

Thirty-nine variables were created from a content analysis of police reports in order to provide a list of elements common to offences. All variables are dichotomous with values based on the presence (1) or absence (0) of each category of behaviour. A description of the categorization scheme in alphabetical order is given below. Occurrence percentages are provided in brackets after the variable name.

1. Address book (4.6%). This variable refers to the offender stealing and using the victim's address book in order to make contact with future victims.

2. Bind (51.9%). This variable refers to the victim's body being bound by, for example, panty hose, rope, electrical cord, etc.

3. Bite (3.4%). This variable refers to evidence of biting the victim's body, including any evidence of chewing on a particular body part.

4. Blindfold (6.4%). This variable refers to the use, at any time during the attack, of any physical interference with the victim's ability to see.

5. Blitz (61.0%). This variable refers to the sudden and immediate use of violence, which may or may not be preceded by a confidence or ploy approach, and which incapacitates the victim.

6. Bludgeon (43.7%). This variable refers to the offender's method of attack being by hand, fist, or the use of a blunt instrument as a weapon (such as a club or statue).

7. Body (bagged) (3.8%). This variable refers to the victim's body being found disposed of in a bag.

8. Body (burned) (3.8%). This variable refers to the victim's body being found burned, either in part or completely.

9. Body (concealed) (56.5%). This variable refers to the victim's body being found concealed or hidden (e.g., the body could not be viewed with ease, the body was covered or buried, visibility was obstructed by trees or other barriers, etc.).

10. Body (moved) (52.1%). This variable refers to the victim's body being moved from the assault or murder site to the disposal site. This could include either moving the body by foot or transport.

11. Body (open) (25.7%). This variable refers to the victim's body, when discovered, being found openly displayed/exposed (e.g., the body could be viewed with ease, and it was not obstructed by any barriers).

12. Body (staged) (10.2%). This variable refers to the offender intentionally staging or posing the victim's body, usually with the intent of shocking those who discover it.

13. Body (weighted) (3.0%). This variable refers to the victim's body being found weighted down by an object such as rocks, metal, cement, tires, or chain.

14. Cannibal (6.3%). This variable refers to the offender engaging in cannibalism and/or drinking of the victim's blood.

15. Captive (38.0%). This variable refers to the victim being held captive for more than eight hours prior to their murder.

16. Carve (1.3%). This variable refers to evidence of writing or carving on the victim's body using various instruments (e.g., a knife, lipstick, pen, etc.).

17. Choke (33.2%). This variable refers to the offender having choked the victim.

18. Con (56.8%). This variable refers to the offender having initiated contact with the victim prior to the attack by use of a con/ploy/ruse or deception.

19. Crime kit (28.8%). This variable refers to the offender possessing a crime kit for torturing his victims. This could include any items to torture the victim, such as pliers, electrical devices, etc. 20. Destroy evidence (37.3%). This variable refers to the offender having destroyed or attempted to destroy physical evidence at the crime scene. This variable is taken as a sign that the offender is showing forensic awareness.

21. Disfigure (24.6%). This variable refers to the offender having committed acts of torture or unusual assaults on the victim's body. This could include any removal of body parts, burns, and/or mutilation of body cavities.

22. Dismember (23.2%). This variable refers to the offender having dismembered the victim's body by any means, including, but not limited to, biting, sawing, and cutting.

23. Foreign object (21.0%). This variable refers to any evidence that a foreign object had been inserted into any of the victim's body cavities.

24. Gag (30.2%). This variable refers to the use, at any time during the attack, of any article to prevent the victim from making noise.

25. Hand/Feet (25.3%). This variable refers to the use of the offender's hands, feet, legs, or arms to strangle or beat the victim.

26. Injury post-mortem (22.8%). This variable refers to the victim's body incurring injuries post-mortem.

27. Photo (12.2%). This variable refers to the offender taking photos or videos of the victim prior to death.

28. Piqueur (10.5%). This variable refers to acts performed on the victim's body with a knife or other sharp instrument that indicates excessive stabbing, cutting, or ripping of the flesh. These wounds are usually inflicted near the genital or breast areas.

29. Restraints (44.9%). This variable refers to the offender bringing restraining device(s) to the crime scene.

30. Sexual assault (ante-mortem) (74.1%). This variable refers to the offender sexually assaulting the victim prior to killing them.

31. Sexual assault (post-mortem) (12.1%). This variable refers to the offender sexually assaulting the victim post-mortem.

32. Souvenir (65.0%). This variable refers to the offender taking from the victim after death personal items (e.g., photos, jewellery, but not clothing).

33. Stab (lower body) (10.1%). This variable refers to the location of the stab wounds on the victim's body. This could include any stabbing or cutting wounds below the waist.

34. Stab (multiple) (28.5%). This variable refers to the victim suffering multiple (10 or more) stab wounds.

35. Stalk (14.3%). This variable refers to the offender stalking the victim for one day or more prior to committing the murder.

36. Torture (40.1%). This variable refers to the offender performing sadistic acts upon the victim's body while he or she was still alive. This may include acts such as cutting, burning, or electric shock. Also included in this variable would be any mental torture inflicted on the victim.

37. Trophy (28.3%). This variable refers to the offender retaining personal items/clothes of the victim's for the purpose of personal gratification.

38. Weapon (brought) (87.2%). This variable refers to the offender pre-selecting the weapon and bringing it to the crime scene.

39. Weapon (opportunity) (12.4%). This variable refers to the offender finding the weapon at the crime scene, or it was brought to the scene by the victim.

Appendix C

Content Dictionary of Serial Burglary Behaviours

Content Dictionary

Twenty-eight variables were created from a content analysis of police reports in order to provide a list of elements common to offences. All variables are dichotomous with values based on the presence (1) or absence (0) of each category of behaviour. A description of the categorization scheme in alphabetical order is given below. Occurrence percentages are provided in brackets after the variable name.

1. Alarm (7.6%). This variable refers to the dwelling being equipped with an operational security system at the time of the burglary.

2. Brick (4.3%). This variable refers to entry to the dwelling being gained by using a brick or stone.

3. Carrier taken (11.4%). This variable refers to the offender taking an item from the dwelling in order to carry other stolen items (e.g., pillow case, bag, etc.).

4. Crowbar (1.9%). This variable refers to entry to the dwelling being gained by using a crowbar.

5. Dog (1.4%). This variable refers to a guard dog occupying the dwelling at the time of the burglary.

6. Detached (48.1%). This variable refers to the type of dwelling as being detached (e.g., house, bungalow, farm building, etc.).

7. Enclosed (46.7%). This variable refers to the dwelling having an enclosed front or rear supporting unobservable access.

8. Exit prepared (0.5%). This variable refers to the exit point from the dwelling being prepared after entry to dwelling was gained.

9. Five items (27.6%). This variable refers to the offender having stolen five or more separate items from the dwelling.

10. Force (44.8%). This variable refers to entry to the dwelling being gained by using force (e.g., bodily force, crowbar, brick, etc.).

11. Gratuitous mess (5.7%). This variable refers to the offender having ransacked the dwelling or making more of a mess than is necessary for the commission of the burglary (e.g., defecating in the dwelling, urinating in the dwelling, causing unnecessary damage, etc.).

12. Jewellery (36.7%). This variable refers to the offender having stolen any form of jewellery (valuable or costume).

13. Maintained (poor) (4.8%). This variable refers to the dwelling appearing to be poorly maintained by the occupants from the outside.

14. Maintained (well) (57.1%). This variable refers to the dwelling appearing to be well maintained by the occupants from the outside.

15. Non-detached (48.1%). This variable refers to the type of dwelling being non-detached (e.g., terraced, flat, or high rise dwelling).

16. Nothing stolen (9.5%). This variable refers to the offender not taking any property from the dwelling.

17. Occupied (car) (11.9%). This variable refers to a car being in the driveway of the dwelling indicating probable occupancy.

18. Occupied (curtains) (12.4%). This variable refers to curtains of the dwelling being drawn indicating probable unoccupancy.

19. Occupied (lights) (6.2%). This variable refers to lights in the dwelling being on indicating probable occupancy.

20. Search (intrusive) (34.3%). This variable refers to the search of the dwelling being deemed intrusive (e.g., drawers or cupboards searched, drawers removed, food consumed, etc.).

21. Search (multiple) (41.9%). This variable refers to multiple rooms being searched.

22. Search (no) (25.2%). This variable refers to the offender making no search of the dwelling (either property at specific sites within the dwelling were targeted (e.g., observable cash) or only property from the immediate area surrounding the point of entry was taken).

23. Search (private) (21.9%). This variable refers to the search of the dwelling targeting most bedrooms, bathrooms, or the master bedroom only.

24. Search (untidy) (36.7%). This variable refers to the search of the dwelling being conducted in an untidy fashion (e.g., drawers removed and tipped out).

25. Secured (11.0%). This variable refers to the dwelling being secured by the offender to exclude occupants or observation by neighbours (e.g., closing curtains, wedging and locking internal doors, pacifying a dog with drugs or food, etc.).

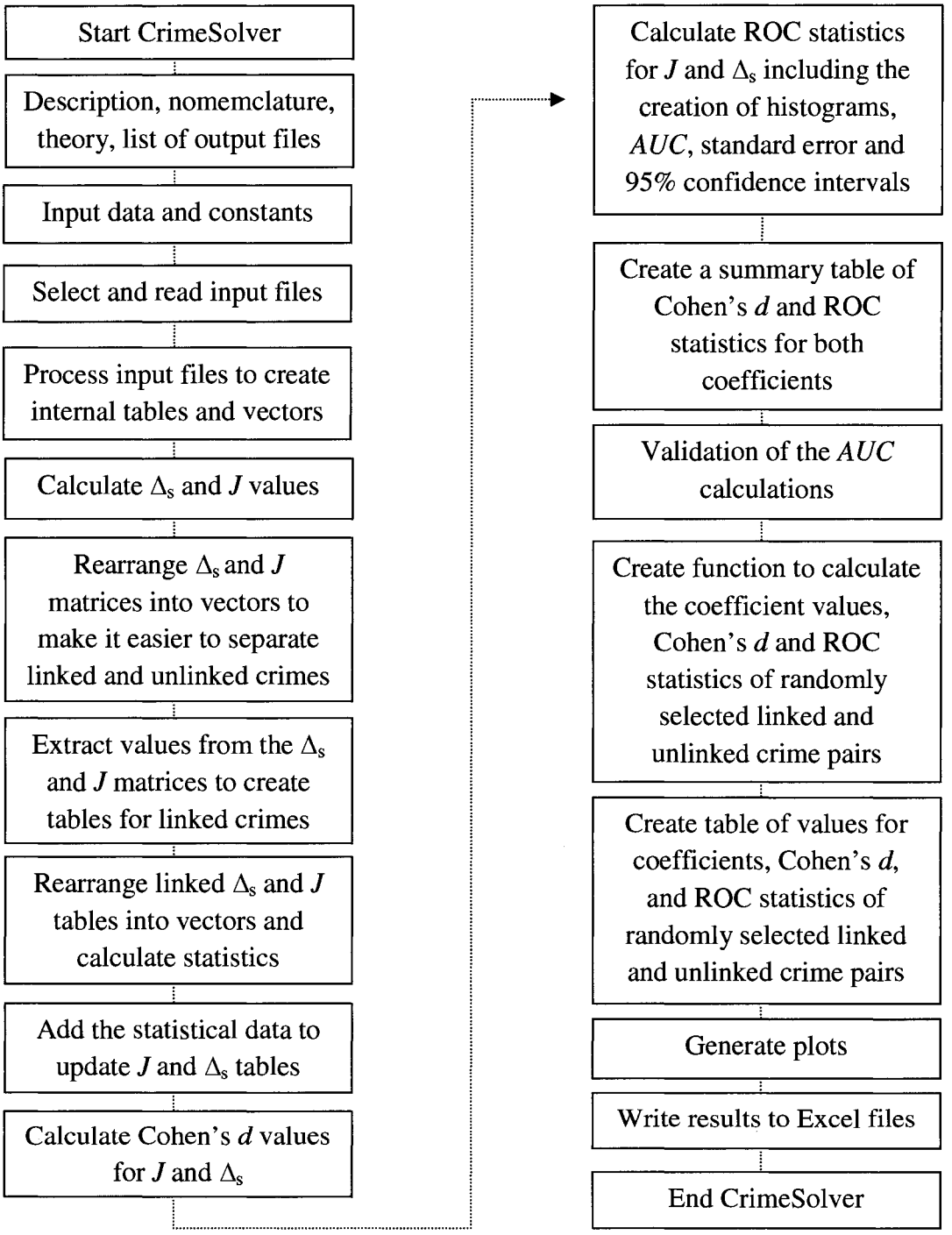
26. Security light (9.0%). This variable refers to the dwelling being equipped with one or more security lights.

27. Sentimental (38.1%). This variable refers to items stolen from the dwelling that were deemed to have sentimental value to the owner (e.g., antiques, clocks and watches, glass or crystal, etc.).

28. Special disposal (25.7%). This variable refers to the offender stealing items that likely require special arrangements to be made in order to sell the items (e.g., bank documents, china and porcelain, paintings, etc.).

Appendix D

CrimeSolver Flow Chart



Note. Results are displayed throughout the program. To reduce clutter, this is not shown on the flow chart. Flow chart is as provided by Gauthier (2008).

Appendix E

Path Lengths Table of Serial Sexual Assault Behaviours

Case A → / ← Case B	1 blitz	2 forcomme	3 forcpartic	4 singlevio	5 multivio	6 clothtorn	7 demeans	8 demandgd	9 stealID	10 stealnoID	11 stealperso	12 binds	13 blindfold	14 disguise	15 displaywea	16 foraware	17 gags	18 IDsvic	19 surpattck	20 conappr	21 implyknow	22 threatnorep	23 verbthreat	24 analpene	25 cunnilin	26 fellatio	27 vagfront	28 vagrear	29 sexcomm	30 apologize	31 asksq	32 complim	33 extends	34 kiss	35 reassures	36 reveals
1 blitz	0	1	1	1	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2 forcomme	1	0	1	1	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3 forcpartic	1	1	0	1	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4 singlevio	1	1	1	0	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5 multivio	1	1	1	1	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
6 clothtorn	1	1	1	1	1	0	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
7 demeans	2	2	2	2	2	2	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
8 demandgd	3	3	3	3	3	3	3	0	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
9 stealID	3	3	3	3	3	3	3	1	0	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
10 stealnoID	3	3	3	3	3	3	3	1	1	0	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
11 stealperso	3	3	3	3	3	3	3	1	1	1	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
12 binds	3	3	3	3	3	3	3	3	3	3	3	0	1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
13 blindfold	3	3	3	3	3	3	3	3	3	3	3	1	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
14 disguise	3	3	3	3	3	3	3	3	3	3	3	1	1	0	1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
15 displaywea	3	3	3	3	3	3	3	3	3	3	3	1	1	1	0	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
16 foraware	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	0	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
17 gags	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	0	1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
18 IDsvic	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	1	0	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
19 surpattck	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	1	1	0	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
20 conappr	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	0	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3
21 implyknow	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	1	0	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3
22 threatnorep	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	1	1	0	1	3	3	3	3	3	3	3	3	3	3	3	3	3
23 verbthreat	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	1	1	1	0	3	3	3	3	3	3	3	3	3	3	3	3	3
24 analpene	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	1	1	1	1	2	3	3	3	3	3	3	3
25 cunnilin	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	0	1	1	1	2	3	3	3	3	3	3	3
26 fellatio	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	0	1	1	2	3	3	3	3	3	3	3
27 vagfront	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	0	1	2	3	3	3	3	3	3	3
28 vagrear	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	0	2	3	3	3	3	3	3	3
29 sexcomm	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	0	3	3	3	3	3	3	3
30 apologize	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	1	1	1	1	1	1
31 asksq	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	0	1	1	1	1	1
32 complim	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	0	1	1	1	1
33 extends	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	0	1	1	1
34 kiss	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	0	1	1
35 reassure	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	0	1
36 reveals	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	1	0

Appendix F

Path Lengths Table of Serial Homicide Behaviours

← Case B / Case A →	1 crimekit	2 restroff	3 weapoff	4 stalked	5 bind	6 blindfol	7 captive	8 destevid	9 gagged	10 con	11 bodybag	12 bodyburn	13 bodyconc	14 bodymove	15 bodyweight	16 sexante	17 photos	18 addbook	19 weaponop	20 bodyopen	21 blitz	22 bludgeon	23 choked	24 handfeet	25 foreobj	26 piqueur	27 sexpost	28 bitemark	29 bodposed	30 cannibal	31 carving	32 disfigur	33 dismember	34 injrpost	35 lowbody	36 multstab	37 tortured	38 trophies	39 souvenir
1 crimekit	0	1	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2 restroff	1	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3 weapoff	1	1	0	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4 stalked	2	2	2	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5 bind	3	3	3	3	0	1	1	1	1	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6 blindfol	3	3	3	3	1	0	1	1	1	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
7 captive	3	3	3	3	1	1	0	1	1	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
8 destevid	3	3	3	3	1	1	1	0	1	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
9 gagged	3	3	3	3	1	1	1	1	0	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
10 con	3	3	3	3	2	2	2	2	2	0	2	2	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
11 bodybag	3	3	3	3	2	2	2	2	2	2	0	1	1	1	1	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
12 bodyburn	3	3	3	3	2	2	2	2	2	2	1	0	1	1	1	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
13 bodyconc	3	3	3	3	2	2	2	2	2	2	1	1	0	1	1	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
14 bodymove	3	3	3	3	2	2	2	2	2	2	1	1	1	0	1	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
15 bodyweight	3	3	3	3	2	2	2	2	2	2	1	1	1	1	0	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
16 sexante	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
17 photos	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	0	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
18 addbook	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	1	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
19 weaponop	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
20 bodyopen	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
21 blitz	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	0	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
22 bludgeon	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	1	0	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
23 choked	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	1	1	0	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
24 handfeet	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	1	1	1	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
25 foreobj	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	0	1	1	2	2	2	2	2	2	2	2	2	2	2	2
26 piqueur	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	1	0	1	2	2	2	2	2	2	2	2	2	2	2	2
27 sexpost	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	1	1	0	2	2	2	2	2	2	2	2	2	2	2	2
28 bitemark	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	0	1	1	1	1	1	1	1	1	1	2	2
29 bodposed	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	0	1	1	1	1	1	1	1	1	2	2
30 cannibal	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	0	1	1	1	1	1	1	1	2	2
31 carving	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	0	1	1	1	1	1	1	2	2
32 disfigur	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	0	1	1	1	1	1	2	2
33 dismember	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	1	0	1	1	1	1	2	2
34 injrpost	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	1	1	0	1	1	1	2	2
35 lowbody	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	1	1	1	0	1	1	2	2
36 multstab	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	1	1	1	1	0	1	2	2
37 tortured	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	1	1	1	1	1	1	1	1	1	0	2	2
38 trophies	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	0	1
39 souvenir	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	1	0

Appendix G

Path Lengths Table of Serial Burglary Behaviours

↓ Case B / Case A →	1 crowbar	2 exitpre	3 multsrch	4 private	5 secured	6 fiveitem	7 jewellry	8 sentimen	9 specdis	10 enclose	11 detach	12 occcurt	13 wellmain	14 force	15 gratmess	16 intrude	17 untidy	18 carrier	19 brick	20 nosrch	21 nilstole	22 alarm	23 dog	24 nondet	25 occcar	26 occlight	27 poormain	28 seclight
1 crowbar	0	2	2	2	2	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2 exitpre	2	0	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3 multsrch	2	1	0	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4 private	2	1	1	0	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5 secured	2	1	1	1	0	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6 fiveitem	2	2	2	2	2	0	1	1	1	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
7 jewellry	2	2	2	2	2	1	0	1	1	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
8 sentimen	2	2	2	2	2	1	1	0	1	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
9 specdis	2	2	2	2	2	1	1	1	0	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
10 enclose	3	3	3	3	3	3	3	3	3	0	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
11 detach	3	3	3	3	3	3	3	3	3	1	0	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
12 occcurt	3	3	3	3	3	3	3	3	3	1	1	0	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
13 wellmain	3	3	3	3	3	3	3	3	3	1	1	1	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
14 force	4	4	4	4	4	4	4	4	4	4	4	4	4	0	2	2	2	2	3	3	3	3	3	3	3	3	3	3
15 gratmess	4	4	4	4	4	4	4	4	4	4	4	4	4	2	0	2	2	2	3	3	3	3	3	3	3	3	3	3
16 intrude	4	4	4	4	4	4	4	4	4	4	4	4	4	2	2	0	2	2	3	3	3	3	3	3	3	3	3	3
17 untidy	4	4	4	4	4	4	4	4	4	4	4	4	4	2	2	2	0	2	3	3	3	3	3	3	3	3	3	3
18 carrier	4	4	4	4	4	4	4	4	4	4	4	4	4	2	2	2	2	0	3	3	3	3	3	3	3	3	3	3
19 brick	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	0	2	2	3	3	3	3	3	3	3
20 nosrch	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	2	0	2	3	3	3	3	3	3	3
21 nilstole	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	2	2	0	3	3	3	3	3	3	3
22 alarm	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	0	1	1	1	1	1	1
23 dog	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	0	1	1	1	1	1
24 nondet	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	1	0	1	1	1	1
25 occcar	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	1	1	0	1	1	1
26 occlight	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	1	1	1	0	1	1
27 poormain	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	1	1	1	1	0	1
28 seclight	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	1	1	1	1	1	1	0

Appendix H

Sexual Assault Behavioural Hierarchy Branch Definitions

1. *Aggression*: Overtly aggressive offence behaviours; violent or otherwise destructive or forceful actions perpetrated by the offender, often motivated by anger and hostility as opposed to being sexual in nature.

1A. *Aggression – Physical*: Aggressive behaviours can be physical, and may thus involve violent/forceful behaviours inflicted for a harmful/destructive purpose, and may include such actions as punching, kicking, or other type of physical force perpetrated by the offender.

1B. *Aggressive – Verbal*: Aggressive behaviours can be verbal. Examples may involve aggressive/hostile language, such as insults, profanity, or verbal attempts made by the offender to humiliate the victim.

2. *Theft*: Instrumental actions by the offender that involve stealing or attempts to steal from the victim.

3. *Control*: Control behaviours are intended to permit or facilitate the commission of the offence, or prevent detection of the crime by others. Control behaviours are thus often, but not exclusively, those that may require some pre-planning or preparation on behalf of the offender.

3A. *Control – Physical*: Control behaviours can be physical. Examples may be those behaviours used to immobilize the victim, any steps taken by the offender to conceal his identity, or any other physical actions used to prevent interruption of the offence during its commission or detection of the crime after the fact. Such

behaviours may also include any physical actions by the offender used to gain personal information about the victim for control-type purposes.

3B. *Control – Verbal*: Control behaviours can be verbal. Examples of such behaviours may include verbal strategies used to gain the trust of the victim or coerce them prior to the offence, or other verbal demands/instructions on behalf of the offender used for control-type purposes.

4. *Sex*: Behaviours performed or forced by the offender during the offence that are explicitly sexual in nature.

4A. *Sex – Physical*: Sexual behaviour can be physical. Examples of such behaviours include defining, core behaviours of sexual assault, or any otherwise explicitly sexual physical action.

4B. *Sex – Verbal*: Sexual behaviour can be verbal. Examples of such behaviours may include the use of sexually explicit language.

5. *Pseudo-Intimacy*: Behaviours of a more involved, intimate nature that reflect the offender's possible desire for social contact with the victim (as opposed to an exclusively sexual encounter as the motivation for the assault). Such behaviours involve the offender interacting with the victim as an individual, as opposed to strictly a sexual object. Examples include attempting to converse, relate to, or comfort the victim, or displaying other signs of affection typical of an intimate partner relationship.