1-1-2013

Go Bayes or Go Home: Algorithms for Improving Predictive Methods of Police Decision Support

Jared C. Allen
Ryerson University

GO BAYES OR GO HOME:
ALGORITHMS FOR IMPROVING PREDICTIVE METHODS OF POLICE DECISION SUPPORT

by

Jared C. Allen

Bachelor of Arts with Honours *magna cum laude* in Psychology, York University, June 2010

A thesis presented to Ryerson University

in partial fulfilment of the requirements for the degree of Master of Arts in the

Program of Psychology

Toronto, Ontario, Canada, 2013

Author's Declaration for Electronic Submission of a Thesis

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

Go Bayes or go home: Algorithms for improving predictive methods of police decision support

Jared C. Allen, Master of Arts, Psychology, 2013, Ryerson University

This thesis tests novel methods of creating advice to assist police with behavioural aspects of investigations. Using a sample of 361 serial stranger sexual offenses, simulated samples, and a sample of 84 serial burglary offences, the paper predicts behavioural characteristics using frequency information and a cross-validation approach. Experiment 1 predicts dichotomous offender characteristics from dichotomous and categorical crime scene characteristics. Experiment 2 predicts continuous behavioural variables from point estimates. Novel Bayesian algorithms are compared to base rate, mean, and point estimate prediction methods. In Experiment 1, Bayes' Theorem (74.6% accurate) predicts with 11.1% more accuracy than base rates (63.5% accurate), and provides improved advising estimates. In Experiment 2, Bayesian algorithms predict more accurately than mean and point estimate methods (this improvement is not always statistically significant). These tests suggest that Bayesian approaches increase predictive power. Advising statements are considered, and suggestions regarding future research for police decision support are discussed.

Acknowledgements

I wish to thank

Ceara Allen, for inspiring and completing me;

Dr. Alasdair Goodwill, for patience, and advocacy;

Norma Esler, for teaching me to learn and value learning;

Charles Allen, for letting questions remain questions;

Dr. Michael Luther, for opening doors and believing in me; and

Chuck and Dorothy Allen, for their love, support, and encouragement.

Dedication

This is dedicated to the late Edith Tresham. She lives in deeds today both remembered and enacted.

Table of Contents

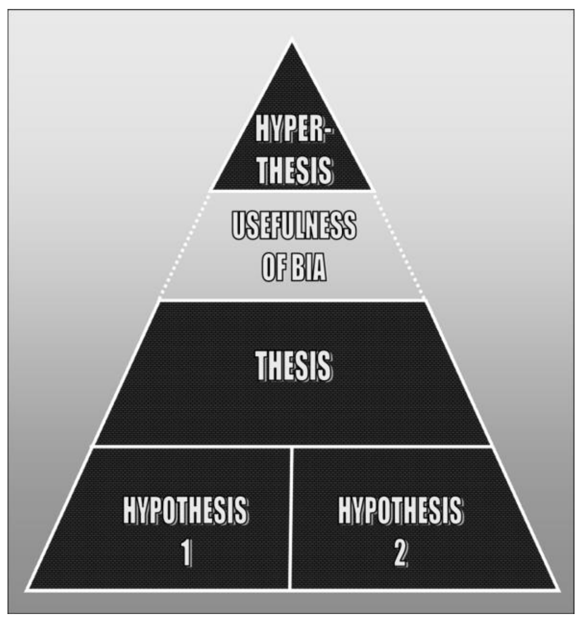
List of Tables

List of Figures

Introduction

The present research attempts to review and improve empirical methods of Behavioural Investigative Advising (BIA). The paper will 1) introduce the modest aims of the present research, 2) introduce the field of BIA, with specific attention to profiling, information use, what investigators want, and "best practices" from the published research, 3) provide a conceptual review of research and empirical methods in the field of BIA, with focus on thematic approaches and regression analysis, 4) introduce Bayes' Theorem, its use in BIA, its differences compared with Fisherian statistical methods, and its role in research related to BIA, 5) perform multiple experiments that introduce novel Bayesian algorithms to predictive BIA analysis, and 6) suggest future directions for BIA research and practice.

Figure 1: Structural diagram of thesis support.
Hyperthesis: Bayesian Behavioural Investigative Advising can help police make better decisions.
Thesis: Bayesian approaches to police decision support provide greater accuracy than estimates from more commonly used statistical approaches.
Hypotheses 1 & 2: See numbered list.
Usefulness of BIA: Whether the strength of the thesis lends empirical support to the hyperthesis depends upon the demonstrable usefulness of BIA.



Aims

This paper aims to explore and test the overall thesis that Bayesian approaches to police decision support can provide greater predictive accuracy than more commonly used statistical approaches. The primary aim is to discover whether current approaches to case decision support, which involves using frequencies and base rates from cases similar to the case under consideration, can be improved. Two hypotheses are tested to find support for the thesis being explored:

1) Using Bayes' Theorem to incorporate secondary information will result in more correct predictions of offender

behavioural characteristics than using base rates alone.

2) Using Bayes' Theorem with a point estimate and vague priors will result in more correct predictions of crime characteristics than using the point estimate alone.

If these hypotheses are correct, then the thesis will be supported (see Figure 1). The thesis is motivated by a hyperthesis, namely, that a Bayesian approach to BIA can help police make better decisions. The hyperthesis, however, requires more support independent of, and not provided in, the following investigations. That is, real improvement in police decision making will not be assessed by this paper (because predictions made in this paper were not tested for their usefulness or accuracy in real police investigations and the overall usefulness of BIA has yet to be measured). Therefore, this paper is capable of demonstrating that Bayesian methods can improve upon the accuracy of other statistical methods used in BIA, but until the usefulness of BIA is itself validated, it cannot be said that the Bayesian method is necessarily more useful. Figure 1 illustrates how, for the thesis to support the hyperthesis, the usefulness of BIA practices must be demonstrated. To use a metaphor: Consider current methods of prediction to be a saw, and Bayesian methods to be a saw with a thinner blade. The thinner blade may provide a more precise cut, but use of the tool (the act of BIA) must be accurate enough to take advantage of the improvement for it to have effect.

Since the pragmatic, real-world accuracy and impact of BIA has, at this time, yet to be empirically determined, this paper can only demonstrate improvements upon the statistical tools used in BIA, namely, the thinner blade which is a Bayesian approach to decision support. The statistical results will not indicate nor quantify the degree to which any improvement will translate to improved police investigations, even where effect sizes and accuracy rates are considerable.

There is lively debate in the field regarding the extent to which BIA is helpful to investigations (Snook et al., 2008). The problem largely lies in determining how to measure the success of BIA. It is difficult to measure, for instance, the extent to which the prediction 'the offender is likely university educated with a knife collection' will have helped an investigation, even if it turns out to be an accurate prediction.

The thesis under consideration deals with improving BIA tools, with the intention that their later use, combined with empirically validated theory, can improve the real-world efficacy of BIA.

Behavioural Investigative Advising

BIA is the process of assisting and advising law enforcement officers in the investigation of difficult to solve cases. Behavioural Investigative Advisers (BIAs) have "the single goal" of supporting and improving investigative decision making (Rainbow & Gregory, 2011, p. 18). The expertise of BIAs may lie in psychology, criminology, policing experience, or a similar field (Alison & Rainbow, 2011). In Canada, a senior peace officer should be pictured when a BIA is referenced, as nearly all professional (i.e., salaried) BIAs are police officers. This, however, is not the case in many other western countries (e.g., Germany and the UK) where it is a recognized "civilian" position.

BIAs will perform multiple tasks to assist Senior Investigating Officers (SIOs). Arguably the most important of these is suspect prioritisation, which is the process of "narrowing down" or ranking an existing pool of crime suspects so that law enforcement officers may concentrate on the most probable or most dangerous suspects (Canter & Youngs, 2009). BIAs may also conduct case linkage analysis, wherein multiple crimes are examined for connections that may indicate a similar perpetrator, or geographical profiling to help create or narrow suspect lists. The latter involves mapping multiple crimes (thought to be related) geographically and estimating, based on simple geometry and a pre-specified error rate, the likely location of the perpetrator's home or workplace (Rossmo, 1999). BIAs may also assist in formulating questions for canvassing police officers, who may attempt to contact all residents or employees within a certain radius of the crime scene or of the victim's home. Based on details of the crime, or of the crime scene, BIAs can advise the officers as to what questions to ask, what responses to watch for, and what features of the person or their home may characterize them as a suspect.

This advising may also extend to the questioning or interrogation of suspects, interacting with news media, and even the administration of psychometric or psychophysiological tests. BIA involves the utilization of expert knowledge and scientific approaches to investigation. It can be said that it "is not an established science," because there are no globally recognized standards for the field (Rainbow & Gregory, 2011, p. 20). Hence, the methods, procedures, and products of one BIA professional may differ quite widely from that of another, and the scientific and empirical standards of their employers may similarly vary by region or personal preference. The practice of BIA, especially in the United Kingdom, is utilizing an increasingly scientific approach (Alison & Rainbow, 2011). However, just as a field should not be judged by its worst contributors, it should not be represented only by its best. The field of BIA is young and still establishing professional standards (Alison & Rainbow, 2011). An important aspect of BIA is whether the predictions made by BIAs are accurate, useful, specific, and falsifiable (Alison, Smith, Eastman, & Rainbow, 2003). These are general standards that can be universally adopted, and applied to evaluation of the field.

Profiling

Part of the advising process is the creation of the most likely description of the perpetrator. Research into this process has the very difficult task of determining how to measure its efficacy. Fujita et al. (2013) have stated that none of the many studies of profiling "have empirically shown that this technique has scientific validity [and] accuracy" (p. 214). However, investigators have reported that the profiles are useful as a second opinion and as a decision support tool (Alison & Rainbow, 2011). A formal description of the offender in fact constitutes a comparatively small part of what professional BIAs do (Canter & Youngs, 2009). Often, aspects of the offence will point to suspects who may have previous convictions, motives, or opportunity. The offence type, possible case linkage, behaviours observed, and the investigative options available may indicate to knowledgeable BIAs appropriate actions or priorities that can be relayed to the supervising investigator (Alison & Rainbow, 2011). The profile itself will likely not be used as evidence, but will be "used to focus the investigative resources on the most likely suspect[s]" (Muller, 2011, p.3). The most useful details of the suspect profile are those that can be used to narrow suspect lists, to help police identify the suspect on the street, and to inform their approach when executing warrants (Alison & Rainbow, 2011). A profile is also less likely to describe the more common personality traits of the offender (e.g., neuroticism, extraversion), and more likely to stress characteristics such as age, gender, and any distinguishing behavioural features that the offender may be expected to have (e.g., tattoos, children's toys, prior offences).

The offender profiling process is often summarized as moving inferentially from the Actions of an offender (A) to the likely Characteristics of the offender (C). This is diagrammed in Figure 2. Canter (2011), the widely recognized progenitor of Investigative Psychology, notes that the recent literature has been focused on finding relationships between A and C within a dataset, but has neglected the "inference processes" being used, which prevents "reverse engineering" of results and claims (p. 6). In essence, much of the literature only allows for reasoning along the lines of "a relationship was found in this dataset, so it should exist in the world." What Canter argues for is clear theoretical explanation of these relationships, so that the reasoning behind them may be applied to future cases. He does not advocate for empirical cross-validation, which would test the generalizability and pragmatic utility of the relationships. Rather, Canter demands researchers open the "black box" of quantitative analysis and see what relationships are at work.

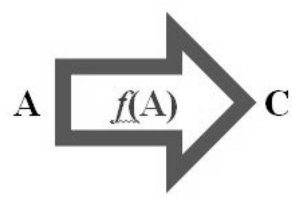


Figure 2: Profiling diagram showing some function f(A) mapping the actions (A) of an offender to characteristics (C) of that offender.

For Canter, what lies between A and C is a powerful theory (often a "theme") that can serve as a function mapping A values to C values. Given the situational nature of crime, however, even such a function mapping A to C for very specific offence types (i.e., those with more differentiability according to Canter's model) would have to be tuned to dynamic local, temporal, and even economic conditions (Canter, 2000). Abstracting a generalizable theory from an analysis of an offender sample may be less pragmatically useful than presenting a clear account of the quantitative steps used to go from A to C and their efficacy in the cases considered. That is, a theoretical explanation of the relationship within a sample may not be as useful as a predictive analysis using an objectively recorded procedure.

Information use

The BIAs that assist in crime investigations have limited information with which to work. The most valuable contributions of BIAs may come from research-based knowledge or relevant databases they possess prior to the investigation (Rainbow & Gregory, 2011). While the volume of academic research in the area of BIA is increasing exponentially, the BIA research literature is still in its adolescence with regard to development of both theory and methodology (Alison & Rainbow, 2011). The field thus requires research directed toward improvement of the scientific foundations of the profession (Hicks & Sales, 2006; Dowden, Bennell, & Bloomfield, 2007; Alison & Rainbow, 2011). For this reason, many new publications in the field simultaneously investigate variables of interest and demonstrate new research methodologies. This frequent burning of the figurative candle at both ends contributes to an increasing need for standards and methods upon which to build a cumulative system of scientific knowledge.

For instance, it still remains to be empirically demonstrated whether the homology assumption, the foundational assumption of behavioural profiling — namely, that consistent relationships exist between offence characteristics and offender characteristics — is an accurate or useful assumption. This is despite many journal articles purporting, with different sample types and methodologies, to have found support for the assumption (Rainbow, Almond, & Alison, 2011; Canter & Youngs, 2009). See Table 1 for specific examples of studies demonstrating some support for the homology assumption.

To professionalize the BIA process and its product, cumulative empirical research must be accrued that BIAs may directly and confidently reference as warranting and backing the claims and advice they put forward (Rainbow, Almond, & Alison, 2011). Due in part to an absence of such research, within the profession, well-supported and reasoned predictive statements have not been as common in the reports of BIAs as they should be (Alison et al., 2003). While the situation is measurably improving (as measured by proportions of falsifiable, specific, and supported statements), the prevalence of unfalsifiable, ambiguous, and unwarranted claims still reflects negatively upon BIAs as professionals and the burgeoning field of BIA research (Almond, Alison, & Porter, 2011). Exploratory, theoretical, actuarial, or quantitative, research with the explicit goal of creating substantive support for statements of investigative advice — and BIAs that make use of it — will be what furthers the BIA profession.

BIA and offender profiling are primarily utilized in serious violent offences (Canter & Youngs, 2009). Cases of murder and rape prompt the need to utilize all available resources to prevent future offending by the perpetrator. Similarly, serial offending, in which a single offender is believed to have committed multiple criminal acts at different times, may prompt the employment of BIAs to link crimes and anticipate likely sites of future offending. In many cases, suspect lists are generated in part by referencing databases of previously convicted criminals. BIAs can assist in prioritizing suspects through “their access to appropriate national datasets and other relevant base rate data” (Rainbow et al., 2011, p. 37). This prioritizing can be based on likely age, gender, offence history, and other probabilistic estimates derived from the crime information available. Databases in Canada (Violent Crime Linkage System, ViCLAS) and the United Kingdom (Serious Crime Analysis Section, SCAS) currently record prodigious amounts of data on violent offenders. Use of these databases to link offenses, prioritise suspects, and generate suspect lists is a great leap forward for quantitative BIA (Fujita et al., 2013). Also, given that the majority of offences are committed by a minority of repeat offenders (referred to as the 80/20 rule, whereby 20% of offenders commit 80% of crimes; Mosher, Miethe, & Hart, 2011), the optimal use of existing prior conviction information is crucial and pragmatically useful for on-going investigations. Using database information in a way that is warranted and supported by empirical research, BIAs may assist in reducing investigation time and cost. This could also save lives in the context of serious violent offence investigations. For these reasons, BIA research and practice, despite not being "established" science, are important.

What Investigators Want

SIOs require advice from BIAs that is specific and supported. Quoting one SIO from Cole and Brown (2011):

> I want someone to say to me, look this is likely to be a kid between 14 and 18 with this sort of background and it's based on the fact that we've dealt with 500 murders in the last 10 years in the database, 100 of which fit this MO and on 86 occasions it was someone who fitted this profile. (Cole & Brown, 2011, p. 198)

This SIO is essentially requesting that BIAs offer estimates of offender characteristics based on base rates of similar cases. The BIA could use case information to obtain "pared down" base rates of offender characteristics from local, state/provincial, or federal databases. The BIA would have to 1) determine which recorded cases are similar (this is the process of "paring down"; here, out of 500 murders in the hypothetical database, 100 "fit" or are similar); 2) find the average incidence of important or distinguishing offender characteristics within these cases (here, the characteristic is age); and 3) report these as likelihoods (here, the BIA could claim it is 86%

“likely” the offender is 14-18 years old). The usefulness and general propriety of this approach is apparent, if rather crude. The SIO above requests specific information with specific support. This is accomplished by referencing appropriate databases and using the data to write quantified, warranted, and backed advising statements (Rainbow, Almond, & Alison, 2011). Data collection and input with the explicit goal of creating substantive support for future investigations would improve the advising product as desired by the above SIO (Gottschalk, 2006).

Best Practices

Alison et al. (2003) have outlined the ideal structure for an advising statement: Following an argument structure put forward by Toulmin (2008), advising statements should start from the evidence (or “Data”), and make some qualified conclusion based on a clearly stated warrant for linking the data to this conclusion.

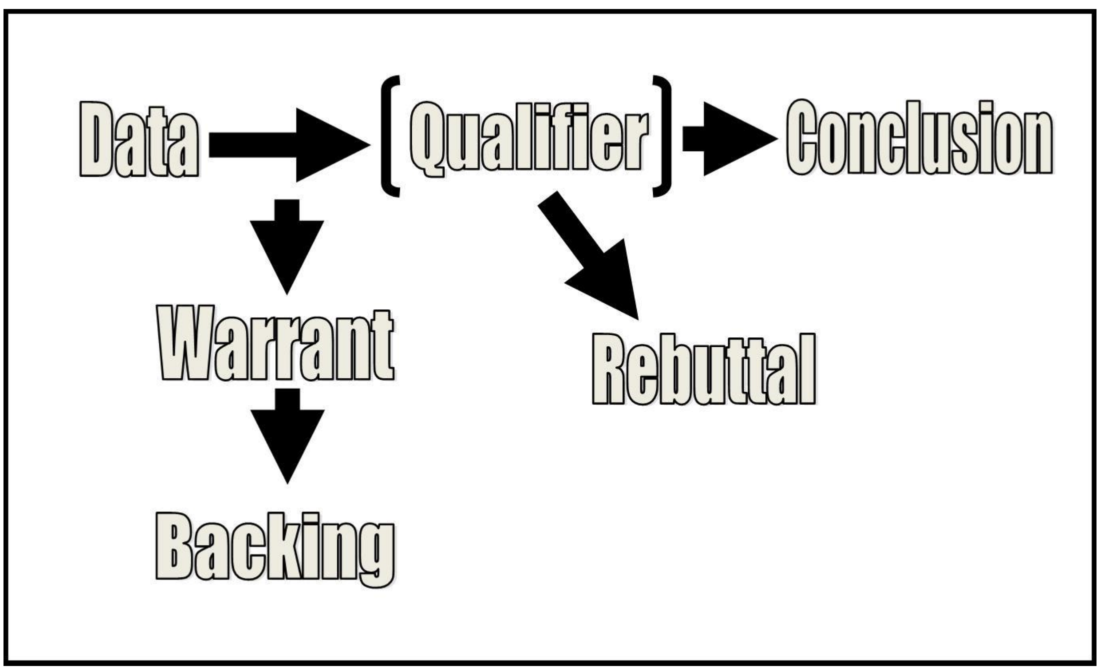


Figure 3: Stephen E. Toulmin’s (2008) proposed structure of argument.

This warrant must moreover be backed by its specific details, and the qualifier should include an explicit “Rebuttal,” meaning some statement indicating when such a conclusion would not be valid. Figure 3, reproduced from Toulmin’s (2008) *Uses of Argument*, illustrates this argumentative structure. Toulmin proposed this structure, a mirror of the form which legal arguments take, as a pragmatic circumnavigation of Hume’s (1748/1993) problem of induction (i.e., a way of “getting around” the fact that one cannot demonstrate the logical certainty of anything one would call a prediction).

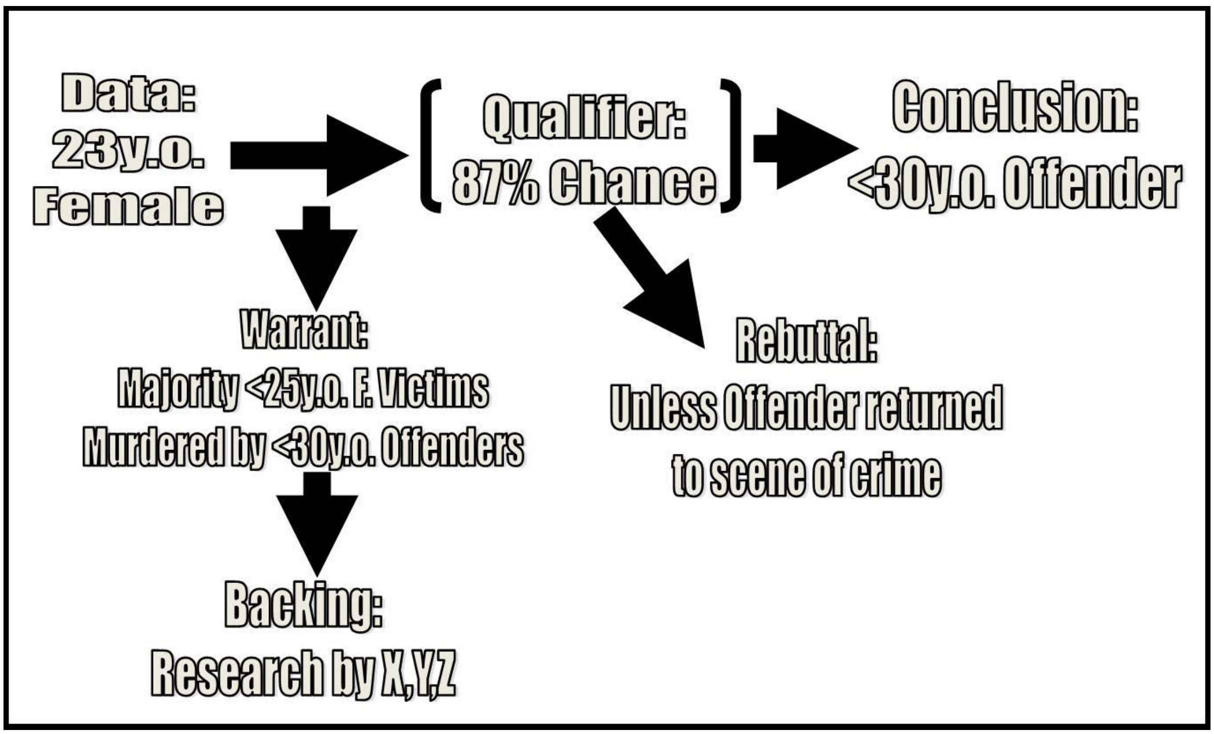


Figure 4: Example of Toulminian advising statement from Alison et al. (2003).

An example advising statement is offered by Alison et al. (2003) similar to the one discussed above. Following Toulmin’s approach, it considers a line of reasoning linking the age and gender of a murder victim to the likely age of the perpetrator. The data states the victim was a 23 year old female, and the qualified conclusion is that there is an 87% chance that the offender is less than 30 years old. The conclusion is warranted by the statement “[t]he majority of offenders who murder women under 25 yrs are themselves <30yrs old” (Alison et al., 2003, p. 175), which is itself backed by a reference to other published empirical research. The qualifier “87% chance” is rebutted by the exception that, if the offender “has returned to the crime scene,” this likelihood may change (Alison et al., 2003, p. 175). Figure 1 reproduces this example. The advice is derived from obtaining a base rate pared down by the crime type and the gender and age of the victim.

The example advice in Figure 4 takes the form desired by the SIO quoted above: specific information with specific support. The rebuttal statement is a first clue as to the conditionality of this base rate estimate. It demonstrates that other information which may be available could change the value of the qualifier, and even perhaps change the conclusion. There are many ways to statistically “model” all of the information that may have bearing on the estimate, and hence incorporate more information. However, constructing a model from a database and interpreting an estimate through the constructed model can be logically and empirically problematic (Gigerenzer, 2004; Cohen, 1994). This will be explored below where the Bayesian approach is described. While the appropriate use of base rates is the “most frequently addressed” issue for BIAs, this paper attempts to demonstrate that use of base rate information can be considerably improved (both in terms of the accuracy of predictions made and correct usage of inverse probability for inference) by a Bayesian approach (Rainbow, Almond, & Alison, 2011, p. 36).

BIA Research

Profiling did not begin as a quantitative science. Although psychiatrists had advised police in investigations prior to the 20th century, empirically based BIA is generally thought to have been conceived in the 1980s by Federal Bureau of Investigation (FBI) investigators (Canter & Youngs, 2009). The professional agents of the FBI worked as profilers on unsolved, high-profile cases, using the insight gained from their experiences to shed light on more difficult investigations. The initial forays of the profilers into research consisted of interviewing convicted offenders, gaining insight and experience with violent (often sexual) crimes, and creating subjective typologies which they believed informed each investigation (Ressler, Burgess, & Douglas, 1988). Hence, their expert recommendations were generally not based on quantitative empirical research, but rather on the interviews, experience, and common sense the profilers had accrued (Canter & Youngs, 2009). This approach leaves practitioners vulnerable to systematic errors and cognitive biases that a more objective, quantitative, empirical approach may help them to avoid.

Thematic approaches

Many of the more recent typologies created for the purpose of criminal profiling are similarly concerned with how crime scene information forms a “theme” from which investigators may predict or narrow the field of possible characteristics of the offender (see Trojan & Salfati, 2008, for a methodological review). Interpreting crime information through the lens of empirically derived themes has a large and recent literature.

Hicks and Sales (2006) claim that what they call “the Canter model” was the first, and up to 2006 the only, scientific approach to profiling (p. 71). The Canterian approach posits a hierarchy of crime and behaviour specificity and distinctiveness (Canter, 2000). Specific elements of a crime that are distinct should differentiate the offender and provide insight into the offender’s characteristics (Canter, 2011). The Canterian approach is to classify these crime elements into a theme, place the offender within that theme (preferably based on “salient” behaviours), and make inferences about the offender based on other aspects of that theme (Canter, 2000, 2011). While elegant and potentially powerful, the difficulty that accompanies this approach is the vagueness of the themes, which are central to and irremovable from this approach. Essentially, the validity, quality, and specificity of inferences made using the Canterian approach are limited by the validity, meaningfulness, and distinctiveness of the themes (Hicks & Sales, 2006). These themes are not falsifiable, not well defined, not mutually exclusive, and occasionally found to be in contradiction with each other. While the approach is elegant, the empirically derived Canterian themes are functionally and scientifically little better than the subjective typologies that preceded them.

The most popular research tool for deriving Canterian offence themes has been Multidimensional Scaling (MDS). Predominantly, the subtype of MDS called Smallest Space Analysis (SSA) has been used. This approach typically takes a large sample of solved cases and uses coefficients of relation (calculated between each variable) to rank and simultaneously display the relation of every variable to every other variable. This is plotted in Cartesian space, using the fewest dimensions in which the complex relationships can be adequately portrayed (Guttman, 1968). When SSA is used in BIA research, crime scene details that co-occur more often are plotted closer to each other (Bloombaum, 1970; see Salfati & Canter, 1999 for a typical example, and Canter, Bennell, Alison, & Reddy, 2003, for a more interesting one). The plot is then divided by the researcher based upon which regions of the plot seem to represent the different themes desired, and the individual cases are classified based on how many of an offender's behaviours fall into a given thematic region. This determines, for example, whether the aggression displayed by the offender was thematically more *expressive* or more *instrumental* (two commonly "found" themes). This popular classification method, first adapted for BIA by

Salfati and Canter (1999), takes thematic classification as its goal, with the assumption that arriving at the theme of a case will inform investigative inference. These thematic classification methods allow for theories of offender traits to be conceptualized, and have the powerful advantage of utilizing all available information (from a database and the case at hand) in a fairly assumption-free statistical procedure. Figure 5 illustrates an SSA approach to BIA analysis.

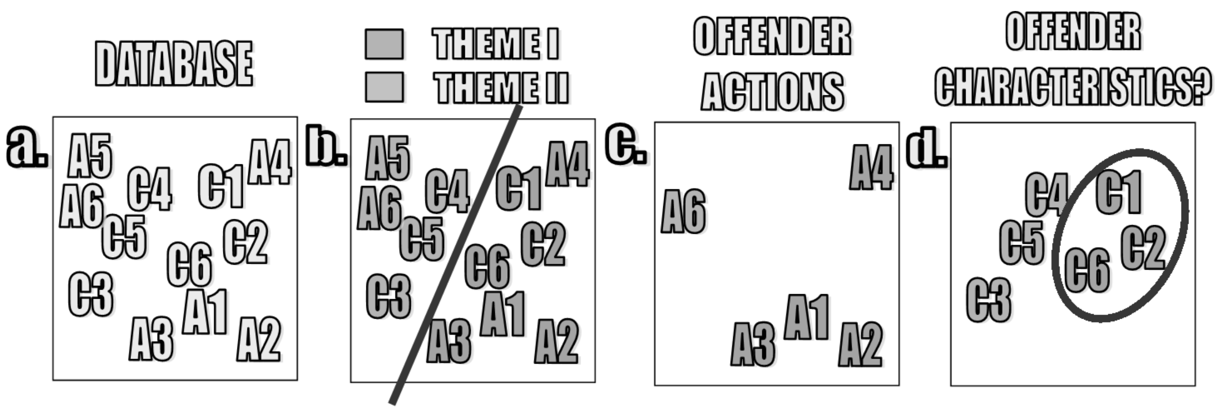


Figure 5: Process of SSA thematic profiling analysis: a) Plot all Action (A) and Characteristic (C) variables from a database of crime information such that their co-occurrence makes them appear closer in the plot, b) divide the resulting graph based on thematic differences observed in graph regions, c) classify an individual case based on whether the available Action variables are mostly of theme I or mostly of theme II, d) infer that the offender is more likely to possess Characteristics within the theme complementary to the Action theme.

While this is generally a good place to start, typically one-third of cases are not classifiable by this SSA method (Trojan & Salfati, 2008), and there is little in the literature to support the assumption that a thematic classification is useful in actual investigations. Moreover, the SSA approach is non-metric, so 1) visualization could be based on five or five thousand database cases without regard to the probabilistic implications of such different sample sizes, and 2) the magnitude of different relationships is ignored in favour of a ranked approach, which may result in ignoring or misrepresenting useful relationships within the data. As with any statistical procedure, the relationships displayed may not be representative or pertinent to the case at hand.

The use of such computational methods to derive very general themes can be very useful in initial theory construction (e.g., Towl, 2007), but may be of little pragmatic or applied use.

In producing general, thematic results, SSA research has the same limitations as that of its FBI predecessors, namely, the themes are too general, too "fuzzy," to be pragmatically useful and empirically validated. It may be believed that trained experts could use their insight into the themes to produce advice, but such an approach is what the field of BIA is attempting to move away from (Alison & Rainbow, 2011; Canter & Youngs, 2009). The goal of BIA research is to create a cumulative knowledge base of method, theory, and fact, that may be referred to specifically and explicitly to support case-specific decision advice (Rainbow & Gregory, 2011). It is doubtful that this could be accomplished using thematic SSA analysis. The desired future of BIA requires reducing subjective elements from practice and published literature, and increasing the scientific quality and content of the work. BIA did not begin as an objective process, and reaching such a point has been, necessarily, a gradual and iterative process. To this end, thematic SSA has provided a stepping stone from subjective typologies to predictive quantitative analysis.

Goodwill and colleagues (2013) have made advancements in the use of SSA-type procedures by using the spatial centre of an offender's plotted actions to obtain *x* and *y* coordinates. These are quantified average scores of an offender on whatever facets the principal axes happen to represent. This approach combines use of MDS to plot the complex multivariate thematic relationships and use of regression analysis to predict offender characteristics from these relationships. This is a departure from previous attempts to simply "classify" cases by theme. Figure 6 illustrates the approach of Goodwill et al. (2013).

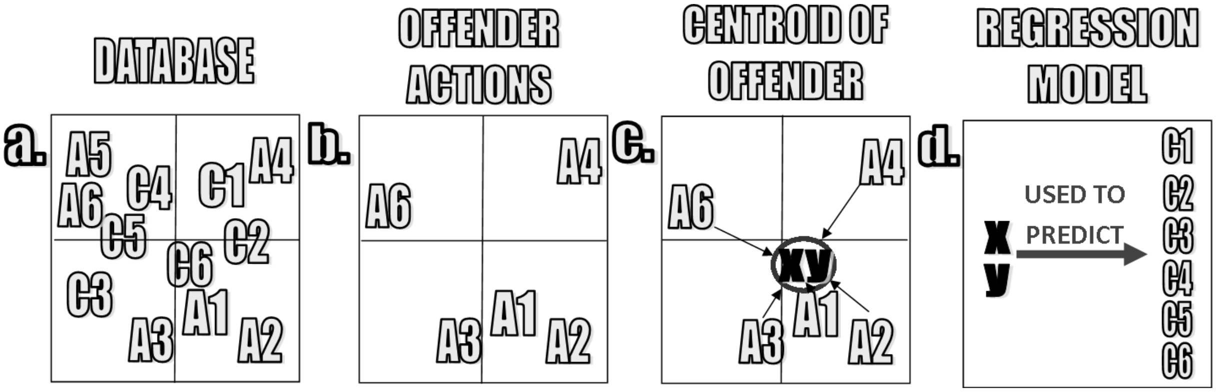


Figure 6: Quantitative advancement of SSA-type analysis by Goodwill and colleagues (2013): a) Plot all Action (A) and Characteristic (C) variables from a database of crime information such that their co-occurrence makes them appear closer in the plot, b) consider the Actions of the offender being profiled, c) obtain *x* and *y* centroid scores representing averages on the axial themes for the offender, and d) use these scores as predictors in a regression model predicting the offender's Characteristics.

The approach does away with much of the subjectivity of SSA methods, and firmly reframes the task of such analysis as one of quantification and prediction. What remains the same between SSA and Goodwill and colleagues' approach, is the assumption that meaningful themes underlie the plot produced. However, one does not need to understand or make explicit these themes in any given database for them to be predictively useful. So long as the database sample is representative of the case being considered, the axial themes can be utilized without opening Canter's (2011) "black box," and the predictive procedure can be utilized without penance for non-adherence to universal theories of criminal "types." That is, the meaningful relationships in local and crime-specific databases need not be explained by grand, vague themes, as these may be different by jurisdiction and crime type, and it is undetermined whether universal criminal themes (should they exist) could be useful in individual investigations.

In response to research on Bayesian approaches to geographical profiling, in which local data is used for local cases, Canter (2009) stated that conceptual approaches and the local Bayesian approach "enshrine different epistemologies" (p. 161), since the conceptual approach attempts to “establish general theories and principles […] independent of any given locality” (p. 161) while the Bayesian approach does not. This characterizes the considerable gap between the Canterian approach and the pragmatic Bayesian one being put forward. From a pragmatic perspective, data can be locally useful without being universally explicable, and this local usefulness should be demonstrated before Canterian (2011) theorizing of what underlies it.

The latest iteration of the thematic approach involves Latent Class Analysis (LCA), a method conceptually similar to cluster analysis and factor analysis, wherein “classes” of related or similar variables are formed using an iterative quantitative approach. As with other approaches, the researcher may determine before analysis the number of classes sought, and is free to name and describe each class. Fox and Farrington (2012) conducted LCA and used the analyses to classify types of burglars similarly to how Canterians classify by themes: The authors created four unique classes (with labels reminiscent of FBI typologies), then reported the frequencies of different types of offenders (e.g., older white males), representing the likelihood that each type belonged in each category. The advantages of an LCA approach over SSA include a) no subjective drawing of thematic lines, b) more available goodness-of-fit statistics, and c) the results of class membership are probabilistic rather than ranked, graphic, or distance-based (Vaughn et al., 2009). Hence, predictions made based on LCA thematic classification may be accompanied or qualified by probabilistic estimates of theme membership. This approach may improve upon the conceptual vagueness associated with Canterian themes, but the real-world value of class-based predictions remains to be determined. Once quantified, however, the thematic score of an offender may alternatively be used as simply a small part of a more sophisticated, explicitly predictive, analysis.

Regression analysis

Other multivariate methods, more amenable to predictive decision support, have yielded potentially useful, and potentially cross-validatable, approaches to BIA. If data is plentiful, prediction of offender characteristics can be performed using multivariate regression techniques. This may involve predicting continuous, count, categorical, or bivariate offender characteristics, from any number of predictor actions or crime scene variables.

Goodwill et al. (2009) utilized multivariate loglinear regression to predict prior offences, comparing the use of thematic variables as predictors versus using raw data as predictors (*n.b.* theory is involved in choosing the data). The multivariate approach "performed best," demonstrating that the thematic and typological approaches did not add predictive value over using the raw data (Goodwill et al., 2009, p. 523). The data outperformed the theory-rich approaches, suggesting that the "black box" function mapping offender Actions to Characteristics has, up to 2009, been sufficiently opaque, at least where predicting prior offences is concerned. This suggestion was further noted and discussed by Alison et al. (2010). Other valuable predictive studies using regression analysis have demonstrated "moderate and sufficient accuracy" in predicting offender characteristics from raw crime scene information (Fujita et al., 2013, p. 214).

Important limitations of this approach include the necessity of having very large samples to model with, and the question of content validity. The latter refers to utilizing and controlling for relevant variables, and eliminating irrelevant ones from the predictive model (e.g., Goodwill & Alison, 2007). These are important elements of regression-based BIA. Fujita and colleagues (2013) distinguish content validity from predictive validity in the context of BIA. However, it is the enhanced predictive validity of the content that must determine the content validity (e.g., Pinizzotto & Finkel, 1990). Hence, content validity need not refer to theory-based contributions or explanations of the black box content, but must, in any case, refer to improvement of the black box predictive function. Content validity is therefore measured by improvement in predictive accuracy. In most cases, this refers to predicting some cases in a database from other cases in the same database. Predicting offender characteristics in novel database cases is likely to provide a fair approximation of the external validity of a prediction method. However, only real-world prediction of on-going cases will provide honest and indubitable estimates of the utility of BIA prediction. Not only is real-world prediction likely to be accompanied by different accuracy scores, it will test whether there is any utility or usefulness in having these accurate predictions. Hence, the apparent support for the homology assumption provided by many of the most recent regression-based studies (often accompanied by area-under-curve analysis of how well the regression model performed) should be interpreted with caution.

Regression studies linking offender Actions to Characteristics have shown that, with sufficient sample size, predictive power may be acquired preceding an analysis of content validity. This initial "uninformed" baseline of predictive power is what BIA as a science must take as its initial point of reference. That is, any theoretical approach attempting to predict offender Characteristics must outperform an atheoretical predictive model given the same raw data. The degree to which the theoretical model matches or outperforms the raw data at predicting Characteristics is an objective measure of what the theory contributes to predictive BIA. Regression analysis is currently the most powerful atheoretical method for such predictive analysis, and its capacity for a priori modification of predictive models (e.g., based on validity of predictors or availability of information) makes regression analysis adaptable for use in theory testing and real investigations. Barriers to implementation of regression analysis as a standard multivariate tool for predictive BIA analyses include 1) the knowledge

Study	Sample Size	Method	Predictors	Outcome Variable	Conclusion
Fujita et al. (2013)	839	Logistic Regression	Crime scene information "police could observe objectively [at] discovery of the crime" (p. 217)	Various offender characteristics	"moderate and sufficient [predictive] accuracy" (p. 214) … "sufficient for police to prioritize lists of criminals" (p. 224)
Goodwill et al. (2013)	72	Logistic Regression	Latent scale scores versus robbery themes	Prior convictions	Score method provided "some improvement" in prediction (p. 90)
Janka et al. (2012)	682	Logistic Regression	Offending behaviour	Sexual recidivism	"characteristics of actual crime scene behavior of sexual offending have a predictive power" (p. 163)
Corovic et al. (2012)	66	Logistic Regression	Offender behaviours	Serial versus single-victim rape offender	Outcome variable predicted with 80% accuracy
Burrell et al. (2012)	166	Logistic Regression	Distance, target selection, temporal proximity, control, property stolen	Case linkage	"distance and target selection emerge as the most useful linkage factors [for robbery cases] with promising results also found for temporal proximity and control" but not property stolen (p. 201)
Goodwill et al. (2009)	85	Logistic Regression	Thematic models versus multivariate approach	Preconvictions	Multivariate approach "performed best" (p. 523)
Goodwill & Alison (2007)	85	Moderated Linear Regression	Victim age moderated by planning and aggression	Offender age	"crime scene factors can have differential moderating effects on predictive outcomes" (p. 823). Decision trees can be used with the regression equations to obtain estimates of age.

Table 1: A selection of recent Behavioural Investigative Advising studies using regression analysis, the results of which tend to support the interpretation that regression methods link offender Actions to Characteristics effectively enough to be utilized in police investigations.

and software required to compute, interpret, and adapt the results for prediction, 2) the quantity of data required to make valid predictions (which increases with the number of predictors being utilized), and 3) use of frequentist logic (further explained below) under which it is incorrect to evaluate a suspect using a model that the suspect's prior offence data may have influenced (the logic supporting prioritization would then be both circular and in violation of the Fisherian assumption of independent or "random" data).

The Bayesian Approach

Bayesian statistical inference is the use of prior and current information to infer the probability of a hypothesized cause. In other words, it is the inferential use of inverse probability, where inverse probability is defined as the use of data to obtain the probability of one or more causes producing the data (de Morgan, 1838). This is different from inferring the simple probability of said data being observed (randomly or otherwise), which is the cornerstone of the more common methods of statistical inference.

Bayes' Theorem

Bayesian inference follows the logic most investigators desire from statistical analyses (Gigerenzer, 2004). It consists of one's old information (a prior), some data (used to compute the likelihood and a normalizing constant), and one's new information (the posterior). Bayes' Theorem is most simply expressed as: The probability of a hypothesis given an observation is equal to the probability of obtaining the observation given the hypothesis is true, multiplied by the prior probability of the hypothesis, divided by the unconditional probability of obtaining the observation. This is expressed in equation 1.1.

$$1.1 \qquad P(\mathrm{H} \mid \mathrm{O}) = P(\mathrm{O} \mid \mathrm{H}) * P(\mathrm{H}) \,/\, P(\mathrm{O})$$

Each element of equation 1.1 is uniquely conceptualized and essential to the computation of inverse probability: *P*(H) is one's prior guess, estimate, or knowledge of the hypothesis or quantity being estimated; *P*(O | H) is the computation of the likelihood of obtaining the observation or data given that the hypothesis under consideration is true; *P*(O) is the probability of obtaining the observation regardless of whether the hypothesis is true, which serves as a normalizing constant to assure the estimate is bounded by a [0,1] interval; and finally *P*(H | O) is one's new ("posterior") guess, estimate, or knowledge of the hypothesis, based on the calculation performed. Figure 7 further illustrates use of these terms.

$$P(\mathrm{H} \mid \mathrm{O}) = P(\mathrm{O} \mid \mathrm{H}) * P(\mathrm{H}) \,/\, P(\mathrm{O})$$

$$\text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Unconditional Chance of Obtaining the Observation}}$$

Figure 7: Bayes' Theorem described in terms of its constituent elements.

Bayes' Theorem and BIA

The Bayesian method of estimation allows investigators to estimate the likelihood of a given hypothesis (or the likely value of some variable) from the evidence at hand (Taroni et al., 2006). It takes the form of computing the probability of the hypothesis x given data y. In the classic Bayes-Laplace equation (referred to as Bayes' Theorem), this is estimated by taking the prior probability of the hypothesis (for example, the base rate) and multiplying it by the likelihood of the data *y* given the hypothesis *x*, divided by a constant representing the probability of obtaining the data *y* (Gill, 2008). Equation 1.2 states this logical structure more formally.

1.2. $P(\text{offender is } x \mid \text{data } y) = \text{base rate offender is } x * P(\text{data } y \mid \text{offender is } x) / P(\text{data } y)$

This equation may include any number of predictor variables (data *y*). Bayes' Theorem is a rigorous method for creating probability statements, updating with new information, and including probability changes that may arise with knowledge of infrequent or non-identical (but "similar") variables (Christensen et al., 2011; Taroni et al., 2006). This sets Bayesian statistical methods apart from more common "frequentist" methods, which assess likelihood under a null hypothesis and have been designed to move inferentially from the frequency of A to a decision about A, but not from the frequency of B to a decision about A (Allen, 2013). With Bayesian results, one may take an obtained probability and use it as a prior in one's next analysis, making it ideal for applied cumulative science.

There are many formulations of Bayes' Theorem other than the one listed above (e.g., expressed for values, odds, proportions, combinations), but what remains constant throughout them is the use of prior information. When modeling complex multivariate relationships, for example, each parameter is given a prior distribution. This prior may be "informative," meaning the prior value is fairly limited and based on a good deal of information, or it may be "non-informative," meaning the prior value has a large range and is less likely to influence the model. Then, for example, Markov Chain Monte Carlo (MCMC) methods may be utilized with Gibbs sampling to iterate all model parameters over samples of all likely prior values. As the model parameters are iteratively optimized across values of the prior and posterior, the variation of the model parameters and of their conditional relationships is quantified (this is somewhat analogous to the confidence intervals of Beta parameters in regression analysis). This approach differs from Fisherian optimization under the Generalized Linear Model (GLM) because the model parameters, rather than the data, are being treated as random and iteratively solved for using data, prior information, and a simulated posterior distribution.

Bayesian versus Fisherian Approaches

The most important distinctions between Bayesian and Fisherian (also called frequentist) approaches to statistics are the use of a null hypothesis and the use of prior information. Bayesian logic involves treating data as constant and modelling one's belief about relationships in the data based on the context of the data (i.e., the prior) and the data itself, whereas Fisherian logic involves treating the data as random, ignoring the context of the information so as to be objective, and evaluating the existence of a relationship from the initial standpoint of the assumption that no relationship exists. Conceptually, Bayesian statistics attempt to discern some cause (usually represented conceptually as "theta"), while Fisherian statistics attempt to discern some effect (conceptualized as *y* values). This distinction is often a blurred one, but the difference in logic between the two approaches is stark. Table 2 details key differences between Bayesian and Fisherian approaches to statistical inference. Note, however, that some exceptions to these differences exist, especially when considering very simple applications of Bayes' Theorem and very complex applications of Fisherian statistics.

Bayesian Research in BIA

Bayesian approaches to inference are increasingly being advocated for in the areas of 1) offender profiling, 2) geographic profiling, 3) case building, and 4) risk assessment. Common reasons for such advocacy include calls for sound inferential logic, explicit and rigorous treatment of probability estimates, and the simple power of the inverse probability equation. The following paragraphs briefly review use of Bayes' Theorem in these areas in order to demonstrate the application of Bayesian reasoning to questions and problems related to BIA, and the results and responses with which these applications are met.

	Bayesian	**Fisherian**
Context	Incorporates context of data using the prior	Decontextualizes data so as to create an objective "experiment" without the inexactness of prior knowledge
Null Hypothesis	Does not assume no effect	Assumes no relationship exists in the data and evaluates chance of obtaining data under this assumption
What is Random	The parameters describing the relationships within the data are treated as random within some distribution. (e.g., in MCMC, the data is treated as constant, but the relationships taking the researcher from the data to a prediction are randomly iterated to optimize the model for each data value and determine how parameter values vary)	The data is treated as random so that the likelihood of obtaining it under the null can be assessed
Logic	Follows "inverse logic," moving from effect to estimation of cause	Follows null logic, using rejection of no effect to infer effect
Philosophy	Probability is a measure of subjective belief based on all available information	Probability is a measure of frequency based on objective (isolated) experiments
Primary Error Measure	Variation of the parameters modeling the relationship (which are treated as random)	Unexplained variance in outcome variable
Summative Statement	"Based on the information, I believe it is 95% probable that x changes y"	"It is only 5% likely that, if x did not change y, this information would be obtained, therefore x changes y"
Primary Difficulty	New information must compete with old, making the process of discovery more conservative and necessarily cumulative	The assumption of no effect is often invalid, and where it is valid, the indirect assessment of effect under the assumption of no effect is an error of formal logic
Pragmatic Difficulty for BIA	Determining priors can be subjective, and Bayesian methods are often perceived as unscientific, especially in legal circles	Does not produce estimates of the form desired (e.g., "a 77% chance"), and results logically pertain to the data itself, not to the prediction of new cases

Table 2: Differences between Bayesian and Fisherian (frequentist) approaches to statistical inference.

Offender Profiling. Offender profiling involves the simultaneous estimation of multiple dependent variables to create a "picture" of the most likely perpetrator. Bayesian Networks have been touted as the knowledge management systems of the future (e.g., Gottschalk, 2006), and an ideal method for simultaneously calculating all of the required conditional probability estimates (Baumgartner et al., 2005; 2008). With adequate data sources, a quantitative decision support tool composed of Bayesian functions could be implemented and utilized by police, with little training required. A dissertation by Zollweg (2012) reports 69.6% prediction accuracy in predicting 42 novel serial sexual offender cases from 270 others. Zollweg (2012) utilized a trained Bayesian Network model, which uses Bayes' Theorem in a complex procedure involving causal, unidirectional modelling. Previous research into the use of complex Bayesian Networks for criminal profiling (e.g., Baumgartner et al., 2005, 2008) show that Bayesian Networking can allow for powerful and flexible inference, as well as incorporation of subjective information, such as expert knowledge (Taroni et al., 2006).

The approach can either be purely empirical, where a stochastic model is "trained" on a dataset by having it causally model all variables until proportion scores relating each variable have been optimized, or a model can be crafted manually according to a guiding theory of the causal relationships between variables. Thus far, the latter has not been attempted in BIA literature. Other descriptions of this Bayesian Network approach in BIA compare it to the people-who-liked-x-also-liked-y programs on commercial websites (Canter, 2011). This comparison is apt, as adaptive Bayesian Network algorithms are behind these programs (and the most utilized website in history: Google). Due to the ability of Bayesian Networks to consider multiple dependent and independent variables simultaneously, this approach is likely to play a large role in the future of quantitative BIA.

Geographical Profiling. One of the hallmarks of a Bayesian approach to inference is sequentially updating one's estimate with new information. For this purpose, the posterior estimate from one's previous estimation becomes one's prior estimate in the next iteration of Bayesian calculations. A special issue in the *Journal of Investigative Psychology and Offender Profiling* on Bayesian "journey-to-crime modelling" exemplified this use of Bayes' Theorem to predict where an offender lives from where the offender commits crimes (Levine, 2009). Local information regarding where previous offenders had committed crimes and lived was used as the initial estimate of where an offender lived (conceptually similar to using a spatial base rate). This estimate was then updated using the information from the individual case under consideration. This represents one of the simplest applications of Bayes' Theorem. Given that no subjective priors were used, this approach is called "empirical Bayes."

Case Building. Empirical use of Bayes' Theorem can be effective both as an algorithm and as an analogue to the logical problems faced by investigators. Taroni et al. (2006) demonstrate that Bayesian analysis is well-suited for all aspects of forensic investigation, and Schneps and Colmez (2013) illustrate the grievous errors that can occur when cases are built based on a frequentist analysis of the evidence. For example, calculating a simple 1 in 6 chance of identifying an offender from a line-up versus a 1 in 12 chance may lead one to believe that having more individuals as foils in a police line-up increases the posterior probability that an accurate match was made. Wells and Turtle (1986) noted that this is not the case, and also shed empirical light, using a Bayesian updating model, on the practice of having all-suspect line-ups, which they found increases the risk of false identification. Previously police had not differentiated between single-suspect and all-suspect models, which do not have equivalent posterior risks of false identification. Employing the simple frequentist approach in this case would involve committing two mathematical errors, which Schneps and Colmez (2013) refer to as Unjustified Estimates and Choosing a Wrong Model. Errors of this “1 in x chance” frequentist variety are common in investigations, and constitute the error type most frequently addressed and faced by professional BIAs (Rainbow, Almond, & Alison, 2011).

Blair and Rossmo (2010) also discuss the problem of correctly assigning probability values for case construction. They argue that a Bayesian approach can help investigators come to a succinct answer regarding the probability that a suspect is guilty of a crime, and suggest assigning probability ranges to single or multiple pieces of evidence. They note that the latter approach does not solve the problem of assigning numerical values to pieces of evidence, but may be a step toward “more systematic assessments and improved investigative decision making” (Blair & Rossmo, 2010, p. 133). Use of Bayes’ Theorem to directly quantify the potential guilt of an offender is not advocated for in the present paper. This paper attempts to test the use of Bayes’ Theorem for producing quantified predictive estimates of likely behavioural characteristics, which may assist police in identifying, prioritizing, finding, and approaching suspects, linking or interpreting crime scenes, and maximizing the use of available information. These two uses of Bayes’ Theorem (i.e., to a: quantify guilt or b: predict offender characteristics) have the similar overarching aim of determining the most likely offender. However, one does not need to quantify an a priori likelihood of “guilt given evidence x” in order to use Bayes’ Theorem to predict characteristics. Objectively quantifying guilt must take into account the potential for erroneous evidence, coincidence, and “unknown unknowns” and any summation of such contingencies may easily mislead investigators and courts. Bayes’ Theorem can and may be used at all stages of investigation. However, the use explored in this paper is to estimate offender or crime characteristics, not to build cases by quantifying probabilities of guilt.

Risk Assessment. Risk assessment concerns itself with forecasting. The probability that a convicted offender will reoffend is forecast over a period of time by assembling relevant data and performing actuarial computations. These calculations involve using a version of Bayes' Theorem (Donaldson & Wollert, 2008). Risk assessment is typically done on convicted offenders, and BIA generally concerns itself only with pre-trial issues. However, risk assessment approaches may be employed to evaluate a suspect before a trial, and new research has been incorporating crime-scene and BIA analysis into risk assessment evaluations.

In the 20th century, insurance companies used inverse probability, contrary to a rabidly Fisherian zeitgeist, without knowing that their computations were Bayesian (McGrayne, 2011). Similarly, courts in the United States have been using Bayesian risk assessments while also lambasting Bayesian approaches (e.g., Doren, 2006). Despite this, Wollert (2007) has advocated for an even more explicitly Bayesian approach to risk assessment and to the evaluation of risk assessment methods. In one study, he found an increased bias toward confirmation of diagnostic criteria in dual-rater risk assessment systems — similar to how Wells and Turtle (1986) found increased risk of false positives in all-suspect line-ups (Wollert, 2007). His findings were criticized for their boldness and lack of "reliability" (Doren & Levenson, 2009), despite the fact that improvement of reliability and validity had been the primary concerns of his paper. In fields such as law, Bayesian analysis can have a negative reputation as a subjectivist and unscientific approach to analysis. This bias can colour interpretations of analyses as simple as empirical Bayesian inverse probability. Regardless of the reputation of Bayesian analysis, the task and field of risk assessment are both fundamentally Bayesian (Fenton & Neil, 2012).

Inverse probability is routinely used in risk assessment, insurance assessment, and biomedical science, among other reputable institutions. Hence, an individual struck by a drunk driver may have the subsequent medical treatment, insurance settlement, and trial result all partially determined by Bayes' Theorem. What these all have in common is that inference regarding risk must be made based on data and evidence. Bayes' Theorem is ideal for this task.

Preamble to Experiments

The two experiments that follow, and their sub-experiments, seek to test for improvement of prediction using Bayesian methods. The metaphysical research paradigm employed is signal detection theory. Bayesian analysis, if it is a superior approach to finding and utilizing relationships within the data, should perform the pragmatic task of predicting novel cases better when some signal is present. In some cases this signal may simply be the central tendency of the data, whereas in others (e.g., more complex models) the signal may be something made distinguishable by the more sophisticated Bayesian approach. In the case of each experiment, prediction of novel cases is the test of the efficacy of the methods. That is, predictive accuracy is the measure of signal detection.

Experiment 1

Methods (Experiment 1)

In the first experiment, offender characteristics will be predicted first using base rates, then using Bayes' Theorem. Results will be compared to ascertain any difference in efficacy of the methods. It is hypothesized that Bayes' Theorem will predict more cases correctly due to its incorporation of additional information other than the base rate. Novel cases (cases not used in computation of the base rate and the Bayesian terms) will be predicted. This "cross-validation" approach should make it less likely that the methods will perform well, and better simulate their real-world use (Cohen, 1990, 1994). It is hypothesized that the Bayesian method, being given more information (though not necessarily more useful information), will outperform the base rate in predicting variable values. The aim of the example is to demonstrate first the statistical efficacy of both methods, and second the pragmatic utility of the Bayes' Theorem method for incorporating additional information. The standard of base rate estimation serves as an ideal comparison for how well Bayes' Theorem performs, because neither method involves creating an abstract "model" from the data. Rather, both methods in this case produce a probability estimate directly from the available data. Moreover, the Bayesian results maintain structural and conceptual similarity with the base rate method, which is ideal for constructing advising statements of the form suggested by Alison et al. (2003).

Sample. From a sample of 1,000 male sexual offenders, serving sentences in a Quebec Correctional Service of Canada penitentiary, data from 69 offenders who committed serial violent sexual offences against stranger victims were selected. These offenders were responsible for 361 stranger sexual offences occurring between 1994 and 2005. Offenders committed a mean of 2.9 (SD=6.3, mode=3) violent sexual offences. Offense information was collected through semi-structured interviews (lasting between 2 and 12 hours) and police reports. In the event of discrepancy between the two sources, police report information was used. Crimes were committed against women, children, or both, and offenders were largely Caucasian (91%), with criminal records prior to the offence under study (90%). The average offender age at first sexual offence was 31 years (SD=9.4). Offenders who participated were not given compensation for their time, as per Correctional Services of Canada guidelines.

Data. Sixteen variables were chosen to be predictor variables based on the assessment that, for many violent sexual offences, the information they contain could likely be known at the time of investigation (Alison & Rainbow, 2011). These variables include the time of the crime, site of the crime, victim activity before the crime, weapon used, etc. For a complete list of predictor and estimated variables, see Table 3. Four offender characteristic variables were chosen to be the predicted (or estimated) values. These were chosen based on their potential usefulness during investigation (Mokros & Alison, 2002). These include offender looks in specific places for victims, offender lives alone, offender abuses drugs, and offender interacts verbally with officers during arrest. The first variable was selected as a differentiating characteristic that could be used to uniquely find and identify the offender (Canter, 2000). The latter three variables were selected for their potential to help inform the strategy for safely and effectively approaching the offender, something SIOs have requested from BIAs in the past (Cole and Brown, 2011).

Predictor Variables:		
Name	**Description**	**Values/Conditions**
Timec	Time of the crime	Day, night, both
Timc	Condition of the crime	Clear, dark, both
Rela	Victim/offender relationship	Stranger, acquaintance, known
Intim	Relationship detail	Stranger, seen, talked to, seen and talked to
Influ	Victim under influence of drugs/alcohol	No, Yes
Active	Victim activity before crime	At home, at work, commuting, walking/jogging, hitchhiking, other travel, visiting friend, outdoor recreation, at bar/nightclub, at other social event, prostitution
Weapo	Use of a weapon during crime	None, knife, firearm, sharpened object, rope/wire/chain, fake weapon/art-craft, other
Restr	Use of restraints during crime	No, Yes
Mutil	Victim mutilated	No, Yes
Mutip	Victim body part mutilated	None, nonsexual parts, sexual parts
Harmv	Physical harm to victim	None, physical harm, death
Conce	Offender tried to conceal identity	No, Yes
Glov	Offender used gloves	No, Yes
Face	Offender prevents his face from being seen	No, Yes
Ebef	Offender encountered victim before crime	No, Yes
Cdes	Crime site description	Residence, hotel/motel, school/educational, business/shopping site, entertainment site, red-light zone, vehicle, public transport, private yard, parking lot, street/sidewalk, alley/lane/path, highway/ditch, park, farm/field, river/lake/marsh, forest/woods, hills/mountains, desert/wasteland
Estimated Variables:		
Name	**Description**	**Values/Conditions**
Pdrug	Offender uses drugs	No, Yes
Look	Offender looks in specific places for victims	No, Yes
Livin	Offender lives alone	No, Yes
Atit	Attitude/Offender speaks during arrest	No, Yes

Table 3: Names and descriptions of variables included in analyses, Experiment 1.

Analysis. The four offender characteristics were predicted first using the base rate of occurrence of each characteristic, then using a Bayesian equation. Each prediction used crime scene characteristics from 300 cases to predict the offender characteristics of 61 other randomly selected cases. The ratio of 300 predictor cases to 61 predicted cases was chosen so as to retain adequate statistical power while making an adequate number of predictions. Random selection was performed four separate times (using Microsoft Excel), so that prediction of each variable involved different randomly selected cases (i.e., different random sets were used for each variable). Base rates for estimated variables were computed by taking the frequency of the condition (e.g., Yes, No) of the variable and dividing by the number of observations (i.e., 300).

Bayesian estimates were computed in a basic C program (Microsoft Visual C Compiler) using the equation $p(\Theta=x \mid y)=p(\Theta=x)*p(y \mid \Theta=x)/p(y)$, where y refers to all predictor variables and their conditions (e.g., timec=night, weapo=knife, …), and theta equals x ("Θ=x") refers to the variable condition being predicted (e.g., pdrug=yes). Use of multi-categorical data is where replication did require coding: to compute p(y), the proportions for each variable were multiplied. That is, if the offender used a knife, the number of incidents of knife use divided by the number of incidents total was the proportion calculated for the knife variable, and these proportions were computed for each variable and multiplied to obtain the singular p(y) value. Note that the Bayesian equation does not require independence of these proportions, only mutual exclusivity, which is satisfied. These proportions are also computed conditionally to obtain the likelihood term, meaning the same computation is made only for cases where theta equals x. If p(y) returned 0, then $p(\Theta=x)$, which is the base rate, was the $p(\Theta=x \mid y)$ estimate. This general approach is called "empirical Bayes," meaning no subjective information informs the equation (whereas other approaches may have an expert "guess" at a prior value).

All offender characteristics were predicted for novel cases, that is, cases not contained in the set of predictor data used. A total of 244 (i.e., 4*61) predictions were made with each method (488 predictions altogether). Base rate predictions were made by acquiring the base rate of the variable condition for the 300 cases and using it to predict the offender characteristic (e.g., a prior likelihood of .33 would predict, for an individual case, that the variable would not be present). A base rate over .5 would predict the characteristic to be present, while a base rate under .5 would predict it to be absent. Note that, for the base rate method, the percentage accuracy is not identical to the base rate (nor its value subtracted from 1) because the base rate was predicting novel cases. A count of the correct predictions divided by the total predictions (i.e., 61 for each variable) provides the percentage accuracy for both the base rate and Bayesian method. Percentage accuracy using Bayes' Theorem minus percentage accuracy of the base rate method provides the accuracy, over and above base rate estimation, of offender characteristic estimation using Bayes' Theorem.

Results (Experiment 1)

The overall percentage accuracy for predictions of all four offender characteristics using the base rate method was 63.5%. The overall percentage accuracy of Bayes' Theorem was 74.6%. Both methods consistently performed better than chance (i.e., for each variable the methods outperformed the null assumption that they would correctly estimate roughly half of the cases), and the Bayesian equation predicted offender characteristics in novel cases with 11.1% greater accuracy than the base rate method. In total, the Bayesian equation correctly predicted 27 cases that the base rate did not. For these cases the additional information it incorporated was clearly useful. Table 4 breaks down prediction accuracy rates of the two methods for each individual behavioural characteristic.

Offender Characteristic	%Accuracy of Base Rate	%Accuracy of Bayes	%Difference
Uses drugs	68.85	78.69	9.84
Looks in specific places for victims	75.41	90.16	14.75
Lives alone	55.74	65.57	9.83
Speaks during arrest	54.10	63.93	9.83
Mean of all four variables	*63.53*	*74.59*	*11.06*

Table 4: Results of variable predictions for Base Rate and Bayes' Theorem methods, Experiment 1.

Results suggest that, especially for the variable "offender looks in specific places for victims," the predictive accuracy of the Bayesian approach is consistently greater and much farther from the 50% mark (where the approach would be predicting no better than chance).

Experiment 2

The second experiment tests the assertion that elementary use of Bayesian priors can improve the predictive utility of point estimates used in criminal profiling. While base rates of the kind used in Experiment 1 are the most commonly utilized type of prior data, point estimates may also be used in investigative decision making. That is, police investigators may ask the investigative question of what an offender's future or past value on a given variable may be, based on a single known instance of this value. It is tested whether Bayes' Theorem can improve the predictive utility of such point estimates by incorporating some estimate of central tendency and the "spread" of possible variable values. Unlike Experiment 1, which uses empirical priors only in the form of calculated base rates, Experiment 2 utilizes an empirical base rate prior and a subjective spread prior, estimating the likely within-offender variation to be half of the between-offender variation (this ratio was chosen for simplicity). The purpose of the latter prior is to control the "pull" of the observed value to the empirical mean estimate, which is the mean of all offenders.

The experiment will consist of three Parts: a simulated experiment, an experiment using data where a strong signal or relationship is very likely, and an experiment using data where a weak signal or relationship is likely. In each of the three Parts, the Point Estimate (PE), Bayesian Estimate (BE), and Mean Estimate (ME) will be compared.

Part 1 will demonstrate using simulated data, Part 2 will apply the method to estimating victim age of serial sexual offenders, and Part 3 will apply the method to estimating number of items stolen in serial burglary cases. The application in Part 2 is chosen based on research suggesting that serial sexual offenders often specialize and prefer victims in specific age ranges (Rossmo, 2009), while the application in Part 3 is chosen based on research suggesting that the items stolen in burglaries is not an empirically useful variable for linking the offences of serial offenders (Burrell et al., 2012). Therefore, Part 2 provides an experiment in which there is likely a strong signal or relationship in the data, whereas Part 3 provides an experiment in which there is likely only a weak one. Part 1 demonstrates the ideal conditions for the BE analysis, namely, a situation where within-offender variation is less than between-offender variation, and where this is adequately captured by the Bayesian prior. The possible need for simulation of different variances will be addressed in the Discussion.

Methods (Experiment 2)

In Part 1, data points are randomly selected from a simulated distribution (simulating the set of all offenders); these points are then given additional variation (simulating within-offender variation) to imitate "obtained" values. Each data point is then used to predict its own initial value (that is, its value before additional variation). For the PE method, the varied point alone is used to predict its initial value; for the BE method, a naïve prior estimate of the (within-offender) variation is incorporated; and for the ME method, the mean from the simulated distribution of all offenders is used as the estimate. The predictive accuracy of PE, BE, and ME are compared using the loss function $|\text{actual value} - \text{estimate}|$, which takes the absolute value of the difference between the original value being predicted and the estimate (i.e., the PE, BE, or ME). This is referred to as the "loss" of the estimate. Given the conditions of the simulation, the BE should yield less estimation loss, demonstrating the advantage of including some prior estimate of additional variation, even when performing simple estimation based on point estimates.

In Parts 2 and 3, this principle is applied to serial sexual offender data and serial burglary data, respectively. An observed value is randomly selected from a distribution of serial offenders and their offences. This value, x3, comes from some offender x. The number of items stolen by offender x in offences [x1, x2, x4, …] are then each predicted. The PE method uses x3 alone; the BE method uses x3 and both empirical and estimated priors; and the ME method estimates each [x1, x2, x4 …] using the mean of the distribution of serial offenders and their offences. This is repeated for each x value for every offender, and the losses of the estimates of the three methods are compared.

It is hypothesized that the Bayesian prior will improve prediction of serial offence variables. This improvement is measured by the loss function. The intent is to test whether what is demonstrated by the experiment simulated in Part 1 applies to real prediction based on single observations in Parts 2 and 3, that is, whether Bayesian estimation incorporating priors can improve upon prediction based on point estimates and means.

Sample and data, Part 1. All data for Part 1 were obtained by simulating distributions in the software program R, and randomly sampling from these distributions.

Sample and data, Part 2. Sample data for Part 2 came from the dataset used for Experiment 1, which is described above. The variable being sampled and predicted was the age of the victim of the sexual assault at the time of the offence. The variable consists of 361 observations from offences committed by 72 serial offenders. It has a range of [4, 68], a mean of 18.7, and sd = 9.6. As will be discussed below, data was transformed to obtain normality.

Sample and data, Part 3. The sample for Part 3 was taken from a database of single-offender burglaries. All offences occurred between May 1998 and May 1999 in Huddersfield, England. The sample consists of 7 serial burglary offenders, each with five or more burglary offences, which make up the sample of 35 cases. Data was collected by the West Yorkshire Police, Huddersfield division.

Items	**Frequencies**	**Description**
Cash	19%	Cash was stolen from the premise.
Coins	16%	Coins were stolen from the premise.
Cash Instrument	8%	Cash Instruments (e.g., credit cards and banking cards) stolen from premise.
Handbag	1%	A handbag was stolen from the premise.
Gold Jewellery	21%	Gold Jewellery was stolen from the premise.
Costume Jewelley	7%	Costume Jewellery was stolen from the premise
Watches	16%	Watch(es) stolen from the premise.
Audio/Visual	65%	Audio visual equipment was stolen from the premise.
Game Console	18%	A game console (i.e. Nintendo system) was stolen from the premise.
Other Electrical	14%	Other electrical goods were stolen from the premise.
Computer	6%	A personal computer was stolen from the premise.
CD/Videos	19%	CD and videos were stolen from the premise.
Small Electronics	7%	Small electrical items were stolen from the premise.
Camera	13%	A camera was stolen from the premise.
Alcohol/Cigs	1%	Alcohol and/or Cigarettes were stolen from the premise.
Ornament	6%	One or more ornaments were stolen from the premise.
Clothing	9%	Clothing was stolen from the premise.
Carrier	12%	The offender took an item from the premise to carry other items stolen (e.g., pillow case, bag, hold-all)

Table 5: Data used for analysis in Experiment 2, Part 3. Frequencies of items stolen in the burglary sample. The sum of the items stolen has a possible range of 0-18.

Eighteen variables recording the different items stolen were selected to create the variable number of items stolen. The frequencies of items stolen appear in Table 5. It has been found that the items stolen in burglaries may not be a differentiating factor useful for linking one case to another (Burrell et al., 2012). This makes the variable an ideal example of prediction where a weak relationship may be present, allowing for a test and comparison of the PE, BE, and ME methods in this instance. The number of items stolen by an offender in a single case was used as the point estimate to predict other cases in that offender's series.

Analysis. Initial analyses are all implemented using the statistical software package R. The simulated analysis in Part 1, as well as the results of Parts 2 and 3, were followed by significance testing using the Statistical Package for the Social Sciences, version 18 (SPSS). Effect size analysis was performed manually, and data organization was generally done in Microsoft Excel.

Analysis: Part 1. In Part 1, a general simulation similar to the real analysis to be performed in Parts 2 and 3 is conducted. A simulated normal distribution with mean = 9.5, sd = 4, and values ranging from 0 to 18 is randomly sampled from 83 times (these values were chosen to resemble those in Part 3). Each sampled value ("original value") is then used as the mean in a second simulated normal distribution with sd = 2.5 (this value was chosen to model roughly half of the variation). This second distribution is then randomly sampled from. This second distribution represents the within-offender variation not accounted for by knowledge of the between-offender distribution. The point estimate used for prediction is the sampled value from the second distribution, which is used to predict the original value. This is illustrated in Figure 8.

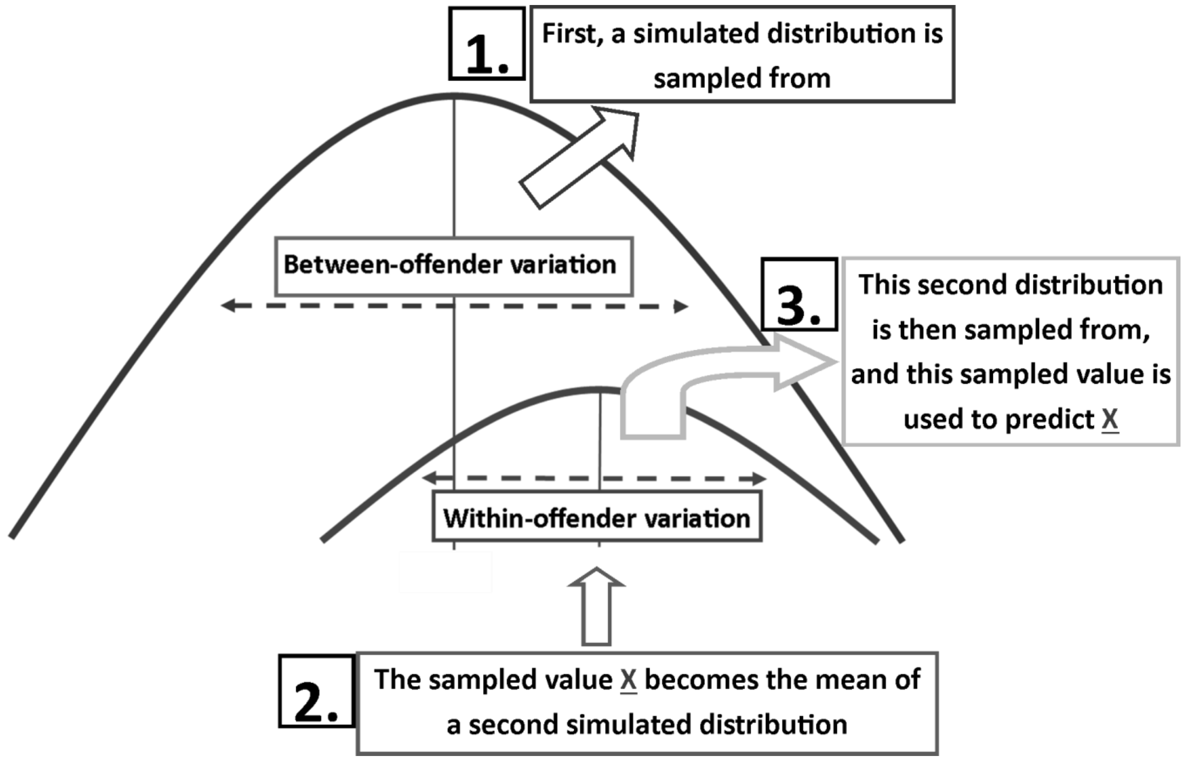


Figure 8: Diagram describing the simulation in Part 1 and conceptualizing the choice of subjective priors in Parts 2 and 3. For the prior spread parameters in Parts 2 and 3, it was estimated that victim age and the number of items stolen, respectively, are likely to vary half as much across an individual offender's offences compared to the variation of these across all offenders' offences.

The predictive accuracy of the PE, BE, and ME, are measured by the loss function. What is here being tested is whether the Bayesian approach results in smaller estimation loss. This would demonstrate the usefulness of Bayes' Theorem in accounting for estimable unknowns such as within-offender variation. The equation used to obtain the BE is given in formula 1.3.

$$1.3. \quad \mathrm{BE} = \frac{9.5*0.0625 + \text{point estimate}*0.16}{0.0625 + 0.16}$$

The arrival at the numbers used for the prior values is further explained in the context of the non-simulated analysis below.

Analysis: Part 2. In Part 2, the age of each offender's victims are predicted using a) the age of one of the offender's victims (the PE method), b) the age of one of the offender's victims and some prior estimated information (the BE method), and c) the mean age (*M* = 18.7) of all victims in the dataset (the ME method). The variable distribution showed significant positive skew (skewness / std. error of skewness = 1.122/.128 = 8.8) and leptokurtosis (kurtosis / std error of kurtosis = 2.335/.256 = 9.1). To obtain a normal distribution, the variable was log-transformed, resulting in skewness=-2.6, kurtosis=0.7 — acceptable levels for the sample size (Field, 2009).

The transformed mean is used as the empirical Bayesian prior. The spread parameter prior (which is called the "precision") is estimated by assuming that the within-offender variation would be roughly half of the between-offender variation. Variance computed after transformation was .05347. For normally distributed data, precision = 1/variance, therefore the between-offender precision is computed as 1/.05347 = 18.7. The subjective prior precision representing the within-offender spread of values of victim age, is then 1/(0.5*.05347) = 37.4. The Bayesian estimate (BE) is calculated from the point estimate using formula 1.4.

$$1.4. \quad \mathrm{BE} = \frac{\text{mean}*\text{between offender precision} + \text{point estimate}*\text{within offender precision}}{\text{between offender precision} + \text{within offender precision}}$$

Inserting the values computed above results in formula 1.5, which is used to obtain the Bayesian estimates for Part 2 with the software package R.

$$1.5. \quad \mathrm{BE} = \frac{1.209264*18.7004 + \text{point estimate}*37.40081}{18.7004 + 37.40081} - 1.209264$$

The PE, BE, and ME are each used to predict the victim's age in the offender's other offences. The absolute value of the true value minus the estimate provides the "loss" of each estimate. Loss values are averaged for each offender and paired comparisons of each offender's PE, BE, and ME losses are computed.

Analysis: Part 3. In Part 3, the number of items stolen by an offender in each of the offender's burglaries is predicted using a) the number of items stolen in one offence (the PE method), b) the number of items stolen in one offence and some prior estimated information (the BE method), and c) the mean number of items stolen in all cases for all offenders (M = 2.6) in the dataset (the ME method). The variable distribution showed significant positive skew (skewness = 3.2) for the sample size. To obtain a normal distribution, the variable was root-transformed, resulting in skewness=-1.09, kurtosis=1.11.

The transformed mean is again used as the empirical Bayesian prior. The spread parameter prior (the precision) is again estimated by assuming that the within-offender variation would be roughly half of the between-offender variation. The between-offender precision is computed as 1/.65634 = 1.5. The prior estimate of the precision of within-offender values is thus estimated at 1/(.5*.65634) = 3.0. The Bayesian estimate (BE) is calculated from the point estimate using formula 1.4. Inserting the values calculated results in formula 1.6. This is used to obtain the Bayesian estimates for Part 2, within the software package R.

1.6. $\mathrm{BE} = \frac{1.3733*1.5236 + \text{point estimate}*0.3.0472}{1.5236 + 0.3.0472}$ - 1.3733

As in Part 2, the PE, BE, and ME, are each used to predict the number of items stolen in each of the offender's offences. The absolute value of the true value minus the estimate provides the "loss" of each estimate. Loss values are averaged for each offender and paired comparisons of each offender's PE, BE, and ME losses are computed.

Results (Experiment 2)

Executive Summary. The simulated result from Part 1 suggests that the estimation loss of point-estimate-based prediction can be considerably reduced, if the data takes a given form, by using the Bayesian approach with prior values. Two examples of this reduction of estimation loss are suggested in Parts 2 and 3. Part 2 suggests that the BE serves as a signal enhancer when a strong signal is present, and Part 3 suggests that the BE can be of added value without such a strong signal. However, results do not necessarily indicate that the Bayesian algorithm has value as an applied BIA estimator, as Parts 2 and 3 demonstrate only very small improvements in predictive accuracy.

Part 1: Simulation. Results suggest significant differences in predictive accuracy between each method (all *p*s $< .01$), with the BE providing the most accurate predictions. Means of the PE ($M = 7.8$, $SD = 1.4$), BE ($M = 5.6$, $SD = 1.2$), and ME ($M = 13.3$, $SD = 2.38$) methods were compared using paired *t*-tests. The PE predicted significantly better than the ME, $t(82) = 3.956$, $p < .001$, two-tailed, Cohen's d =0.43, Pearson's $r = 0.40$. The BE predicted significantly better than the ME $t(82) = 5.414$, $p < .001$, two-tailed, Cohen's d =0.59, Pearson's $r = 0.51$, and the BE also predicted significantly better than the PE, $t(82) = 2.803$, $p = .006$, two-tailed, Cohen's d =0.31, Pearson's $r = 0.30$, with all values well below the family-wise error criterion.

As shown in Figure 9, results from Part 1 suggest the relative strength of the BE in the simulated situation. In practical terms, the BE is estimating on average 2 units closer to the correct answer than the PE and 8 units closer than the ME.

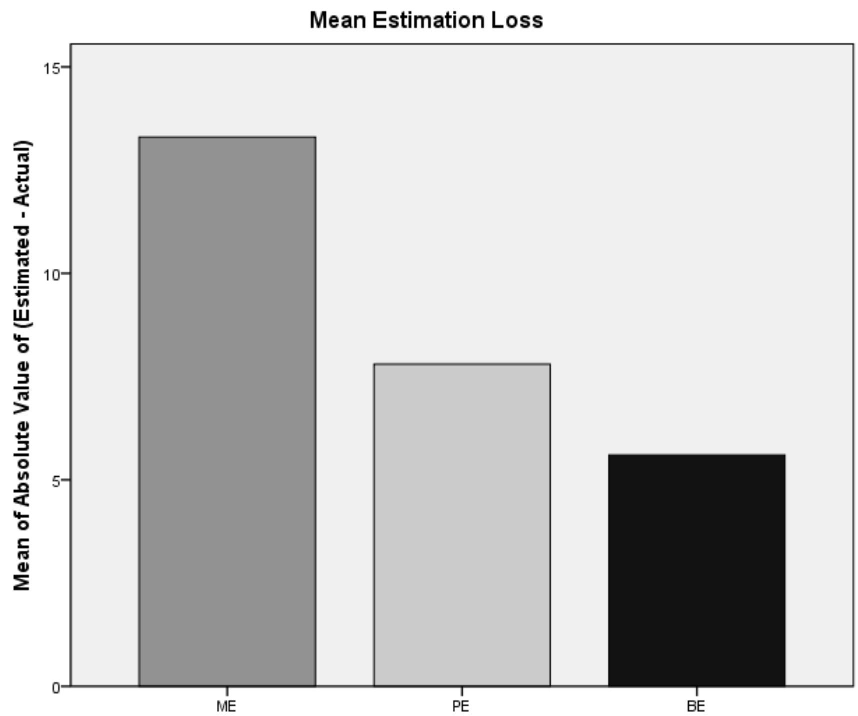


Figure 9: Results from simulation in Experiment 2 Part 1, showing reduction in estimation error. The mean estimate (ME) is outperformed by the point estimate (PE), which is in turn outperformed by the Bayesian estimate (BE), all *p*s < .01.

Part 2: Victim Age. Tests of within-subjects effects suggest a significant main effect of prediction method, $F(2, 578) = 90.258$, $p < .001$. Mean loss estimates from prediction of the transformed values were obtained for the PE ($M = 0.13$, $SD = 0.11$), BE ($M = 0.12$, $SD = 0.10$), and ME ($M = 0.19$, $SD = 0.11$) methods. The BE obtained less estimation loss than the ME, $t(360) = 12.897$, $p < .001$, two-tailed, Cohen's d =0.68, Pearson's $r = 0.32$. The PE also predicted better than the ME $t(360) = 8.873$, $p < .001$, two-tailed, Cohen's d =0.47, Pearson's $r = 0.18$. Despite a lower average estimation loss (see Figure 10), the BE did not obtain statistically less estimation loss than the PE, $t(360) = 0.647$, $p > .05$, two-tailed, Cohen's d =0.03, Pearson's r = 0.00.

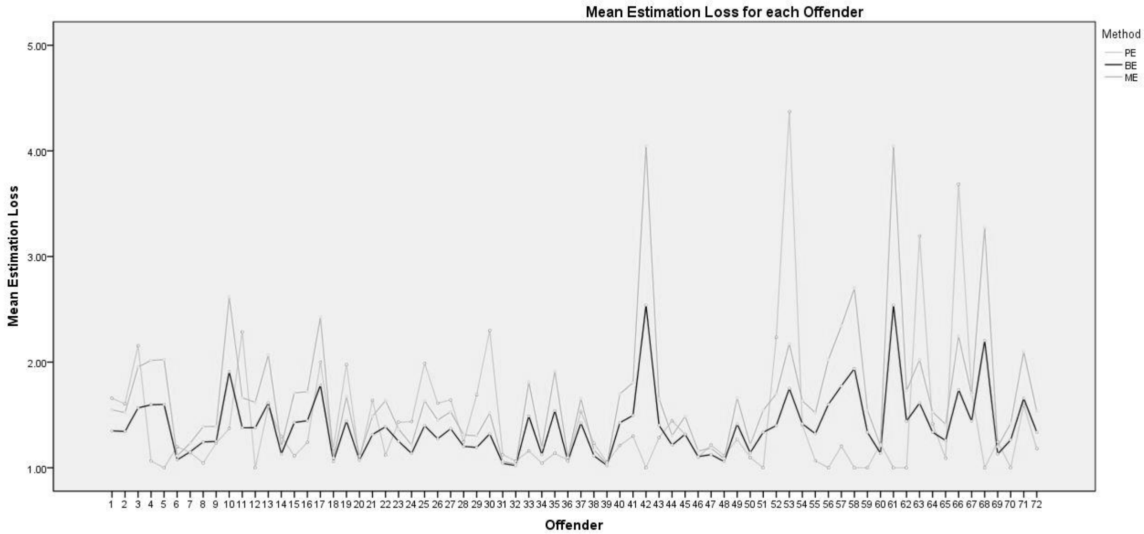


Figure 10: Mean estimation loss results by offender. The abscissa represents perfect accuracy (zero estimation loss). Note the relative consistency of the Bayes estimator (dark line) as it holds consistently closer to the axis.

In practical terms, the results indicate that the untransformed average loss values for the PE (*M* = 1.35, *SD* = 1.3), BE (*M* = 1.32, *SD* = 1.2), and ME (*M* = 1.55, *SD* = 1.3), are very similar, indicating the pragmatic difference between the estimates may be negligible. That is, all estimates tend to be incorrect in estimating victim age by (the transformed equivalent of) approximately 1.5 years on average. This small number indicates a very strong prior relationship between the age of one victim of an offender and the age of another victim of the same offender (for example, median within-offender variance is 7 years, with a mode of 0).

Part 3: Items Stolen. Results of the analysis in Part 3 suggest a within-subjects main effect of prediction method, $F(2, 166) = 98.291$, $p < .001$, partial eta squared = .542. Pairwise comparisons with Bonferroni adjustment suggest statistically significant differences between all three methods (all *p*s < .001). Means of the PE (*M* = 0.902, *SD* = 0.34), BE (*M* = 0.41, *SD* = 0.32), and ME (*M* = 0.65, *SD* = 0.18) methods were compared. Results suggest the ME predicted more accurately than the PE, $t(83) = 6.44$, $p < .001$, two-tailed, Cohen's d =0.70, Pearson's r =

0.33; the BE predicted more accurately than the ME $t(83) = 5.94$, $p < .001$, two-tailed, Cohen's d =0.65, Pearson's $r = 0.30$; and the BE predicted considerably more accurately than the PE, $t(83) = 21.33$, $p < .001$, two-tailed, Cohen's d =2.33, Pearson's $r = 0.84$. Means estimation losses by offender are displayed in Figure 11.

These results appear to suggest the relative strength of the BE over the PE and ME in situations where a weak relationship (i.e., a wide within-offender spread) is believed to exist. However, similar to the differences in Part 2, the pragmatic difference between the results may be negligible, as each of the estimation methods improves in accuracy by (the transformed equivalent of) only a fraction of an item.

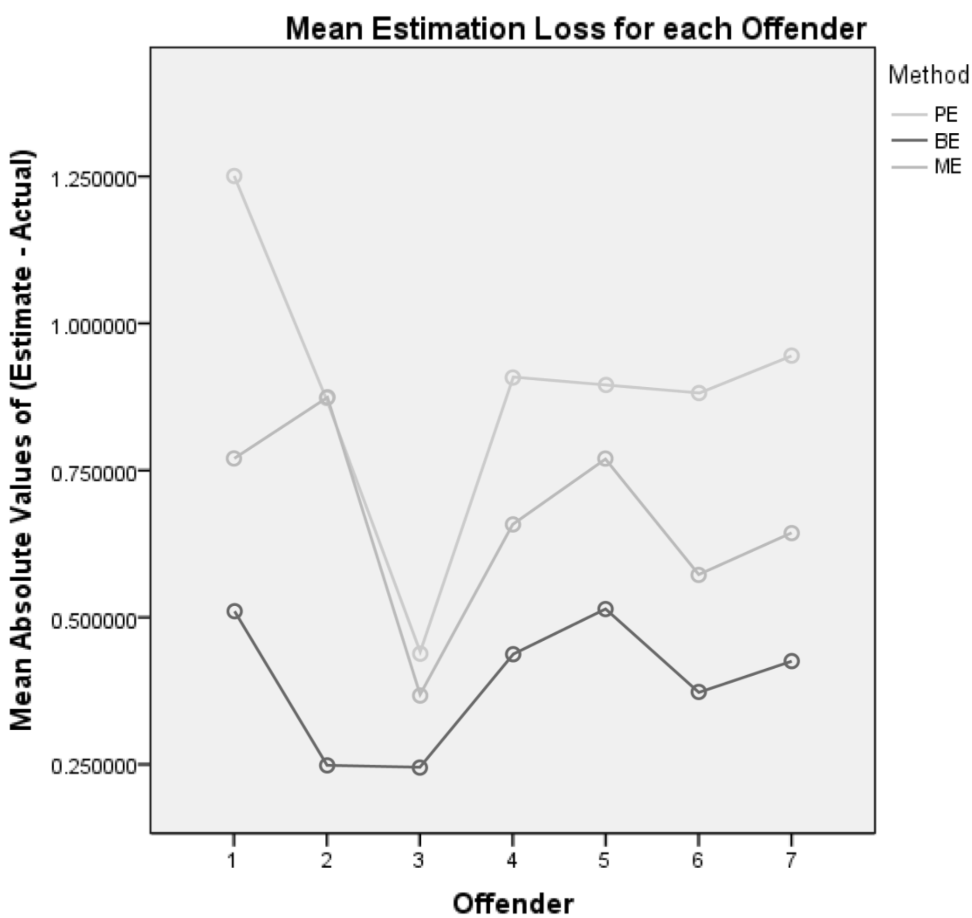


Figure 11: Mean estimation loss for PE, BE, and ME methods predicting items stolen. Note the visible

and consistent improvement of BE over the PE and ME methods for all offenders.

Discussion

The purpose of the experiments in this paper is not to suggest that the relationships explored in the sample data are generalizable. That is, the goal is not to show, as in Experiment 2 Part 2, that victim age is generally consistent across serial offences. Rather, the aim is to test the usefulness of the Bayesian tool for enhancing any signal within the dataset such that prediction of novel cases improves. The features of the samples used for the analyses, namely, the similarity of the cases in terms of offence type, the fact that the samples are generally local to a town, city, or province, and the generally small samples (especially in the case of Experiment 2 Part 3), are what make them representative of the relevant data that a local police force may have available for use. This representativeness suggests generalizability of the Bayesian predictive advantage to cases where such data is used, despite that fact that the predictive relationships within a given dataset may be different.

Signals (and noise) within any police database may differ. That is, local cases from one area may have different dependencies between variables than other localities (e.g., socioeconomic status may be related to drug use differently in different populations), different local approaches may be taken to coding data, different crime prevalence rates and motivations may be present in different areas and subpopulations, and even unique variables of interest (e.g., social group membership) may be important. While frequencies and analyses from other localities and from different crime types may provide useful estimates and priors for prediction, these will likely not be as effective as analyses of more relevant and representative data. In statistical prediction, the most predictively useful sample will generally be the most representative one.

There may exist, for example, universal Canterian crime themes, but the goal of the paper is not to find them. Rather, the goal is to test algorithms which may better find any signal present in a dataset and apply them to prediction of a case represented by that dataset, regardless of how well the signal in the dataset matches the signal hypothesized by a grand theory. The thematic approach, previously discussed, favours the application of a grand theory to the interpretation of local databases (e.g., as displayed in an SSA plot), but the signal represented by the universal theory may or may not be the predictively useful signal in the dataset considered. While such an approach can inform the use of Bayesian algorithms, it should not replace or countermand the local, representative Bayesian approach, or local applications of other less theme-dominated approaches such as the quantified centroid approach proposed by Goodwill and colleagues (2013).

The field of BIA requires a standard empirical method that a) yields specific, probabilistic results, b) is amenable to updating and change in light of new information, c) is compatible with case-specific availability of evidence, and d) outperforms mean-based (or "baseline") estimates from previous data. A Bayesian approach, rather than the typological or thematic approaches typically taken, can satisfy these requirements. Moreover, the situation presented in Experiment 1, wherein a choice is made between predicting the presence or absence of a characteristic, is an ideal application of both Bayesian theory and signal detection (or "information") theory within the field of psychology (Luce, 2003). The binary approach reduces the non-random noise created by the inter-dependencies of the psychological dependent variables, so the signal can be better differentiated from what is random in the data (Luce, 2003). In effect, the signal and the noise are both simplified. Luce (2003) argues that this simplification is necessary for the application of information theory to psychological research, whereas this paper simply acknowledges it is ideal for signal detection.

The work above has tested the use of Bayesian statistical methods to move inferentially from investigative data to quantified estimates of variables of interest, such as predicting where an offender may go to commit another crime or how the offender may behave during arrest from crime scene or witness information. Selection of this approach is pursuant of claims in recent literature that a) the future of police data management and forensic decision support lies in Bayesian Networks (BNs) and Bayes-Laplace decision processes (Gottschalk, 2006; Baumgartner et al., 2005, 2008; Taroni et al., 2006; Sullivan & Mieczkowski, 2008), and b) Bayesian methods present an advancement generally in cumulative and applied empirical research methods (Wagenmakers et al., 2011; Kruschke et al., 2012).

The type of statistical analyses conducted in this paper may potentially be misused. By using databases of convicted criminals, both the Bayesian and Fisherian statistical approaches may perpetuate biases inherent in a system of justice. That is, using the “usual suspects” to predict characteristics of offenders could lead to further focus on these individuals at the expense of other potential investigative leads. The Bayesian approach is not immune to this criticism, but it is less vulnerable to the specific claim that its inherent logic is biased to this conclusion. Frequentist approaches assume the validity of a null hypothesis, that is, it assumes the predictor and outcome variables may legitimately be thought to not be related. When this logic is used to evaluate a candidate suspect whose prior offenses are used in the model quantifying his candidacy, this assumption is grossly violated and the logic of the frequentist test is circular. That is, the offender's statistical relationship to himself is used as evidence against him because the test showed he was related to himself. In Frequentist approaches, this is a violation of the logic of the method. In Bayesian approaches this is not a logical violation (largely because the context of the information is adequately incorporated into the estimate and no null assumption is required). However, the potential for an offender's resemblance to himself to make his candidacy as a suspect more likely still remains. Ultimately, use of data from solved cases may unwittingly narrow the profiles created, limiting the types of predictions made, and in the worst case potentially perpetuating a biased system of justice wherein only the usual suspects are suspected, pursued, tried, convicted, and added to an increasingly biased database.

Signal detection is assessed by the correct prediction of novel cases. Therefore, usefulness can and must be validated through assessing predictive accuracy of the methods. There can be no guarantee that in some local cases a given method may be "interpreting" noise, but if the method is tested often enough, a measure of how frequently the method finds useful relationships can be determined. Such measures are missing from much of the BIA literature, and if BIA is to follow the field of Risk Assessment in achieving the status of a widely respected field, it must follow the lead of the latter and begin correctly making and recording falsifiable predictions.

Experiment 1 Discussion

In the first experiment, dichotomous and categorical crime scene data were used to predict dichotomous offender characteristics. The 63.5% predictive accuracy of the base rate suggests a notable benefit to creating advising statements from base rates of offender characteristics. The 74.6% accuracy obtained using Bayes' Theorem suggests that a considerable advantage may be gained by using Bayes to incorporate more information into these base rate advising estimates. Both results are encouraging given that a cross-validation method of predicting novel cases was used.

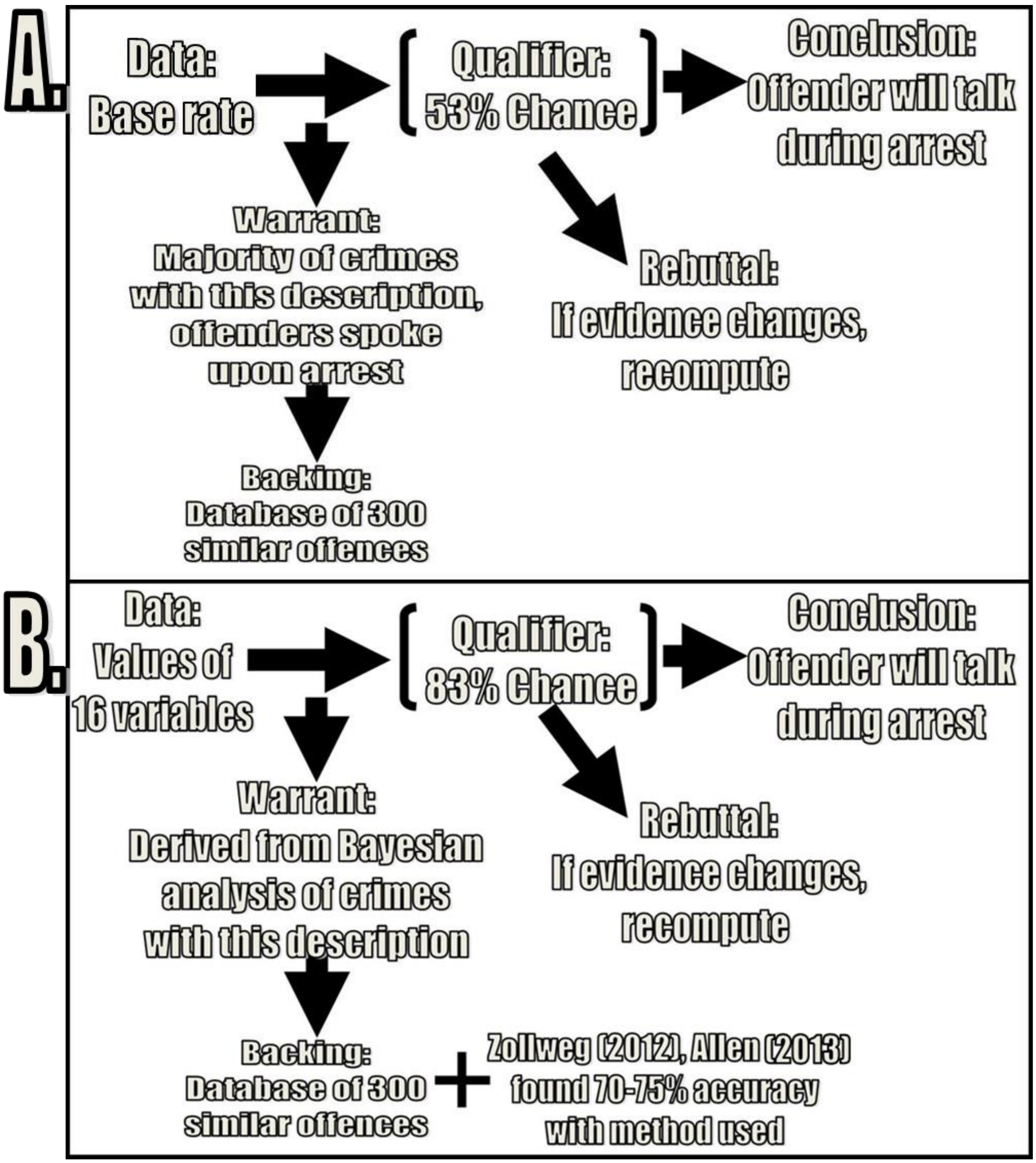


Figure 12: A closer look at case #22 (from Experiment 1) as a Toulminian advising statement. Both the base rate (A.) and the Bayesian method (B.) predict the case correctly, yet the Bayesian method produces a stronger estimate for decision support.

To demonstrate how Bayes' Theorem produces pragmatic results, in Figure 12 are two specific examples from the 488 predictions made. The results are translated into Toulminian advising statements. The figure illustrates the predictions of case number 22, which was correctly predicted by both models (note, however, that Bayes' Theorem, using more information, was able to estimate a greater probability of the correct outcome). Figure 12A illustrates the result from base rate estimation; the form that the warrant takes is similar to the example from Alison et al. (2003). Figure 12B illustrates prediction using Bayes' Theorem.

The example suggests an advantage, beyond predictive accuracy, of using Bayes' Theorem. The base rate has correctly predicted that the offender would speak during arrest, but the confidence (or, in Bayesian terms, the credibility) of the estimate is only 53%, which is not a strong figure to support a decision. The Bayesian method is here more pragmatically useful due to its higher estimate of 83%. Both methods predicted this case correctly and provided a quantified, warranted, and backed estimate, but Bayes' Theorem produced a more actionable estimate by easily incorporating more information, while maintaining the simple output structure of a single probability estimate.

The performance of Bayes' Theorem suggests one potential future of actuarial offender profiling and BIA methods. Bayesian networks have been touted as the knowledge management systems of the future (e.g., Gottschalk, 2006). With adequate data sources, a quantitative decision support tool composed of Bayesian functions could be implemented and utilized by police, with little training required. Results of the present study are somewhat similar to those obtained by Zollweg (2012), whose dissertation reports 69.6% prediction accuracy in predicting 42 novel serial offender cases from 270 others. Zollweg (2012) utilized Bayesian Networking, which similarly uses Bayes' Theorem, but is a more complex procedure involving causal modelling, not the simple Bayesian function used in the above analysis.

Previous research into the use of complex Bayesian Networks for criminal profiling (e.g., Baumgartner et al., 2005, 2008) could make the present use of the simple Bayes' Theorem algorithm appear unsophisticated or underpowered. Bayesian Networking can allow for more powerful and flexible inference than was done in the present study, as well as for incorporation of subjective information, such as expert knowledge (Taroni et al., 2006). The purpose of the study in Experiment 1, however, was to incorporate additional information without modelling a complex causal theory of behaviour, and without having to "train" a model in how to interpret the evidentiary data. The present study also utilized a sample size small enough to prevent the use of these more complex methods (Baumgartner et al., 2008). For example, Zollweg (2012) noted that his sample of 270 offences should be increased to improve validity, reliability, and predictive accuracy, perhaps decreasing the moderately high false negative (Type II) error rate of his analysis. In many real-world cases relevant data may similarly be scarce.

Limitations. The use of a dataset consisting of serial cases may be considered "cheating" due to relationships in the data being stronger than that of a non-serial sample. It is the nature of sexually violent offences, however, that they are often serial cases (Rossmo, 2009). The goal of the present study was not to generalize from this sample to the whole population, but rather to exemplify how Bayes' Theorem may be utilized in real world advising, using local, similar cases to produce estimates. Ideally, local similar data would be used to compute estimates for local similar cases, as was done here (i.e., Quebecois serial sexual offence data were used to predict offender characteristics in Quebecois serial sexual offence cases).

Conclusions. The overall performance of the base rate model is promising for BIA generally. Results of this simple approach satisfy the requirements of Toulminian argumentation and the advising standards of Alison et al., (2003). The overall predictive accuracy of 63.5% demonstrates some value in using base rates from a database sample to predict novel similar cases. This likely provides enough of an advantage to be useful, especially when the multiple predictions that would be made for a given investigation are considered in aggregate. Also, the clarity of a single likelihood estimate (a percentage value of how likely it is that an offender has a given characteristic) is not to be underestimated. One caveat of base rate prediction is that very high or very low base rates will predict with greater accuracy overall, but this does not necessarily represent the usefulness of the base rate predictor. For example, if something occurs in only 5% of cases, a prediction of non-occurrence will be correct 95% of the time. Always predicting non-occurrence, however, will not allow investigators to take advantage of the possible presence of the variable. Moreover, the 95% accuracy rate can be misleading if it is interpreted as an ability to discriminate rather than a measure of overall accuracy.

The notable advantage, over and above the base rate, conferred by incorporating more information using Bayes' Theorem, is promising for quantitative decision support. In general, estimation using Bayes' Theorem fits the Toulminian framework of argument both logically and substantively, and the 74.6% predictive accuracy of the method demonstrates BIAs and SIOs can use Bayes' Theorem with local data to predict offender characteristics in novel local cases. In this example, the simple Bayes' Theorem function efficiently produced advising statements with predictive accuracy and pragmatic utility. The two approaches maintain simplicity in their numerical results and in how they are obtained, and provide incrementally valuable predictions for investigative decision support.

Experiment 2 Discussion

In Experiment 2, point-based prediction of the continuous crime variables victim age and items stolen was conducted. The lower overall estimation loss of the Bayesian method in all three Parts is promising for the algorithm. The ideal situation for the Bayesian algorithm used, and simulated in Part 1, however, does not appear to be present in Parts 2 and 3. That is, in Part 2 the signal appears very strong, with victim age being a very consistent within-offender variable that is highly predictable by point estimate, and in Part 3 the signal is very weak, with items stolen being so inconsistent within-offenders that the mean is predicting comparably to the Bayesian method.

Figure 13 illustrates the different within-offender variances in Parts 1-3. These were all modelled with the same prior, which is most representative of the within-offender variance in the simulation in Part 1. Different priors more tailored to these variances may improve prediction.

The derivation of the prior for within-offender spread may be improved by taking into account the likely relationship within the data. That is, the prior spread for victim age, which has strong consistency, should likely be narrower than the prior spread for items stolen, which has weak consistency. In the above experiments the spread prior was estimated similarly for all three Parts (i.e., by first assuming within-offender variance to be half of between-offender variance, then computing the precision). Hence, Parts 3, 2, and 1 may show the algorithm with this prior when it is too narrow, too wide, and "just right," respectively. Initial variable selection was based in part on research by Burrell et al. (2012), which suggested the strength of target selection variables and weakness of items stolen variables for linking cases to a single offender. That article suggested variables that may fall in between in terms of relative signal strength: temporal proximity and control. The latter is a construct combining several variables related to the degree of control the offender exercises over the situation (e.g., whether the offender uses a weapon or threats). Variables such as this, which combine information somewhat subjectively based on an established construct, may be ideal for the BE approach presented. That is, the relationships that such variables represent may be ideally not-too-strong and not-too-weak, but rather optimal for the algorithm presented.

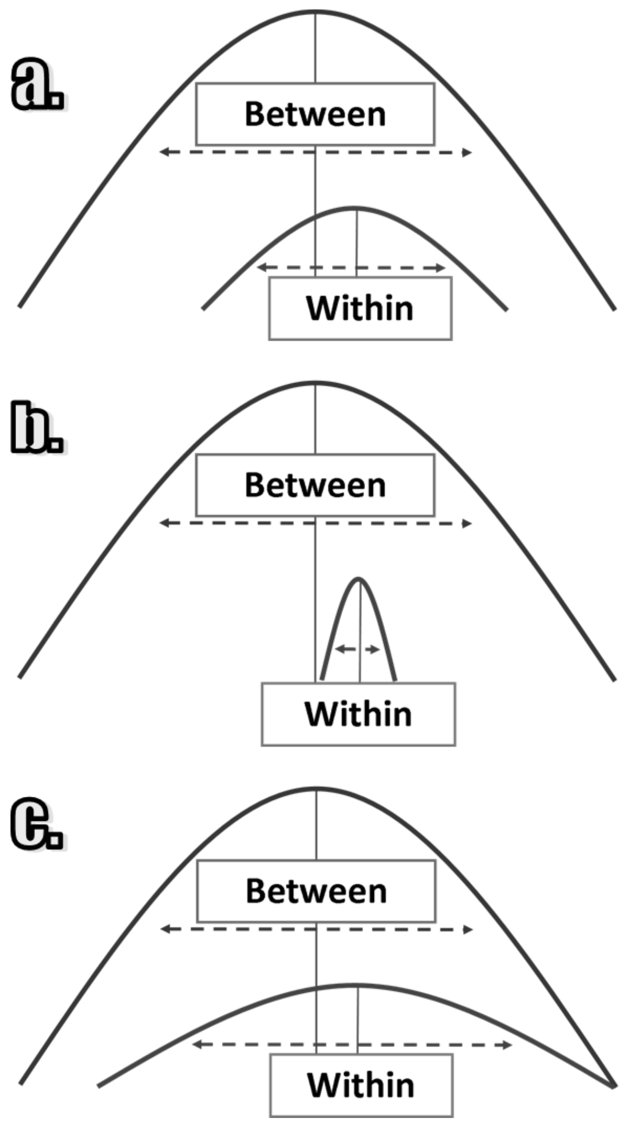


Figure 13: Visualization of within-offender variances for simulation in Part 1 (a.), victim age in Part 2 (b.), and items stolen in Part 3 (c.). Differences in variance likely account for good relative performance of the Bayesian method, point estimate, and mean estimate in parts a., b., and c., respectively. While the Bayesian method yielded better prediction in all cases, the method was not significantly better than the point estimate in Part 2.

Limitations. The aim of Experiment 2 was to create a useful algorithm (as in Experiment 1) that can be applied to police datasets. However, more validation of the predictive usefulness of the algorithm is required (e.g., predicting other variables) before suggesting any advantage of the method. Also, some investigation into the predictive usefulness of alternative priors that may be chosen based on prior knowledge of the likely strength of the signal or relationship in the data may improve use of the algorithm in cases of different signal strengths. The requirement in Parts 2 and 3 of transforming the data to obtain normal distributions is a further impediment to the use of the method, as this requires assessing, transforming, evaluating, and then back-transforming the variable of interest. The precisions calculated in the Bayesian algorithm assume a normal distribution. These can be adapted to other distributions, which tend to have more complex relationships between the variance and precision. In cases of variables with a positively skewed distribution (e.g., victim age), it is possible that this more sophisticated approach may provide better predictive power than transforming to a normal distribution. Running the experiments in Part 2 and 3 without transforming the variables, but retaining precisions which assumed normality, resulted in somewhat similar results as those with transformed variables. However, with untransformed variables the BE method did not demonstrate statistically significant predictive advantage over the PE in Part 2 and the ME in Part 3, instead yielding similar results to those estimators.

Conclusions. The experiments comprising Experiment 2 suggest that the algorithm being tested holds promise as a method of signal enhancement to improve decision support. Further study with different variables of interest, different priors, and different assumed distributions, will better assess when and where use of such an algorithm is appropriate, and which form the algorithm should take.

The pragmatic difference between the predictive accuracy of the BE and other methods does not appear to be substantial in many cases, but some predictive improvement is clearly suggested in all three Parts of Experiment 2. This warrants further investigation, and indicates the Bayesian algorithm may potentially be incrementally useful wherever values are used predictively in BIA analyses.

Future Directions

Future modifications to the specific experiments conducted have been suggested above. The future of research into Bayesian analysis for predictive BIA, however, is likely to involve more complex models than those considered above. This is in part due to the need for multivariate approaches that consider several dependent variables simultaneously and in part due to greater demand or desire in the published literature for demonstrations of complex, cutting-edge, and sophisticated statistical models.

Regression Analysis. The experiments conducted in this paper test Bayesian algorithms that may be used with datasets of police information to improve prediction. None of the experiments pit Bayesian approaches against the "gold standard" of predictive BIA: regression analysis. There are many approaches one could take to improving regression analysis with a Bayesian approach. 1) The first is to conduct a regression analysis with Bayesian parameters. That is, structure a model similar to that used in Fisherian regression analysis, but iterate all parameter values using MCMC methods with Gibb sampling to incorporate prior estimates of the distributions. This approach is Bayesian in construction but not Bayesian in use, assuming the model would be used as an equation, with evidentiary values "plugged in" to the model once parameter means have been calculated. 2) The second approach is to treat the complete model calculated in 1 as a prior value (the previous priors then become "hyperpriors"). Its credibility estimate provides a prior proportion x. A model using the case evidence can then be used as the "data" y, with a normalizing constant z to incorporate an estimate of the probability of getting the evidence if the hypothesized (or estimated) value were not observed. Equation 1.7 contains this formula. 3) A third approach is to conduct the analysis in 2 as a singular iterative model in an MCMC paradigm, optimizing parameters to obtain the estimate from the observed value.

$$1.7. \quad \mathrm{BE} = \frac{x*y}{x*y+z(1-x)}$$

The approach in 2 is likely to strike the best balance between providing incrementally useful predictive Bayesian estimates and maintaining computational parsimony. 4) A fourth approach is to set coefficients of the regression equation a priori, which has been suggested as a method for correcting "overfitting," which occurs when a regression model is so well fitted to the sample used to create it that any peculiarities in the sample reduce the predictive accuracy of the model when it is used to predict novel (non-sample) cases (Babyak, 2004).

It may be the case that regression analysis holds the most promising results in the field of BIA precisely because of the overfitting phenomenon. That is, the power of the method to model peculiarities in a sample may account for the high variance explained and the high level of predictive accuracy obtained when predicting values within the particular sample. The problem this poses is that the model and the method may be significantly less predictively useful than the results suggest. This is less likely to pose a problem when the predictions being made are of cases actually local to and represented by the data. However, this is often not how results in the literature are framed. That is, results are often framed in terms of the generalizability and usefulness of the regression model. The approach, especially when used with large datasets, may be accomplishing the finding of generalizable relationships as suggested. However, only cross-validation methods utilizing different samples will provide adequate tests of this assertion.

This cumulative approach, assimilating results from multiple samples and instances, is required when applying the signal-detection paradigm to questions of complex human psychology (Luce, 2003). That is, for more complex signals, more data is proportionally required. Incorporating the new observations from diverse samples into a single summative assessment of the method is an inherently Bayesian task (Silver, 2012). That is, predicting the efficacy of a method from an amalgamation of results of non-identical studies is moving probabilistically from an effect to a cause in the manner of Bayesian inverse probability. Therefore, assessment of the overall usefulness of predictive regression analysis for the field of BIA should be cumulatively assessed by a Bayesian analysis of the overall predictive accuracy of the method, the summation of which is continuously updated with incorporation of the results of each subsequent predictive study. This will provide a measure of how well regression analysis utilizes the predictive signals in police datasets.

Bayesian Networks. A more sophisticated Bayesian approach than the one used in the above analyses is Bayesian networking. Bayesian Networks (BN) are used to probabilistically model the relationship of every variable of interest (i.e., crime scene and offender variables) to every other variable simultaneously. Similar to the SSA approach in this regard, BN also accomplishes the simultaneous consideration of the dependent relationships of the variables metrically, and can provide estimates of any variables of interest based on values of the network nodes.

The logical structure of BN is a simple extension of the Bayesian algorithm used in Experiment 1, with the value of each node of a network determined by a unique variant of formula $p(\Theta=\mathrm{x} \mid y)=p(\Theta=\mathrm{x})*p(y \mid \Theta=\mathrm{x})/p(y)$, where *y* refers to all predictor variables and their conditions. A purely data-driven approach can be taken, yielding valuable decision thresholds for certain pre-specified cases and contingencies; and an instructed modelling approach can be taken, in which case-specific estimates are entered to yield probability estimates specific to the investigation being conducted. Data-driven Bayesian analysis has been proposed for applied BIA use (Baumgartner et al., 2005, 2008). The incorporation of subjective estimates in an instructed modelling approach has also been touted as a potentially useful BN analysis method for investigations (Taroni et al., 2006). However, in building a data-driven Bayesian model of sexual homicide, Stahlschmidt, Tausendteufel, and Härdle (2011) noted that experts often avoid detailing the "exact relationships" between variables for fear of being misleadingly exact (p. 3). That is, the officers or experts whose quantified estimates are needed for the instructed model may be unwilling or unable to quantify their belief. It is one of the strengths of the Bayesian approach that prior beliefs must be quantified, but it may not always be possible to adequately quantify one's information. This is a limitation of the all-inclusive BN approach, wherein one may be considering the whole of the evidence rather than a part. Large samples are needed to quantify reliable relationships in multivariate BN networks (Baumgartner et al., 2005, 2008; Stahlschmidt et al., 2011), making them less feasible for local use. However, this need not preclude smaller-scale use of the BN structure to estimate multiple dependent variables from several others. That is, one could conceivably model predictively useful relationships between a small number of variables with only a small database. The field of BIA will likely benefit a great deal from further investigation of predictive use of BN.

Conclusion

This paper has tested the thesis that Bayesian approaches to police decision support can provide greater accuracy than more commonly used statistical approaches. The results have suggested that the use of frequencies from similar cases can be improved by a Bayesian approach. In Experiment 1, using Bayes' Theorem to incorporate secondary information resulted in more correct predictions of offender behavioural characteristics than using base rates alone. In Experiment 2, using Bayes' Theorem with a point estimate and vague priors resulted in more correct predictions of crime characteristics than using the point estimate alone, although this difference was not statistically significant when comparing Bayesian prediction of victim age with point estimation of the same.

Both of the Bayesian algorithms are further adaptable to different investigative situations. The empirical Bayes approach taken in Experiment 1 can be replaced by an approach using subjective prior estimates (i.e., expert knowledge) or incorporating research outside of the database of information being utilized. Similarly, the heuristic prior estimation approach taken in Experiment 2 can be differently "tuned" and applied to different types of datasets and distributions. It is precisely the increase in "moving parts" in Experiment 2 that makes the algorithm require further study and more varied experimental application to ascertain its pragmatic usefulness for different distributions. Importantly, both algorithmic approaches can provide a structured approach, wherein prior knowledge is quantified and systematically used to obtain posterior estimates from the information available. Such an approach is inherently Bayesian.

The pragmatic intent of Experiments 1 and 2 has been to create simple algorithms that may be used by police and BIAs in behavioural science units to improve decision support. While the hyperthesis that the algorithms can be effective in real investigations has not been directly tested, the experiments performed support the conclusion that they improve the accuracy of prediction in novel cases. It is therefore possible that this improvement may enhance the efficacy of BIA and police decision support. Further research, testing the real-world predictive usefulness of BIA methods in the cumulative Bayesian manner suggested throughout this paper, should be conducted to establish the usefulness of BIA methods in providing decision support for police investigations.

References

Alison, L. & Canter, D. (1999a). Professional, legal, and ethical issues in offender profiling. In D. Canter & L. Alison (Eds.), *Offender profiling series: II. Profiling in policy and practice* (pp. 21-54). Aldershot, England: Ashgate.

Alison, L. & Canter, D. (1999b). Profiling in policy and practice. In D. Canter & L. Alison (Eds.), *Offender profiling series: II. Profiling in policy and practice* (pp. 1-19). Aldershot, England: Ashgate.

Alison, L., Goodwill, A., Almond, L., van den Huevel, C., & Winter, J. (2010). Pragmatic solutions to offender profiling and behavioural investigative advice. *Legal and Criminological Psychology*, *15*, 115-132.

Alison, L., & Rainbow, L. (Eds.) (2011). *Professionalizing offender profiling: Forensic and investigative psychology in practice*. London: Routledge.

Alison, L., Smith, M. D., Eastman, O., & Rainbow, L. (2003). Toulmins philosophy of argument and its relevance to offender profiling. *Psychology, Crime & Law*, *9*(2), 173-183.

Allen, J.C. (2013). Significance Testing: What It Tells Us and What It Doesn't. In M.G. Luther, J.C. Allen, J. Gerber, & J. Luther (Eds.), *Meeting the Human Challenge: Transformation, Healing, and Recognizing the Zombies Living Amongst Us, Revised Edition* (pp. 145-150). Concord, Ontario: Captus Press.

Almond, L., Alison, L., & Porter, L. (2011). An evaluation and comparison of claims made in behavioural investigative advice reports compiled by the National Policing Improvement Agency in the United Kingdom. In L. Alison & L. Rainbow (Eds.), *Professionalizing* *offender profiling: Forensic and investigative psychology in practice* (pp. 250-263). London: Routledge.

Babyak, M.A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine 66*, 411-421.

Baumgartner, K., Ferrari, S., & Salfati, C.G. (2005). Bayesian network modeling of offender behavior for criminal profiling. In *Proceedings of the 44th IEEE Conference of Decision and Control, and the European Control Conference*, Seville, Spain, pp. 2702-2709.

Baumgartner, K., Ferrari, S., & Palermo, G. (2008). Constructing bayesian networks for criminal profiling from limited data. *Knowledge-Based Systems, 21*(7), 563-572.

Bloombaum, M. (1970). Doing smallest space analysis. *Conflict resolution*, 14, 3, 409-416.

Burrell, A., Bull, R., & Bond, J. (2012). Linking personal robbery offences using offender behaviour. *Journal of Investigative Psychology and Offender Profiling, 9*(3), 201-222.

Canter, D. (1994). *Criminal shadows*. London: HarperCollins.

Canter, D. (2000). Offender profiling and criminal differentiation. *Legal and Criminological Psychology*, 5, 23-46.

Canter, D. (2009). Developments in geographical offender profiling: Commentary on bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling, 6*(3), 161-166.

Canter, D. (2011). Resolving the offender "profiling equations" and the emergence of an investigative psychology. *Current Direction in Psychological Science, 20*(1), 5-10.

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

Canter, D., & Youngs, D. (2009). *Investigative psychology: Offender profiling and the analysis of criminal action*. New York: John Wiley & Sons Ltd.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T.E. (2011). *Bayesian ideas and data analysis: An introduction for scientists and statisticians*. New York: CRC Press

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304-1312

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*(12), 997-1003

Cole, T. & Brown, J. (2011). What do Senior Investigating Police Officers want from Behavioural Investigative Advisers? In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 191-205). London: Routledge.

Corovic, J., Christianson, S. Å., & Bergman, L. R. (2012). From crime scene actions in stranger rape to prediction of rapist type: Single-victim or serial rapist? *Behavioral Sciences & the Law, 30*(6), 764-781.

Davis, J.A. (1985). *The logic of causal order*. Newbury Park, CA: Sage.

De Morgan, A. (1838). *An Essay on Probabilities and Their Application to Life Contingencies and Insurance Offices*. London: Longman, Orme, Brown, Green, & Longmans.

Donaldson, T., & Wollert, R. (2008). A mathematical proof and example that bayes's theorem is fundamental to actuarial estimates of sexual recidivism risk. *Sexual Abuse: Journal of Research and Treatment, 20*(2), 206-217.

Doren, D.M. (2006). Battling with Bayes: When statistical analyses just won't do. *Sex Offender Law Report, 7*(4), 49-50, 60-61.

Doren, D.M. & Levenson, J.S. (2009). Diagnostic reliability and sex offender civil commitment evaluations: A reply to Wollert (2007). *Sexual Offender Treatment, 4*(1).

Dowden, C., Bennell, C., & Bloomfield, S. (2007). Advances in offender profiling: A systematic review of the profiling literature published over the past three decades. *Journal of Police and Criminal Psychology, 22*, 44-56.

Fenton, N. & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian Networks*. London: CRC Press.

Field, A. (2009*). Discovering statistics using SPSS, third edition*. London: Sage.

Fox, B. H., & Farrington, D. P. (2012). Creating burglary profiles using latent class analysis: A new approach to offender profiling. *Criminal Justice and Behavior, 39*(12), 1582-1611.

Fujita, G., Watanabe, K., Yokota, K., Kuraishi, H, Suzuki, M., Wachi, T., & Otsuka, Y. (2013). Multivariate models for behavioral offender profiling of Japanese homicide. *Criminal Justice and Behavior*, *40*(2), 214-227.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*, 587-606

Gill, J. (2009). *Bayesian Methods, a social and behavioural sciences approach, second edition*. London: CRC Press.

Goodwill, A. M., & Alison, L. J. (2007). When is profiling possible? offense planning and aggression as moderators in predicting offender age from victim age in stranger rape. *Behavioral Sciences & the Law, 25*(6), 823-840.

Goodwill, A. M., Alison, L. J., & Beech, A. R. (2009). What works in offender profiling? A comparison of typological, thematic, and multivariate models. *Behavioral Sciences & the Law, 27*(4), 507-529.

Goodwill, A.M., Stephens, S., Oziel, S., Sharma, S., Allen, J.C., Bowes, N., & Lehmann, R. (2013). Advancement of Criminal Profiling Methods in Faceted Multidimensional Analysis. *Journal of Investigative Psychology and Offender Profiling, 10*(1), 71-95

Gottschalk, P. (2006). Stages of knowledge management systems in police investigations. *Knowledge-Based Systems, 19*(6), 381-387.

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

Guttman, L. (1968) A general non-metric technique for finding the smallest co-ordinate space for a configuration of points. *Psychometrika*, *33*, 469-506.

Hicks, S.J. & Sales, B.D. (2006). *Criminal profiling: Developing an effective science and practice*. Washington, DC: American Psychological Association.

Hume, D. (1748/1993). *An enquiry concerning human understanding.* Indianapolis: Hackett Publishing Company, Inc.

Janka, C., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K. (2012). The significance of offending behavior for predicting sexual recidivism among sex offenders of various age groups. *International Journal of Law and Psychiatry, 35*(3), 159-164.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722-752.

Levine, N. (2009). Introduction to the special issue on Bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling, 6*(3), 167-185.

Lindley, D.V. (1971). *Making decisions*. Toronto, ON: Wiley Interscience.

Luce, D.R. (2003). Whatever happened to information theory in psychology? *Review of General Psychology, 7*(2), 183-188.

McGrayne, S.B. (2011). *The theory that would not die: How Bayes' Rule cracked the Enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New York: Yale University Press.

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

Mosher, C.J., Miethe, T.D., & Hart, T.C. (2011). *The mismeasure of crime, second edition*. Los Angeles: Sage.

Muller, D.A. (2011). Qualitative approaches to criminal profiling as ways of reducing uncertainty in criminal investigations. *Policing*, *5*(1), 33-40.

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

Rainbow, L. & Gregory, A. (2011). What behavioural investigative advisers actually do. In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 18-34). London: Routledge.

Rainbow, L. Almond, L., & Alison, L. (2011). BIA support to investigative decision making. In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 35-50). London: Routledge.

Ressler, R. K., Burgess, A. W., & Douglas, J. E. (1988). *Sexual homicide: Patterns and motives*. Lexington, MA, England: Lexington Books/D. C. Heath and Com.

Rossmo, K. (1999). *Geographic Profiling*. New York: CRC Press.

Rossmo, D.K. (2009). Geographic profiling in serial rape investigations. In R.R. Hazelwood & A.W. Burgess (Eds.), *Practical aspects of rape investigation: A multidisciplinary approach (4th ed.)* (pp.139-170). Boca Raton: CRC Press.

Salfati, C. G., & Canter, D. V. (1999). Differentiating stranger murders: Profiling offender characteristics from behavioral styles. *Behavioral Sciences & the Law, 17*(3), 391-406

Schneps, L. & Colmez, C. (2013). *Math on trial: How numbers get used and abused in the courtroom*. New York: Basic Books.

Silver, N. (2012). *The signal and the noise; why so many predictions fail – but some don't*. New York: The Penguin Press.

Snook, B, Cullen, R.M., Bennell, C., Taylor, P.J., & Gendreau, P. (2008). The criminal profiling illusion: What's behind the smoke and mirrors?. *Criminal Justice and Behavior, 35,* 1257-1276.

Stahlschmidt, S., Tausendteufel, H., & Härdle, W.K. (2011). Bayesian networks and sex-related homicides. Discussion paper for Humboldt-University Collaborative Research Center 649: Economic Risk. Retrieved: 12-12-2011, from http://sfb649.wiwi.hu-berlin.de./papers/pdf/SFB649DP2011-045.pdf

Sullivan, C. J., & Mieczkowski, T. (2008). Bayesian analysis and the accumulation of evidence in crime and justice intervention studies. *Journal of Experimental Criminology*, *4*(4), 381-402.

Tartoni, F., Aitken, C., Garbolino, P., & Biedermann, A. (2006). *Bayesian networks and probabilistic inference in forensic science*. New York: John Wiley & Sons, Ltd.

Toulmin, S.E. (2008). *The uses of argument*, *updated edition*. New York: Cambridge University Press.

Towl, G.J. (Ed.). (2007). *Psychological research in prisons*. London: Wiley-Blackwell

Trojan, C., & Salfati, C. G. (2008). Methodological considerations of determining dominance in multidimensional analyses of crime scene behaviours and offender characteristics. *Journal of Investigative Psychology and Offender Profiling*, *5*(3), 125-145.

Vaughn, M. G., DeLisi, M., Beaver, K. M., & Howard, M. O. (2009). Multiple murder and criminal careers: A latent class analysis of multiple homicide offenders. *Forensic Science International*, 183, 67-73.

Wagenmakers, E., Wetzels, R., Borsboom, D., & van, d. M. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426-432.

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*(3), 320-329.

Wollert, R. (2007). Poor diagnostic reliability, the null-Bayes logic model, and their implications for sexually violent predator evaluations. *Psychology, Public Policy, and Law, 13*(3), 167-203.

Wortley, R. (2011). *Psychological criminology: An integrative approach.* New York: Routledge.

Zollweg, B. (2012). Using Bayesian networking to develop a profile for serial offenders, available at: http://search.proquest.com/docview/1036891399?accountid=13631 (accessed 13 March 2013).