The Precision, Accuracy, and Efficiency of Geographic Profiling Predictions: A Simple Heuristic versus Mathematical Algorithms

Craig Bennell and Karla Emeno
Department of Psychology, Carleton University

Brent Snook
Department of Psychology, Memorial University of Newfoundland

Paul J. Taylor
Department of Psychology, Lancaster University

Alasdair M. Goodwill
Department of Psychology, University of Birmingham

Keywords: Crime mapping; journey-to-crime; *CrimeStat*; bayesian; heuristics

Acknowledgement: We would like to express our thanks to Phil Canter of the Baltimore County Police Department for allowing us access to the data used in this study and to Ned Levine for his many helpful comments on earlier drafts of this article.
Correspondence can be sent to Craig Bennell, Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada, Telephone: 613-520-2600 ext. 1769, Email: cbennell@connect.carleton.ca.

Abstract

This study compared the precision, accuracy, and efficiency of geographic profiles made by students to those made by mathematical algorithms. After making predictions on 20 maps, each depicting a different offence series, nearly half of the sampled students were instructed that "the majority of offenders commit offences close to home". All of the students were then asked to make predictions on a different set of 20 maps. Seven different mathematical algorithms, several derived from a new Bayesian journey-to-crime estimation method, were also applied to the 40 maps. Results showed that informing students about the "distance decay heuristic" increased the precision of their predictions, but these predictions were not as accurate or efficient as those made by most of the algorithmic procedures. Implications of these results for the field of geographic profiling are discussed.

The purpose of this article is to compare the precision, accuracy, and efficiency of geographic profiling predictions made by students, some of whom were informed of a simple profiling heuristic, to those made using a range of mathematical algorithms, several derived from a new Bayesian journey-to-crime (JTC) estimation method (Levine, 2009). The article begins with a brief review of previous research that has compared the accuracy of human subjects to computerized profiling systems that utilize traditional JTC functions. We then argue that there is a need for research to move beyond these basic comparisons to include more sophisticated modelling approaches, such as calibrated JTC functions or Bayesian prediction models that incorporate more than just JTC distances. In the last part of the paper, the results from an empirical study that examines this issue are presented and discussed. The implications of the results for the field of geographic profiling are also discussed.

A BRIEF REVIEW OF PREVIOUS RESEARCH

Experimental research has demonstrated that individuals who are made aware of fundamental geographic profiling principles (e.g., when they are informed of the distance decay heuristic that the majority of offenders live close to their crimes) can often make predictions that are comparable to those made by computerized profiling systems (e.g., Bennell, Snook, Taylor, Corey, & Keyton, 2007; Bennell, Taylor, & Snook, 2007; Paulsen, 2006; Snook, Canter, & Bennell, 2002; Snook, Taylor, & Bennell, 2004; Taylor, Snook, & Bennell, 2009). Indeed, a recent meta-analytic review demonstrated that participants who use appropriate profiling heuristics can make predictions that are as accurate as those made by computerized profiling systems across a wide range of conditions (e.g., different participant groups, crime types, map characteristics, etc.; Bennell, Taylor, et al., 2007). The results from this body of research have both theoretical and practical implications.

Theoretically, the findings conflict with studies demonstrating the general superiority of mathematical algorithms over human judgments when attempting to maximize decision accuracy (e.g., Grove & Meehl, 1996). The predictive prowess of humans in this domain is somewhat surprising given the fact that humans are obviously more prone than algorithms are to experience limitations that negatively impact decision making (e.g., information overload, reliance on prior expectations, overconfidence, etc.; Jacob, Gaultney, & Salvendy, 1986; Kahneman, Slovic, & Tversky, 1982; Kleinmuntz, 1990). In contrast, however, such results are consistent with a growing body of research highlighting the value of simple heuristics for making both trivial and consequential decisions when those heuristics match the structure of the decision making environment (i.e., ecologically rational heuristics; Gigerenzer, Todd, & The ABC Research Group, 1999). Under these circumstances, the use of heuristics can result in levels of decision accuracy that equal, and sometimes exceed, the accuracy achieved by mathematical algorithms (e.g., Czerlinski, Gigerenzer, & Goldstein, 1999). Matching can presumably occur in the geographic profiling context when individuals use the decay heuristic because this heuristic exploits the reality that most serial offenders commit most of their crimes close to their home (e.g., Snook, 2004).

From a practical perspective, the heuristics identified through geographic profiling experiments appear to provide a cost-effective, easy-to-implement, and relatively accurate alternative to computerized profiling systems, which can potentially be used in circumstances where these systems are not viable or desired (Petrocelli, 2007). For example, simple profiling heuristics may be desirable when the police are confronted with high volume serial crime cases where geographic profiling may be useful (e.g., for determining a canvassing strategy), but the seriousness of the crimes does not warrant the involvement of a professional geographic profiler who would employ a computerized profiling system. The use of simple profiling heuristics might also be desirable in police jurisdictions that lack the financial resources or necessary infrastructure to support the use of computerized profiling systems (e.g., police forces in developing countries).

BOUNDARY CONDITIONS FOR THESE IMPLICATIONS

These theoretical and practical implications emerge directly out of previous comparisons between the predictive accuracy of human judges and computerized profiling systems (or, more precisely, comparisons between human judges and certain JTC functions incorporated into computer systems that capture the relationship between the residential location of offenders and the locations of the crimes they commit). In these studies, the focus has been on a limited number of JTC functions found within two computer systems (*Dragnet* and *CrimeStat*), each of which follows a distance decay pattern. However, these JTC functions were never calibrated to the spatial data being examined in those studies (Canter, Coffey, Huntley, & Missen, 2000; Levine, 2007). Thus, although the aforementioned theoretical and practical implications appear to hold true across various conditions, there remains the possibility that different mathematical algorithms will produce better predictions than those made when using simple heuristics. This could be the case, for example, if the algorithms were able to model offender spatial behaviour in a more accurate fashion than the default (i.e., uncalibrated) functions used in previous research.

One algorithm that may outperform human predictions is the calibrated JTC function. In contrast to functions we have examined in the past, these functions are derived empirically for a particular jurisdiction and are used to predict home locations *within* that jurisdiction (Levine, 2007). Calibrated functions should capture the unique characteristics of the sample (e.g., with respect to offence, offender, and/or environmental characteristics) upon which predictions are going to be made to a greater extent than uncalibrated functions. Theoretically, these functions should result in more accurate predictions than uncalibrated functions and human judges that do not take these characteristics into account. More recently, a range of other alternatives to calibrated functions have been proposed, many of which are based on Bayes theorem (Levine, 2009).

The Bayesian approach to geographic profiling goes beyond the use of traditional JTC functions by incorporating information about the origin (i.e., residence) and destination (i.e., crime) points of offenders who have committed crimes in the same general area as the offender who is being profiled. Specifically, the Bayesian approach relies on the link between the origin and destination points of other offenders (represented in the form of an origin-destination matrix) to refine the search for a serial offender's home location. Essentially, if data shows that previous offenders who committed assaults in area A tend to live in area B, then the Bayesian method will prioritize area B as a potential site containing the offender's residence, even if other areas are closer to area A (Block & Bernasco, 2009).

Five specific risk (i.e., probability) surfaces can be calculated using the Bayesian JTC estimation method (see Block & Bernasco, 2009; Levine, 2009, or Levine & Block, 2009 for a more detailed discussion):

1. A risk surface based on a traditional distance decay function (either a default function or a calibrated function) being applied to the locations of crimes attributed to the serial offender of interest. In the current study, this risk surface is referred to as P(dJTC) or P(cJTC) depending on whether a default or calibrated function was used.
2. A risk surface based on the general distribution of offenders' residences who have previously committed crimes in the area of interest, without considering where they committed their crimes or the locations of the crimes committed by the serial offender of interest. In the current study, this risk surface is referred to as P(O).
3. A risk surface based on the distribution of other offenders' residences given the locations of crimes committed by the serial offender of interest; or in other words, the probability that other offenders lived at a particular location when they committed crimes in the same location as the serial offender of interest. In the current study, this risk surface is referred to as P(O|JTC).
4. A risk surface based on the product of the JTC function, P(JTC), and the conditional JTC function (P(O|JTC), which takes into account both the distance decay phenomenon and the journey-to-crime histories of other offenders. In the current study, this risk surface is referred to as P(JTC)*P(O|JTC).
5. A risk surface based on the application of Bayes formula, namely P(JTC)*P(O|JTC)/P(O).

Preliminary evidence indicates that some of these new procedures (e.g., P(JTC)*P(O|JTC)) can result in better profiling predictions than traditional JTC modelling procedures (e.g., see Levine, 2009 and other contributions to that special issue).

THE CURRENT STUDY

The current experiment was designed to compare the predictive accuracy of students, some of whom were instructed to use the decay heuristic, to a range of mathematical algorithms, several of which are based on the new Bayesian prediction model. Based on previous studies, we expect that students armed with knowledge of this heuristic will make more accurate predictions than students without this knowledge. We also predict that the predictions made by these informed students will be as accurate as those produced by the algorithmic procedures. In the current experiment, accuracy refers to the degree of closeness between the predicted home location for a specific offender and their actual home location.

Given that the goal of any geographic profiling procedure is to produce predictions that are accurate, precise, *and* efficient (Levine, 2009), we extend past research by determining how the use of the decay heuristic impacts profile precision and efficiency. Precision refers to the variability in predictions made for a crime series when using a particular profiling procedure. Based on previous studies (e.g., Snook et al., 2002), we expect that informing students about the decay heuristic will significantly increase precision, but not to a point where the variability in predictions is zero (as it will be when an algorithmic procedure is applied to a crime series on separate occasions). Efficiency refers to the effort needed to be exerted by the police (e.g., the area they will need to search to find offenders' home locations) when using a particular profiling procedure. Efficiency has previously been operationalized as the percentage of a risk surface (produced using some algorithmic procedure) that must be searched before locating the offender's anchor point (e.g., Rossmo, 2000). However, because some of the profiling procedures examined in the current experiment (e.g., participant predictions) do not directly result in a risk surface, we measure efficiency as the percentage of offenders' home locations found within a certain distance from their predicted home location.[10] We expect that efficiency will significantly improve for students who are instructed to use the decay heuristic. However, due to a lack of research, we refrain from making a prediction about how the students' efficiency will compare to that of the algorithmic procedures.

METHODS

PARTICIPANTS

Fifty-seven undergraduate students participated in the study. Participants were randomly assigned to a Control ($n = 27$) or Decay ($n = 30$) condition where the difference in n across the conditions resulted from true random assignment. The average age of the students was 22.5 years ($SD = 6.2$) and there were 26 men and 31 women. None of the students had ever been employed as police officers or produced a geographic profile.

MATERIALS

A set of 40 maps, each depicting the crime locations of a different offence series, were selected randomly from a larger sample of 88 serial offenders that was made available for research purposes by the Baltimore County Police Department. The crimes were committed in the mid-1990s in Baltimore County, MD and the number of incidents committed by each of the 40 offenders ranged from 3 to 24 crimes and included a range of different crime types (e.g., larceny, burglary, robbery, etc.).[11] For each offender, the data consisted of geo-coded x-y coordinates indicating the location of the offender's residence at the time the crimes were committed and x-y coordinates for each of the offender's crime locations. There was no identifying information included in the data set.

The 40 maps were scaled from actual maps to fit onto a regular sheet of paper (map size = 190 mm x 253 mm). To remain consistent with the information used by algorithmic procedures, the maps were presented to students in black and white and without any topographical features (e.g., road systems, land use indicators, city

---

[10] Ned Levine (personal communication, August 26, 2008) has also suggested that it will be important in the future to examine the workload efficiency of the different profiling procedures (i.e., the amount of time/energy that needs to be invested by a police force in order to put together a geographic profile using the various procedures). We agree with this suggestion and look forward to future research that examines this issue.

[11] Although geographic profiling research to date has tended to focus on crime series of a single crime type (e.g., burglary), Leitner and Kent (2009) indicate that profile accuracy and precision actually increase when focusing on series of crimes including multiple crime types.

boundaries). The 40 maps were split randomly into two groups and integrated into an experimental booklet that contained, in order, and on separate pages: (a) an informed consent form, (b) instructions to indicate, by marking an 'X' on each of the first set of 20 maps, a place where they thought the offender's home was most likely to be located, (c) the first set of 20 maps, (d) a page consisting of instructions to either take a break before moving on to the second set of maps (Control group) or to use the decay heuristic to make predictions on the second set of maps, which read, "The majority of offenders commit offences close to home" (Decay group), (e) the second set of 20 maps, and (f) a debriefing form. For approximately half the students, the order of the two sets of 20 maps was reversed to check for order effects (none were found).

PROCEDURE

Students were presented with the experimental booklet and asked to read the instruction sheet. They were given the opportunity to ask questions and were reminded that they were free to leave the experiment at any time. On confirming that they understood the experimental task, the students were asked to work through the booklet at their own pace. Once they had made predictions on all 40 maps, they were thanked for their participation and debriefed as to the purpose of the study. The students' predictions, precision, accuracy, and efficiency were then measured by hand using the hard copies of their maps. See Figure 1a for an example of how the maps appeared once students had made their predictions (the home location was obviously not indicated on the map during the profiling task).

To derive the algorithmic predictions, we used procedures available in *CrimeStat* (v. 3.1; Levine, 2007). Specifically, we tested seven different algorithms on the 40 maps. They were:

1. The centre of minimum distance, CMD, which is "the location from which the summed distances to all crime locations is minimal" (Block & Bernasco, 2009, p. 191).
2. A default JTC function, P(dJTC), which took the form, $f(d_{ij}) = a*e^{-c*d_{ij}}$, with constants set at $a = 1$ and $c = 1$ to be consistent with our previous research (Snook et al., 2004).
3. A calibrated JTC function, P(cJTC), which took the form, $f(d_{ij}) = a*e^{-c*d_{ij}}$, with constants set at $a = 5.58$ and $c = 0.23$.
4. The general probability function, P(O).
5. The conditional JTC function, P(O|cJTC).
6. The product of the calibrated JTC function and the conditional JTC function, P(cJTC)*P(O|cJTC).
7. The Bayesian risk estimate, P(cJTC)*P(O|cJTC)/P(O).

For the last five algorithmic procedures in this list, we utilized a calibrated JTC function and origin-destination matrix provided by Ned Levine. Both the function and the matrix were derived from a much larger sample of serial and non-serial offenders from the Baltimore County area (consisting of approximately 42,000 cases). For all procedures, the *x-y* coordinates for each of the offender's crimes were submitted to *CrimeStat* as a primary file. The Bayesian JTC routine was then used to interpolate the input data into a likelihood estimate of offender home location for each grid cell. Each of the algorithmic procedures was applied to each of the 40 offence series. For the algorithmic predictions, measures of precision, accuracy, and efficiency were generated automatically using data from *CrimeStat*.

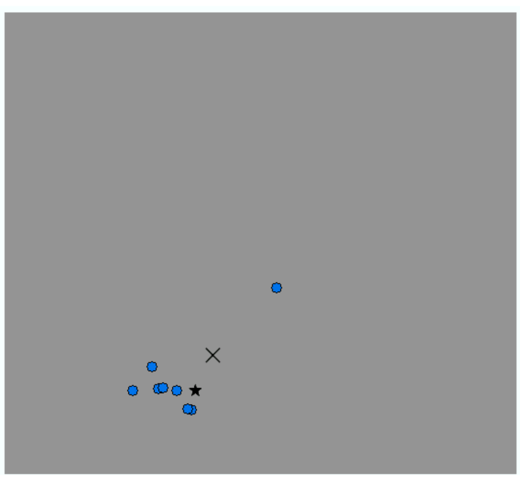
MEASURING PERCISION, ACCURACY, AND EFFICIENCY

For student predictions, precision was measured by calculating the standard deviation of the error distances associated with each map; the lower the standard deviation, the more precise the predictions (i.e., the higher the level of agreement among participants with respect to their predictions). Accuracy was measured in miles as the straight-line distance between the predicted and actual home location (i.e., error distance); the smaller the error distance, the more accurate the predictions. Efficiency was examined by calculating the percentage of offenders living within several specified distances (< 1 mile, < 3 miles, and < 5 miles) from the predicted home location (i.e., the location where the students marked an 'X', or the cell with the highest probability for the algorithmic procedures); the higher the percentage of offenders found within a particular distance threshold, the more efficient the profiling procedure.

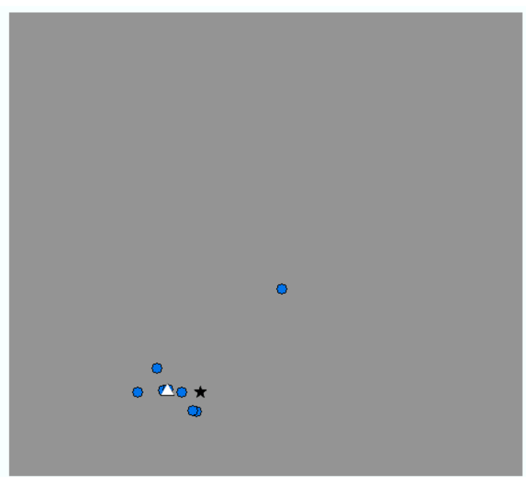
RESULTS

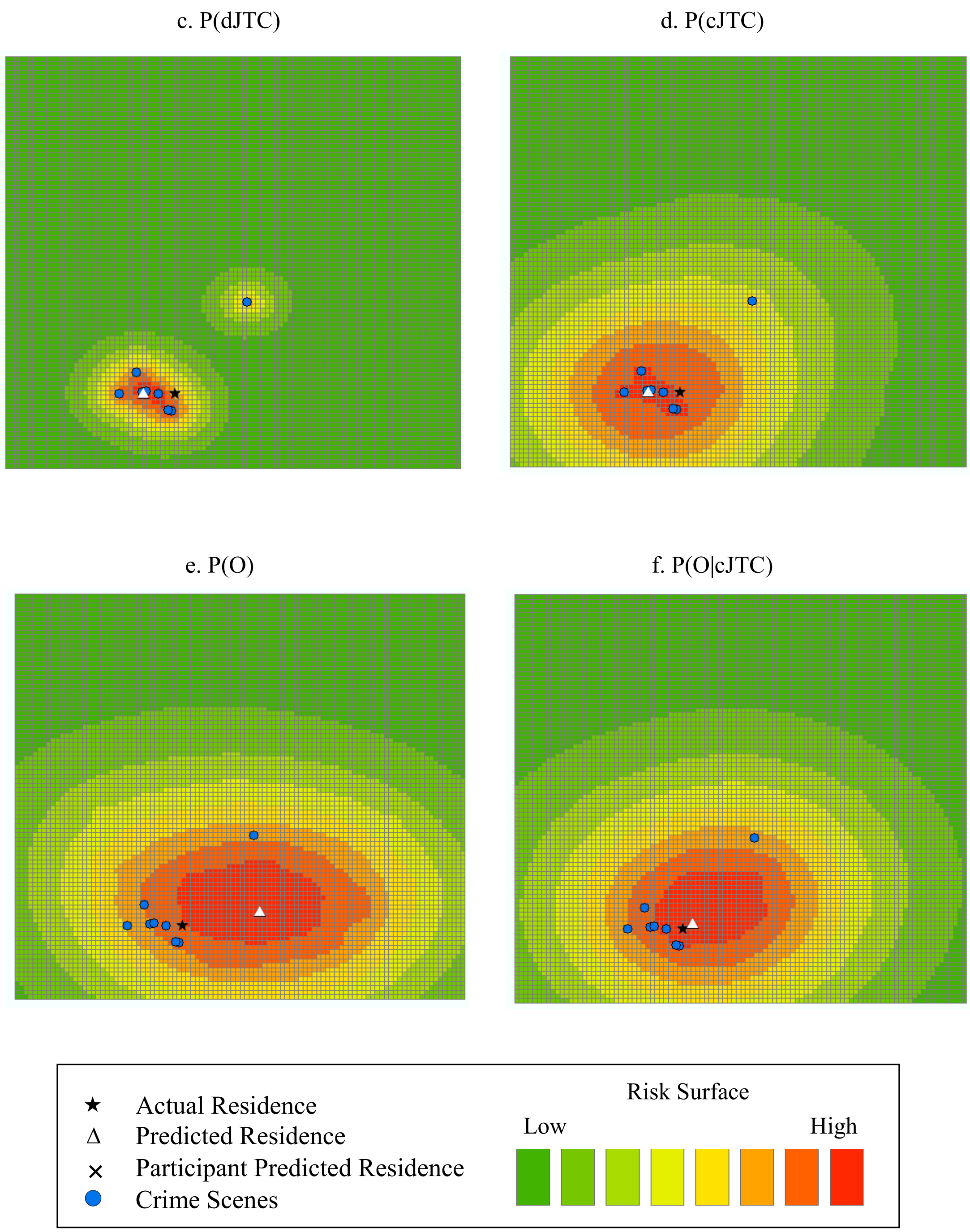
For illustrative purposes, an example output from each of the profiling procedures is presented in Figure 1 (a-h) for one specific offender included in the sample.

a. Participant



b. CMD




c. P(dJTC)
d. P(cJTC)
e. P(O)
f. P(O|cJTC)
★ Actual Residence
Δ Predicted Residence
× Participant Predicted Residence
● Crime Scenes
Risk Surface
Low
High

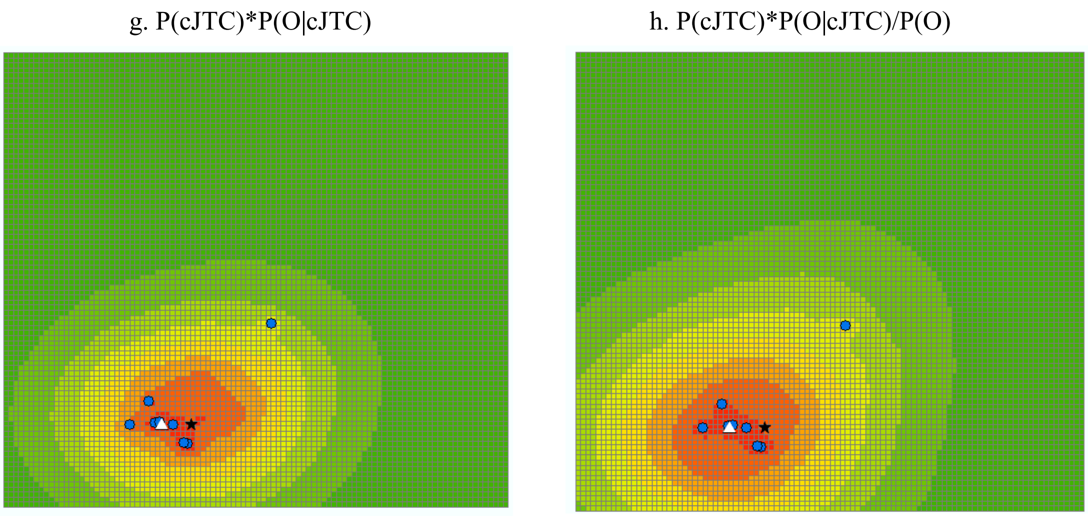


Figure 1 (a-h). Outputs from each profiling procedure for one specific offender included in the sample. The outputs were derived from: (a) a participant, (b) the CMD, (c) P(dJTC), (d) P(cJTC), (e) P(O), (f) P(O|cJTC), (g) P(cJTC)*P(O|cJTC), and (h) P(cJTC)*P(O|cJTC)/P(O). The scale of these maps is 1:8 the size of the original maps.

CONTROL vs. DECAY: PRECISION

Table 1 contains the average standard deviations associated with predictions in the Control and Decay conditions as a function of experimental phase. After confirming that these data met the conditions necessary for general linear modelling, they were submitted to a 2 (Condition: control and decay) x 2 (Phase: pre-instruction and post-instruction) mixed-design analysis of variance (ANOVA). Condition was the between-subjects measure, phase was the within-subjects measure, and standard deviation was the dependent variable.

Table 1. Average Standard Deviations (in miles) for the Control and Decay Groups as a Function of Experimental Phase.

Group	*M*	*SD*	CI_{95}
Pre-Instruction:			
Control	4.61	1.11	4.16 - 5.06
Decay	5.62	.61	5.34 - 5.91
Overall	5.12	.95	4.86 - 5.38
Post-Instruction:			
Control	3.03	.96	2.63 - 3.42
Decay	1.54	1.07	1.10 - 1.97
Overall	2.28	1.15	2.00 - 2.57

The ANOVA revealed a main effect for Phase, $F(1,38) = 188.68$, $p < .001$, $\eta^2 = .83$, with pre-instruction precision being significantly worse than post-instruction precision. A significant Condition x Phase interaction was also found, $F(1,38) = 36.83$, $p < .001$, $\eta^2 = .49$, indicating that the increase in precision for the Decay group across phases was greater than for the Control group. No main effect was found for Condition.

CONTROL vs. DECAY: ACCURACY

Table 2 contains the average error distances associated with predictions in the Control and Decay conditions as a function of experimental phase. Again, these data were submitted to a 2 (Condition: control and decay) x 2 (Phase: pre-instruction and post-instruction) mixed-design ANOVA. Condition was the between-subjects measure, phase was the within-subjects measure, and error distance was the dependent variable.

The ANOVA revealed a main effect for Phase, $F(1,55) = 14.32$, $p < .001$, $\eta^2 = .21$, with post-instruction accuracy being significantly better than pre-instruction accuracy. In contrast, there was no main effect for Condition and no Condition x Phase interaction.

STUDENTS vs. ALGORITHMS: ACCURACY

To compare the accuracy of students in the post-instruction phase with the algorithmic procedures, a series of one-sample t-tests were conducted (with a Bonferroni correction applied to each set of tests). Mean differences between the students' predictions and those of the algorithmic procedures are presented in Table 2, along with corresponding effect sizes. As can be seen in Table 2, participants in the Control group were significantly less accurate than all of the algorithms, with the exception of P(O). In fact, the Control group made predictions that were significantly more accurate than P(O). Participants in the Decay group were also significantly less accurate than the following five algorithmic procedures: CMD, P(dJTC), P(cJTC), P(O|cJTC), and P(cJTC)*P(O|cJTC). However, there was no significant difference in accuracy between participants in the Decay group and the Bayesian risk estimate, and participants in the Decay group made predictions that were significantly more accurate than those produced by P(O).

A COMPARISON OF THE ALGORITHMS: ACCURACY

To compare the accuracy of the seven algorithms, Friedman's test for k-related samples was conducted because many of the algorithms being tested are interdependent. A significant difference in error distance was found between the algorithmic procedures, $\chi^2 = 49.13$, $df = 6$, $p < .001$, $W = .21$. A follow-up analysis using Wilcoxin's Signed Ranks Test (with and without a Bonferroni correction) indicated that this difference was due to the fact that the predictions produced by P(O) were significantly less accurate than all of the other algorithmic procedures (all p's < .001). No other significant differences were found.

Table 2. Average Error Distances (in miles) for the Control and Decay Groups as a Function of Experimental Phase. Also Shown are the Comparison Data for the Algorithmic Procedures. The Bolded Value Indicates the Procedure Resulting in the Highest Accuracy.

				Comparisons[a]	
Profiling Procedure	*M*	*SD*	CI_{95}	Control (post)[b]	Decay (post)[c]
Pre-Instruction:					
Control	5.41	3.61	3.99 - 6.84	Ñ	Ñ
Decay	5.78	4.9	3.95 - 7.61	Ñ	Ñ
Overall	5.61	4.3	4.43 - 6.79		
Post-Instruction:					
Control	4.22	2.05	3.41 - 5.03	Ñ	Ñ
Decay	2.9	0.44	2.73 - 3.06	Ñ	Ñ
Overall	3.52	1.58	3.11 - 3.93		
Algorithmic Procedures:					
CMD	**2.35**	2.92	1.45 - 3.26	1.87[d], .41	.55[d], .53
P(dJTC)	2.67	3.32	1.64 - 3.70	1.55[d], .35	.23[e], .25
P(cJTC)	2.55	3.24	1.54 - 3.55	1.67[d], .38	.35[d], .37
P(O)	7.48	3.79	6.31 - 8.66	-3.26[d], .62	-4.58[d], .98
P(O|cJTC)	2.62	2.74	1.77 - 3.47	1.60[d], .36	.28[e], .30
P(cJTC)*P(O|cJTC)	2.49	2.99	1.57 - 3.42	1.73[d], .39	.41[d], .42
P(cJTC)*P(O|cJTC)/P(O)	2.88	3.28	1.87 - 3.90	2.05[e], .31	.02, .02

[a] Mean differences are in miles and effect sizes are η^2.
[b] $df = 26$ for a one-sample t-test.
[c] $df = 29$ for a one-sample t-test.
[d] $p < .001$ for a one-sample t-test.
[e] $p < .01$ for a one-sample t-test.

STUDENTS vs. ALGORITHMS: EFFICIENCY

To explore efficiency, we calculated the percentage of offenders living within several distance thresholds from the predicted home location. Three different criteria were tested: < 1 mile, < 3 miles, and < 5 miles. Cochran's Q test was used determine if there were significant differences between the procedures. Follow-up tests using Cochran's Q were also conducted to determine how the participants in our experiment compared to specific profiling procedures. All of the analyses had to be conducted separately for the Control and Decay groups because the frequency of exposure to specific maps in the post-instruction phase of the experiment varied across conditions (due to sample size variations).

For the analysis of maps from the Control group, there was a significant difference in the efficiency of profiling procedures across each of the distance thresholds (see Table 3). In each case, the percentage of offenders located within the various distance thresholds was lowest for P(O). Based on Cochran's *Q* test, participants in the Control group made predictions that were more efficient than those produced using P(O)

Table 3. Percentage of Offenders Living < 1 mile, < 3 miles, and < 5 miles From the Cell with the Highest Probability for the Various Profiling Procedures. Bolded Values Indicate the Procedure Resulting in the Highest Efficiency.

	Control Maps		
Profiling Procedure	< 1 mile[a]	< 3 miles[b]	< 5 miles[c]
Control	9.28	47.31	68.83
CMD	45.3	63	76.5
P(dJTC)	53.06	72.91	80.33
P(cJTC)	**53.25**	**75.14**	**82.56**
P(O)	2.3	8.8	22.7
P(O|cJTC)	41	56.7	76.5
P(cJTC)*P(O|cJTC)	53.06	67.72	**82.56**
P(cJTC)*P(O|cJTC)/P(O)	45.45	70.13	79.96

	Decay Maps		
Profiling Procedure	< 1 mile[d]	< 3 miles[e]	< 5 miles[f]
Decay	14.36	66.78	84.31
CMD	48.5	69.2	84.2
P(dJTC)	**50.75**	71.45	79.3
P(cJTC)	**50.75**	**74.29**	82.14
P(O)	2.2	10.7	24.3
P(O|cJTC)	43	60.8	**84.8**
P(cJTC)*P(O|cJTC)	**50.75**	67.11	82.14
P(cJTC)*P(O|cJTC)/P(O)	43.57	69.95	79.97

[a]Across all procedures, $Q = 1{,}187.00$, $df = 7$, $p < .001$; across algorithmic procedures, $Q = 967.50$, $df = 6$, $p < .001$.

[b]Across all procedures, $Q = 958.60$, $df = 7$, $p < .001$; across algorithmic procedures, $Q = 1{,}028.00$, $df = 6$, $p < .001$.

[c]Across all procedures, $Q = 945.70$, $df = 7$, $p < .001$; across algorithmic procedures, $Q = 1{,}030.00$, $df = 6$, $p < .01$.

[d]Across all procedures, $Q = 1{,}023.00$, $df = 7$, $p < .001$; across algorithmic procedures, $Q = 922.80$, $df = 6$, $p < .001$.

[e]Across all procedures, $Q = 980.80$, $df = 7$, $p < .001$; across algorithmic procedures, $Q = 1{,}048.00$, $df = 6$, $p < .001$.

[f]Across all procedures, Q = 1,128.00, $df = 7$, $p < .01$; across algorithmic procedures, $Q = 1{,}110.00$, $df = 6$, $p < .01$.

across all three distance thresholds (all *p*'s < .001). However, participants in the Control group made predictions that were significantly less efficient than those produced by the most efficient procedure across all three distance thresholds (all *p*'s < .001).

For the analysis of maps from the Decay group, there was also a significant difference in the efficiency of profiling procedures across each of the distance thresholds (see Table 3). In every case, the percentage of offenders located within the various distance thresholds was lowest for P(O). Based on Cochran's *Q* test, participants in the Decay group made predictions that were significantly more efficient than those produced using P(O) across all three distance thresholds (all *p*'s < .001). However, participants in the Decay group made predictions that were significantly less efficient than those produced by the most efficient procedure at < 1 mile and < 3 miles (each *p* < .01). There was no significant difference in efficiency between participants in the Decay group and the most efficient profiling procedure for the < 5 miles threshold.

A COMPARISON OF THE ALGORITHMS: EFFICIENCY

To compare the seven algorithmic procedures with respect to their efficiency, Cochran's *Q* test was used to determine if there were significant differences across the algorithms at each distance threshold. Given that our purpose is to identify the most efficient method, we also conducted Cochran *Q* tests between the most efficient method and the second most efficient method at each distance threshold. We were not interested in whether the second best method was significantly better than the third best, and so forth.

Drawing on the maps used in the Control condition, significant differences were found between the algorithmic procedures across each of the distance thresholds (see Table 3). When focusing on the most efficient procedure and the second most efficient procedure, significant differences were found at < 3 miles and < 5 miles (each *p* < .01). At < 3 miles, P(cJTC) maximized efficiency. At < 5 miles, P(cJTC) and P(cJTC) * P(O|cJTC) maximized efficiency.

Drawing on the maps used in the Decay condition, significant differences were also found between the algorithmic procedures across each of the distance thresholds (see Table 3). Once again, Cochran *Q* tests were conducted to compare the most efficient method and the second most efficient method across each of the distance thresholds. These tests revealed a significant difference but only at < 3 miles, where P(cJTC) maximized efficiency.

DISCUSSION

This study compared the precision, accuracy, and efficiency of geographic profiling predictions made by students, some of whom were instructed to use the distance decay heuristic, with seven algorithmic procedures: the CMD, a default JTC function, a calibrated JTC function, and four procedures based on the new Bayesian prediction model.

THE IMPACT OF THE DECAY HEURISTIC ON THE PRECISION OF PROFILING PREDICTIONS

Consistent with previous studies (e.g., Bennell, Snook, et al., 2007; Snook et al., 2002, Snook et al., 2004), the level of agreement amongst participants in the instructional group was substantially modified by providing them with the decay heuristic, such that they tended to carry out the task with greater similarity after they were informed of the heuristic. Based on this result, it is clear that most students were able to implement the decay heuristic after receiving very brief instructions. Having said this, no amount of instructions or training will ever allow participants to achieve the level of precision that can be accomplished by an algorithmic procedure because algorithmic procedures will make the same prediction every time they encounter the same crime series. This fact must be taken into account when making decisions about whether a human judgement should be used (Bennell, Taylor, et al., 2007). In addition, just because knowledge of the decay heuristic increases precision does not necessarily mean that prediction accuracy will also be enhanced. Precision only indicates whether participants are able to implement a heuristic, whereas accuracy indicates whether the heuristic being implemented is appropriate for the given task (i.e., whether or not it is ecologically rational).

THE ACCURACY OF PROFILING PREDICTIONS MADE BY STUDENTS AND ALGORITHMIC PROCEDURES

Also consistent with previous studies (e.g., Bennell, Snook, et al., 2007; Paulsen, 2006; Snook et al., 2002; Snook et al., 2004) was the finding that providing participants with information about a simple profiling heuristic increased prediction accuracy. This result suggests that the decay heuristic is an ecologically rational (and relatively robust) heuristic that can be used to improve the accuracy of predicting serial offender home locations in Baltimore County, MD. That being said, different heuristics may lead to different levels of accuracy on this task and it is unclear if the decay heuristic is the optimal heuristic for the purpose of making geographic profiling predictions.

In contrast to previous research, the Condition x Phase interaction in the accuracy analysis was found to be non-significant, indicating that participants in the Control condition also showed a significant increase in prediction accuracy across the two phases of the experiment (confirmed by a paired-samples t-test, $t(26) = 2.18$, $p < .05$, $\eta^2 = .21$). While a similar trend has been reported elsewhere (e.g., Bennell, Snook, et al., 2007; Snook et al., 2002; Snook et al., 2004), participants in the Control condition have never shown significant improvement. Why this occurred in the current study is unclear. Based on additional analyses, this finding does not appear to be the result of differing proportions of marauding versus commuting offenders across the maps in the pre- and post-instruction phases, nor does this finding appear to be the result of any participant characteristic.[12] For us, the important point is that participants in the Control group did not improve to the same extent as participants in the Decay group.

Also in contrast to previous research was the finding that students in the Decay group did not achieve a level of accuracy comparable to all of the algorithmic procedures.

[12] Consistent with Canter and Larkin (1993), an offender was defined as a marauder when their residence was located within a circle that was just large enough to encompass all of their crimes and as a commuter when their residence was located outside of this circle.

Although the level of accuracy they achieved was certainly respectable, these students performed significantly worse than five of the seven algorithmic procedures: the CMD, P(dJTC), P(cJTC), P(O|cJTC), and P(cJTC)*P(O|cJTC). These students did, however, perform at a level that was comparable to the Bayesian risk estimate and they significantly outperformed P(O). It is difficult to explain the departure from previous research with respect to the default JTC function, P(dJTC). Given the fact that individuals using the decay heuristic and default JTC functions have tended to produce comparable profile predictions in previous research, we believe a replication of the current finding is important before undue weight is put upon it.

It is perhaps less surprising that the students in the Decay group performed significantly worse than P(cJTC), P(O|cJTC), and P(cJTC)*P(O|cJTC). All of these functions were derived empirically from a sample of Baltimore offences and the calibration process presumably captures important characteristics of the crimes that are being profiled (Levine, 2007). These characteristics are not captured by P(dJTC) or by the students in the Decay group. In addition, P(O|cJTC) and P(cJTC)*P(O|cJTC) incorporate information about the origin points of offenders who committed crimes in similar locations to the offenders being profiled. This information is obviously not incorporated into P(dJTC) and was not provided to the students in the Decay group, who may have been capable of using that information to improve their predictions.

Despite the potential value of this additional information for geographic profiling purposes, and the small “jumps” in performance that were observed when moving from P(dJTC) to these other algorithmic procedures, it is important to note that neither the calibration process nor the process of taking the predisposition of other offenders into account significantly improved the accuracy of predictions. Indeed, the *CI*s for all of the algorithmic procedures almost completely overlap, with the exception of P(O), and the CMD was found to result in the most accurate predictions. Given that other studies also find that the CMD occasionally results in quite accurate predictions (e.g., Leitner & Kent, 2009; Levine & Lee, 2009; Paulsen, 2006; Snook, Zito, Bennell, & Taylor, 2005), we must carefully consider the value that is added by using more complex mathematical algorithms for the purpose of making geographic profiling predictions.

Also of interest is the finding that the Bayesian risk estimate and P(O) performed poorly compared to both the other algorithmic procedures and the Decay group. Concerning P(O), one might have expected that information about the baseline density of previous offenders’ residences within a particular area would allow more accurate predictions to be made for offenders committing crimes within that area. We found the opposite. Predictions based solely on P(O) were very poor. The finding that P(O) does not perform poorly across all studies (e.g., Block & Bernasco, 2009) suggests that there may be something unique about the environment of Baltimore County, MD, perhaps in the way that it influences the distribution of offenders’ residences and their spatial behavior.

Although the sub-par performance of the Bayesian risk estimate was not totally unexpected (see Levine & Block, 2009), this finding is somewhat counterintuitive to the findings relating to P(O). The evidence from predictions based solely on P(O) suggests that incorporating this information as a denominator of the Bayesian risk estimate will weight away from previous offender’s residences, since a high P(O) will result in a lower predicted likelihood for the area in question. We might therefore expect the inclusion of

P(O) within the risk estimate to increase prediction accuracy. The complexities of why P(O) appears to hinder performance in both scenarios is an issue that requires further study.

ACHIEVING EFFICIENCY THROUGH THE USE OF ALGORITHMIC PROCEDURES

Of particular importance in the current study was the finding that the predictions made by students were much less efficient than those produced using the majority of algorithmic procedures, at least when a strict criterion was used (i.e., < 1 mile). The only exception to this was P(O). It was not until relatively lenient thresholds were tested (i.e., < 3 miles and < 5 miles) that students were able to "catch up" to the algorithmic procedures. This is an important finding because strict criteria are likely the most operationally relevant in a policing context (i.e., the police will want to minimize the area they need to search for an offender). Notwithstanding some indication that the calibrated distance decay function often resulted in the most efficient profiles, no obvious front runner emerged as the most efficient profiling procedure when using the strictest, most operationally relevant threshold. Of note is the fact that the simplest algorithmic procedure (i.e., the CMD) was found to be as efficient as the more complex procedures when using this strict threshold. Although we must wait to see if this result can be replicated, when combined with the results pertaining to accuracy, this surely begs the question of whether or not complex algorithmic procedures are in fact needed in the field of geographic profiling.

IMPLICATIONS FOR POLICING

We have suggested that if the use of simple heuristics can be shown to result in geographic profiles that are similar to those produced by algorithmic procedures, they may provide a low-cost, easy-to-implement alternative to computerized geographic profiling systems. While our previous research has confirmed the potential value of simple heuristics, the current study has shown that the use of heuristics will not always result in profiles that are as precise, accurate, or efficient as those produced through the use of algorithmic procedures, especially algorithms based on the Bayesian JTC procedure. While replicating this result is important, we must ask where this finding leaves police agencies that may benefit from the use of simple heuristics for geographic profiling purposes.

There are two general points that can be made here. First, although participants in the Decay group did not produce profiles that were as precise, accurate, or efficient as those produced by most of the algorithmic procedures, informing people of the decay heuristic did significantly improve their performance. In addition, performance was improved to a level that may be operationally useful. For example, the most accurate algorithmic procedure (i.e., CMD) had an average error distance of 2.35 miles, whereas participants in the Decay group achieved an average error distance of 2.90 miles – a difference of about half a mile. Although these error distances were found to be statistically different from one another, it is certainly possible that they are not practically different from one another, at least in some jurisdictions, which means that the decay heuristic would still have value. Similar arguments could be made for the precision and efficiency analyses.

The second point relates to the value of a very simple algorithmic procedure, namely, the CMD. As demonstrated elsewhere (e.g., Leitner & Kent, 2009; Levine, 2007; Levine & Lee, 2009; Paulsen, 2006; Snook et al., 2005), the CMD appears to be a useful geographic profiling method that could be implemented (and automated) easily by all police forces. Based on results from the current study, the CMD is more precise, accurate, and efficient than the decay heuristic, and if automated, it has many of the advantages associated with more complex geographic profiling systems (e.g., an ability to communicate with other databases in order to manage investigative information and prioritise suspects; Bennell, Taylor, et al., 2007). Police organizations that are not in a position to use a complex geographic profiling system, and are uncomfortable relying on a simple heuristic, could use the CMD as a useful alternative.

Of course, neither simple heuristics nor the CMD result directly in a risk (i.e., probability) surface that can be systematically searched by the police; an issue that has been raised as a concern by numerous individuals and a potential limitation of these simple geographic profiling approaches (e.g., Gorr, 2004; Rossmo, 2005). However, as shown in a recent study (Taylor et al., 2009), search strategies can be derived from single point predictions, even by students implementing a simple heuristic. Furthermore, these search strategies can be as efficient as those derived from more complex algorithmic procedures (Taylor et al., 2009). While we encourage replication of this research, these findings suggest that there is some potential value in using simple profiling approaches. This may be useful for some police organizations, especially those that lack the financial resources or necessary infrastructure to support the use of computerized profiling systems.

LIMITATIONS AND FUTURE RESEARCH

There are at least five reasons to be cautious when interpreting and integrating the results of this study into our understanding of performance on the geographic profiling task. Firstly, there is a possibility that the results can be attributed in part to the different measurement methods used for assessing the performance of the participant groups and the algorithmic procedures.[13] Recall that participants made their predictions on hard copies of maps and all measurements were made manually by one of the authors, whereas the measurements for the algorithmic procedures were derived through an automated process. While great care was obviously taken to determine the measurements for our participant groups, and many of the measurements were double-checked, clearly these measurements cannot be as precise as those that were generated for the computer system. While this difference is unlikely to have influenced the results in a very significant way, it is possible that some of the differences between the participant-based predictions and the algorithmic-based predictions are larger or smaller than what was reported here.[14] Future research would be improved by automating the process for measuring the performance of the participant groups.

---

[13] We would like to thank one of our anonymous reviewers for raising this possibility.

[14] Consider, for example, the inherent noise associated with the fact that the performance of students was initially measured in millimeters on the hard copies of the maps, but the distance analysis was conducted in miles. Slight measurement errors on the hard copies of the maps may have impacted the measurement of overall accuracy.

Secondly, when considering the inferior performance of the students in the Decay group compared to the majority of algorithmic procedures in this study, it is important to keep in mind the impoverished nature of the instructions that the participants received. Recall that these participants were not provided with information that was used by many of the algorithmic procedures (e.g., the origin locations of other offenders). Although it seems unlikely, given that these students were unable to outperform even the most basic algorithms in the current study, it is possible that students may also be able to capitalize on this information by using it to adjust their own predictions. Related to this is the possibility that other heuristics, beyond the simple decay heuristic examined here, could lead to improved performance, and that other participants (e.g., police personnel) may be better equipped to implement profiling heuristics. Further studies testing these possibilities are needed before firm conclusions can be reached about the value of simple heuristics in the profiling context.

A third point has to do with a potential confound in the current study that may have widened the gap in accuracy between students in the Decay group and some of the algorithmic procedures. Specifically, the calibrated function used by many of the algorithmic procedures (and the origin-destination matrix) was developed on a sample of crimes that included the 40 test cases examined in the current study, thus potentially biasing the accuracy of those procedures. The solution to this problem involves testing the algorithmic procedures on a data set that is independent from the one used to develop the models. In other words, none of the same crimes should be included in the development and test samples. Although we believe that this issue had a negligible impact on the results of the current study (because the 40 cases made up less than 1% of the original development sample) it is true that prediction models tend to lose some of their predictive power when applied to novel data (Efron, 1982). Given the possibility of over-fitting, appropriately cross-validated algorithmic procedures should be examined in the future to see if the results of the current study can be replicated.

Fourthly, it is important to stress the fact that accuracy was measured solely by error distance because this was the only measure that could be used to evaluate all of the profiling procedures. This measure may not be the most appropriate method for assessing accuracy (Rich & Shively, 2004), especially when examining techniques that can produce probability surfaces (Rossmo, 2005). In addition, the determination of what profiling procedure is most accurate appears to depend to some extent on the accuracy measure chosen (e.g., Paulsen, 2006). Given these concerns, one should be cautious when interpreting the results of this study until a more thorough examination of accuracy is undertaken. In this regard, it will also be useful in the future to alter the instructional set provided to students so that other measures of accuracy can be calculated (see Taylor et al., 2009, for a recent example).

Lastly, it is important to consider the possibility that even more accurate and efficient algorithmic procedures may be developed in the future. Therefore, it will be necessary for researchers who study geographic profiling to examine these updated procedures, to compare them with one another, and to compare them with simpler procedures. The product function, $P(cJTC)*P(O|cJTC)$, examined in the current study may hold particular promise given that additional information can be systematically added to this prediction model (e.g., target attractiveness; Levine, 2009). Such revisions to this procedure should increase its predictive accuracy/efficiency, perhaps to a point where the procedure consistently and significantly outperforms other algorithmic procedures, including the CMD. When improvements are made to algorithmic procedures, studies like the current one should be replicated, taking into account the issues discussed above.

REFERENCES

Bennell, C., Snook, B., Taylor, P. J., Corey, S., & Keyton, J. (2007). It's no riddle, choose the middle: The effect of number of crimes and topographical detail on police officer predictions of serial burglars' home locations. *Criminal Justice and Behavior, 34,* 119-132.

Bennell, C., Taylor, P. J., & Snook, B. (2007). Clinical versus actuarial geographic profiling approaches: A review of the research. *Police Practice and Research, 8,* 335-345.

Block, R., & Bernasco, W. (2009). Finding a serial burglar's home using distance decay and conditional origin-destination patterns: A test of empirical Bayes journey to crime estimation in the Hague. *Journal of Investigative Psychology and Offender Profiling, 6,* 187-211.

Canter, D. V., Coffey, T., Huntley, M., & Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology, 16,* 457-478.

Canter, D. V., & Larkin, P. (1993). The environmental range of serial rapists. *Journal of Environmental Psychology, 13,* 63-69.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97-118). Oxford: Oxford University Press.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia, PA: Society for Industrial Mathematics.

Gigerenzer, G., Todd, P., & The ABC Research Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gorr, W. (2004). *Framework for validating geographic profiling using samples of solved crimes.* Retrieved January 14, 2010 from http://www.heinz.cmu.edu/research/156full.pdf.

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.

Jacob, V. S., Gaultney, L. D., & Salvendy, G. (1986). Strategies and biases in human decision making and their implications for expert systems. *Behaviour and Information Technology, 5*, 119-140.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York, NY: Cambridge University Press.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107*, 296-310.

Leitner, M., & Kent, J. (2009). Bayesian journey to crime modeling of single- and multiple-crime type series in Baltimore County, MD. *Journal of Investigative Psychology and Offender Profiling, 6*, 213-236.

Levine, N. (2007). *CrimeStat: A spatial statistics program for the analysis of crime incident locations* (version 3.1). Washington, DC: National Institute of Justice.

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

Levine, N., & Block, R. (2009). Bayesian journey-to-crime estimation: An improvement in geographic profiling methodology. *Unpublished manuscript.*

Levine, N., & Lee, P. (2009). Bayesian journey to crime modelling of juvenile and adult offenders by gender in Manchester. *Journal of Investigative Psychology and Offender Profiling, 6*, 237-251.

Paulsen, D. (2006). Human versus machine: A comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling, 3,* 77-89.

Petrocelli, J. (2007). Geo-profiling. Technology isn't always the best solution. Retrieved October 25, 2008, from http://www.officer.com/web/online/Editorial-and-Features/Geo-Profiling/19$36899.

Rich, T., & Shively, M. (2004). *A methodology for evaluating geographic profiling software*. Washington, DC: National Institute of Justice.

Rossmo, D. K. (2000). *Geographic profiling*. Boca Raton, FL: CRC Press.

Rossmo, D. K. (2005). Geographic heuristics or shortcuts to failure: Response to Snook et al. *Applied Cognitive Psychology, 19*, 651-654.

Snook, B. (2004). Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling, 1*, 53-66.

Snook, B., Canter, D. V., & Bennell, C. (2002). Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system. *Behavioral Sciences and the Law, 20,* 1-10.

Snook, B., Taylor, P. J., & Bennell, C. (2004). Geographic profiling: The fast, frugal, and accurate way. *Applied Cognitive Psychology, 18,* 105-121.

Snook, B., Zito, M., Bennell, C., & Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology, 21*, 1-26.

Taylor, P. J., Snook, B., & Bennell, C. (2009). The bounds of cognitive heuristic performance on the geographic profiling task. *Applied Cognitive Psychology, 23*, 410-430.