Spatial Signatures and Mental Maps: Using offenders' activity locations to enhance geographic profiling for crime investigations

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Psychology

at

The University of Waikato

by

Sophie Curtis-Ham

KO TE TANGATA

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2022

Abstract

In crime investigations, geographic profiling involves inferring information about 'whodunit' from information about where and when the crime occurred. Such inferences are possible because people usually commit crime in places they know, rather than seeking opportunities elsewhere. The 'spatial signature' of a crime location thus reflects the 'mental map' of the offender: the locations they are familiar with from their everyday activities, such as where they live, work, go to school or visit family and friends. Previous empirical research suggests that people are more likely to commit crime near some activity locations in their mental map than others. However, these differences have not been accounted for in theory or in geographic profiling methods—particularly methods that infer the most likely suspect to have committed a crime, given the proximity of their activity locations to the crime. This thesis aimed to enhance geographic profiling by developing, validating, and applying a theoretical model that explains how people's different activity locations influence their crime locations.

To achieve this aim, I first drew on existing theoretical and empirical literature to propose a theoretical model that identifies the attributes of people's activity locations that influence whether they will commit crime nearby and the psychological mechanisms through which this influence occurs. Next, I tested hypotheses derived from the model to add to its empirical support, using New Zealand Police data on approximately 60,000 offenders who committed burglaries, robberies and extra-familial sex offences between 2009 and 2018. This national dataset included details of the offenders' crime locations and their pre-crime activity locations such as home addresses, family members' home addresses, schools, and the locations of prior crimes and other police interactions. I then applied the theoretical model in a new geographic profiling algorithm: the Geographic Profiling: Suspect Mapping and Ranking Technique (GP-SMART). For a given crime, GP-SMART predicts suspects' probability of committing the crime given the proximity of their activity locations to the crime and the attributes of those activity locations that influence whether they will commit crime nearby—as described in the theoretical model—and prioritises the suspects

accordingly. I tested GP-SMART's accuracy by investigating how often it placed the actual offender among the top ranked suspects.

The theoretical model proposed that people are more likely to commit crime in locations where they have reliable knowledge that is relevant to the future crime. Attributes of people's activities—such as their frequency and similarity to the future crime—affect the development of reliable and relevant knowledge. The empirical tests confirmed that offenders were generally more likely to commit crime closer to the kinds of activity locations that people visit more frequently (e.g., home versus family homes) or likely to impart more relevant knowledge about crime opportunities (e.g., prior crimes versus prior victim or witness location). They were also more likely to commit crime near prior activity locations that would generate both reliable and relevant knowledge, than near prior activity locations lower on either or both dimensions. By accounting for the theoretically salient attributes of suspects' activity locations, GP-SMART ranked the offender at or near the top of the suspect list at rates exceeding chance and those produced by baseline methods—approximating existing algorithms—that use proximity alone.

This thesis makes significant theoretical, methodological and practical contributions. The theoretical model explicitly extends crime pattern theory, and the empirical studies add to the evidence base for this extension and core tenets of both the environmental criminology and investigative psychology theoretical approaches to geographic profiling. The police data included a wider range of activity locations than used in past crime location choice studies based on administrative data, and the findings highlight their utility for signalling offenders' mental maps. Indeed, the size of the dataset required developing a novel method for sampling when using Discrete Spatial Choice Modelling—a major methodological contribution to the growing discrete crime location choice literature. But of most significance in light of the motivation for this thesis, the GP-SMART results demonstrate that enhancing geographic profiling by incorporating the attributes of people's various activity locations that influence their crime locations, could help solve crime in practice.

Acknowledgements

This PhD was never a matter of if, only when. But good things take time. Ideas needed to percolate; creativity needed to spark; stars needed to align. I'm chuffed to have finally accomplished this longstanding goal, but it would not have been possible without the help of many people along the way. I'm eternally grateful to everyone acknowledged below, and eternally apologetic to anyone I have missed from this list.

In the academic realm, thank you first to Devon and Oleg for your skilled supervision and your enthusiasm in joining me on this learning journey. My gratitude also extends to my fellow 'labsters' at the University of Waikato for your camaraderie and contributions, and to the University itself for the doctoral scholarship that funded this research. Further afield, I'm deeply appreciative of the input and guidance of Wim Bernasco, whose expertise was critical to the completion of this research. Additional thanks to Stijn and others at Vrije University Amsterdam for the warm welcome, help with r code and thoughtful discussions during my all too brief visit. Likewise, thanks to other ECCA members, especially Spencer and Kim, whose influence on this research goes back much further than our discussions about it in 2019. Countless others in the #EBP, #rstats and #phdlife twitterverse helped by way of sage soundbites, technical tips, and motivational morsels.

Thanks also to my NZ Police colleagues, particularly those at the Evidence Based Policing Centre—past and present—for your 'pastoral care', interest and insights, which helped improve this research. Special thanks to Darren, Rhi and Ross for your advice in the early days, setting me off on the right trajectory, and to Erena and Jess the for time and resources to assume my 'student cat' persona. Further thanks to Cameron, Gavin K, Gavin R, Lana, Obert, Sean, and Tori for sharing insights and concerns about the data (and for listening to mine); to Rob, for being a Bayesian sounding board; and to many others in Intel and beyond for providing practitioner perspectives. Thanks, too, to NZ Police as an organisation, for the data, and the opportunity to seek signal in its noise.

On the personal front, huge thanks to Sim, for riding this roller coaster with me (and all life's others). Thanks to my friends, for forgiving my social absences IRL, and to my new-found 'funk fam' for cheering me on online. And last but far from least, thanks to my family, for being such great role models of lifelong learning, and for the driving force of (my perceptions of) your expectations. As self-indulgent and self-fulfilling as this PhD has been, it also brings me joy to think I've done you proud. A final, special thanks to Dad, for the gifts of pedantry and pragmatism that helped me in equal measure with this work; I wish you were still here to for me to share it with you.

Table of Contents

List of Abbreviations

ANZSOC	Australia New Zealand Standard Offence Classification
CI	Confidence interval (95%)
Com. Rob.	Commercial robbery
CP Theory	Crime Patten Theory
DIS	Distance importance sampling
DSCM	Discrete spatial choice modelling
EM Bail	Electronic monitoring bail
Family: immed	Family: Immediate family
Family: IP	Family: Intimate Partner
GP-SMART	Geographic Profiling: Suspect Mapping and Ranking Technique
IIA	Independence of irrelevant alternatives assumption
IQR	Inter-quartile range
MO	Modus operandi
MOJ	Ministry of Justice (New Zealand)
NIA	National Intelligence Application (New Zealand Police crime and intelligence database)
Non-res. Burg.	Non-residential burglary
NZ	New Zealand
OR	Odds ratio
Pers. Rob.	Personal robbery
Res. Burg.	Residential burglary
SA1	Statistical Area 1 (New Zealand Census geographic unit)
SA2	Statistical Area 2 (New Zealand Census geographic unit)
SIS	Simple importance sampling
SRS	Simple random sampling
TA	Territorial Authority (New Zealand local authority boundary)

List of Publications

Peer-reviewed journal articles

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information, 10*(2), 47. https://doi.org/10.3390/ijgi10020047

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Relationships between offenders' crime locations and different prior activity locations as recorded in police data.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Familiar locations and similar activities: how offenders' past experiences predict their crime locations.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART. Journal of Investigative Psychology and Offender Profiling. https://doi.org/10.1002/jip.1585

Conference presentations

Curtis-Ham, S. (2020, 2 July). People who offend near to each other tend to share other activity locations. University of Waikato & University of Canterbury Forensic Psychology conference, online.

Curtis-Ham, S. (2020, 3 September). People who offend near to each other tend to share other activity locations. Evidence Based Policing Centre Intern Conference, online.

Curtis-Ham, S. (2021, 13 May). People are more likely to commit crime near some routine activity locations than others. Evidence Based Policing Centre and University of Waikato Conference, Wellington, New Zealand.

Curtis-Ham, S. (2021, 12 August). Geographic profiling: how using police data on suspects' activity locations might help solve crime. Waikato Waitaha Inter-lab Conference, online.

Curtis-Ham, S. (2021, 6 September). Geographic profiling: how reconstructing suspects' mental maps might help solve crime. New Zealand Psychological Society Conference, online.

Curtis-Ham, S. (2021, 24 October). Geographic profiling: how using police data on suspects' activity locations might help solve crime. Australia New Zealand Society for Evidence Based Policing Conference, online.

CHAPTER 1

Introduction

The Problem: Getting from Where to Who in Crime Investigations

When investigating a 'whodunit' crime, what if we[1] could tell who committed it, simply by comparing the location of the crime with the locations potential suspects frequent in their everyday lives? In the same way that people who commit crime may leave traces of themselves in their fingerprints, the marks of their shoes, or a distinctive behavioural 'signature' (Canter, 1995; Douglas & Olshaker, 1998), their crime locations also tell us something about them. They are an indication of their mental map, their internal representation of the world they inhabit and of the opportunities for crime within it. Each of us builds an internal map as we go about our daily and other habitual routines, formed from the places we frequent ('activity locations' or 'activity nodes') and the paths between them (Golledge & Stimson, 1997; Gould, 1973; Gould & White, 1986), and just as this map informs our decisions about where to carry out activities of our daily life, it also informs offenders' decisions about where to commit crime (P. L. Brantingham & Brantingham, 1991, 1993a; Canter & Youngs, 2008b). In a crime investigation, therefore, information about the mental maps of potential suspects, as evident from the location and nature of their known activity locations, may provide clues as to which of those people is more likely to have chosen the location of the crime.

The problem of how to infer who committed a crime from where it was committed is not new. Moreover, a range of analytic methods exist to assist police in making this inference, collectively described as 'geographic profiling' (Rossmo, 2000; Rossmo &

---

[1] In this this thesis, I use the word "we" in a general sense to refer to what is known, or not known, in the wider scientific community, what is commonly thought or done in the policing community, or what is experienced by people in general (as in this first instance). And although the research in this thesis is my own, I also use "we" in reference to research activities, to recognise the contributions of my supervisors and co-authors to those activities.

Rombouts, 2008) or ‘geographical profiling’ in the UK (Canter & Youngs, 2008b, 2008a). These methods involve using geographic information—such as the location and timing of crimes, or the direction of travel of the offender before and after the offence—to infer where the offender might live or where (and when) they might strike next (Knabe-Nicol & Alison, 2011; Rossmo & Rombouts, 2008; Rossmo & Summers, 2015). However, geographic profiling methods to date have not included a means of systematically harnessing information about suspects’ *various* activity locations—beyond their home address—to identify the suspects more likely to have committed a given crime. The research presented in this thesis therefore aimed to develop such a means to enhance geographic profiling.

The thesis achieved this aim through first creating and validating a theoretical model that explains how people’s different activity locations influence their crime locations, then applying the model in a geographic profiling algorithm that uses suspects’ activity locations to predict which suspects are more likely to have committed a given crime. This introductory chapter expands the argument that there is potential to improve on present means of inferring who from where in crime investigations, then outlines the steps taken in this thesis to realise that potential.

The Context: Geographic Profiling and its Underpinning Theory

In crime investigations where the offender is not known, investigators can draw not only on witness and physical evidence to help identify the offender but behavioural clues such as how, where and when the crime was committed. Spatial and temporal clues can be especially telling because people are constrained by their routine non-criminal activities in where and when they can identify and act on crime opportunities (P. L. Brantingham & Brantingham, 1991; Ratcliffe, 2006). The location of the crime—its ‘spatial signature’—can thus indicate characteristics of the offender that are particularly useful to an investigation such as where they live or where they might commit further crimes. Geographic profiling, typically carried out by specially trained police analysts (Rossmo, 2012), is the process of identifying the spatio-temporal clues in a crime investigation and using them to make inferences about the offender that inform investigative strategies (Emeno et al., 2016; Knabe-

Nicol & Alison, 2011; Rossmo, 2014). For example, inferring the likely area of the offender's residence can inform database searches for and prioritisation of suspects, door to door enquiries, or targeted media campaigns (Rossmo, 2008; Rossmo & Rombouts, 2008; Rossmo & Velarde, 2008). Similarly, inferring the likely location of the offender's next crime in a series can inform surveillance or 'sting' operations aimed at catching the offender in the act (e.g., Casady, 2008; Daniell, 2008).

The inferences made in geographic profiling, as developed in North America, are explicitly grounded in core theories of environmental criminology (Rossmo, 2000; Rossmo & Rombouts, 2008). Rational choice theory tells us that people weigh risk, costs and reward in deciding whether, when, where and how to commit crime (Cornish & Clarke, 1986). This decision-making process means, for example, that people tend not to travel farther—a higher cost—than necessary to achieve a given criminal objective. Routine activity theory tells us that crime occurs where people motivated to commit a crime converge in space and time with suitable and available targets for the crime (Cohen & Felson, 1979). Crime pattern theory tells us about where this convergence is likely: that is, within the offender's 'awareness space'—the places they are aware of, or mental map—as shaped by their activity locations or 'nodes'—where they live, work, socialise, shop and so on (P. L. Brantingham & Brantingham, 1991, 1993a). Empirical research examining the relationship between offenders' activity locations and crime locations has consistently confirmed that people are generally more likely to commit crime closer to these locations than farther away (Bernasco, 2019; Menting et al., 2020; Rengert & Wasilchick, 1985; Townsley, 2016).

In the United Kingdom, 'geographical' profiling evolved from environmental psychology (Canter, 1977) and is a sub-domain of investigative psychology (Canter & Youngs, 2009). The investigative psychology approach places a greater emphasis on offenders' mental maps as internal representations of their environment and the criminal opportunities it affords (Canter & Youngs, 2008b). Nonetheless, it is clearly grounded in the same principles of human spatial behaviour: people will generally not travel farther than needed to take advantage of an opportunity to achieve a goal, and their knowledge of opportunities derives from their past interactions with the environment in the places they live, work and play (Canter & Youngs, 2008b). In this thesis the term ‘geographic’—rather than ‘geographical’—profiling encapsulates both the North American and British approaches.

Geographic profiling can be described as the structured application of these theories to case material and the environmental context of the crime(s)[2] to gain insight into the decision-making of the offender and hence who they may be or how to identify them (Knabe-Nicol & Alison, 2011; Rossmo & Rombouts, 2008). This process involves the use of quantitative and qualitative methods (Daniell, 2008; Rossmo & Rombouts, 2008), analogous to the use of statistical and clinical judgment in other forensic settings such as offender risk assessment (Otto, 2000). The remainder of this section describes the specific statistical and clinical inferences that motivated the research undertaken in this thesis, which aimed to provide further theoretical and empirical bases for these inferences.

Statistical methods commonly applied in geographic profiling include descriptive analyses to identify spatio-temporal patterns (e.g., do the crimes occur at certain times of the day or week and at regular intervals, enabling an inference about when the next crime might occur?) and the use of predictive algorithms implemented in specialist software (Daniell, 2008; Knabe-Nicol & Alison, 2011; Rossmo & Rombouts, 2008). In practice the most commonly used predictive software predict the likely location of the offender’s home, or other activity node (Emeno et al., 2016; Perry et al., 2013). The research literature also includes algorithms that predict the location of the next crime in a series (e.g., O’Leary, 2009; Porter & Reich, 2012), or who, out of a list of suspects, is more likely to have committed the crime(s) given their known activity locations (e.g., Frank, 2012; Tayebi et al., 2017), though these algorithms appear not to have been commercialised or otherwise implemented in

[2] Geographic profiling is most commonly applied to investigations of multiple crimes believed to have been committed by the same offender because they involve multiple location data points from which to derive patterns and draw inferences, but it can also be applied to single crime investigations (Knabe-Nicol & Alison, 2011).

practice.[3] This thesis focuses on the first and third of these predictions: predicting the offender's activity locations from crime locations and predicting the offender from suspects' activity locations. These 'crime-based' and 'suspect-based' predictions are described in more detail next, along with clinical considerations that complement or aid interpretation of the software outputs.

Rigel (Environmental Criminology Research Inc., n.d.; Rossmo, 1995, 2000), CrimeStat (Levine, 2013) and Dragnet (Canter et al., 2000) are the most common software used by analysts for crime-based predictions of offender activity locations (Emeno et al., 2016; Rich & Shively, 2004). With all of these programs the analyst inputs the addresses of a series of crimes believed to have been committed by the same offender and the software applies an algorithm that uses the locations of the crimes to make a prediction, visualised on a map, of the most likely home or other activity location of the offender. Some algorithms are ideographic: they make simple predictions using solely the geometric pattern of the crime locations, such as using the geometric centre of the crime locations (Canter et al., 2013). Others are nomothetic: their predictions are based on the observed offence and home locations of offenders who have committed similar offences, which provide a 'distance decay function' that describes the distribution of the number of offences committed at increasing distances from home.[4] The distance decay function is applied to the distance between each crime in the input crime series and each possible activity location within a set distance of the crimes, to predict how likely the offender is to have an activity location at that distance from

---

[3] Given their lack of mention in surveys or interviews of geographic profiling practitioners (e.g., Emeno et al., 2016; Öhrn, 2019).

[4] The software vary in the extent to which the distance decay function can be calibrated by the user. Rigel uses a pre-set decay function, resembling a truncated negative exponential curve, based on theory and empirical research on offenders' home-crime distances (Rossmo, 1995, 2000). Dragnet permits the user to select from a range of functions such as linear, negative exponential and truncated negative exponential (Hammond & Youngs, 2011). CrimeStat enables users to import historical calibration data on offender home and crime locations then fits decay functions to the data to identify which function best fits the distribution of home to crime distances, and the function's parameters such as the intercept and slope coefficient (Levine, 2013).

the crime.[5] The predictions for all the input crimes are then aggregated for each possible activity location to give a total prediction score, visualised as a heatmap that shows the high scoring locations in a prominent colour.

A common next step in geographic profiling is to search available databases to identify potential suspects who live nearby and rank them based on the proximity of their home addresses to the predicted location.[6] An analyst might also include other activity locations in the search, on the grounds that the prediction could indicate not the offender's home location but another activity location (see Hauge et al., 2016 for an example of this approach in using geographic profiling to identify the artist Banksy). Police databases can include activity locations such as home addresses, places of employment, prior crimes, victimisations, arrests, and 'stop and searches', where these have been recorded for operational purposes. But if multiple suspects have activity locations at an equal distance to the predicted location, what can the analyst infer about which suspect to rank as a higher priority for investigative attention?

To answer this question, the analyst could turn to qualitative considerations based on details of the crimes and the environment in which they were committed. For example, if the crimes were mostly committed during typical working hours the analyst might prioritise a

---

[5] Despite that crime-based geographic profiling tools apply distance decay functions derived from observed *home*-crime distances, the predicted location may not be the offender's home location, but some other activity location. This is potentially problematic because the distance decay pattern may be different for other activity locations. However, even if different decay functions for predicting different activity locations were implemented in the algorithm, one would not know at the point of investigation which activity location one is predicting and thus which decay function to use (although some inferences are possible, as discussed below).

[6] With the software Rigel, the user can import suspects' home addresses (or other activity nodes) and the software will automatically rank the suspects by the prediction score at their known activity nodes (Environmental Criminology Research Inc., n.d.; Rossmo & Velarde, 2008). Similarly, the software IOPS (Interactive Offender Profiling System) developed by UK geographic profiling pioneer David Canter ranks suspects by proximity of their home locations to the location predicted by applying the Dragnet algorithm and the extent to which the modus operandi of their prior crimes match that of the input crimes (Canter & Youngs, 2008c).

suspect with a nearby workplace over a suspect with a nearby home address. If the crime locations appear to relate more to where specific targets are available, rather than where the offender lives or works, the analyst might prioritise a suspect who was recently stopped and searched for behaving suspiciously nearby over suspects with home or work addresses nearby.

Illustrating the latter scenario, Daniell (2008) described a case where seven females had been sexually assaulted near their homes in various suburbs in Bath, England. The geographic profiling software output predicted a street in a central nightlife area as most likely to include an activity location of the offender. However, considering the lack of residential housing in the predicted area and information about the victims' whereabouts before the attacks, the analyst interpreted the software output as more likely to indicate not the offender's home address but a base where he searched for potential victims whom he then followed as they walked home. Recommended database searches therefore included a search for people stopped by police in the predicted area. Ultimately a 'sting' operation successfully caught the offender in the act when he attempted to follow an undercover police officer as she left a bar in the predicted area.

Beyond these clues in the crimes and their 'environmental backcloth' (P. J. Brantingham & Brantingham, 2016; P. L. Brantingham & Brantingham, 1993a), it could also help to know which activity locations, in general, people are more likely to commit crime nearby. Analysts could then prioritise suspects with the most salient activity locations near the location predicted by the software.[7] In the past decade, empirical research has started to explore the question of which activity locations are more or less salient to offenders' crime location choices, suggesting differences between different activity locations (e.g., Bernasco, 2010; Menting et al., 2016; van Sleeuwen et al., 2018). However, there is no theoretical

---

[7] Van der Kemp (2021) suggests that suspects with the strongest connection to the area should be prioritised, but he does not elaborate how this should be defined or determined. In contrast, salient activity locations are those that through connecting a person 'strongly' to a location, or through other factors, make it more likely that they would commit crime there—which is the basis for prioritisation.

model systematising these differences, on which analysts can draw to inform the inference of which suspects to prioritise, given a crime-based activity location prediction and a range of suspects' activity locations. This thesis provides one.

This suspect prioritisation inference can alternatively be approached from the suspects' point of view. Suspect-based geographic profiling algorithms start with the activity locations of suspects falling within a set distance of a crime being investigated (only one crime, not a crime series, is required). They predict how likely each suspect is to have offended at the crime's location given its proximity to their activity locations, or to a wider area of 'activity space' (P. L. Brantingham & Brantingham, 1991) around their activity locations. Suspects are then ranked by their relative likelihood of offending at that location. Examples of such algorithms in the literature are sparse, and evidence of their use in practice is even sparser. Bache et al., (2008), Frank et. al., (2012), Gore et al., (2005), Snook et al., (2006) and Tayebi et al., (2017) presented and tested the predictive accuracy of suspect-based algorithms, but reference to the use of these algorithms is noticeably absent in the few studies that have surveyed or interviewed geographic profiling specialists about their methods (Emeno et al., 2016; Knabe, 2008; Knabe-Nicol & Alison, 2011; Öhrn, 2019). Chapter 7 describes the operation of these algorithms and their accuracy in detail, but two key limitations of them warrant mention here. First, they have only included a limited range of activity locations, being at most: home, prior crimes and co-offenders' homes (Tayebi et al., 2017). Second, they have not differentiated between different types of activity location in making their predictions, despite the evidence of differences between activity locations in their effects on people's crime locations.

There is thus potential for a greater understanding of the relative salience of different activity locations to also aid the suspect-based inference process. In principle, including a fuller set of suspects' activity locations, and weighting them by how strongly they influence crime location choice, should lead to more accurate inferences about who among the suspects is the most likely culprit. Further, enhancing the accuracy of suspect-based predictions in this way could translate into greater use of them in geographic profiling practice. In practice such predictions could be made statistically, through a suspect-based algorithm, or clinically, by mapping the activity locations of a shortlist of suspects and comparing these to crime locations, assigning greater priority to suspects with more salient activity locations near the crime(s). This thesis therefore sought to enhance both crime- and suspect-based geographic profiling methods through developing and applying a more detailed theoretical understanding of the relative influence of different activity locations on offenders' crime locations. The next section describes the steps taken in the thesis to do so.

The Thesis: Structure and Scope

Thesis Structure

The thesis moves from developing theory, to testing theory, to the application of that theory in the form of a new geographic profiling algorithm. It begins with the question: what information about offenders' activity locations (their mental maps) indicates where they are most likely to commit crime? In police investigative practice, we might then be able to assess the likelihood of suspects having committed a crime, given the location of that crime (its spatial signature) and salient information about the suspects' previous activity locations. As signalled above, and more fully detailed in Chapter 2, previous theories linking offenders' routine activity locations to their crime locations have not systematically considered this question, despite its applied importance. As a first step, therefore, we developed a theoretical framework that identified from existing literature the attributes of prior activity locations that predict offenders' future crime locations, and systematised the links between these attributes and crime locations in a theoretical model, as described in Chapter 2 (Curtis-Ham et al., 2020).

Chapters 3 through 6 collectively form the theory testing phase of the thesis, which expands the empirical basis for the proposed theoretical model. Chapter 3 (Curtis-Ham et al., 2021a) introduces the police dataset used in the thesis' empirical studies and presents initial exploratory analyses of offenders' pre-offence activity locations as revealed in this dataset. These analyses explored how many pre-offence activity locations were recorded for people who had committed a burglary, robbery or extra-familial sex offence in New Zealand between 2009 and 2018 and the geographic span of these offenders' activity locations. They also included a test of an important assumption underpinning the practical problem addressed in this thesis: that offenders' pre-offence activity locations are sufficiently distinctive to differentiate between offenders. If offenders' activity locations overlapped to a large extent, it would be hard to prioritise among the many potential suspects whose activity locations were equally consistent with them having committed an offence in a given location.

Chapter 4 (Curtis-Ham et al., 2021b) focuses on methodology, introducing the discrete spatial choice modelling (DSCM) method used in subsequent chapters to test hypotheses about the relative associations between different activity locations and offenders' crime location choices. It explores a technical issue presented by our 'big data' in the context of DSCM research. We investigated sampling methods to overcome computational challenges inherent in the application of DSCMs to large datasets and proposed a solution that not only enabled the present research but will facilitate further growth of the crime location choice DSCM literature.

The next two chapters then describe the application of this method to investigate relationships between offenders' pre-offence activity locations (mental maps) and their crime locations (spatial signatures). We first compared relationships between proximity to broad types of activity locations (e.g., offenders' homes, family members' homes, schools, prior crimes, and other police interactions) and the locations of the most recent offences of 17,054 residential burglars, 10,353 non-residential burglars, 1,977 commercial robbers, 4,315 personal robbers and 4,421 extra-familial sex offenders, in New Zealand (Chapter 5: Curtis-Ham et al., under review-b). In comparing types of activity location, we tested hypotheses derived from the theoretical model about whether offenders were more likely to commit crime closer to the kinds of activity locations that are visited more frequently (e.g., home versus family homes) or likely to impart more relevant knowledge about crime opportunities (e.g., prior crime locations versus prior victim or witness locations).

We then compared relationships between the specific attributes of offenders' activity locations proposed in the model and their subsequent crime locations (Chapter 6: Curtis-Ham et al., under review-a). The model proposes that offenders are more likely to commit crime near to activity nodes visited (a) more frequently, (b) more recently, (c) for longer time periods, or involving more similar (d) behaviour, (e) timing (hour of day, day of week, season of year), or (f) types of location, by comparison to activity nodes lower on these factors. The data used for this research presented some challenges in measuring these attributes, but we were able to construct composite variables and test the interaction between these variables specified in the theoretical model. Specifically, we tested whether offenders were more likely to commit crime where they were more familiar with the location (having been there more frequently, recently or over a longer time) *and* their mental maps included crime-relevant information (having conducted activities there that were more similar to, thus generating knowledge transferable to, the present crime). In the language of the theoretical model, the analysis examined whether crime was more likely near prior activity locations that were likely to have generated both reliable and relevant knowledge, and less likely near prior activity locations lower on either or both dimensions.

The final empirical study of the thesis moved from hypothesis testing to prediction, and from theory to practice. As described earlier in this chapter, in police investigations geographic profiling involves applying analytical methods that use information about the location and timing of offences to predict where or how the offender might be found. The final study (Curtis-Ham et al., 2022) addresses the issue that existing geographic profiling methods have not included a means of systematically harnessing salient information about suspects' routine activity locations to predict the most likely offender among them. We applied the findings from the preceding chapters to develop a novel solution to the practical problem of getting from where a crime was committed to who committed it. This solution takes the form of a geographic profiling algorithm that is distinct from, but complementary to, existing crime-based algorithms that automate the inference of where the offender is more likely to live given the locations of a series of crimes. As a suspect-based algorithm, it implements the reverse of this inference: given the location of potential *suspects'* home and other activity locations, and the nature of those activity locations, who among those suspects is more likely to have offended in the location of a crime. This process is described by the acronym GP-SMART: Geographic Profiling Suspect Mapping and Ranking Technique.

Chapter 7 (Curtis-Ham et al., 2022) first describes the few existing geographic profiling algorithms that have similarly automated suspect prioritisation and highlight limitations in their methods for prioritising suspects and measuring accuracy in identifying the actual offender in the top ranked suspects. It then describes GP-SMART in detail, including the method by which it identifies potential suspects with activity nodes nearby and estimates each suspect's likelihood of offending at the location of an input crime. In short, this likelihood is a function of the distance between the suspects' activity nodes and the input crime, and attributes of those activity nodes that make it more or less likely (as established by the preceding research in this thesis) that the suspect would have offended nearby. The chapter then presents and discusses the results of analysis of the predictive accuracy of GP-SMART that examined whether it identifies the actual offender in the top ranked suspects sufficiently frequently to have potential utility in police investigations.

The final chapter of this thesis (Chapter 8) presents an overall discussion of the research findings and their implications. It first recaps and synthesises the results of the studies presented in Chapters 2 through 7 in light of the aims of the thesis. It next reviews the thesis' theoretical contributions and highlights its methodological contributions in relation to: the use of police data on offenders' activity locations; the analysis of crime location choice using discrete choice models; and the testing of geographic profiling algorithms' accuracy. The chapter then examines the thesis' practical implications, elaborating the implications for geographic profiling practice—how the findings can be used to support qualitative and quantitative inferences from 'wheredunit' to 'whodunit'—and highlighting implications for data recording and offender management practices. Next, the chapter discusses the limitations of the research and opportunities for future research. The thesis concludes with a brief recap and reflection on the benefits of improving geographic profiling.

Thesis Scope

Some caveats about the scope of this research are worth clarifying from the outset. First, this thesis makes no claims as to the efficacy or utility of geographic profiling as a professional enterprise. Geographic profiling is a cluster of analytic methods that apply a spatial lens to crime investigation; it involves much more than just running an algorithm

(Fumagilli & Johnson, 2020; Knabe-Nicol & Alison, 2011; Öhrn, 2019; Rossmo & Rombouts, 2008). Algorithms and tools that provide 'decision-support' are exactly that—a means of supporting a wider decision-making process. Their outputs should always be interpreted by analysts and investigators in the context of all other case information. Historical, and heated, debates about the fruitfulness of geographic profiling, fuelled by comparisons of the accuracy of humans versus machines in making a single geographic profiling inference—predicting the offender's residence location from crime locations—(e.g., Bennell, Taylor, et al., 2007; Dern et al., 2009; Paulsen, 2006b; Snook et al., 2005), appear to have missed this point (Canter, 2005; Daniell, 2008; Goodwill et al., 2014). In this thesis the important role of the human element in geographic profiling is acknowledged. GP-SMART automates a complex calculation process that enables many thousands of possible suspects to be assessed and prioritised. Such processes are inherently more speedily and reliably performed by computers than the human brain (Canter, 2008; Canter et al., 2013), but they do not replace the analyst. Chapter 8 provides some concrete examples of how GP-SMART can be used in practice, to help demystify the context in which such algorithms operate.

A related point is that neither geographic profiling algorithms nor clinical geographic profiling judgments directly solve crime (Daniell, 2008; Öhrn, 2019; Rossmo, 2000). Analysts conduct geographic profiling in the intelligence phase of an investigation to provide probabilistic information about where and on whom an investigation might focus its search for evidence, but this information is not itself evidence of an offender's identity (Rossmo, 2021). The admissibility in court of geographic profiling outputs from either an algorithm or an analyst is beyond the scope of this research to investigate or comment on.

In sum, geographic profiling involves a range of inferences, made by analysts with algorithmic support, for intelligence, not evidential, purposes. Within this context, this thesis focuses on two inferences: working from either suspect locations or crime locations to infer which suspects to prioritise. As the first step towards broadening the theoretical and empirical basis for these inferences, Chapter 2 addresses this question: what systematic links exist between people's activity locations and their crime locations that could support these geographic profiling inferences from where to who in crime investigations?

CHAPTER 2

A Theoretical Model Linking Offenders' Activity and Crime Locations

If we are to move from where a crime was committed to who committed it, we need to first understand how offenders choose their crime locations. Naturally, crime locations are determined by the availability of crime opportunities, but offenders need to be aware of these opportunities—they need to be part of offenders' mental maps. These mental representations vary between individuals, forming an important determinant of whether and how they identify crime opportunities and hence of their crime locations. This research therefore focused on the links between individuals' mental maps and their crime locations, and further, it focused on the observable elements of offenders' mental maps: the physical locations where they have carried out their routine activities. Within this scope, the first study of this thesis addressed a theoretical question: how do offenders' different routine activity locations influence their crime location choices? The following paper, published in *Crime Science*, presents a theoretical model—derived from existing theoretical and empirical literature—that explains how offenders' prior activities in locations affect the development of knowledge about those locations that is brought to bear on future crime location decisions.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Note: A paper cited in this article (Bernasco & Van Dijke, 2020) has been retracted. An amendment to the article removing this citation is pending.

Curtis-Ham *et al. Crime Sci (2020) 9:23*
https://doi.org/10.1186/s40163-020-00132-7

Crime Science

THEORETICAL ARTICLE

Open Access

A framework for estimating crime location choice based on awareness space

Sophie Curtis-Ham[1*], Wim Bernasco[2,3], Oleg N. Medvedev[1] and Devon Polaschek[1]

Abstract

This paper extends Crime Pattern Theory, proposing a theoretical framework which aims to explain how offenders' previous routine activity locations influence their future offence locations. The framework draws on studies of individual level crime location choice and location choice in non-criminal contexts, to identify attributes of prior activities associated with the selection of the location for future crime. We group these attributes into two proposed mechanisms: reliability and relevance. Offenders are more likely to commit crime where they have reliable knowledge that is relevant to the particular crime. The perceived reliability of offenders' knowledge about a potential crime location is affected by the frequency, recency and duration of their prior activities in that location. Relevance reflects knowledge of a potential crime location's crime opportunities and is affected by the type of behaviour, type of location and timing of prior activities in that location. We apply the framework to generate testable hypotheses to guide future studies of crime location choice and suggest directions for further theoretical and empirical work. Understanding crime location choice using this framework could also help inform policing investigations and crime prevention strategies.

Keywords: Awareness space, Crime location choice, Crime pattern theory, Rational choice theory, Routine activity nodes

The location of crime is not random; as we elaborate in this paper, offenders' decisions about where they commit crime follow predictable patterns, that reflect decision-making processes common to human spatial behaviour more generally. In the context of criminology, understanding these processes at the individual level enables predictions that can inform policing strategies: where might a given person offend next? Who is more likely to have committed crime in that location? Much has already been done to advance our understanding of these processes. Foundationally, Crime Pattern Theory explains that offenders commit crime where crime opportunities coincide with their 'awareness space' around 'activity nodes'; the places they learn to know during everyday activities (Brantingham & Brantingham, 1991, 1993). Further, a growing body of empirical research (discussed below) reveals that the associations between activity nodes and crime locations vary for different activity nodes. These variations hint at systematic mechanisms that mediate the relationship between activity nodes, opportunities, and crime. But these mechanisms have not yet been articulated in a coherent framework that explains how and when different activity nodes influence crime location choice. Drawing on both criminological scholarship and a broader literature in geography and psychology, this paper proposes a theoretical framework to systematise our current understanding and guide future research. The paper begins by setting the theoretical context of the framework. We then summarise the framework in a formal model and introduce its elements, before discussing the empirical support for each element in turn. The paper concludes by teasing out testable hypotheses for empirical exploration and suggesting further directions in which to expand the framework.

*Correspondence: SC398@students.waikato.ac.nz
[1] Institute of Security and Crime Science and School of Psychology, University of Waikato, Knighton Road, Hamilton 3240, New Zealand
Full list of author information is available at the end of the article

Context: awareness space, activity nodes and location choice

Our framework is set in the context of Crime Pattern (CP) and Rational Choice explanations of crime location choice. CP theory holds that offenders commit crime where crime opportunities coincide with their awareness space around and between routine activity nodes, such as their homes, schools, workplaces, shopping and recreation locations (Brantingham & Brantingham, 1991, 1993). Awareness space is, most literally, the places of which a person is aware[1] and the related term 'activity space' refers to the subset of locations that people directly experience during their activities (Brown & Moore, 1970; Horton & Reynolds, 1971), consisting of activity nodes where people spend nontrivial amounts of time carrying out activities (Golledge, 1978; Golledge & Stimson, 1997) and the paths (routes) between them.

Awareness space encompasses more than activity space. First, it includes the area normally within visual range of activity space. Second, awareness space exceeds activity space where it is generated from sources other than direct experience; for example, through word of mouth, news or other media (Brown & Moore, 1970; Golledge & Stimson, 1997; Horton & Reynolds, 1971). When deciding where to carry out a future activity (i.e., making a location choice), we can either return to somewhere in our activity space, or explore somewhere new that is already in our awareness space from secondary sources, or of which we have no prior knowledge.

CP Theory views decisions about where to commit crime as products of individuals' activity and awareness spaces, and the structural backcloth of opportunities and environmental features that impact their attractiveness and accessibility (Brantingham & Brantingham 1993; Brantingham et al. 2008). We supplement the former element of CP Theory in describing systematically the attributes of individuals' activity nodes which may make crime more or less likely in their vicinity.

Rational Choice (Clarke & Cornish, 1985; Cornish & Clarke, 1986) provides the psychological decision-making model on which our framework rests. It holds that the choice of crime location follows a rational decision-making process; the framework considers how prior activities contribute to that process. In the decision process, benefits, risks and costs associated with alternative locations inform a calculation of their usefulness for a given activity; locations with high perceived utility are more likely to be chosen.[2] These decisions are not necessarily objectively optimal; offenders' knowledge of the alternatives is incomplete, limited to their awareness space. And, as our framework details, offenders' judgments of locations' crime utility are affected by the extent and nature of their knowledge of these locations, based on their prior activities at them and indirect information sources.

A model of activity node influence on crime location choice

Within this theoretical context, we focus on *how* knowledge of locations generated by previous activities shapes the decision about where to commit crime. To a lesser extent, we discuss the role of awareness space derived from indirect sources, and exploratory behaviour. In this section we briefly introduce the framework, before elaborating on its elements in their own respective sections. From the crime location choice literature, we identify a range of factors associated with people's prior activities that predict their subsequent offending in the same locations as, or near to, those activities. We group these factors into two clusters, reflecting quantitative attributes of prior activities in a place (e.g., their frequency), and the quality or nature of those activities (e.g., what the activity was). These groupings lead us to suggest two mechanisms through which prior activities influence (crime) location choice. First, the frequency, recency, and duration of prior activities affect how well a location is known, and thus how much the knowledge of that location can be relied on in calculating (crime) utility. Second, the nature of those activities affects what is known about a location, and the type of activities it affords. Further, the more similar the prior activity to the activity involved in the current choice scenario, the more likely it is to have generated knowledge suggesting the location's usefulness for the activity. We attach the terms 'reliability' and 'relevance' to describe the two mechanisms: offenders are more likely to commit crime where they have knowledge that is both reliable and relevant to the particular crime.

Although we provide more detailed explanation of these concepts below, a concrete example may help at this point to illustrate the distinction, and interaction,

[1] The terms 'action space' (Horton & Reynolds, 1971; Wolpert, 1965), 'awareness space' (Brown & Moore, 1970), 'mental map' (Gould, 1966, 1973; Gould & White, 1986) and 'cognitive map' (Downs & Stea, 1973; Tolman, 1948) are often used synonymously in this literature. Herein, only 'awareness space' is used for consistency with the environmental criminological literature. Similarly, the term 'activity node' or just 'node' is used in preference to the synonym 'anchor point' that also appears in the literature.

[2] We use the term utility as defined in the descriptive and predictive (as opposed to normative) version of expected utility theory. Expected utility theory describes and predicts how people make choices in situations where they are uncertain about the outcomes of their actions (Lattimore & Witte, 1986). Their uncertainty is a function of information constraints. In the case of the decision where to commit an offense, offenders can only be aware of a subset of all potential offense locations and can have only limited information on their characteristics, and thus on the possible benefits and costs of offending in each of these locations.

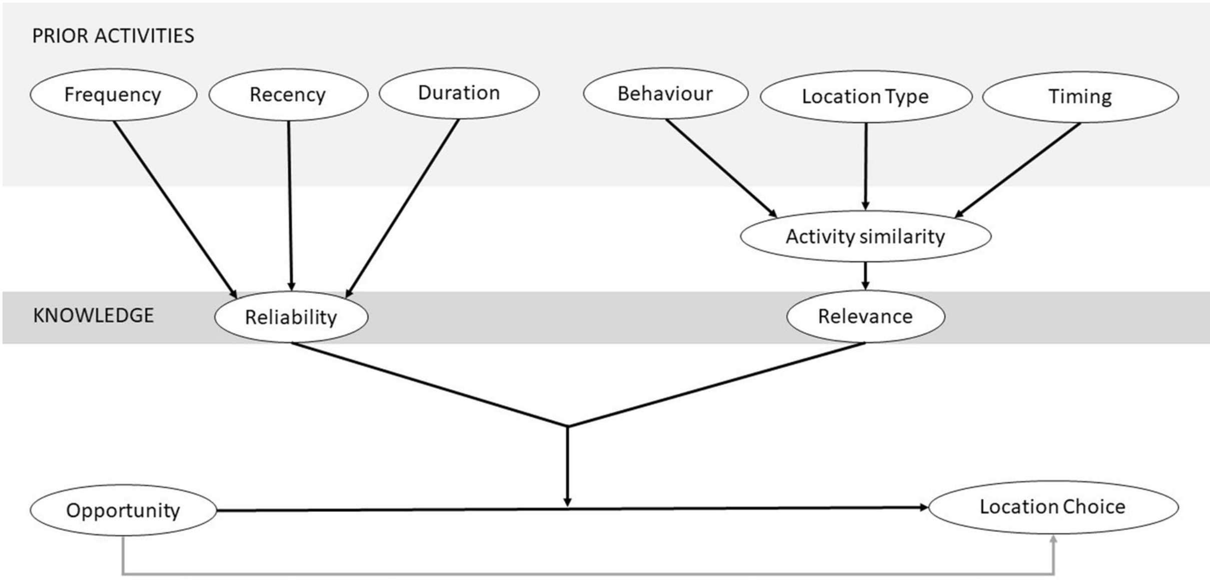


Fig. 1 Model of the influence of prior activities on location choice. Location choice is a product of both opportunity and knowledge of the opportunity generated during prior activities. The frequency, recency and duration of prior activities affect the reliability of that knowledge; the behaviour (e.g. crime, recreation, work), type of location, and timing involved in prior activities affect the relevance of that knowledge to the location decision

between the two mechanisms. A burglar might have a high level of familiarity with a neighbourhood, A, built up over many previous visits to relatives who live in the area. They have reliable knowledge of the location, but not having committed a burglary there before, they do not have specific knowledge of its potential for burglary. Neighbourhood A has high reliability but low relevance. Neighbourhood B, in contrast, the burglar has only visited once, and committed a burglary on that occasion. It definitely has burglary potential, but they might have just been lucky; the burglar does not know it well enough to be confident returning. Neighbourhood B has high relevance but low reliability. Neighbourhood C is just as familiar as neighbourhood A, as the burglar lived there until recently, and they have committed burglaries there in the past. Knowing it well, and knowing from direct experience that it has attractive burglary opportunities, the burglar more likely considers neighbourhood C more of a sure thing, and therefore the better option. Neighbourhood C has high reliability and high relevance. Neighbourhood D, which has abundant opportunities but lies outside of awareness space, is not considered. The offender has no reliable or relevant knowledge of Neighbourhood D.

Figure 1 summarises the framework in a formal model illustrating the theorised pathways from attributes of prior activities to crime location choice. In the model, knowledge generated by prior activities moderates the relationship between crime opportunities and location choice; opportunity is a necessary condition for a location to be chosen but is not sufficient, because offenders need to have knowledge of the opportunity (Brantingham & Brantingham, 1991). Opportunity is also directly linked to location choice in the model (grey line in Fig. 1), to reflect that people sometimes choose to explore new locations rather than exploit previously visited ones. It captures situations where crime occurs outside of offenders' existing activity space, either as a result of indirect sources of awareness, or exploratory behaviour. Since our focus is on the role of offenders' prior activities in generating awareness of criminal opportunities, the framework does not elaborate on what makes situations crime opportunities. For the purposes of our model, the paths between activity nodes are also locations in which people have previously conducted activities (travel between activity nodes) and their influence on crime location choice is affected by the same factors.

We suggest these factors predict the likelihood that an offender will return to a given activity location and commit an offence, whether that offence is planned or opportunistic (see Brantingham & Brantingham, 2012; Cromwell et al., 1990; Elffers, 2004 for discussions of such 'target search' typologies). Planned offences typically involve first a decision of whether to offend, then where. Knowledge of prior activity nodes informs the assessment of where. Offences committed opportunistically or provoked spontaneously generally involve a decision whether to offend, given a current location. The likelihood of an offender being at that location, at that time, and identifying the crime opportunity, is also a product of their previous activities there, and knowledge generated by those activities will inform the assessment of the risks and rewards associated with the offence. The more reliable and relevant that knowledge, the more likely a future crime opportunity will be identified and acted on.

The next two sections review the empirical literature which underpinned the development of the reliability and relevance pathways in the framework, drawing on crime-specific research supplemented with research in non-criminal domains. The crime specific evidence comes primarily from a series of studies using Discrete Spatial Choice Models (DSCMs), a form of Discrete Choice Model. Discrete Choice Models apply a form of logistic regression to decision-makers' choices from a set of alternatives, to identify the attributes of alternatives that are associated with increased likelihood of being chosen (Ben-Akiva & Lerman 1985; McFadden, 1984). Non-spatial discrete choices can include such decisions as between cereal brands (e.g., Nevo, 2001), travel modes (e.g., Nguyen et al., 2017) or service providers (e.g., Ida and Kuroda, 2006). In DSCMs the choice set contains alternative locations where the decision-maker can choose to carry out an activity; for example, commit a crime (Ruiter, 2017), shop (e.g., Hillier et al., 2017), or move residence (e.g., Ben-Akiva & Bowman, 2016). In crime location choice, DSCMs calculate the relative probability of a location being selected for offending based on its proximity to a given activity node, controlling for proximity to other activity nodes and other features of the location (Ruiter, 2017).

Reliability factors

Three quantitative attributes of prior activity nodes have strong empirical support for their association with crime location choice and are thus included in the framework under the 'reliability' pathway. *Frequency* (how often) refers to the number of visits per time period (e.g., "daily", "twice a week" or "four times per year"). *Recency* refers to amount of time elapsed since the last visit (e.g., "since yesterday" versus "since last month"). *Duration* refers to the length of a period during which the individual has been visiting an activity node (e.g., "for three months" versus "for five years"), as opposed to the average length of time spent per visit.[3]

[3] We note average visit duration as an additional potential activity attribute but choose to focus on those attributes that have more empirical evidence relative to crime location choice. Future research could explore its contribution to the model.

No DSCM study to date has examined frequency, recency and duration of activity nodes simultaneously to separate out their individual and combined contributions. Considering all activity nodes (home, others' homes, school/work and leisure) combined, Menting, Lammers, Ruiter and Bernasco (2020) and Bernasco (2019) found that offences were more likely to occur near activity nodes that were visited more frequently. Further, the more recently and longer that offenders have resided at an address, the more likely they are to offend nearby, with recency and duration producing an additive effect (e.g., Bernasco, 2010; Bernasco & Kooistra, 2010; Menting et al., 2016). The same pattern holds for close family members' homes and recency of residence (Menting et al., 2016). There are similar findings for both the number (frequency) and recency of prior crimes (Bernasco et al., 2015; Lammers et al., 2015; Long et al., 2018).

Studies revealing differences between different *types* of activity node are also informative. All other things being equal, we would expect activity nodes that are typically visited more frequently, recently or over longer periods to have higher odds of crime occurrence nearby than other activity nodes. In line with this expectation, the odds of crimes being committed near offenders' home nodes are consistently higher than near their family members' home nodes (Menting, 2018; Menting et al., 2016). Further, more offences occur along the home-work path than other paths: this path being the most well-trodden (Rengert & Wasilchick, 1985; Ruiter & Davies, 2018).

The roles of frequency, recency and duration in crime location choice are reflective of their roles in location choice in general. The more often people have visited a location, the more likely they are to return (Pappalardo et al. 2015), even if similar opportunities exist closer to home (Hannes et al.. 2008). The compounding effect of frequency is also evident from the habitual, automated nature of day to day travel choices (Gärling & Axhausen. 2003; Hannes et al., 2009). Recency increases the odds of choosing a location, whether frequently visited previously or not (Alessandretti et al., 2018; Barbosa et al., 2015). These patterns have been confirmed across a range of countries and geographic scales (Yan et al., 2017). Further, incorporating activity nodes that have been in awareness space for a longer time, but are less frequently visited, improves predictions of future location choice (Song et al., 2010).

We suggest that frequency, recency and duration operate on location choice insofar as they affect how well we know locations and thus how much we can rely on that knowledge making a location decision. In short, the more often, recently, or longer we have visited a location, the more reliable our knowledge of it. Reliability, as we use the term, refers to perceived or subjective reliability, because one can be wrong about the accuracy of one's knowledge about a location, leading to over-reliance on inaccurate knowledge or under-reliance on accurate knowledge.[4] This perception, however, may be unconscious. Spatial decisions are frequently habitual or automatic and not experienced as conscious deliberations, with assessments of reliability thus informing these decisions implicitly (Downs & Stea, 1973; Hannes et al., 2006, 2008, 2009).

Evidence from interviews designed to elicit offenders' implicit (or explicit) decision-making process highlights the importance of having reliable knowledge of potential crime locations. Rengert and Wasilchick's (1985) seminal research found that US burglars favour familiar areas, tending to extend out from those familiar areas when searching for crime opportunities. In two more recent studies, burglary and prolific property offenders in the US and UK respectively were asked to rate their level of familiarity with areas on a map, along with how likely they were to offend in each area (Rengert & Wasilchick, 2000) or which areas were the best (to worst) for them to offend in (Summers et al., 2010). Both studies found large overlaps between locations rated as more familiar and more likely or attractive for offending (see similarly, Costello & Wiles 2001; Wiles & Costello 2008). The importance of knowing an area well, through either non-criminal activities or through multiple reconnaissance visits before offending is also evident from offenders' narrative accounts reported by Summers et al. (2010, pp. 266–267):

> *"...you tend to stick to the areas that you know well. (Offender RP07).*
> *... you want to know the best routes to get out quick, and, you know, so you trawled it for a couple of days and then you'll go back... (Offender RC02)".*

There is also more general evidence that frequency, recency and duration affect the reliability of people's location knowledge, supporting our proposed mechanism. For example, Golledge (1978) compared people's self-assessed familiarity with different areas, judgments of the relative spatial layout of various locations within those areas, and the actual spatial layout of those locations. Both familiarity (i.e., perceived reliability) and accuracy of spatial judgments were highly correlated with how long participants had resided in an area. Likewise, accuracy of spatial knowledge is higher for locations which have been visited before, and home nodes of longer duration (Spector, 1977 cited in Golledge, 1978). More recent studies confirm positive relationships between self-reported familiarity with areas and both duration and recency of exposure to the area, and with proximity to high frequency activity nodes such as home, work, and commercial areas (Zhang et al., 2016, 2019). Both self-assessed way-finding ability and accuracy of spatial knowledge are also positively related to length of residence in a location and the number of trips to a location per week (Chorus & Timmermans, 2010).[5] Consistent with our model, therefore, more frequent, recent and enduring activity nodes produce more accurate knowledge that is, generally, more likely to be assessed as reliable to inform future location choices.

Relevance factors

We find empirical support for three qualitative attributes of prior activity nodes that appear to affect crime location choice. We can think about prior activities in terms of the behaviour involved (e.g., criminal vs non-criminal, or the specific type of crime), the type of location involved (e.g., residential vs commercial), and their timing (e.g., time of day, day of week). The more similar a prior activity is in any of these respects to the activity involved in the location choice, the more likely we are to choose to return to that location.

The clearest example, where crime location choice is concerned, is prior crime. Prior crimes potentially constitute the closest degree of similarity, *behaviourally*, between a prior activity in a location and a future choice scenario, thereby providing the most relevant knowledge of the location's crime opportunities. Correspondingly, prior crime nodes are generally more likely to be chosen than home or family home nodes (e.g., Frith, 2019; Menting et al. 2016; Vandeviver & Bernasco, 2019) and other non-home nodes (Bernasco 2019).[6] The odds are even higher if the prior crime is of the same broad crime

[4] For example, information processing errors can occur during encoding, storage and retrieval of information about the environment (Chorus & Timmermans, 2010; Golledge, 1999; Golledge & Stimson, 1997; Lloyd & Cammack, 1996; Mark et al., 1999). But, as decades of research on heuristics and biases have shown, people are often not aware of the fallibilities produced during information processing (Kahneman, 2003).

[5] Interestingly, the average duration of individual trips to a location, although predictive of accuracy, was not related to self-assessed spatial knowledge. Individual trip duration is distinguished from our concept of duration (total length of a person's association with a given node); see note 3 above.

[6] One exception is when violent offences are considered separately, in which case the odds are higher at home or family home nodes (Menting, 2018). In our model's terms, the family violence offences within this category would involve similarity in activity between the past family dynamics in a home location, and the future situation giving rise to the violence. If family related violence offences were removed, we predict that home nodes would become less influential relative to prior crime nodes, in line with the pattern for property and other crimes that are less likely to involve family members.

category as the future crime (van Sleeuwen et al., 2018).[7] Future studies could leverage research into the similarity between specific crime types (Kuang et al., 2017) in exploring the relationship between crime behaviour similarity and crime location choice.

There is also evidence that although prior crime success increases the likelihood of returning to its location for future crime, negative outcomes of prior crimes are not a lasting deterrent to returning to their location. Long et al., (2018) found that offenders were less likely to commit a robbery in the same location as a prior robbery they were arrested in the act of, than if they (initially) got away with it. However, they were still more likely to offend in a prior crime location where they were arrested, than a non-prior crime location. The observation that often crime returns to previous levels following the deterrent effects of police presence (Banerjee et al., 2019; Sorg et al., 2013), further suggests that offenders return to prior crime locations once they believe 'the coast is clear'.

We can also look outside prior criminal behaviour to a range of prior activities that could be more or less similar to any given crime activity. For example, youth 'hangout' nodes could be seen as a step removed from prior crime nodes in terms of the behavioural similarity of prior and future activities. Studies using space–time budget surveys to collect data on the locations and contexts of young offenders routine and delinquent activities have found that they are more likely to offend at or near hangout locations which involve unsupervised and unstructured peer-group activities, even compared to higher frequency nodes such as school and home (Bernasco et al., 2013; Miller, 2013; Wikström et al., 2010; see also Bichler et al., 2012). Further, the odds of young offenders committing crime near any activity node are almost four times greater than the odds near home nodes alone (Menting et al., 2020). 'Hangout' activities are more likely to involve delinquent and boundary pushing behaviour with greater similarity to criminal activity than activities in constrained settings; they are also more likely to involve crime attracting/generating location types such as malls, entertainment facilities and other commercial hubs (Bichler et al., 2014). In a similar vein, interviews with drug dependent residential burglars revealed that their offences tended to cluster around places they purchased drugs (Rengert, 1996). Reconnaissance activities, where offenders seek to develop knowledge of crime opportunities to which they can later return (Rengert & Wasilchick 1985; Summers et al., 2010; van Daele et al., 2012), are a further example of prior activities that could be considered behaviourally similar to the crime itself by comparison to other routine activities.

Identifying the types of prior non-criminal activities that are associated with crime location choice, based on the similarity between particular non-criminal behaviours and specific crimes, requires further investigation. We speculate, for example, that locations where a person has been involved in a crime as a victim or witness may be 'similar' in the sense of involving the same crime, but in a different role. For example, a teenager witnessing a friend shoplift may return to that shop if later motivated to shoplift.

Turning to *location type*, at the aggregate level it is well-established that different types of crime are correlated with different types of location, at scales ranging from specific premises such as bars and shops to areal level land use (e.g., Tillyer & Walter, 2019; Weisburd et al., 2016). Isolating the influence of prior activities in different types of location therefore requires studying specific crime types separately. The only DSCM study comparing activity node influence across specific crime types[8] found that residential nodes (current and prior homes) had a stronger relationship with location choice for residential burglary than for thefts of cars and robbery (Bernasco, 2010).[9] Similarly, offenders tend to travel farther (on average) to commit commercially than residentially focused crimes, suggesting lower home node influence in commercial crime location choices (Ackerman & Rossmo 2015; Townsley, 2016). These results support the suggestion that offenders are more likely to offend near activity nodes of the same location type (residential versus commercial) as that targeted by the offence. Locations can also be similar in terms of their social rather than built

[7] Here the discrete choice research converges with literature on the near repeat phenomenon (e.g., Bernasco, 2008; Johnson et al., 2009) and crime linkage (e.g., Tonkin et al., 2011, 2012), which confirms that offences that are close in space and time are more likely to have been committed by the same offender. This pattern is often explained in terms of a 'boost' mechanism, whereby successful crime commission causes a follow-up crime by motivating the offender to return to the same location, and a 'flag' mechanism, whereby repeated offending at the same location (possibly but not necessarily by the same offenders) is merely a symptom indicating that the location continuously provides criminal opportunities (Johnson et al., 2009; Lantz & Ruback, 2017; Pease, 1998). These explanations are consistent with our model: the boost effect is causal, and a product of offenders' highly recent (reliable) and relevant (prior crime) knowledge; the flag mechanism is only a product of opportunity.

[8] Other DSCM studies either use aggregate groupings (e.g., all felonies, all violence, all property offences, or combining residential and commercial burglary) which mask variation between more specific crimes; or they focus on one crime (predominantly residential burglary), with cross-crime comparisons confounded by cross-study differences in jurisdiction, nodes and other co-variates included, and modelling methods.

[9] Assaults had an even stronger association with home and prior homes, likely due to the inclusion, and prevalence, of domestic violence (see also Menting, 2018). Combining domestic and other violence is likely to have masked differences that our model predicts, due to their typical settings (home/public places).

features, such as the degree of place management or guardianship (Felson, 2008). Similarity in these respects may also be salient to crime location choice: offenders would be more likely to return to places with similar social environments to those which facilitate a particular crime.

Similarity of prior activity *timing* also matters in crime location choice, at least for prior crimes. Offenders were approximately 46 times more likely to choose a location where they previously committed a crime of the same type, on the same weekend day, at the same hour of the day, than a location with no prior crime (26 times for weekdays; van Sleeuwen et al., 2018). These odds decayed steeply with decreasing timing similarity (e.g., within 2, 3 h, and so on, and different weekend or weekdays).

The pattern of returning to locations where prior activities match the current intended activity is seen in studies of non-criminal spatial behaviour. Most of people's activities occur in a small set of locations that are visited recurrently for the same purpose, at the same time of day, using the same transport mode (Hanson & Huff, 1988). A prior visit to a shopping location is a far larger predictor of future shopping location choice than the mere presence of shopping opportunities in a location (Sivakumar & Bhat, 2007; see also Arentze et al., 2008).

We suggest that prior activity similarity operates on location choice insofar as it reflects whether prior activities have generated knowledge that is relevant to the activity involved in the location choice. The more similar a prior activity is to the activity involved in a future choice, the more likely the prior activity is to generate knowledge of its location's utility that is generalisable to the future activity (i.e., more relevant knowledge). To use the language of CP Theory, similar prior activities are more likely to generate knowledge that matches an offender's mental template of a 'good' opportunity for future crime (P. L. Brantingham and Brantingham 1993). The generalisability of learning from past to future activities is a product of their similarity; broadly speaking, the more similar two situations or stimuli are, the more likely we are to generalise knowledge of one to the other (Gentner & Medina, 1998; Howard, 2000; Tenenbaum & Griffiths, 2001). To give a simple example, in determining where to dine out, we think first of places we have dined before, not the places we have purchased groceries; our knowledge from previous dining experiences is more relevant to future restaurant choices. Relevance, in our usage, refers to how much the knowledge of a location favours its usefulness for a given activity. In contrast to lay definitions of 'relevance', we use the word to indicate knowledge of good opportunities; knowledge of a lack of opportunities is less relevant knowledge. In the dining example, a restaurant where we previously experienced poor service would have less relevance to our decision than one where we previously experienced excellent service.

Relevant knowledge acquisition requires both the presence of opportunities and the generation of awareness of them through prior activities. The presence of crime opportunities can, to an extent, indicate an awareness of them (relevant knowledge). For example, considering the interaction of opportunity and home node proximity, Menting (2018) found that the odds of crime near home nodes were lower when there were fewer bars, restaurants and hotels—premises which can act as crime generators or attractors (Brantingham & Brantingham, 1995)—in the vicinity. Further, interview studies confirm that prior non-criminal activities can generate knowledge of crime opportunities (e.g., Clare, 2011; Cromwell et al., 1990; Wiles & Costello, 2008). For example, some burglars identify suitable neighbourhoods, or specific targets, through employment or social activities (Wright & Decker, 1994). But to our knowledge, no study has yet explicitly explored the link between prior activity *similarity* and offenders' knowledge of locations' opportunities for particular crimes.[10]

We can, however, interpret the studies discussed above linking activity similarity to location choice, in terms of how similarity of behaviour, location and timing generates relevant knowledge. Prior crimes naturally generate highly relevant knowledge applicable to decisions about where to commit future crimes, particularly if they involve the same crime. We also saw that experience of disutility (through arrest, or increased risk of arrest) leads to reduced odds of location choice, by comparison with successful offences. Consistent with our restaurant example, this negative experience produces knowledge that is less well matched to the ideal template for the crime, and thus less relevant than had the crime succeeded. But because the prior unsuccessful crime behaviour is more similar (to future crime) than prior non-criminal behaviour, the knowledge gained from unsuccessful prior crimes may still be more relevant than that gained during non-criminal activities, as reflected in the preference for prior unsuccessful crime locations over locations with no prior crimes (Long et al., 2018).

In terms of location type, locations are typically designed with specific uses (behaviours) in mind, so if the location type is the same, so is the behaviour. Exceptions exist when locations are designed for multiple purposes

[10] Several studies have quantified offenders' crime-relevant knowledge by asking them to rate how attractive locations are for a given crime, or how likely it is they would offend there (Rengert & Wasilchick, 2000; Summers et al., 2010) but did not compare these measures with the kind of activities carried out in those locations.

or when they afford activities, such as crime, for which they were not intended. Thus, knowledge of shops visited previously for legitimate purposes will be more relevant to a decision about where to shoplift than, say, a home or workplace node, despite the different behaviour involved in the activities (shopping/shoplifting). And knowledge of a residential area gained during prior house burglaries will be less relevant to a decision about where to burgle a commercial property, despite the behaviour (burglary) being similar. The greater association of residential nodes with residential burglary location choice (Bernasco, 2010), and apparent lesser influence of home nodes on crimes for which commercial areas present more opportunities (Ackerman & Rossmo, 2015; Bernasco, 2010; Townsley, 2016), is consistent with the role of location similarity in generating relevant knowledge.

Lastly, if the timing of previous exposure to a location does not match the timing of its opportunities for a given activity, we are less likely to identify those opportunities and thus gain relevant knowledge. For example, having only visited a shopping precinct during the day, we would be less aware of its nightlife affordances. Likewise, a prior burglary committed overnight provides knowledge of the location's overnight burglary opportunities, and less information of relevance to its daytime utility. The burglar is more likely to choose that location for a subsequent overnight burglary, than a subsequent daytime burglary (as seen in van Sleeuwen et al., 2018). Additionally, as argued by Ratcliffe (2006), paths connecting to the home node are traversed with greater temporal variability than other paths, generating greater exposure to differently timed opportunities. Home nodes are thus more likely to generate relevant knowledge of nearby crime opportunities than nodes visited only at particular times or on particular days, potentially explaining why even locations near home with few crime opportunities are more likely to be chosen than locations further afield with more crime opportunities (Menting, 2018).[11] However, there are exceptions to the tendency to return to prior activity nodes, which we now consider.

Crime location choice outside of activity space

With increasing distance from prior activity locations, activity space transitions to awareness-only space, then to exploratory, unknown space. Absent direct measurement of awareness space, we do not know where these lines are drawn, and thus which crimes fall into each space.[12] But we do know that crimes are most likely to occur in the immediate vicinity of prior activity nodes, and that this likelihood declines with distance from these activity nodes (Bernasco, 2019; Menting et al., 2020) and their connecting paths (Reid et al., 2014; Rengert & Wasilchick, 1985; Ruiter & Davies, 2018), as the likelihood of being in exploratory space increases. Crime in novel locations is, therefore, a novelty.

As Bernasco (2018) points out, this 'distance decay' pattern for crime reflects non-criminal spatial behaviour patterns. Most people exhibit a predominant pattern of returning rather than exploring, basing their activities around a few highly frequented nodes (Alessandretti et al., 2018; González et al., 2008; Pappalardo et al., 2015) and displaying distance decay, with new activity nodes chosen with decreasing frequency at increasing distances from existing activity nodes (Hasan et al., 2013; Kang et al., 2012; Sivakumar and Bhat, 2007).

In our framework, distance decay in location choice probability could reflect distance decay in the reliability and relevance of location knowledge. The closer a location is to a prior activity location, the more likely we are to have reliable knowledge of it, having been exactly at, near or passed through it en route. Further, applying Tobler's first law of geography (Tobler, 1970), whereby the similarity of locations increases the closer they are to each other, the closer a location is to a prior activity location, the more likely it is to afford the same opportunities and involve the same risks, and thus to generate relevant knowledge. Correspondingly, empirical research confirms that the closer it is to a node, the more likely a location is to be in self-reported awareness space (Horton and Reynolds, 1971; Zhang et al., 2019). But on those occasions where offending occurs away from activity space, what drives location choice?

Co-offending generates awareness of crime opportunities away from an offender's activity space through the sharing of knowledge between current or prior co-offenders. DSCM studies confirm the former: crimes committed in groups are more likely to occur near at least one of the current co-offenders' present or past homes (Bernasco, 2006; Lammers, 2018) or prior crimes (Lammers, 2018; Vandeviver & Bernasco, 2019) than at locations in no group member's activity space. Confirming the latter, Lantz and Ruback (2017) found that repeat burglaries of the same property were more likely to be committed by a previous co-offender of the initial offender, than by a burglar with no co-offending connection.

These co-offending effects exemplify the influence of location knowledge generated through social networks in general. Toole et al. (2015) demonstrated using cell phone data that when people travelled to locations they had not

[11] The high reliability of knowledge around home nodes also likely contributes to this finding.

[12] Most studies *estimate* awareness space from activity locations, rather than measuring it directly. See Summers et al. (2010) for a rare exception, where offenders' awareness space was directly measured via ratings of familiarity.

previously visited (during the data collection period), their novel destinations could be predicted from the locations frequented by their social contacts (connected via their phone calls).

Exploration outside existing awareness space to commit crime is also evident.[13] Offenders may deliberately seek to expand their awareness space, venturing away from familiar locations in search of offending opportunities (Brantingham and Brantingham, 1991, 1993; Rengert & Wasilchick, 1985). Offenders new to a city or country may engage in exploratory behaviour prompted by a lack of opportunities within their limited awareness space (van Daele & Vander Beken, 2011). Conversely, exploration might occur as a product of confidence associated with criminal expertise (Clare, 2011; Nee, 2015) or of the need to 'forage' for criminal opportunities elsewhere when prior crimes have resulted in depleted opportunities or increased risk of capture (Johnson et al., 2009; Johnson & Bowers 2004). Opportunistic offences that occur 'on the spur of the moment' in the context of non-criminal activities in places not previously visited will also appear exploratory.

Criminal exploration driven by a lack of opportunity knowledge is consistent with general human mobility patterns: people explore more when they have fewer return-worthy nodes in their activity space, and less as the number of nodes in their activity space grows (Pappalardo et al., 2015). Further, where people explore can be predicted from the popularity of a location amongst the population (Hasan et al., 2013; Pappalardo et al., 2015; Wang et al., 2019), suggesting the potential to predict individual offenders' crime locations in exploratory space from locations' aggregate popularity amongst offenders (i.e., where similar crimes concentrate).

Examining the circumstances when offenders' location choices—based on secondary sources, or deliberate exploration—cannot be predicted from prior activity nodes presents an opportunity for further theoretical development. In the final section, we highlight some additional avenues for both empirical and theoretical exploration.

[13] Long home-crime distances do not necessarily imply exploration outside awareness space. Evidence of long distances between offenders' homes and their crimes from 'Journey to Crime' studies (as summarised in Ackerman & Rossmo, 2015; and Xiao et al., 2018 for example), or of offenders who 'commute' from homes located outside of the area in which they offend (Canter & Larkin, 2008), is uninformative on this point, insofar as it fails to account for the presence of other activity nodes (see for example, Wiles & Costello, 2008).

Future research directions

To help guide future crime location choice studies, we apply the framework to generate some examples of new (or newly framed) predictions as to the relative influence of individual nodes and of different activity node types (home, work, etc.), thus demonstrating its fertility (Ward et al., 2005). We also make other suggestions for future research and research methods. Turning first to the hypotheses, research testing these could measure reliability and relevance through offenders' self-report (e.g., ratings of familiarity and locations' crime utility) or use the attributes of prior activities (frequency, recency, duration and activity similarity) as proxies.

H1: Nodes (and paths) with high reliability and relevance (regardless of what *type* of node they are) will have stronger associations with crime location choice than those lower on either or both of these dimensions.

H2: Given the likely abundance of residential targets around residential nodes, residential nodes (i.e., the homes of offenders and their family/friends), through a combination of high reliability and high relevance, will have a stronger association with crime location choice for crimes targeting residential properties, than for non-residential crimes.

H3: For crimes targeting or typically occurring near commercial properties, nodes such as work and recreation that tend to be in commercial areas will have a stronger association with crime location choice than residential nodes.

H4: The degree of association between prior crime nodes and crime location choice will reflect the similarity between the prior and future crime. For example, the location of a prior domestic assault is unlikely to have much bearing on where an offender will shoplift; a prior theft location will be more predictive; a prior shoplifting location even more so.

H5: The more reliable and relevant nodes offenders have, the less likely they will be to offend in places they have not been before.

H6 Offences occurring outside of activity space are more likely to occur near co-offenders' and other associates' activity nodes, or where similar crimes concentrate.

Opportunities for further theoretical development arise from several limitations worth mentioning. First, completeness: we focused on activity attributes explored in previous literature, which may simply represent those that are more easily measured. There is opportunity to expand the framework to identify, and quantify, additional activity attributes and variables which moderate the effects of reliability and relevance. As noted, other environmental and individual factors impact the processing of spatial information and thus reliability. Additionally, variation in offenders' ability to generalise past activities to future crime is likely, given differences between novice and expert offenders' recognition of cues to target attractiveness (Clare. 2011; Nee. 2015; Nee and Meenaghan. 2006) and variation in learning generalisation more generally (McDaniel et al.. 2014). Several DSCM studies have found individual differences in the extent to which prior activity nodes are associated with location choice (Frith. 2019; Frith et al.. 2017; Townsley et al.. 2016), which may reflect individual differences in the acquisition of reliable and relevant knowledge or in reliability/relevance preference thresholds.

Further elaborative work might also formulate how steeply the probability of crime location choice declines with distance for different nodes. The shape of the distance decay function is typically non-linear for the home node (Bichler et al., 2011; Hammond & Youngs, 2011; Smith et al., 2009), but just as the peak height of this curve (i.e., probability of location choice) varies between nodes, so may its gradient (as demonstrated by Brantingham and Tita, 2008 in a simulation study). Likewise, future work might explore whether, and when, a 'buffer zone' of reduced crime probability, as sometimes appears in the immediate vicinity of home nodes (Bernasco & Dijke 2020), appears around non-home nodes.[14]

A final theoretical avenue to highlight relates to crime location choice outside of activity space. For example, how do indirect sources such as online maps and other location information aid the identification of criminal opportunities?[15]

Lastly, on a methodological note, future empirical research should corroborate the findings on which the framework was based, using a wider range of measurement and modelling methods. DSCM studies to date either measured small subsets of activity nodes using large administrative datasets, or measured all activity nodes in small survey samples, which precluded comparison of different node types. Future studies could explore 'big data' sources (such as those used by the human mobility studies cited here), expand survey samples, identify additional administrative data on activity nodes (e.g., Authors, under review), and innovate using existing offender location-monitoring data (e.g., Rossmo et al., 2012). Each data source provides potential to operationalise variables in this framework, perhaps partially in isolation, but painting a fuller picture when combined. DSCM involves assumptions, such as decision-makers considering all choice alternatives and deciding based on utility maximisation, which some argue are not applicable to all location choices (Arentze & Timmermans, 2005; Golledge & Stimson, 1997; Hannes et al., 2012; Ruiter, 2017). As with any theory building endeavour, its predictions need to be robust to testing via a range of methods.

Conclusion

This paper has presented a systematic framework within which to consider the causal relations between activity nodes, opportunity, and crime location choice. It contributes to the ongoing elaboration of environmental criminology theories. But its contribution is more than theoretical. From a practical perspective, the framework enables predictions about which of an offender's prior activity nodes is more likely to be near a given offence. Such predictions can be used in the context of geographic profiling (Knabe-Nicol & Alison, 2011; Rossmo, 2000, 2014) in police investigations. For example, identifying which prior activity nodes may be more salient to a given crime can inform the prioritisation of suspects, given knowledge of their different activity nodes. Understanding the relative influence of different activity nodes can also help in formulating sentenced offenders' supervision conditions and risk management planning, by identifying places of higher risk for individual offenders, which should be avoided. For example, offenders on electronic monitoring might be restricted from entering areas identified as likely offending locations based on their prior activities there. Of further significance, since crime location choice exemplifies general location choice processes, this framework has potential wider application to non-criminal behaviour studied in the fields of human mobility and urban planning.

Acknowledgements
We are grateful to all reviewers who provided feedback on earlier versions of this paper, which helped to refine our ideas and improve the manuscript.

[14] Buffer zones may appear where offenders need to minimize the risk of recognition by victims or witnesses (Rossmo, 2000) but can also reflect a lack of crime opportunities or specific target search patterns. In the former case, there could be a threshold at which high frequency nodes become less relevant due to the higher risk of identification, suggesting the location's disutility for offending. We also note that buffer effects appear when considering the frequency of travel to *non-criminal* activity nodes at different distances from home (Bichler et al., 2010). However, we suggest a buffer zone would be unlikely for prior crime nodes, given that near-repeat offences against neighbouring targets are often committed by the same offender (Bernasco, 2008; Johnson et al., 2009).

[15] Evidence of the use of Google Maps to aid crime location choice has been documented, but its prevalence is not known (Vandeviver, 2014).

Authors' contributions
SCH developed the theoretical framework and drafted and revised the manuscript. WB, OM and DP reviewed and provided feedback on all drafts of the manuscript. All authors read and approved the final manuscript.

Funding
This research forms part of the first author's PhD thesis, which is funded by a University of Waikato doctoral scholarship.

Availability of data and materials
Not applicable.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Institute of Security and Crime Science and School of Psychology, University of Waikato, Knighton Road, Hamilton 3240, New Zealand. [2] Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands. [3] School of Business and Economics, Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

Received: 24 June 2020 Accepted: 18 October 2020
Published online: 04 November 2020

References

Ackerman, J. M., & Rossmo, D. K. (2015). How far to travel? A multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology, 31*(2), 237–262. https://doi.org/10.1007/s10940-014-9232-7.

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility. *Nature Human Behaviour, 2*(7), 485–491. https://doi.org/10.1038/s41562-018-0364-x.

Arentze, T. A., Dellaert, B. G. C., & Timmermans, H. J. P. (2008). Modeling and measuring individuals' mental representations of complex spatio-temporal decision problems. *Environment and Behavior, 40*(6), 843–869. https://doi.org/10.1177/0013916507309994.

Arentze, T. A., & Timmermans, H. J. P. (2005). Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. *Transportation, 32*(4), 321–340. https://doi.org/10.1007/s11116-004-7964-1.

Banerjee, A., Duflo, E., Keniston, D., & Singh, N. (2019). *The efficient deployment of police resources: Theory and new evidence from a randomized drunk driving crackdown in India (No. w26224)*. Cambridge, MA : National Bureau of Economic Research. https://doi.org/10.3386/w26224.

Barbosa, H., de Lima-Neto, F., Evsukoff, A., & Menezes, R. (2015). The effect of recency to human mobility. *EPJ Data Science, 4*(1), 1–14. https://doi.org/10.1140/epjds/s13688-015-0059-8.

Ben-Akiva, M. E., & Bowman, J. L. (2016). Integration of an activity-based model system and a residential location model. *Urban Studies*. https://doi.org/10.1080/0042098984529.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

Bernasco, W. (2006). Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling, 3*(3), 139–155. https://doi.org/10.1002/jip.49.

Bernasco, W. (2008). Them again?: Same-offender involvement in repeat and near repeat burglaries. *European Journal of Criminology, 5*(4), 411–431. https://doi.org/10.1177/1477370808095124.

Bernasco, W. (2010). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology, 48*(2), 389–416. https://doi.org/10.1111/j.1745-9125.2010.00190.x.

Bernasco, W. (2018). Mobility and location choice of offenders. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford handbook of environmental criminology*. Oxford : Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.001.0001/oxfordhb-9780190279707-e-17

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLoS ONE, 14*(1), e0210733. https://doi.org/10.1371/journal.pone.0210733.

Bernasco, W., & van Dijke, R. (2020). Do offenders avoid offending near home? A systematic review of the buffer zone hypothesis. *Crime Science, 9*(1), 1–10. https://doi.org/10.1186/s40163-020-00118-5.

Bernasco, W., Johnson, S. D., & Ruiter, S. (2015). Learning where to offend: Effects of past on future burglary locations. *Applied Geography, 60*(Supplement C), 120–129. https://doi.org/10.1016/j.apgeog.2015.03.014.

Bernasco, W., & Kooistra, T. (2010). Effects of residential history on commercial robbers' crime location choices. *European Journal of Criminology, 7*(4), 251–265. https://doi.org/10.1177/1477370810363372.

Bernasco, W., Ruiter, S., Bruinsma, G. J. N., Pauwels, L. J. R., & Weerman, F. M. (2013). Situational causes of offending: A fixed-effects analysis of space–time budget data. *Criminology, 51*(4), 895–926. https://doi.org/10.1111/1745-9125.12023.

Bichler, G., Christie-Merrall, J., & Sechrest, D. (2011). Examining juvenile delinquency within activity space: Building a context for offender travel patterns. *Journal of Research in Crime and Delinquency, 48*(3), 472–506. https://doi.org/10.1177/0022427810393014.

Bichler, G., Malm, A., & Christie-Merrall, J. (2012). Urban backcloth and regional mobility patterns as indicators of juvenile crime. In M. A. Andresen & J. B. Kinney (Eds.), *Patterns, prevention, and geometry of crime* (pp. 118–136). London, UK: Routledge.

Bichler, G., Malm, A., & Enriquez, J. (2014). Magnetic facilities: Identifying the convergence settings of juvenile delinquents. *Crime and Delinquency, 60*(7), 971–998. https://doi.org/10.1177/0011128710382349.

Bichler, G., Schwartz, J. A., & Orosco, C. A. (2010). Delinquents on the move: Examining subgroup travel variability. *Crime Patterns and Analysis, 3*(1), 14–37.

Brantingham, P. J., & Tita, G. (2008). Offender mobility and crime pattern formation from first principles. In L. Liu & J. Eck (Eds.), *Artificial crime analysis systems: Using computer simulations and geographic information systems* (pp. 193–208). Pennsylvania: IGI Global.

Brantingham, P. L., & Brantingham, P. J. (1991). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (2nd ed., pp. 27–54). Long Grove: Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993). Environment, routine, and situation: Toward a pattern theory of crime. In R. V. Clarke & M. Felson (Eds.), *Routine activity and rational choice* (pp. 259–294). Piscataway, NJ: Transaction Publishers.

Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research, 3*(3), 5–26. https://doi.org/10.1007/BF02242925.

Brantingham, P. J., & Brantingham, P. L. (2012). The theory of target search. In F. T. Cullen & P. Wilcox (Eds.), *The Oxford Handbook of Criminological Theory*. Oxford : Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199747238.013.0028.

Brantingham, P. J., Brantingham, P. L., & Andresen, M. A. (2008). The geometry of crime and crime pattern theory. In R. Wortley & M. Townsley (Eds.), *Environmental Criminology and Crime Analysis* (pp. 98–117). Milton Park: Taylor and Francis.

Brown, L. A., & Moore, E. G. (1970). The intra-urban migration process: A perspective. *Geografiska Annaler Series B, Human Geography, 52*(1), 1–13. https://doi.org/10.2307/490436.

Canter, D., & Larkin, P. (2008). The environmental range of serial rapists. In D. Canter & D. Youngs (Eds.), *Applications of geographical offender profiling* (pp. 57–68). Farnham, UK: Ashgate.

Chorus, C. G., & Timmermans, H. J. P. (2010). Determinants of stated and revealed mental map quality: An empirical study. *Journal of Urban Design, 15*(2), 211–226. https://doi.org/10.1080/13574801003638095.

Clare, J. (2011). Examination of systematic variations in burglars' domain-specific perceptual and procedural skills. *Psychology, Crime and Law, 17*(3), 199–214. https://doi.org/10.1080/10683160903025810.

Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and Justice, 6*, 147–185.

Cornish, D. B., & Clarke, R. V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. Berlin: Springer-Verlag.
Costello, A., & Wiles, P. (2001). GIS and the journey to crime: An analysis of patterns in South Yorkshire. In K. J. Bowers & A. Hirschfield (Eds.), *Mapping and analysing crime data: Lessons from research and practice* (pp. 27–60). Milton Park: Taylor and Francis.
Cromwell, P. F., Olson, J. N., & Avary, D. W. (1990). *Breaking and entering: An ethnographic analysis of burglary* (1st ed.). Newbury Park, CA: Sage.
Downs, R. M., & Stea, D. (1973). Cognitive maps and spatial behavior: Process and products. In R. M. Downs & D. Stea (Eds.), *Image and environment: Cognitive mapping and spatial behavior* (pp. 8–26). Aldine, CA: Aldine.
Elffers, H. (2004). Decision models underlying the journey to crime. In G. J. N. Bruinsma, H. Elffers, & J. W. De Keijser (Eds.), *Punishment, places and perpetrators: Developments in criminology and criminal justice research* (pp. 182–197). London: Willan.
Felson, M. (2008). The routine activity approach. In R. Wortley & L. Mazerolle (Eds.), *Environmental criminology and crime analysis* (pp. 87–97). London: Willan.
Frith, M. J. (2019). Modelling taste heterogeneity regarding offence location choices. *Journal of Choice Modelling, 33,* 100187. https://doi.org/10.1016/j.jocm.2019.100187.
Frith, M. J., Johnson, S. D., & Fry, H. M. (2017). Role of the street network in burglars' spatial decision-making. *Criminology, 55*(2), 344–376. https://doi.org/10.1111/1745-9125.12133.
Gärling, T., & Axhausen, K. W. (2003). Introduction: Habitual travel choice. *Transportation, 30*(1), 1–11. https://doi.org/10.1023/A:1021230223001.
Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*(2), 263–297. https://doi.org/10.1016/S0010-0277(98)00002-X.
Golledge, R. (1978). Representing, interpreting and using cognized environments. *Papers in Regional Science, 41*(1), 169–204. https://doi.org/10.1111/j.1435-5597.1978.tb01046.x.
Golledge, R. (1999). Human wayfinding and cognitive maps. In R. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 5–45). Baltimore: Johns Hopkins University Press.
Golledge, R., & Stimson, R. (1997). *Spatial behavior: A geographic perspective*. New York, NY: Guilford Press.
González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782. https://doi.org/10.1038/nature06958.
Gould, P. (1966). *On mental maps*. Ann Arbor, MI: University of Michigan.
Gould, P. (1973). On mental maps. In R. M. Downs & D. Stea (Eds.), *Image and environment: Cognitive maps and spatial behavior* (pp. 182–220). Aldine, CA: Aldine.
Gould, P., & White, R. (1986). *Mental maps* (2nd ed.). London: Routledge.
Hammond, L., & Youngs, D. (2011). Decay functions and criminal spatial processes: Geographical offender profiling of volume crime. *Journal of Investigative Psychology and Offender Profiling, 8*(1), 90–102. https://doi.org/10.1002/jip.132.
Hannes, E., Janssens, D., and Wets, G. (2006, August). *Proximity is a state of mind: Exploring mental maps in daily activity travel behaviour*. 11th International Conference on Travel Behaviour Research, Kyoto.
Hannes, E., Janssens, D., & Wets, G. (2008). Destination choice in daily activity travel: Mental map's repertoire. *Transportation Research Record, 2054*(1), 20–27. https://doi.org/10.3141/2054-03.
Hannes, E., Janssens, D., & Wets, G. (2009). Does space matter? Travel mode scripts in daily activity travel. *Environment and Behavior, 41*(1), 75–100. https://doi.org/10.1177/0013916507311033.
Hannes, E., Kusumastuti, D., Espinosa, M., Janssens, D., Vanhoof, K., & Wets, G. (2012). Mental maps and travel behaviour: Meanings and models. *Journal of Geographical Systems, 14*(2), 143–165. https://doi.org/10.1007/s10109-010-0144-2.
Hanson, S., & Huff, O. J. (1988). Systematic variability in repetitious travel. *Transportation, 15*(1), 111–135. https://doi.org/10.1007/BF00167983.
Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013). Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics, 151*(1), 304–318. https://doi.org/10.1007/s10955-012-0645-0.
Hillier, A., Smith, T. E., Whiteman, E. D., & Chrisinger, B. W. (2017). Discrete choice model of food store trips using National Household Food Acquisition and Purchase Survey (FoodAPS). *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph14101133.
Horton, F. E., & Reynolds, D. R. (1971). Effects of urban spatial structure on individual behavior. *Economic Geography, 47*(1), 36–48. https://doi.org/10.2307/143224.
Howard, R. W. (2000). Generalization and transfer: An interrelation of paradigms and a taxonomy of knowledge extension processes. *Review of General Psychology, 4*(3), 211–237. https://doi.org/10.1037/1089-2680.4.3.211.
Ida, T., & Kuroda, T. (2006). Discrete choice analysis of demand for broadband in japan. *Journal of Regulatory Economics, 29*(1), 5–22. https://doi.org/10.1007/s11149-005-5124-y.
Johnson, S. D., & Bowers, K. J. (2004). The stability of space-time clusters of burglary. *The British Journal of Criminology, 44*(1), 55–65. https://doi.org/10.1093/bjc/44.1.55.
Johnson, S. D., Summers, L., & Pease, K. (2009). Offender as forager? A direct test of the boost account of victimization. *Journal of Quantitative Criminology, 25*(2), 181–200. https://doi.org/10.1007/s10940-008-9060-8.
Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392.
Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and Its Applications, 391*(4), 1702–1717. https://doi.org/10.1016/j.physa.2011.11.005.
Knabe-Nicol, S., & Alison, L. (2011). The cognitive expertise of Geographic Profilers. In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 126–159). Milton Park: Taylor and Francis.
Kuang, D., Brantingham, P. J., & Bertozzi, A. (2017). Crime topic modeling. *Crime. Science, 6*(1), 1–20. https://doi.org/10.1186/s40163-017-0074-0.
Lammers, M. (2018). Co-offenders' crime location choice: Do co-offending groups commit crimes in their shared awareness space? *The British Journal of Criminology, 58,* 1193–1211. https://doi.org/10.1093/bjc/azx069.
Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology, 53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071.
Lantz, B., & Ruback, R. B. (2017). A networked boost: Burglary co-offending and repeat victimization using a network approach. *Crime and Delinquency, 63*(9), 1066–1090. https://doi.org/10.1177/0011128715597695.
Lattimore, P., & Witte, A. (1986). Models of decision making under uncertainty: The criminal choice. In D. B. Cornish & R. V. Clarke (Eds.), *The reasoning criminal* (pp. 129–155). Berlin: Springer-Verlag. https://doi.org/10.4324/9781315134482-2.
Lloyd, R., & Cammack, R. (1996). Constructing cognitive maps with orientation biases. In J. Portugali (Ed.), *The construction of cognitive maps* (pp. 187–213). Netherlands: Springer. https://doi.org/10.1007/978-0-585-33485-1_9.
Long, D., Liu, L., Feng, J., & Zhou, S. (2018). Assessing the influence of prior on subsequent street robbery location choices: A case study in ZG city China. *Sustainability, 10*(6), 1818. https://doi.org/10.3390/su10061818.
Mark, D. M., Freksa, C., Hirtle, S. C., Lloyd, R., & Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science, 13,* 747–774. https://doi.org/10.1080/136588199241003.
McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*(2), 668–693. https://doi.org/10.1037/a0032963.
McFadden, D. L. (1984). Econometric analysis of qualitative response models. In P. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 105–142). Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4412(84)02016-X.
Menting, B. (2018). Awareness × opportunity: Testing interactions between activity nodes and criminal opportunity in predicting crime location choice. *The British Journal of Criminology, 58,* 1171–1192. https://doi.org/10.1093/bjc/azx049.
Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2016). Family matters: Effects of family members' residential areas on crime location choice. *Criminology, 54*(3), 413–433. https://doi.org/10.1111/1745-9125.12109.
Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: Findings

from an online self-report survey. *The British Journal of Criminology, 60*(2), 303–322. https://doi.org/10.1093/bjc/azz044.
Miller, J. (2013). Individual offending, routine activities, and activity settings: Revisiting the routine activity theory of general deviance. *Journal of Research in Crime and Delinquency, 50*(3), 390–416. https://doi.org/10.1177/0022427811432641.
Nee, C. (2015). Understanding expertise in burglars: From pre-conscious scanning to action and beyond. *Aggression and Violent Behavior, 20*(Supplement C), 53–61. https://doi.org/10.1016/j.avb.2014.12.006.
Nee, C., & Meenaghan, A. (2006). Expert decision making in burglars. *The British Journal of Criminology, 46*(5), 935–949. https://doi.org/10.1093/bjc/azl013.
Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica, 69*(2), 307–342.
Nguyen, H. T. A., Chikaraishi, M., Fujiwara, A., & Zhang, J. (2017). Mediation effects of income on travel mode choice: Analysis of short-distance trips based on path analysis with multiple discrete outcomes. *Transportation Research Record, 2664*(1), 23–30. https://doi.org/10.3141/2664-03.
Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications, 6*, 8166–8173. https://doi.org/10.1038/ncomms9166.
Pease, K. (1998). *Repeat victimisation: Taking stock*. London: Home Office.
Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency, 43*(3), 261–291. https://doi.org/10.1177/0022427806286566.
Reid, A. A., Frank, R., Iwanski, N., Dabbaghian, V., & Brantingham, P. L. (2014). Uncovering the spatial patterning of crimes: A criminal movement model (CRiMM). *Journal of Research in Crime and Delinquency, 51*(2), 230–255. https://doi.org/10.1177/0022427813483753.
Rengert, G. (1996). *The geography of illegal drugs*. Boulder: Westview Press.
Rengert, G., & Wasilchick, J. (1985). *Suburban burglary: A time and a place for everything*. Springfield: C.C Thomas.
Rengert, G., & Wasilchick, J. (2000). *Suburban burglary: A tale of two suburbs* (2nd ed.). Springfield: C.C Thomas.
Rossmo, D. K. (2000). *Geographic profiling*. Boca Raton, FL: CRC Press.
Rossmo, D. K. (2014). Geographic profiling. In G. J. N. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 1934–1942). Berlin: Springer.
Rossmo, D. K., Lu, Y., & Fang, T. B. (2012). Spatial-temporal crime paths. In M. A. Andresen & J. B. Kinney (Eds.), *Patterns, prevention, and geometry of crime* (pp. 3–15). London: Routledge.
Ruiter, S. (2017). Crime location choice. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook of offender decision making* (pp. 398–420). Oxford: Oxford University Press.
Ruiter, S., and Davies, T. (2018, July). *BTW, a test of crime pattern theory*. Environmental Criminology and Crime Analysis Symposium, Spain.
Sivakumar, A., & Bhat, C. R. (2007). A comprehensive, unified framework for analyzing spatial location choice. *Transportation Research Record, 2003*(1), 103–111. https://doi.org/10.3141/2003-13.
Smith, W., Bond, J. W., & Townsley, M. (2009). Determining how journeys-to-crime vary: Measuring inter- and intra-offender crime trip distributions. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting crime in its place* (pp. 217–236). Berlin: Springer.
Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170.
Sorg, E. T., Haberman, C. P., Ratcliffe, J. H., & Groff, E. R. (2013). Foot patrol in violent crime hot spots: The longitudinal impact of deterrence and posttreatment effects of displacement. *Criminology, 51*(1), 65–101. https://doi.org/10.1111/j.1745-9125.2012.00290.x.
Summers, L., Johnson, S. D., & Rengert, G. (2010). The use of maps in offender interviewing. In W. Bernasco (Ed.), *Offenders on offending: Learning about crime from criminals* (pp. 246–272). London: Willan.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629–640. https://doi.org/10.1017/S0140525X01000061.
Tillyer, M. S., & Walter, R. J. (2019). Busy businesses and busy contexts: The distribution and sources of crime at commercial properties. *Journal of Research in Crime and Delinquency, 56*(6), 816–850. https://doi.org/10.1177/0022427819848083..
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46,* 234–240. https://doi.org/10.2307/143141.
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189–208. https://doi.org/10.1037/h0061626.
Tonkin, M., Santtila, P., & Bull, R. (2012). The linking of burglary crimes using offender behaviour: Testing research cross-nationally and exploring methodology. *Legal and Criminological Psychology, 17*(2), 276–293. https://doi.org/10.1111/j.2044-8333.2010.02007.x.
Tonkin, M., Woodhams, J., Bull, R., Bond, J. W., & Palmer, E. J. (2011). Linking different types of crime using geographical and temporal proximity. *Criminal Justice and Behavior, 38*(11), 1069–1088. https://doi.org/10.1177/0093854811418599.
Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., & González, M. C. (2015). Coupling human mobility and social ties. *Journal of the Royal Society Interface, 12*(105), 20141128–20141128. https://doi.org/10.1098/rsif.2014.1128.
Townsley, M. (2016). Offender mobility. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (pp. 142–161). London: Routledge.
Townsley, M., Birks, D., Ruiter, S., Bernasco, W., & White, G. (2016). Target selection models with preference variation between offenders. *Journal of Quantitative Criminology, 32*(2), 283–304. https://doi.org/10.1007/s10940-015-9264-7.
van Daele, S., & Vander Beken, T. (2011). Out of sight, out of mind? Awareness space and mobile offenders. *European Journal of Crime, Criminal Law and Criminal Justice, 19*(2), 125–137. https://doi.org/10.1163/157181711X566326.
van Daele, S., Vander Beken, T., & Bruinsma, G. J. N. (2012). Does the mobility of foreign offenders fit the general pattern of mobility? *European Journal of Criminology, 9*(3), 290–308. https://doi.org/10.1177/1477370812440065.
van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A time for a crime: Temporal aspects of repeat offenders' crime location choices. *Journal of Research in Crime and Delinquency, 55*(4), 538–568. https://doi.org/10.1177/0022427818766395.
Vandeviver, C. (2014). Applying Google Maps and Google Street View in criminological research. *Crime Science, 3*(1), 13. https://doi.org/10.1186/s40163-014-0013-2.
Vandeviver, C., & Bernasco, W. (2019). "Location, location, location": Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-019-09431-y.
Wang, J., Dong, L., Cheng, X., Yang, W., & Liu, Y. (2019). An extended exploration and preferential return model for human mobility simulation at individual and collective levels. *Physica A: Statistical Mechanics and Its Applications, 534,* 121921. https://doi.org/10.1016/j.physa.2019.121921.
Ward, T., Polaschek, D. L. L., & Beech, A. R. (2005). Theory construction, development and evaluation. *Theories of sexual offending* (pp. 3–16). Hoboken: Wiley. https://doi.org/10.1002/9780470713648.ch1.
Weisburd, D., Eck, J. E., Braga, A. A., Telep, C. W., Cave, B., Bowers, K., et al. (2016). Place matters: Criminology for the twenty-first century. *Cambridge University Press*. https://doi.org/10.1017/CBO9781139342087.
Wikström, P.-O., Ceccato, V., Hardie, B., & Treiber, K. (2010). Activity fields and the dynamics of crime. *Journal of Quantitative Criminology, 26*(1), 55–87. https://doi.org/10.1007/s10940-009-9083-9.
Wiles, P., & Costello, A. (2008). The 'road to nowhere': The evidence for travelling criminals. In D. Canter & D. Youngs (Eds.), *Principles of geographical offender profiling* (pp. 165–175). Farnham: Ashgate.
Wolpert, J. (1965). Behavioral aspects of the decision to migrate. *Papers in Regional Science, 15*(1), 159–169. https://doi.org/10.1111/j.1435-5597.1965.tb01320.x.
Wright, R., & Decker, S. H. (1994). *Burglars on the job: Streetlife and residential break-ins*. Boston, MA: Northeastern University Press.
Xiao, L., Liu, L., Song, G., Ruiter, S., & Zhou, S. (2018). Journey-to-crime distances of residential burglars in China disentangled: Origin and destination effects. *ISPRS International Journal of Geo-Information, 7*(8), 325. https://doi.org/10.3390/ijgi7080325.

Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., & Lai, Y.-C. (2017). Universal model of individual and population mobility on diverse spatial scales. *Nature Communications, 8*(1), 1–9. https://doi.org/10.1038/s41467-017-01892-8.

Zhang, W., Ahmad Termida, N., & Susilo, Y. O. (2019). What construct one's familiar area? A quantitative and longitudinal study. *Environment and Planning B: Urban Analytics and City Science, 46*(2), 322–340. https://doi.org/10.1177/2399808317714798.

Zhang, W., Susilo, Y. O., & Ahmad Termida, N. (2016). Investigating the interactions between travellers' familiar areas and their multi-day activity locations. *Journal of Transport Geography, 53,* 61–73. https://doi.org/10.1016/j.jtrangeo.2016.04.012.

Publisher's Note

CHAPTER 3

Offenders' Activity Locations Recorded in Police Data

With the benefit of a theoretical model systematising the links between offenders' mental maps and their crime locations, the research presented in the next four chapters examined these links empirically. The research leveraged a national dataset extracted from New Zealand police records to examine whether the associations between offenders' mental maps—as evident from their activity locations in police records—and their crime locations are consistent with the theoretical model. Using data readily available to police in an investigation maximised the translatability of the findings to geographic profiling practice, but it was important to first establish whether these data capture enough activity locations to yield insight into their associations with offenders' crime locations. The following paper, published in the *ISPRS International Journal of Geo-Information*, describes the dataset and the results of preliminary analyses exploring its potential to tell us about offenders' mental maps, for both research and practical purposes.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information, 10*(2), 47. https://doi.org/10.3390/ijgi10020047

International Journal of
Geo-Information

MDPI

Article

A National Examination of the Spatial Extent and Similarity of Offenders' Activity Spaces Using Police Data

Sophie Curtis-Ham [1,*], Wim Bernasco [2,3], Oleg N. Medvedev [1] and Devon L. L. Polaschek [1]

[1] Te Puna Haumaru NZ Institute of Security and Crime Science, Te Kura Whatu Oho Mauri School of Psychology, Te Whare Wānanga o Waikato University of Waikato, Hamilton 3240, New Zealand; oleg.medvedev@waikato.ac.nz (O.N.M.); polascde@waikato.ac.nz (D.L.L.P.)
[2] Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), 1081 HV Amsterdam, The Netherlands; WBernasco@nscr.nl
[3] Department of Spatial Economics, School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
* Correspondence: sc398@students.waikato.ac.nz

Abstract: It is well established that offenders' routine activity locations (nodes) shape their crime locations, but research examining the geography of offenders' routine activity spaces has to date largely been limited to a few core nodes such as homes and prior offense locations, and to small study areas. This paper explores the utility of police data to provide novel insights into the spatial extent of, and overlap between, individual offenders' activity spaces. It includes a wider set of activity nodes (including relatives' homes, schools, and non-crime incidents) and broadens the geographical scale to a national level, by comparison to previous studies. Using a police dataset including n=60,229 burglary, robbery, and extra-familial sex offenders in New Zealand, a wide range of activity nodes were present for most burglary and robbery offenders, but fewer for sex offenders, reflecting sparser histories of police contact. In a novel test of the criminal profiling assumptions of homology and differentiation in a spatial context, we find that those who offend in nearby locations tend to share more activity space than those who offend further apart. However, in finding many offenders' activity spaces span wide geographic distances, we highlight challenges for crime location choice research and geographic profiling practice.

Keywords: homology assumption; geographic offender profiling; offender activity space; police data; routine activity nodes

Citation: Curtis-Ham, S.; Bernasco, W.; Medvedev, O.N.; Polaschek, D.L.L. A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 47. https://doi.org/10.3390/ijgi10020047

Academic Editor: Marco Helbich and Wolfgang Kainz

Received: 4 December 2020
Accepted: 18 January 2021
Published: 23 January 2021

1. Introduction

Offenders' routine activity locations—where they live, work, and carry out other non-criminal activities—play a significant role in shaping their crime locations [1–3]. Understanding the nature and spatial distribution of individual offenders' activity locations thus has important implications for crime and policing policy and practice: for example, identifying places that individuals are at higher risk of offending, or identifying, in police investigations, suspects who would be more likely to have committed a given crime in a given location [4]. Yet little research to date has studied the full range of offenders' routine activity locations, or addressed practically important questions such as the extent to which offenders' activity locations are shared or can be differentiated, as we elaborate below.

Many police jurisdictions maintain databases which store the details of calls for service, criminal investigations, intelligence reports, arrests, stops/searches and other routine police activities that involve interacting with members of the public [5–8]. The details of such records can include information about the locations of offenses, incidents, police interactions, the home addresses of the parties involved and even where they work or attend school. Provided the data contains location and timing information with sufficient specificity, and can be collated if stored across multiple databases, it could be a rich source of information about offenders' activity spaces (This paper focuses on locations where offenders have carried out routine activities, or travelled between them, collectively making up their 'activity space' [1]. The related concept of 'awareness space' additionally includes places known through sources other than direct experience [1], which are not identifiable from the present data.), particularly for those who have frequently come into contact with police as offenders, or as victims, witnesses, or members of the community. However, such data have not yet been used to study offenders' activity locations other than their residential addresses and prior crime locations [3,9].

The aims of this paper are therefore two-fold: first, from a methodological perspective, to consider the range of routine activity locations that can be identified using the kind of police data described above; second, to answer several exploratory research questions relating to the nature and distribution of offenders' activity locations as revealed by these data. We examine the distributions of activity locations of offenders identified for a burglary, robbery or sexual offense over timeframes of varying lengths prior to the most recent offense and geographic units of analysis of varying sizes. We focus on crimes that typically involve some form of search for a suitable victim or target (rather than the targeting of a specific victim already known to the offender) which is influenced by the offenders' prior activity locations. The subset of crime types included here reflect a mixture of environments and land uses (residential, business/retail), motivations (material/non-material) and planning (strongly versus weakly premeditated) and thus are likely to be representative of a wider set of crime types. We also explore the extent to which the available activity locations of these offenders differentiate between them. The question of differentiation is particularly important for geographic profiling methods that seek to predict, given the locations of unsolved crimes, who may have committed them or where the offender might be found [10,11]. If activity spaces were shared by many offenders, they would be of little use for prioritizing among suspects.

We begin with a review of the small extant literature that has examined offenders' routine activity spaces and expand on the concept of differentiation. We then provide details of the data used in this research and the methods used in the analyses before presenting and discussing the results. We conclude by highlighting the benefits and limitations of the data for both research and practice.

1.1. Offenders' Activity Spaces

Understanding the nature of offenders' activity spaces beyond their homes is an important research endeavor, given well established links between their activity spaces and offense locations. As Routine Activities Theory asserts, crime requires a convergence of motivated offenders and potential targets in time and space [12]. Furthermore, as Crime Pattern Theory explains, this convergence happens when and where the routine activities of offenders (forming their activity space) overlap with crime opportunities [1,2]. Offenders' non-criminal and criminal activities alike equip them with knowledge of possible crime locations that is brought to bear on future decisions about where to commit crime [1,2]. Studies of aggregate crime patterns confirm that crime concentrates in or near routine activity locations likely to be common to many offenders, such as central business districts, shopping precincts and transit hubs [13,14]. At an individual level, many studies demonstrate that people are more likely to commit crime closer to their home than further away [15]. Furthermore, interview studies of small samples of offenders have highlighted that offenders tend to commit crimes at or near a range of other routine activity locations such as work, friends and family members' homes, recreation sites, or prior crimes [16–22]. Studies using Discrete Spatial Choice Modelling (DSCM) have even quantified the increase in odds of a location being chosen based on its proximity to offenders' previous addresses, their family members' current or previous addresses, and the locations of their previous crimes, when controlling for proximity to their home [3,9,23–26]. Two further DSCM studies, both on young offenders, incorporated additional types of activity nodes to confirm that the odds of crime increase with proximity to any activity node [27,28].

However, few studies have described, quantitatively, offenders' routine activity spaces in terms of the number of locations they frequent, or the geographic extent of these locations, over a given time period. Those that have vary in location, cohort, and method, but provide a rough baseline with which to compare the present study. For example, location tracking data from fourteen 19- to 44-year-old parolees on GPS monitoring in Florida revealed that offenders visited 4 non-home nodes (specific sites) on average (range 2 to 6) during the week prior to re-offending [29]. Their activity spaces covered an average of 27 miles squared (median 12, range 0.2 to 70).

In the Netherlands, 78 young offenders aged 18-26 who participated in an online survey reported visiting 6 nodes (neighborhoods) on average (range 1 to 15) in the month preceding their offenses [28]. These included home, school/work, friend/family residences and leisure activity nodes, with most offenders reporting at least one of each. The maximum distance between any individual offender's nodes was 200 km (of a possible maximum distance between any two neighborhoods in the Netherlands of 300 km). In another Dutch interview-based study [27], 70 13- to 16-year-olds who offended in the subsequent 4 years reported an average of 7 activity nodes (200 m × 200 m cells in a map grid covering the Hague) in a 4-day period preceding the interview (median 5, range 1 to 15). Almost 75% of their time was spent at just two nodes (home and typically school), producing a relatively short (average 3 km) "radius of gyration", a measure of the size of an individual's activity space weighted by time spent in each node.

Considering not the number or geographic range of nodes but their overlap between individuals, a UK study found that there was a high correlation between the aggregated self-reported awareness spaces of 17 prolific property offenders and 13 non-offenders from the same county [20]. In other words, the nodes common to many offenders were also common to the general population in the study area. We expand on the issue of commonality versus differentiation in activity space in the next section.

1.2. Homology and Differentiation in Activity Spaces

In geographic profiling for criminal investigations, differentiation between suspects' activity spaces would enable prioritization among suspects when comparing suspects' known activity locations with the predicted base(s) of the offender [11,30]. Differentiation would mean that suspects fit the "geographic profile" to varying degrees. Conversely, if offenders' activity spaces are relatively homogenous, it would be difficult to prioritize among the many potential suspects whose activity spaces are all equally consistent with them having committed an offense(s) in a given location.

This issue has been raised in the behavioral profiling literature. The ability to infer an offender's characteristics from attributes of the crime relies on core assumptions of homology and differentiation [31–33]. Homology means offenders sharing certain crime attributes will also share certain personal characteristics; differentiation means differences in crime attributes indicate differences in offender attributes. The more distinctive (i.e., differentiating) the attribute(s)—of crime and criminal—the more reliably suspect pools can be narrowed [31,33].

With respect to the spatial dimension of offending, homology means that offenders sharing the locations of their crimes will also share other parts of their activity spaces. For example, if two burglars commit an offense in the same street, homology means they are more likely to live in the same neighborhood, attend the same school, or visit the same shopping center, as compared to two burglars offending in different streets. The DSCM studies discussed above are suggestive of homology between offenders' crime locations and activity spaces. Since activity node locations predict crime locations [27], offenders with similar activity node locations would, logically, offend in similar locations, having gained awareness of the same crime opportunities. However, the reverse does not logically follow, though geographic profiling relies on inferring potential activity node locations from offense locations. Illustrative of this point, Costanzo et al. [34] found that offenders with nearby home addresses tended to travel in a similar direction to offend, but that offenders with nearby offenses had not necessarily come from similar directions to offend.

Exploring the link between co-offenders' activity spaces and crime locations, Lammers [35] found that offenses committed at the same place and time (i.e., by co-offenders) were more likely to be committed in a neighborhood where at least two of the co-offenders had lived or committed a previous crime, than elsewhere. However, the co-offending pairs shared less than 50% of their residential or prior offense neighborhoods. Tayebi et al. [36] found that offenders who were more closely connected in a co-offending network lived closer together and shared more home and past offense neighborhoods (however, Malm et al. [37] found no correlation between network proximity and home proximity in a cannabis production network). Thus, past or present co-offending may entail similarity of some, but not all, elements of co-offenders' activity spaces.

There is also evidence that offenders might be differentiated by their crime and routine activity locations. Nearby crimes are more likely to have been committed by the same offender than different offenders [38–42]. That all people exhibit highly distinctive spatio-temporal routine activity patterns [43], suggests that among offenders, activity spaces may also be distinctive.

However, neither the extent to which individual offenders' activity spaces are shared (other than with co-offenders), nor the extent to which homology and differentiation are present in the relationship between offenders' activity spaces and their crime locations, has been directly examined by any studies to date.

1.3. Present Study

We therefore aimed to explore the range of routine activity locations that can be identified in police data and to answer the following research questions relating to the nature, distribution and spatial similarity (or differentiation) between offenders' activity spaces as revealed by these data. (1) What is the distribution of the number activity nodes per offender? (2) What is the distribution of activity space size per offender? (3) What proportion of offenders' activity nodes are shared with other offenders? (4) Do offenders who offended in closer proximity to each other have similar activity spaces preceding the offense, relative to offenders who offended in different locations? (The present research considers purely spatial similarity, rather than the overlap of offenders' activity space in both time and space. Offenders who have frequented the same location years apart could identify the same criminal opportunities to the extent that those opportunities reflect stable features of the environment. Future research might consider whether offenses close together in both time and space are associated with activity space overlapping in both time and space, which was not possible with the present data.)

2. Materials and Methods

2.1. Data

All data used in this study were extracted from the New Zealand Police national crime database (NIA; National Intelligence Application). First, a cohort of *offenders* was identified: as those who had committed a *residential* or *non-residential burglary*, *commercial* or *personal robbery*, or *extra-familial sexual offense* between 2009 and 2018 (detailed definitions of these and all other italicized parameters are provided in Supplementary Material S1). From these offenses, the most recent ("reference") offense was identified for each offender and each offense type. Second, a range of activity locations associated with the offenders, recorded by police where required for operational purposes, were extracted as follows. (Herein 'offender' means the person who committed the reference offense. This is not meant to imply that they were the offender in reference to prior offenses where they were a victim or witness, or prior non-crime incidents. Further, the labels 'burglar', 'robber' and 'sex offender' in this paper are used for expediency and refer solely to the nature of the reference offense, not to prior offense histories.)

Offense data included *non-traffic offenses* committed by the offender. Incident data included non-traffic offenses involving the offender in a *non-offending role* (e.g., witness or victim) and *non-traffic non-crime incidents* involving the offender (e.g., domestic disputes, suspicious behavior, drunk and disorderly). Offenses and incidents dated back to 2004 (older records are patchy as not all were transferred during a change in databases).

Address data included past and present addresses of various types such as home, work, "spoken to at", "seen at", and "arrested at". Manual inspection of NIA records confirmed the latter categories included a range of nodes such as acquaintances' homes, commercial or public facilities, and on-street encounters with police.

Family home address data included past (during the offender's lifetime) and present home addresses of family members, including intimate partner relationships and family relationships. A broader definition of family was included than used in previous studies [24,25], which were restricted to parents, children and siblings. The inclusion of extended family reflects the important role of wider family in many New Zealand (NZ) communities (e.g., Māori, Pacific and Asian cultures [44]). Although maximizing the range of activity nodes extracted was desirable, the inclusion of intimate partner relationships, which can be temporary, will also have introduced some error: home locations of partners may pre- or post-date the relationship and therefore may not have been visited by the offender.

Lastly, data on the *school and other educational institutions* attended by the offenders was also extracted, where recorded. Employment details recorded in NIA typically only included a company name with no specific details of location, branch or any other address, precluding sufficiently complete or accurate geocoding to enable their inclusion.

Reference offense and all activity location data except school/education included the geographic coordinates of the locations' addresses. Addresses are automatically geocoded (allocated coordinates) within the NIA system when addresses are entered. School/education records included only the institution name, not address details, requiring these locations to be geocoded by the researchers. This was completed in R [45] using a Google Maps API to search for the institution name and return its address and coordinates (then transformed in ArcGIS), as described in Supplementary Material S3. The data were then pre-processed in R to exclude records considered too imprecise spatially or temporally to include in the analysis.

Pre-processing steps and filters are detailed in Supplementary Material S2, which also reports the number and percent of records from each dataset that were excluded with each filter. The exclusion of records with missing or imprecise spatial or temporal data resulted in the retention of between 86.2% and 97.0% of the activity locations in each dataset. These proportions exceed the minimum geocoding hit rates (completeness) suggested by recent studies to maintain the spatial distribution of the data, considering the types of offenses and size of the dataset [46,47]. The results of checks on geocoding accuracy and precision in the data following the exclusions are detailed in Supplementary Material S3. Over 95% of records from each dataset were geocoded to the correct location.

The activity locations were then filtered to those which pre-dated the reference offense. Supplementary Material S4 provides details as to how the dates of offenses, incidents, (family) address and education records were treated in identifying whether they were "prior" to the reference offense. This filter reflects our aim of exploring the data's potential for crime location choice research and investigative use. Crime location choice research would limit offenders' activity space to the period before they commit the reference offense to strengthen causal interpretations (e.g., exclude reverse causation). In police investigations, only activity locations known to police prior to the reference offense would be available to inform investigative decisions. Table 1 shows the size of each dataset relative to each reference offense category following the above filters, expressed as the number of unique person-locations, where a location is either a discrete event in space and time (offenses/incidents) or an address without a specific related event record (offender and family addresses, education addresses). It also shows the proportion of offenders for whom there were no prior activity nodes in the data, and who were therefore excluded from the analysis. Table 2 provides basic demographic statistics for the offenders with activity nodes included in the analysis, for each offense type.

Table 1. Sample size in each dataset relative to each reference offense category.

Reference Offense	Offenders [1]	Prior Offenses	Prior Incidents	Addresses	Family Addresses	Education	% with no Nodes
Res. Burg.	34,532	171,973	76,257	516,506	899,147	16,032	1.15
Non-res. Burg.	21,155	106,265	40,314	295,262	493,834	10,897	1.60
Com. Rob.	3975	22,450	8446	61,023	106,880	2761	0.49
Pers. Rob.	8737	43,393	18,809	144,423	285,510	5274	0.92
Sex Offenses	9749	17,546	14,194	85,413	114,677	1971	8.94

[1] There were 60,229 offenders, of whom 11,459 appeared in two offense categories, 2540 appeared in three, 424 appeared in four, and 27 appeared in all five offense categories.

Table 2. Offender demographic statistics

Reference offense	Median Age (IQR)	% Male	% Female
Res. Burg.	21 (13)	83.3%	16.6%
Non-res. Burg.	18 (12)	88.0%	12.0%
Com. Rob.	19 (8)	87.7%	12.3%
Pers. Rob.	19 (10)	80.7%	19.3%
Sex Offenses	27 (24)	96.8%	3.2%

For the purposes of the present analyses, all prior activity nodes were then compiled for each offender and any locations (i.e., specific coordinates) that were duplicated (e.g., both a home address and prior offense location) were removed such that each location was only counted once per offender.

2.2. Analytic Approach

The research questions were addressed via a range of descriptive and inferential statistical analyses, using the software R, as described in turn below.

2.2.1. Number of Activity Nodes

We measured the number of activity nodes per offender at two spatial resolutions: distinct address coordinates and distinct small census units. The former enabled identification of how many exact locations are typically recorded for offenders, which would allow for finer grained analysis and mapping of suspects' activity locations during investigations, where the aim is to identify and locate offenders. Census or other administrative units are, however, a more useful unit of activity space when assessing variability in the activity spaces of an offender population [48]. They can also be considered a proxy for unmeasured activity space: we do not visit places in isolation but tend to cluster our activities together; places immediately around or in between activity nodes are more likely to be in our activity space than places further away [49–52].

To measure activity nodes defined as small census units, for comparability with other studies of offender activity space we used the NZ Statistical Area 1 (SA1). SA1s typically contain 100–200 residents. Outliers include remote regions and bodies of water with no residential population; industrial, commercial or rural areas with residential populations under 100; and densely populated areas containing apartment blocks, retirement villages and residential facilities with populations above 500. The SA1 shapefile and metadata were downloaded from https://datafinder.stats.govt.nz/layer/98761-statistical-area-1-2019-generalised/. There were 29,879 SA1s, with a median area of 0.067 km^2 (mean 8.941 km^2, range 0.001 km^2–5758.384 km^2). (251 SA1s representing bodies of water (inlets, coastlines and lakes) were excluded from the area statistics as outliers with no land area. As a small number of offender node points fell within some of these SA1s they were retained in the analysis.) The median distance between the center of an SA1 and the center of its nearest neighbor SA1 was 199 m (mean 760 m, range 11 m–53.106 km). SA1s thus lie between the spatial units used by Menting et al. [28]—census units averaging 0.68 km^2 and 675 residents—and Bernasco's [27] 200 m × 200 m grid cells. The SA1s of reference offense and activity locations' coordinates were identified using the sf package in R [53].

2.2.2. Activity Space Size

The number of SA1s in an offender's activity space is an indicator of relative activity space size, but it does not capture its geographical extent. For example, were offenders' nodes confined to a small town, or part of a city, or an entire metropolitan area? Did they span multiple town/cities nearby or at opposite ends of the country? To gauge the geographic *range* of offenders' activity spaces, we used a simple indicator of the distance over which their activity nodes extended: the length of the diagonal of a rectangle (bounding box) encompassing all an offender's activity nodes (specific coordinates). This distance measure enabled faster processing (as calculable from Easting and Northing coordinates in meters, without the need to use spatial objects, in R) and simpler interpretation in relation to the above questions than area measures of activity space such as minimum bounding polygons and standard deviational ellipses [54] or the radius of gyration [27].

2.2.3. Shared Activity Space

To identify the distribution of shared activity space, we identified which SA1s were shared by multiple offenders, and then calculated the proportion of each offender's SA1 nodes that were shared. We tested two definitions of "shared": two or more offenders per SA1, and the top 25% most "popular" SA1s. Checks of the SA1s shared by the highest numbers of offenders showed these were, unsurprisingly, SA1s containing prisons, courts, police stations and city centers.

2.2.4. Homology and Differentiation

Previous studies examining homology and differentiation between offense attributes and offender characteristics have framed the question as a test of association: whether similarity/difference in offense attributes corresponds to similarity/difference in offender characteristics [32,33,55]. We applied the same approach, testing whether offenders who offended in nearby locations were more similar in their activity space than offenders whose offenses were further apart (i.e., whether reference offense proximity correlated with activity space similarity). We first identified pairs of offenders who had at least one activity node each and whose reference offenses fell in the same Territorial Authority (TA) boundary (i.e., the same city or region). TAs are administrative units reflecting the jurisdictions of New Zealand's 67 local authorities. We then took random samples of pairs, separately for urban, city-based TAs and rural TAs that cover large areas containing smaller settlements. Stratifying the analysis by urban and rural groups ensured that the correlation results were not an artefact of urban offenders' crimes and activity locations being in closer proximity and rural offenders' crimes and activity locations being further apart, due to the difference in density of crime opportunities and routine activity locations in urban and rural environments. (The Urban group included all City Council TAs. Wellington was treated as a single TA made up of Wellington City, Lower Hutt City, Upper Hutt City, Porirua City and Kapiti Coast District. Although administered by different Councils, these areas make up a single conurbation reflected in commuting and routine activity patterns. The Rural group included all other TAs (District Councils).) For each analysis we sampled 10,000 pairs and computed correlations between the distance between the pairs' reference offenses and three measures of the similarity of the pairs' activity spaces. Kendall's tau-b was used for all correlations because exploratory analyses showed the assumption of linearity was not met.

The similarity measures were as follows. First, we calculated the percentage of the first offender's nodes (SA1s) that were shared with the second offender. This measured the degree of direct overlap between the offenders' activity spaces, with percentage scores above 0 indicating at least one SA1 in common. Second, we calculated the minimum distance between any of the pair's activity nodes, using specific coordinates. Third, we took the median of the distances between each of the first offender's nodes and the nearest of the second offender's nodes. The latter two measures capture how close in space the offenders' activity nodes were, without necessarily falling into the same SA1s. The minimum distance indicates whether any of the pair's nodes were close together; the median nearest neighbor distance provides an indication of overall spatial proximity, while reducing the influence of outliers which could be at great distances apart. The maps in Figure 1 illustrate the complementarity of these measures. Using three anonymized examples from the data, they show for three pairs of offenders the distribution of activity nodes in relation to their crime locations and the measures of spatial similarity of activity space between each pair. Please note that in Example B the pair's reference offenses are so close together that their symbols overlap; likewise, in Example C, offender 2 has an activity node (a prior crime) in the same location as their reference offense thus the symbols overlap.

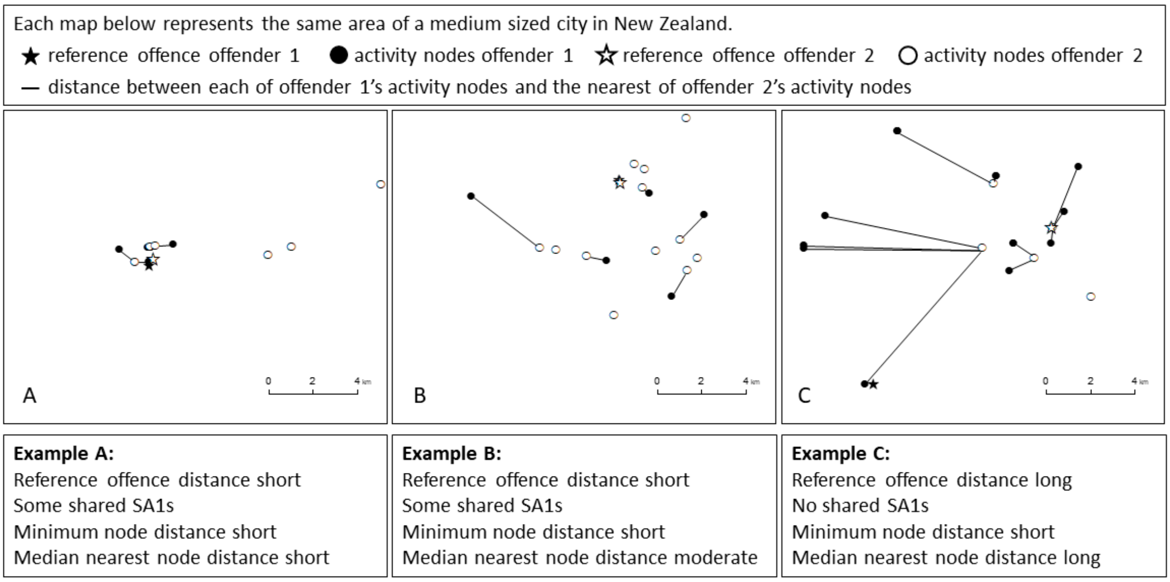


Figure 1. Illustration of activity space similarity measures for three pairs of offenders: (**a**) with close reference offenses and high activity space similarity on all three measures, (**b**) with close reference offenses and some activity space similarity, depending on measure, and (**c**) with distant reference offenses and less activity space similarity

We also investigated whether homology was related to co-offending (i.e., the offender pair committed the reference offense together). Since very few of the random pairs were co-offenders (see Supplementary Material S5 for proportions), we took additional samples of offender pairs who had offended in the same location (within 100 m) as co-offenders, and as independent offenders (n = 1000 per sample). (Samples of 8,000 were taken with replacement and reduced to unique pairs. For offenses with small numbers of co-offending pairs (smallest N = 221) this meant all pairs were included in the sample. This analysis was repeated with pairs of offenders who offended at the exact same coordinates but were not co-offenders. Although the sample sizes were much smaller (smallest N = 156), the results (not reported) were comparable to those for offenders who offended within 100m of each other: all differences remained statistically significant; effect sizes were within ±0.16 of the 100m results.) We then tested whether reference offense co-offenders displayed greater activity space similarity than those who offended in the same location independently. Mann-Whitney U tests were used as the assumption of normality was not met for any of the measures.

2.2.5. Analysis Dimensions

All analyses were conducted for each offense type separately. Given that crime locations are the product of the convergence of offenders' activity space with crime opportunities, and opportunity structures differ for different offenses (e.g., residential and non-residential burglary), the relationship between crime locations and activity space can also differ between offense types.

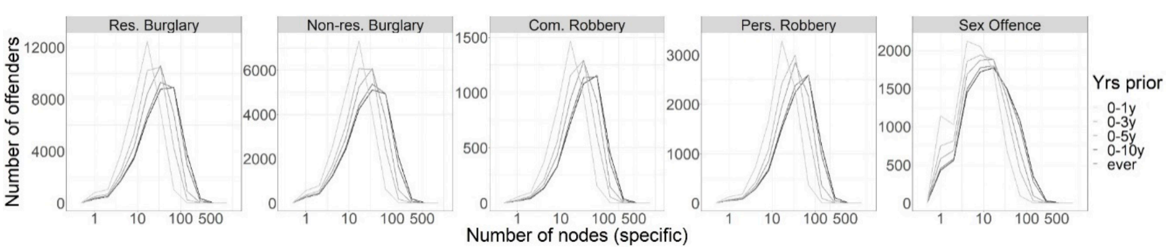
We also considered different timeframes of varying duration preceding the reference offense. Activity space is dynamic: peoples' activity space changes as places are added or stop being visited [56]. Less recently visited places have lower odds of crime location choice [3,9,25]. However, the temporal limit on node influence is not known, and some historical nodes may still have a bearing: childhood nodes could feature prominently in offenders' "mental maps"; offenders might have returned to live at or near a past family home without this being known to police [57,58]. We therefore calculated activity space measures using activity nodes that were active (i.e., occurred or were end-dated) within the year, three years, five years, ten years, and more than ten years preceding the reference offense. Homology analyses were repeated across 1 year and "ever" timeframes.

3. Results

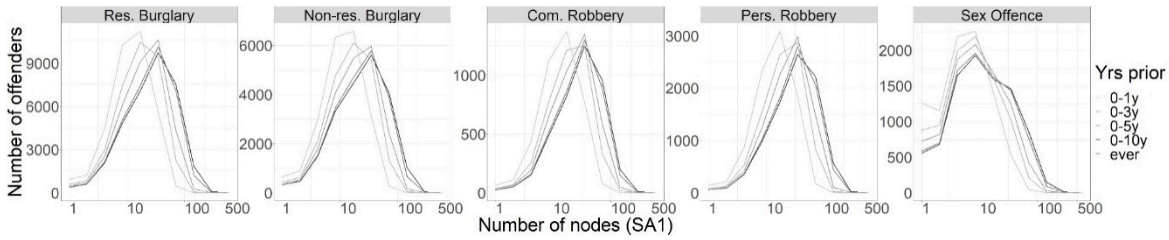
3.1. Number of Activity Nodes

Figure 2 presents the distributions of the number of activity nodes per offender, by offense type and pre-offense timeframe, for specific coordinates (top) and SA1s (bottom). The number of specific locations per offender ranged from 1 to 581 and the distributions are highly skewed: most offenders have a small number of nodes, and a small number of offenders have many nodes. The number of nodes increases with the length of the pre-offense period: as is to be expected since the measure is cumulative. The medians were comparable for burglary (from 14 for one year to 36 "ever") and robbery (16 to 43) but smaller for sex offenses (6 to 13).

The number of SA1s per offender followed much the same distributions, except in the upper extremes. The median number of SA1s per offender ranged from 13/12 (0–1y) to 29/29 ("ever") for residential/non-residential burglary, 14/15 to 32/24 for commercial/personal robbery and 5 to 11 for sex offenses. The maximum number of SA1s per offender was 481. As shown in Figure 1, distributions are very similar when comparing the top and bottom graphs, though the tails at the upper end of the distribution are truncated when specific locations are aggregated to SA1 resolution. This suggests a) that for most offenders, their few nodes are distributed among different SA1s and b) for the few offenders with many nodes, many of those nodes are clustered within SA1s.



(**a**)



(**b**)

Figure 2. Distribution of activity nodes per offender by reference offense type and timeframe preceding the reference offense based on specific coordinates (**a**) and SA1 census units (**b**).

3.2. Activity Space Size

To put the number of SA1 nodes into a geographic context, 49.3% of offender SA1 nodes were located in the most populated urban TAs in NZ: Auckland, Wellington, Christchurch, and Hamilton. The numbers of SA1s per offender represent very small proportions of these urban areas, which contain between 1449 (Hamilton) and 13680 (Auckland) SA1s. For example, 13 SA1s represents 0.89% of Hamilton's SA1s and 0.09% of Auckland's; 29 represents 2.00% and 0.12% respectively; and the maximum of 421 (if they were all in the same city) represents 29.05%. and 3.08%. The activity spaces of offenders as recorded in police data, therefore, are typically constrained to a small portion of any given urban area in which they conduct their routine activities.

The distributions of activity space ranges, however, reveal that these activities are frequently distributed across multiple urban areas (while limited to small areas within each), as shown in Figure 3. The median distance range was 267 km/256 km for residential/non-residential burglary for 1 year (420 km/395 km ever), 268 km/319 km for commercial/personal robbery for 1 year (417/457 ever) and 78 km for sex offenses for 1 year (202 km ever). The maximum distance range was between 1400 and 1600 km for all offense types and activity node timeframes, representing nodes at opposite ends of the country (lengthwise). Interestingly, these maxima emerged early, appearing in the one-year pre-offense activity space timeframe. However, with increasing pre-offense time spans, there were increasing numbers of offenders whose activity spaces spanned multiple cities at greater distances apart.

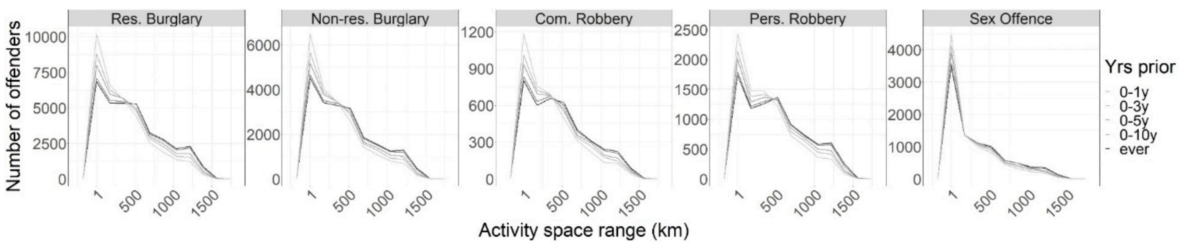
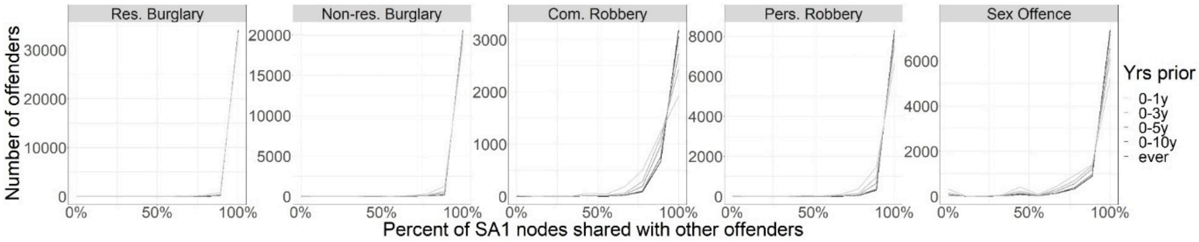


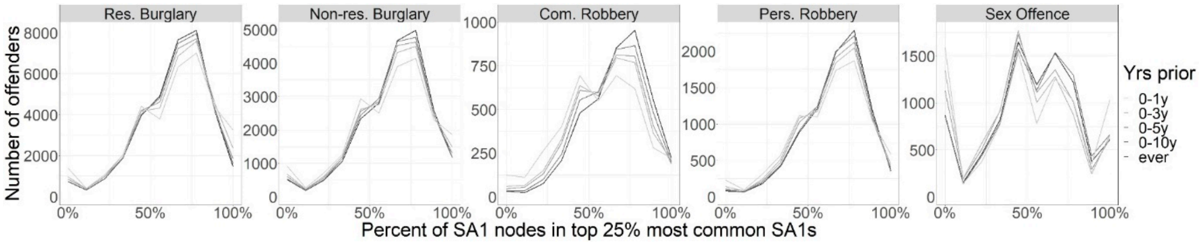
Figure 3. Distribution of activity space range per offender by reference offense type and timeframe preceding the reference offense.

3.3. Shared Activity Space

Most offenders shared most or all of their activity nodes with at least one other offender (Figure 4, top). This is perhaps unsurprising given the volume and geographic spread of activity nodes, which would increase the likelihood of shared SA1s. Correspondingly, residential burglars, with the highest volumes of activity nodes, had larger proportions of SA1s shared with other residential burglars, while commercial robbers and sex offenders, with the lowest volumes of activity nodes, had smaller proportions of shared SA1s. Also as expected, the proportion of shared activity space increased as more activity nodes were included with increasing time prior to the reference offense (illustrated in Figure 4 by the darker lines extending higher on the Y axis than the lighter lines, at upper limit of 100% on the X axis).



(**a**)



(**b**)

Figure 4. Distribution of proportion of activity space shared with any other offenders (**a**) and in the top 25% most commonly shared SA1s (**b**), by reference offense type and timeframe preceding the reference offense.

Even the proportion of activity space falling in the 25% most common SA1s was greater than 50% for most offenders (Figure 3, bottom). The medians were 68%/67% for one year to 68%/68% for residential/non-residential burglary, 58%/67% to 69%/70% for commercial/personal robbery and 52%–56% for sex offenses. The upturns of the distributions at 0% and 100% shared reflect cases where offenders only have one activity node, meaning either all or no activity space is shared. Note also that the number of offenders sharing the 25% most common SA1s is relative to offense type. For example, the top 25% of SA1s for residential burglary offender nodes "ever" were shared by 55 or more offenders; the top 25% of SA1s for sex offender nodes "ever" were shared by 9 or more offenders.

Overall, these results point to a lack of distinctiveness in offender activity spaces recorded in police data, in that any given node is likely to be shared. The following analyses provide insight into whether the shared portions of individuals' activity space are shared with the same or different offenders.

3.4. Homology and Differentiation

There were statistically significant but small correlations between the proximity of offenders' reference offenses and the spatial similarity of their activity spaces. As shown by the fitted regression lines in red in Figure 5, broadly, the closer the reference offenses, the more shared activity space; the farther away the reference offenses, the less shared activity space. The correlations ranged from −0.07 for urban sex offenders' 1-year pre-offense activity space to −0.24 for rural burglars' "ever" (see Supplementary Material S5 for full results). The higher correlations for rural offenders on this measure are likely a product of the larger SA1s in rural areas, increasingly the likelihood of offenses and activity nodes falling into the same SA1.

Notably, the proportion of activity space shared between specific offender pairs is much smaller than that shared with offenders in general (see previous section), suggesting heterogeneity in offenders' activity spaces. In Figure 5 about half of the offender pair points appear at 0% on the Y axes. The median percentage shared activity space was 0% for all groups except rural residential burglars (2.1%), rural commercial robbers (3.0%)

and rural personal robbers (4.5%). Almost no offender pairs shared 100% of their activity space.

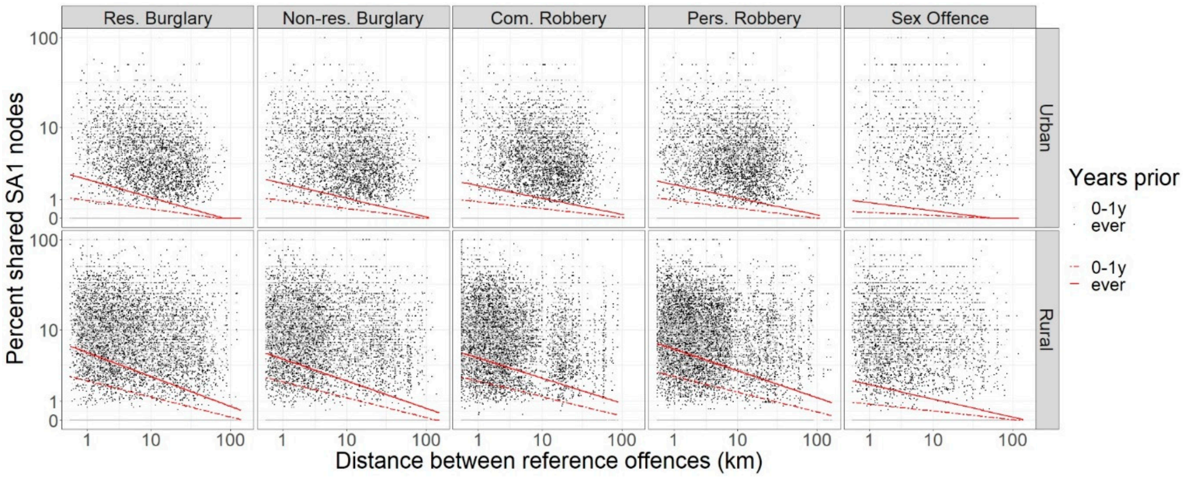


Figure 5. Proportion of activity space shared between pairs of offenders by distance between their reference offenses.

As Figure 6 shows, to some extent, the closer any two offenders' reference offenses, the closer together their nearest activity nodes; the farther away their offenses, the farther apart their nearest nodes. The correlations ranged from 0.14 for urban commercial robbers' pre-offense activity space "ever", to 0.29 for rural burglars' 1-year pre-offense activity space (see Supplementary Material S5 for all results).

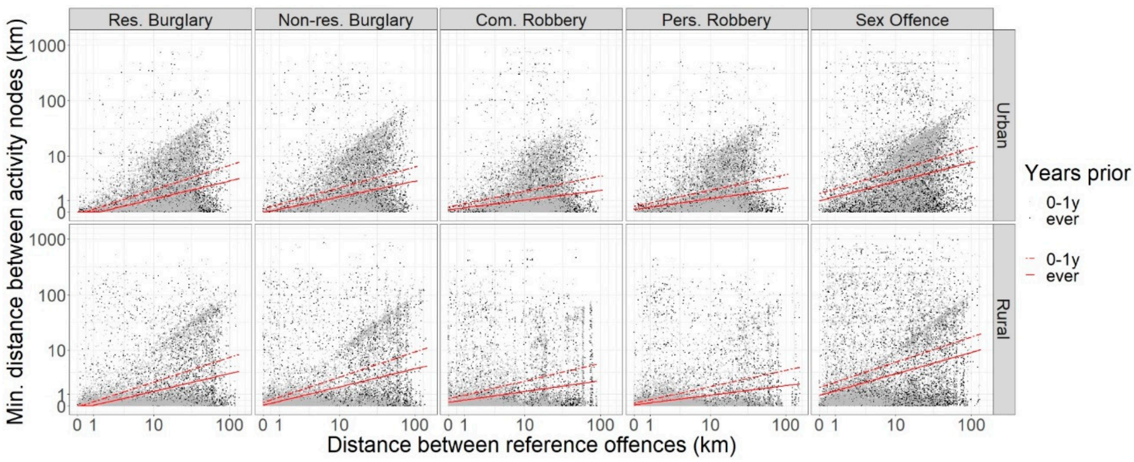


Figure 6. Minimum distance between nearest activity nodes by distance between reference offenses.

As shown in Figure 7, there was a general tendency that the closer any two offenders' reference offenses, the closer their activity nodes on average; the farther away their offenses, the farther apart their nodes. The correlations ranged from 0.15 for rural commercial robbers' pre-offense activity space "ever", to 0.28 for urban burglars' 1-year pre-offense activity space (see Supplementary Material S5 for all results).

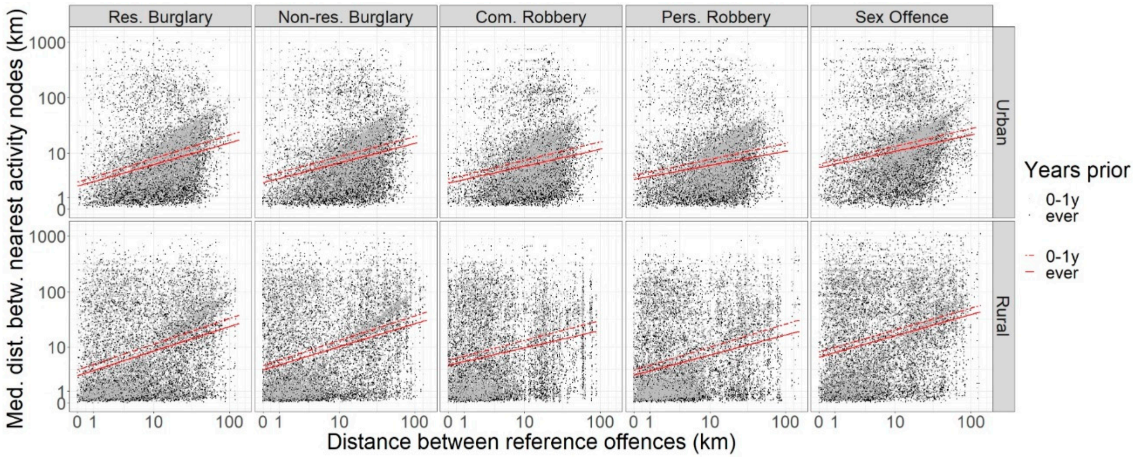


Figure 7. Median distance between nearest activity nodes by distance between reference offenses.

Correlations were slightly weaker for the median nearest node distance than the minimum node distance for rural offenders. In contrast, correlations were slightly stronger for the median nearest node distance for urban burglary and robbery offenders. These small differences are potentially explained by the longer distances between nodes for rural offenders; urban offenders' nodes would be closer together.

Burglars showed the strongest homology between offense proximity and activity space similarity, with very similar results for residential and non-residential burglars. Likewise, there were near identical results for commercial and personal robbers, both displaying less homology than burglars. Sex offenders showed the least homology as measured by percent shared activity space but more than robbers in terms of the distance between activity nodes.

Of pairs of offenders who offended in the same location, co-offenders shared more activity space, and had more proximal activity nodes than independent offenders. All tests of difference between co-offenders and other same location offenders were statistically significant, with small to moderate effect sizes (see Supplementary Material S6 for full results). On average, robbery and residential burglary co-offenders shared at least one exact location node (specific coordinates) in the "ever" timeframe, as indicated by median minimum node distances of 0. The minimum distance between nodes was under 200 m on average for burglary and robbery co-offenders, and under 500 m for independent burglars and robbers. Co-offending and independent sex offenders were less likely to share any nodes (specific coordinates or SA1s); their nearest nodes were also further apart on average (co-offenders: 50–540 m, independent: 280 m–1.86 km). Notably, even among co-offenders, the median distance from each of one co-offender's nodes to the nearest of the other co-offender's nodes was much longer, on average, than the minimum distance between their nodes.

Collectively these results suggest that offenders who offend in spatially similar locations tend to be spatially similar in parts of their activity space, though not necessarily in its entirety. Furthermore, these homology results are not solely attributable to co-offending. However, we also confirmed that co-offenders displayed greater activity space similarity than offenders who offended in the same location separately.

4. Discussion

We used routinely collected police data to provide insight into offenders' activity spaces in terms of the number of activity nodes, their geographic range, and the extent to which they are shared between or differentiate offenders. We believe this to be the largest and most comprehensive study of offender activity space to date, considering the frequency and types of activity nodes included.

On the surface the one-year activity node distributions appear comparable to the results of Rossmo et al.'s [29] study of US parolees and Menting [28] and Bernasco's [27] studies of young Dutch offenders, considering the different timeframes involved. Our medians were 13 to 15 for burglary/robbery and 5 for sex offenses, for nodes that were current during the year prior to the reference offense, compared with 4-7 nodes in the above studies over much shorter timeframes. However, our distributions were wider and appeared more skewed: a minority of offenders had no prior activity nodes on record; many offenders had few nodes; a few had very many. This result is likely a combination of both missingness and reality: we have more complete data for some individuals than for others, and some individuals genuinely have more activity nodes than others, as found in criminal cohorts in the 3 studies above and in the general population [51,56,59,60]. Future research could help establish the extent of (in)completeness by comparing police data at an individual level with alternative data sources such as surveys or GPS data as used in the studies above.

That the activity spaces frequently spanned multiple urban areas is consistent with the few NZ studies of the distances between offenders' home addresses and their offenses, and of the mobility of New Zealanders in general. In NZ, the home-crime distances of sex offenders tend to be longer on average than overseas, with higher proportions of "commuter" offenders whose homes are outside the radius of their offenses [61–64]. Davidson's [65] study of Christchurch burglars' home-crime distances found more geographically constrained patterns but may not represent contemporary trends, given changes in societal travel patterns and mobility. In 21st Century New Zealand, people move frequently [66], family and friends can be widely dispersed across different towns and cities, and domestic travel by car or plane is common, despite the long distances involved [67–72]. The inclusion of prisons as activity nodes, sometimes recorded in the address data, would also have contributed to these distances because offenders may be transferred to prisons a long way from their community; they may also not return to the same community on release [73] (Prison addresses were not always readily identifiable in the data, precluding a comprehensive investigation into the extent to which they accounted for wide activity space ranges.). It is no surprise, therefore, that NZ offenders' activity spaces spanned much longer distances than those of the offenders studied by Rossmo et al. [29], Menting [28] and Bernasco [27], given these studies were more geographically and temporally constrained, and involved cohorts with likely more limited mobility (offenders on GPS monitoring and young people). However, the extent to which our results are unique to New Zealand or indicate that offender populations have wider activity ranges than captured by studies with smaller study areas warrants further investigation. We encourage future studies of offender activity space in other countries to widen their spatial scope.

The analyses of the number and geographic range of activity nodes reveal potential and problems for the use of this data in research and practice. In terms of volume, the data are promising: there were a variety of activity nodes beyond home addresses available for the burglary, robbery and—to a lesser extent—sex offenders included in this study. There was also considerable information gain when extending the timeframe to less recent activity nodes (burglary median 29 nodes, robbery 34, sex offenses 11), though it remains for future research to explore whether this additional information yields any signal: we do not yet know whether these "older" activity nodes bear on the locations of future offenses. In contrast, the distributions were very similar regardless of the unit of analysis, which may partly reflect the small size of SA1s in the urban areas in which activity nodes were concentrated. Research and analysis may therefore benefit from the use of aggregate spatial units of SA1 size without much loss of information, at least in urban areas.

The geographic span of the offenders' activity spaces indicates a potential limitation of this data. Given that offenders' activity nodes were often widely dispersed, a high proportion of these nodes are likely to have little bearing on their choice of crime locations at a micro-geographic or neighborhood level. There may even be no activity nodes known to police in proximity to offenders' latest offense locations. In crime location choice research, if the data do not include—with sufficient frequency—the more proximal activity nodes likely to have influenced a given crime location choice, models may fail to identify relationships between activity nodes and crime locations. In crime investigations, there is potential to miss possible suspects by narrowly focusing on those with local nodes in police data, highlighting the importance of supplementing police information with other sources of information on suspects' activity nodes, be it through data sharing agreements with other agencies or on an individual basis for named suspects in a given investigation.

Any given activity node (SA1) was likely to be shared with other offenders who had committed the same reference offense, though these nodes were not necessarily "active" at the same time. This result is consistent with Hart and colleagues' [60] finding that 80% of the paths between young Australians' activity nodes were shared. The most frequently shared nodes reflected places expected to have high numbers of offenders residing or visiting (prisons, police stations, court houses) or high numbers of people in general (consistent with the findings of Menting et al. [20]). The latter would generate more activity node records in police data through higher levels of crime opportunity and higher odds of encountering police during proactive patrols.

Comparing any two offenders, however, showed much less overlap between their particular activity spaces, signaling considerable individual differences in offenders' routine activity patterns as captured in this data. The homology and differentiation results suggest that those routine activity spaces were—marginally—more likely to converge the closer together offenders' latest offenses were. Our results thus provide evidence of a small degree of spatial homology and differentiation, and some insights into its causes.

For example, reference offense co-offenders had greater activity space similarity than offenders who offended in the same location but independently. One potential explanation for this finding is that it reflects co-offenders who have also committed other crimes together in the past, and who thus share prior crime nodes. Other potential explanations reflect possible familial or social connections between co-offenders. On occasion, co-offenders may be family members [74,75], who may therefore share home addresses, family home address, or school nodes. However, more frequently, co-offenders are connected socially [76,77]. Human mobility studies have shown that people more closely connected in a social network have more similar activity spaces than those who are not socially connected [78–81]. Co-offenders' social ties could be a product of living in close proximity or attending the same school, and both a cause and effect of sharing "hangout" nodes [77,82,83]. In NZ, as elsewhere, co-offending can occur as a part of membership of gangs, from the fluid, loosely connected structures of youth gangs, to more hierarchical organized crime groups [84–88]. Many youth gangs align themselves to particular neighborhoods reflective of the shared activity space of their members, and more formalized gangs are arranged into local "chapters" [84–86]. Consistent with Lammers' study [35], however, co-offenders did not share most of their activity nodes.

However, since co-offenders only made up a small proportion of offender pairs used for the homology correlations (and excluding them made no difference to the results), other mechanisms must be at play. For example, those who share activity space are likely to become aware of the same criminal opportunities; those with different activity spaces are exposed to different opportunities. Social networks also play a role: *previous* co-offenders (who share prior crime nodes) share information with each other about crime opportunities that influences where they offend in the future not only together but separately [89,90].

Robbery offenders displayed less homology than burglars. This could be expected since robbery tends to concentrate in commercial areas with high numbers of potential targets [13,91,92], which would attract potential offenders with disparate residential nodes [2,34]. That homology was still present may reflect the range of nodes in the data: offenders committing robberies in the same commercial area may also have committed prior offenses or have had prior interactions with police in the vicinity.

Those committing sex offenses appear to be a more spatially heterogenous group, also displaying less spatial overlap or proximity in their activity spaces. This might be an effect of there being fewer nodes in the dataset for these offenders, which would mean less chance of finding overlapping or proximal nodes. It might also be an effect of the heterogenous nature of these offenses. The sex offenses included a wide range of behaviors, from indecent exposure to rape, and included offenses against children and adults. They could have occurred at the offender's home, the victim's home, or elsewhere [93–97]. Sex offenders' strategies for identifying, approaching, and attacking victims vary widely from extended grooming to opportunistic "blitz" attacks [98–102]. Offenders might identify their victims through social networks (online or offline) or search for victims in target rich environments such as near schools or night-time economy districts [94,96,97,102–104]. All these factors would lead to variation in the relationship between offenders' activity spaces and their crime locations. Furthermore, social connection and sharing of information between sex offenders, such as about the locations of criminal opportunities, appears to be less common than with property crime [97], which would reduce network driven homology effects. Future research might seek to isolate spatial homology effects for different subtypes of sex offenders.

Several limitations of the data are also worth considering in interpreting our results. First, the data only identifies co-offenders where there was sufficient evidence to proceed against each offender. In some cases, multiple offenders may have been involved in a crime but not recorded as offenders due to a lack of evidence. Furthermore, offender pairs were only identifiable as co-offenders if the offense was both offenders' reference (i.e., latest) offense. Offender pairs would not be identified as co-offenders if one had committed a subsequent offense. To the extent that any co-offenders were thus treated as having offended independently, the differences between co-offenders and independent offenders will have been under-estimated.

Second, the data were not systematically recorded; they were only on file where collected for operational purposes. As noted, some offenders had more complete records, representing more of their routine activity space. Our conclusions are therefore limited to offenders' activity spaces to the extent identifiable by routinely collected police data.

An additional caveat is that not all activity space is equal with respect to crime location choice. Activity nodes that are more recent, more frequently visited, and have been in activity space for a longer time have stronger associations with crime locations [27,28,105], as do activity nodes that provide the most relevant knowledge of criminal opportunity: prior crimes of the same nature [26,106]. We could therefore expect greater spatial similarity in those parts of offenders' activity spaces that have had the greatest influence in producing similarity of their crime locations. Indeed, our results suggest that offenders who offend in spatially similar locations are spatially similar in *parts* of their activity space. Further research would be needed to determine, with the present dataset, which activity nodes are most predictive of crime locations, and whether homology effects are larger when activity space is isolated to, or weighted by, these most influential activity nodes.

5. Conclusions

Our exploration of police data on a wider range of activity nodes than included in previous studies suggests that geographic crime analysis and research could benefit from the use of such datasets, though with some caution. Our findings in relation to the distances within and between offenders' activity nodes—and the implications we have highlighted for research and practice—are likely to generalize to other countries with similar levels of internal mobility and dispersed populations (at both micro and macro geographic scales). In addition, our findings in relation to the sparsity, and heterogeneity, of sex offenders' activity nodes in this data are also likely to apply to equivalent datasets in other jurisdictions. However, despite these limitations, the data provided new direct evidence of the applicability of the homology and differentiation "profiling assumptions" in the geographic profiling context of offenders' offense locations and activity spaces. We encourage further research, with data from police and alternative sources, into the specific elements of activity space that display greater homology and differentiation with respect to offenders' crime locations, and which could further support geographic profiling by enabling specific nodes of potential suspects to receive more weight in prioritization decisions.

Supplementary Materials: The following are available online at www.mdpi.com/2220-9964/10/2/47/s1, Document S1: Data parameter definitions; Document S2: Data filters and sample attrition; Document S3: Geocoding process, accuracy and precision; Document S4: Use of date information in identifying 'prior' activity nodes; Document S5: Offense distance and activity space similarity correlations; Document S6: Co-offender and other same location offender comparisons.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, Sophie Curtis-Ham; methodology, Sophie Curtis-Ham, Wim Bernasco, Oleg N. Medvedev, Devon L. L. Polaschek; software, Sophie Curtis-Ham; validation, Sophie Curtis-Ham; formal analysis, Sophie Curtis-Ham; investigation, Sophie Curtis-Ham; resources, Sophie Curtis-Ham; data curation, Sophie Curtis-Ham; writing—original draft preparation, Sophie Curtis-Ham; writing—review and editing, Sophie Curtis-Ham, Wim Bernasco, Oleg N. Medvedev, Devon L. L. Polaschek; visualization, Sophie Curtis-Ham; supervision, Devon L. L. Polaschek, Oleg N. Medvedev; project administration, Sophie Curtis-Ham; funding acquisition, Sophie Curtis-Ham All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a University of Waikato doctoral scholarship. The APC was funded by the University of Waikato.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the School of Psychology Human Research Ethics Committee of the University of Waikato (reference #19:13, 6 June 2019). Approval of access to data for this study was obtained from the NZ Police Research Panel (reference EV-12-462, 11 July 2019).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are not publicly available due to them containing information that could compromise privacy. The data are available from New Zealand Police, but restrictions apply to the availability of these data. The results presented in this paper are the work of the authors and the authors take full responsibility for the outputs.

Acknowledgments: We gratefully acknowledge the assistance of the NZ Police staff who provided access to and advice on the data used in this research and who reviewed the manuscript prior to submission. We also thank Professor Stijn Ruiter of the Netherlands Institute for the Study of Crime and Law Enforcement for providing the geocoding R script used for the school locations, and the anonymous reviewers, whose insights helped to improve this paper.

Conflicts of Interest: Sophie Curtis-Ham is employed as a researcher at New Zealand Police. This study was not conducted as a part of that employment. Sophie Curtis-Ham is assistant guest editor for the Special Issue in which this paper appears. This paper was processed by an editorial team independent from the Special Issue guest editors. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Brantingham, P.L.; Brantingham, P.J. Notes on the geometry of crime. In *Environmental Criminology*; Waveland Press: Prospect Heights, IL, USA, 1991; pp. 27–54; ISBN 978-0-8039-1678-4.
2. Brantingham, P.L.; Brantingham, P.J. Environment, routine, and situation: Toward a pattern theory of crime. In *Routine Activity and Rational Choice-Advances in Criminological Theory*; Clarke, R.V., Felson, M., Eds.; Transaction Publishers: Piscataway, NJ, USA, 1993; pp. 259–294; ISBN 978-1-56000-087-7.
3. Ruiter, S. Crime location choice. In *The Oxford Handbook of Offender Decision Making*; Bernasco, W., Van Gelder, J.-L., Elffers, H., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 398–420.

4. Curtis-Ham, S.; Bernasco, W.; Medvedev, O.N.; Polaschek, D. A Framework for Estimating Crime Location Choice Based on Awareness Space. *Crime Sci.* **2020**, *9*, 1–14, doi:10.1186/s40163-020-00132-7.
5. Chesher, C. Digitising the Beat: Police Databases and Incorporeal Transformations. *Convergence* **1997**, *3*, 72–81, doi:10.1177/135485659700300209.
6. Garicano, L.; Heaton, P. Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments. *J. Labor Econ.* **2010**, *28*, 167–201, doi:10.1086/649844.
7. Home Office National Law Enforcement Data Programme. *Law Enforcement Data Service (LEDS)–Privacy Impact Assessment Report*; Home Office: London, UK, 2018.
8. Schellenberg, K. Police Information Systems, Information Practices and Individual Privacy. *Can. Public Policy* **1997**, *23*, 23–39, doi:10.2307/3552129.
9. Long, D.; Liu, L.; Feng, J.; Zhou, S. Assessing the Influence of Prior on Subsequent Street Robbery Location Choices: A Case Study in ZG City, China. *Sustainability* **2018**, *10*, 1818, doi:10.3390/su10061818.
10. Canter, D. Geographical profiling of criminals. In *Principles of Geographical Offender Profiling*; Canter, D., Youngs, D., Eds.; Ashgate: Aldershot, UK, 2008; pp. 259–270; ISBN 978-0-7546-2547-6.
11. Rossmo, D.K. *Geographic Profiling*; CRC Press: Boca Raton, FL, USA, 2000; ISBN 0-8493-8129-0.
12. Cohen, L.E.; Felson, M. Social Change and Crime Rate Trends: A Routine Activity Approach. *Am. Sociol. Rev.* **1979**, *44*, 588–608, doi:10.2307/2094589.
13. Weisburd, D.; Eck, J.E.; Braga, A.A.; Telep, C.W.; Cave, B.; Bowers, K.; Bruinsma, G.J.N.; Gill, C.; Groff, E.R.; Hibdon, J.; et al. *Place Matters: Criminology for the Twenty-First Century*; Cambridge University Press: Cambridge, UK, 2016.
14. Weisburd, D. The Law of Crime Concentration and the Criminology of Place. *Criminology* **2015**, *53*, 133–157, doi:10.1111/1745-9125.12070.
15. Townsley, M. Offender mobility. In *Environmental Criminology and Crime Analysis*; Wortley, R., Townsley, M., Eds.; Routledge: London, UK, 2016; pp. 142–161; ISBN 978-1-317-48710-4.
16. Canter, D.; Hodge, S. Criminals' mental maps. In *Atlas of Crime: Mapping the Criminal Landscape*; Turnbull, L., Hendrix, E., Dent, B., Eds.; Oryx Press: Phoenix, AZ, 2000; pp. 186–191.
17. Davies, A.; Dale, A. Locating the Stranger Rapist. *Med. Sci. Law* **1996**, *36*, 146–156, doi:10.1177/002580249603600210.
18. Pettiway, L.E. Copping Crack: The Travel Behavior of Crack Users. *Justice Q.* **1995**, *12*, 499–524, doi:10.1080/07418829500096111.
19. Rengert, G.; Wasilchick, J. *Suburban Burglary: A Time and a Place for Everything*; C.C. Thomas: Springfield, IL, USA, 1985; ISBN 978-0-398-05142-6.
20. Summers, L.; Johnson, S.D.; Rengert, G. The use of maps in offender interviewing. In *Offenders on Offending: Learning about Crime from Criminals*; Bernasco, W., Ed.; Willan: Collumpton, Devon, UK, 2010; pp. 246–272.
21. van Daele, S. Itinerant crime groups: Mobility attributed to anchor points? In *Contemporary Issues in the Empirical Study of Crime;* Governance of Security Research Paper Series; Pauwels, L., Ponsaers, P., Vande Walle, G., Vander Beken, T., Vander Laenen, F., Vermeulen, G., Cools, M., De Kimpe, S., De Ruyver, B., Easton, M., Eds.; Maklu: Antwerp, Belgium, 2009; Volume 1, pp. 211–225.
22. Wiles, P.; Costello, A. The "road to nowhere": The evidence for travelling criminals. In *Principles of Geographical Offender Profiling*; Canter, D., Youngs, D., Eds.; Ashgate: Aldershot, UK, 2008; pp. 165–175.
23. Frith, M.J. Modelling Taste Heterogeneity Regarding Offence Location Choices. *J. Choice Model.* **2019**, *33*, 100187, doi:doi.org/10.1016/j.jocm.2019.100187.
24. Menting, B. Awareness × Opportunity: Testing Interactions between Activity Nodes and Criminal Opportunity in Predicting Crime Location Choice. *Br. J. Criminol.* **2018**, *58*, 1171–1192, doi:10.1093/bjc/azx049.
25. Menting, B.; Lammers, M.; Ruiter, S.; Bernasco, W. Family Matters: Effects of Family Members' Residential Areas on Crime Location Choice. *Criminology* **2016**, *54*, 413–433, doi:10.1111/1745-9125.12109.
26. Van Sleeuwen, S.E.M.; Ruiter, S.; Menting, B. A Time for a Crime: Temporal Aspects of Repeat Offenders' Crime Location Choices. *J. Res. Crime Delinq.* **2018**, *55*, 538–568, doi:10.1177/0022427818766395.
27. Bernasco, W. Adolescent Offenders' Current Whereabouts Predict Locations of Their Future Crimes. *PLoS ONE* **2019**, *14*, e0210733, doi:10.1371/journal.pone.0210733.
28. Menting, B.; Lammers, M.; Ruiter, S.; Bernasco, W. The Influence of Activity Space and Visiting Frequency on Crime Location Choice: Findings from an Online Self-Report Survey. *Br. J. Criminol.* **2020**, *60*, 303–322, doi:10.1093/bjc/azz044.
29. Rossmo, D.K.; Lu, Y.; Fang, T.B. Spatial-temporal crime paths. In *Patterns, Prevention, and Geometry of Crime*; Andresen, M.A., Kinney, J.B., Eds.; Routledge: London, UK, 2012; pp. 3–15.
30. Rossmo, D.K.; Rombouts, S. Geographic profiling. In *Environmental Criminology and Crime Analysis*; Wortley, R., Mazerolle, L., Eds.; Willan: Collumpton, Devon, UK, 2008; pp. 136–149.
31. Canter, D. Offender Profiling and Criminal Differentiation. *Leg. Criminol. Psychol.* **2000**, *5*, 23–46, doi:10.1348/135532500167958.
32. Doan, B.; Snook, B. A Failure to Find Empirical Support for the Homology Assumption in Criminal Profiling. *J. Police Crim. Psychol.* **2008**, *23*, 61–70, doi:10.1007/s11896-008-9026-7.
33. Woodhams, J.; Toye, K. An Empirical Test of the Assumptions of Case Linkage and Offender Profiling with Serial Business Robberies. *Psychol. Public Policy Law* **2007**, *13*, 59–85, doi:10.1037/1076-8971.13.1.59.
34. Costanzo, C.; Halperin, W.; Gale, N. Criminal mobility and the directional component in journeys to crime. In *Metropolitan Crime Patterns*; Figlio, R., Hakim, S., Rengert, G., Eds.; Criminal Justice Press: Monsey, NY, USA, 1986; pp. 73–96.

35. Lammers, M. Co-Offenders' Crime Location Choice: Do Co-Offending Groups Commit Crimes in Their Shared Awareness Space? *Br. J. Criminol.* **2018**, *58*, 1193–1211, doi:10.1093/bjc/azx069.
36. Tayebi, M.A.; Frank, R.; Glässer, U. Understanding the Link between Social and Spatial Distance in the Crime World. In Proceedings of the 20th International Conference on Intelligent User Interfaces, Redondo Beach, CA, USA, 6 November 2012.
37. Malm, A.; Kinney, J.B.; Pollard, N. Social Network and Distance Correlates of Criminal Associates Involved in Illicit Drug Production. *Secur. J.* **2008**, *21*, 77–94, doi:10.1057/palgrave.sj.8350069.
38. Bernasco, W. Them Again?: Same-Offender Involvement in Repeat and near Repeat Burglaries. *Eur. J. Criminol.* **2008**, *5*, 411–431, doi:10.1177/1477370808095124.
39. Haginoya, S.; Hanayama, A.; Koike, T. Linkage Analysis Using Geographical Proximity: A Test of the Efficacy of Distance Measures. *J. Criminol. Res. Policy Pr.* **2020**, doi:10.1108/JCRPP-01-2020-0006.
40. Johnson, S.D.; Summers, L.; Pease, K. Offender as Forager? A Direct Test of the Boost Account of Victimization. *J. Quant. Criminol.* **2009**, *25*, 181–200, doi:10.1007/s10940-008-9060-8.
41. Tonkin, M.; Woodhams, J.; Bull, R.; Bond, J.W.; Palmer, E.J. Linking Different Types of Crime Using Geographical and Temporal Proximity. *Crim. Justice Behav.* **2011**, *38*, 1069–1088, doi:10.1177/0093854811418599.
42. Tonkin, M.; Santtila, P.; Bull, R. The Linking of Burglary Crimes Using Offender Behaviour: Testing Research Cross-Nationally and Exploring Methodology.*Leg. Criminol. Psychol.* **2012**, *17*, 276–293, doi:10.1111/j.2044-8333.2010.02007.x.
43. De Montjoye, Y.-A.; Hidalgo, C.; Verleysen, M.; Blondel, V. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Sci. Rep.* **2013**, *3*, 1376, doi:10.1038/srep01376.
44. Yeoman, A.; Cook, L.W. *The Kiwi Nest: 60 Years of Change in New Zealand Families*; Families Commission: Wellington, New Zealand, 2008.
45. R Core Team. *R: A Language and Environment for Statistical Computing.*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
46. Andresen, M.A.; Malleson, N.; Steenbeek, W.; Townsley, M.; Vandeviver, C. Minimum Geocoding Match Rates: An International Study of the Impact of Data and Areal Unit Sizes. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 1306–1322, doi:10.1080/13658816.2020.1725015.
47. Briz-Redón, Á.; Martinez-Ruiz, F.; Montes, F. Re-Estimating a Minimum Acceptable Geocoding Hit Rate for Conducting a Spatial Analysis. *Int. J. Geogr. Inf. Sci.* **2019**, *34*, 1283–1305, doi:10.1080/13658816.2019.1703994.
48. Bernasco, W. Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places. *J. Quant. Criminol.* **2010**, *26*, 113–138, doi:10.1007/s10940-009-9086-6.
49. Hanson, S.; Huff, O.J. Systematic Variability in Repetitious Travel. *Transportation* **1988**, *15*, 111–135, doi:10.1007/BF00167983.
50. Pappalardo, L. The Origin of Heterogeneity in Human Mobility Ranges. In Proceedings of the CEUR Workshop Proceedings CEUR-WS, Bordeaux, France, 15 March 2016; Volume 1558.
51. Pappalardo, L.; Simini, F.; Rinzivillo, S.; Pedreschi, D.; Giannotti, F.; Barabási, A.-L. Returners and Explorers Dichotomy in Human Mobility. *Nat. Commun.* **2015**, *6*, 8166–8173, doi:10.1038/ncomms9166.
52. Schönfelder, S.; Axhausen, K.W. *Measuring the Size and Structure of Human Activity Spaces: The Longitudinal Perspective*; ETH: Zurich, Switzerland, 2002.
53. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439, doi:10.32614/RJ-2018-009.
54. Patterson, Z.; Farber, S. Potential Path Areas and Activity Spaces in Application: A Review. *Transp. Rev.* **2015**, *35*, 679–700, doi:10.1080/01441647.2015.1042944.
55. Mokros, A.; Alison, L.J. Is Offender Profiling Possible? Testing the Predicted Homology of Crime Scene Actions and Background Characteristics in a Sample of Rapists. *Leg. Criminol. Psychol.* **2002**, *7*, 25–43, doi:10.1348/135532502168360.
56. Alessandretti, L.; Sapiezynski, P.; Sekara, V.; Lehmann, S.; Baronchelli, A. Evidence for a Conserved Quantity in Human Mobility. *Nat. Hum. Behav.* **2018**, *2*, 485–491, doi:10.1038/s41562-018-0364-x.
57. Canter, D.; Shalev, K. Putting crime in its place: A psychological process in crime site selection. In *Principles of Geographical Offender Profiling*; Canter, D., Youngs, D., Eds.; Ashgate: Aldershot, UK, 2008; pp. 259–270; ISBN 978-0-7546-2547-6.
58. Marchment, Z.; Bouhana, N.; Gill, P. Lone Actor Terrorists: A Residence-to-Crime Approach. *Terror. Political Violence* **2018**, 1–26, doi:10.1080/09546553.2018.1481050.
59. González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding Individual Human Mobility Patterns. *Nature* **2008**, *453*, 779–782, doi:10.1038/nature06958.
60. Hart, T.C.; Birks, D.; Townsley, M.; Ruiter, S.; Bernasco, W. Activity nodes, activity spaces, and awareness spaces: Measuring geometry of crime's constructs with smartphone data. In *Space, Time, and Crime*; Hart, T.C., Lersch, K.M., Chataway, M., Eds.; Carolina Academic Press: Durham, NC, USA, 2020; pp. 156–176; ISBN 978-1-5310-1540-4.
61. Hammond, L. Geographical Profiling in a Novel Context: Prioritising the Search for New Zealand Sex Offenders. *Psychol. Crime Law* **2014**, *20*, 358–371, doi:10.1080/1068316X.2013.793331.
62. Lundrigan, S.; Czarnomski, S.; Wilson, M. Spatial and Environmental Consistency in Serial Sexual Assault. *J. Investig. Psych. Offender Profil.* **2010**, *7*, 15–30, doi:10.1002/jip.100.
63. Lundrigan, S.; Czarnomski, S. Spatial Characteristics of Serial Sexual Assault in New Zealand. *Aust. N. Z. J. Criminol.* **2006**, *39*, 218–231, doi:10.1375/acri.39.2.218.
64. Scott, D. *The Travelling Distances of Stranger Intruder Sex Offenders*; New Zealand Police: Wellington, New Zealand, 2012.

65. Davidson, R.N. Patterns of Residential Burglary in Christchurch. *N. Z. Geogr.* **1980**, *36*, 73–78, doi:10.1111/j.1745-7939.1980.tb00919.x.
66. Superu. *Residential Movement within New Zealand: Quantifying and Characterising the Transient Population*; Social Policy Evaluation and Research Unit: Wellington, New Zealand, 2018.
67. Ministry of Transport. *25 Years of New Zealand Travel: New Zealand Household Travel 1989–2014*; Ministry of Transport: Wellington, New Zealand, 2015.
68. Ministry of Transport. *New Zealand Household Travel Survey 2015–2017*; Ministry of Transport: Wellington, New Zealand, 2017.
69. Ministry of Transport. *Transport Outlook Current State 2016: A Summary of New Zealand's Transport System*; Ministry of Transport: Wellington, New Zealand, 2017.
70. Ministry of Transport. Inter-Regional Ground Travel by Residents: Data and Documentation Available online: https://www.transport.govt.nz/mot-resources/research-papers/inter-regional-ground-travel-data-from-qrious/ (accessed on 1 April 2020).
71. Vuletich, S.; Becken, S. *The Tourism Flows Model: Summary Document*; Ministry of Tourism: Wellington, New Zealand, 2007.
72. Witten, K.; Huakau, J.; Mavoa, S. Social and Recreational Travel: The Destinations, Travel Modes and CO2 Emissions of New Zealand Households. *Soc. Policy J. N. Zealand* **2011**, *37*, 172–184.
73. Robson, S.; Yesberg, J.A.; Wilson, M.S.; Polaschek, D.L.L. A Fresh Start or the Devil You Know? Examining Relationships between Release Location Choices, Community Experiences, and Recidivism for High-Risk Parolees. *Int. J. Offender Ther. Comp. Criminol.* **2019**, *64*, 35–653, doi:10.1177/0306624X19877589.
74. Goldsmith, A.; Halsey, M. Cousins in Crime: Mobility, Place and Belonging in Indigenous Youth Co-Offending. *Br. J. Criminol.* **2013**, *53*, 1157–1177, doi:10.1093/bjc/azt039.
75. Reiss, A.J.Jr.; Farrington, D.P. Advancing Knowledge about Co-Offending: Results from a Prospective Longitudinal Survey of London Males. *J. Crim. Law Criminol.* **1991**, *82*, 360–395, doi:10.2307/1143811.
76. Mcgloin, J.M.; Povitsky Stickle, W. Influence or Convenience? Disentangling Peer Influence and Co-Offending for Chronic Offenders. *J. Res. Crime Delinquency* **2011**, *48*, 419–447, doi:10.1177/0022427810393019.
77. Weerman, F. Co-Offending as Social Exchange: Explaining Characteristics of Co-Offending. *Br. J. Criminol.* **2003**, *43*, 398–416, doi:10.1093/bjc/43.2.398.
78. Liu, Z.; Qiao, Y.; Tao, S.; Lin, W.; Yang, J. Analyzing Human Mobility and Social Relationships from Cellular Network Data. In Proceedings of the 13th International Conference on Network and Service Management (CNSM), Tokyo, Japan, 26–30 November 2017; Volume 2017, pp. 1–6.
79. Toole, J.L.; Herrera-Yaqüe, C.; Schneider, C.M.; González, M.C. Coupling Human Mobility and Social Ties. *J. R. Soc. Interface* **2015**, *12*, 20141128, doi:10.1098/rsif.2014.1128.
80. Wang, D.; Pedreschi, D.; Song, C.; Giannotti, F.; Barabasi, A.-L. Human Mobility, Social Ties, and Link Prediction. In Proceedings of the Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, CA, USA, 21–24 August 2011; pp. 1100–1108.
81. Xu, Y.; Belyi, A.; Bojic, I.; Ratti, C. How Friends Share Urban Space: An Exploratory Spatiotemporal Analysis Using Mobile Phone Data. *Trans. GIS* **2017**, *21*, 468–487, doi:10.1111/tgis.12285.
82. Felson, M. The process of co-offending. In *Theory and Practice in Situational Crime Prevention*; Smith, M.J., Cornish, Derek. B., Eds.; Criminal Justice Press: Monsey, NY, USA, 2003; pp. 149–167.
83. Schaefer, D.R. Youth Co-Offending Networks: An Investigation of Social and Spatial Effects. *Soc. Netw.* **2012**, *34*, 141–149, doi:10.1016/j.socnet.2011.02.001.
84. Centre for Social Research and Evaluation. *From Wannabes to Youth Offenders: Youth Gangs in Counties Manukau*; Ministry of Social Development: Wellington, New Zealand, 2008.
85. Eggleston, E.J. New Zealand Youth Gangs: Key Findings and Recommendations from an Urban Ethnography. *Soc. Policy J. N. Z.* **2000**, *14*.
86. Gilbert, J. *Patched: The History of Gangs in New Zealand*; Auckland University Press: Auckland, New Zealand, 2013; ISBN 978-1-86940-729-2.
87. New Zealand Parliament. *Youth Gangs in New Zealand*; New Zealand Parliamentary Service: Wellington, New Zealand, 2019.
88. Office of the Minister of Police. *Whole-of-Government Action Plan to Reduce the Harms Caused by New Zealand Adult Gangs and Transnational Crime Groups*; New Zealand Cabinet: Wellington, New Zealand, 2014.
89. McAndrew, D. The structural analysis of criminal networks. In *The Social Psychology of Crime : Groups, Teams, and Networks*; Canter, D., Alison, L., Eds.; Ashgate: Aldershot, UK, 2000; pp. 53–94; ISBN 978-1-84014-435-2.
90. Lantz, B.; Ruback, R.B. A Networked Boost: Burglary Co-Offending and Repeat Victimization Using a Network Approach. *Crime Delinq.* **2017**, *63*, 1066–1090, doi:10.1177/0011128715597695.
91. Browning, C.R.; Byron, R.A.; Calder, C.A.; Krivo, L.J.; Kwan, M.-P.; Lee, J.-Y.; Peterson, R.D. Commercial Density, Residential Concentration, and Crime: Land Use Patterns and Violence in Neighborhood Context. *J. Res. Crime Delinq.* **2010**, *47*, 329–357, doi:10.1177/0022427810365906.
92. Tillyer, M.S.; Walter, R.J. Busy Businesses and Busy Contexts: The Distribution and Sources of Crime at Commercial Properties. *J. Res. Crime Delinq.* **2019**, *56*, 816–850, doi:10.1177/0022427819848083.
93. Chopin, J.; Caneppele, S. The Mobility Crime Triangle for Sexual Offenders and the Role of Individual and Environmental Factors. *Sex Abus. J. Res. Treat.* **2018**, *31*, 812–836, doi:10.1177/1079063218784558.

94. Chopin, J.; Caneppele, S. Geocoding Child Sexual Abuse: An Explorative Analysis on Journey to Crime and to Victimization from French Police Data. *Child Abus. Negl.* **2019**, *91*, 116–130, doi:10.1016/j.chiabu.2019.03.001.
95. Leclerc, B.; Wortley, R.; Smallbone, S. Investigating Mobility Patterns for Repetitive Sexual Contact in Adult Child Sex Offending. *J. Crim. Justice* **2010**, *38*, 648–656, doi:10.1016/j.jcrimjus.2010.04.038.
96. Leclerc, B.; Felson, M. Routine Activities Preceding Adolescent Sexual Abuse of Younger Children. *Sexual Abus. J. Res. Treat.* **2016**, *28*, 116–131, doi:10.1177/1079063214544331.
97. Smallbone, S.W.; Wortley, R.K. *Child Sexual Abuse in Queensland: Offender Characteristics & Modus Operandi*; Queensland Crime Commission: Queensland, Australia, 2000.
98. Balemba, S.; Beauregard, E. Where and When? Examining Spatiotemporal Aspects of Sexual Assault Events. *J. Sexual Aggress.* **2013**, *19*, 171–190, doi:10.1080/13552600.2012.703702.
99. Beauregard, E.; Rebocho, M.F.; Rossmo, D.K. Target Selection Patterns in Rape. *J. Investig. Psych. Offender Profil.* **2010**, *7*, 137–152, doi:10.1002/jip.117.
100. Beauregard, E.; Busina, I. Journey "during" Crime: Predicting Criminal Mobility Patterns in Sexual Assaults. *J. Interpers. Violence* **2013**, *28*, 2052–2067, doi:10.1177/0886260512471084.
101. Deslauriers-Varin, N.; Beauregard, E. Investigating Offending Consistency of Geographic and Environmental Factors among Serial Sex Offenders: A Comparison of Multiple Analytical Strategies. *Crim. Justice Behav.* **2013**, *40*, 156–179, doi:10.1177/0093854812467948.
102. Mogavero, M.C.; Hsu, K.-H. Sex Offender Mobility: An Application of Crime Pattern Theory among Child Sex Offenders. *Sex. Abus. J. Res. Treat.* **2017**, *30*, 908–931, doi:10.1177/1079063217712219.
103. Aslan, D. Critically Evaluating Typologies of Internet Sex Offenders: A Psychological Perspective. *J. Forensic Psychol. Pr.* **2011**, *11*, 406–431, doi:10.1080/15228932.2011.588925.
104. Beauregard, E.; Rossmo, D.K.; Proulx, J. A Descriptive Model of the Hunting Process of Serial Sex Offenders: A Rational Choice Perspective. *J. Fam. Violence* **2007**, *22*, 449–463, doi:10.1007/s10896-007-9101-3.
105. Bernasco, W. A Sentimental Journey to Crime: Effects of Residential History on Crime Location Choice. *Criminology* **2010**, *48*, 389–416, doi:10.1111/j.1745-9125.2010.00190.x.
106. Lammers, M.; Menting, B.; Ruiter, S.; Bernasco, W. Biting Once, Twice: The Influence of Prior on Subsequent Crime Location Choice. *Criminology* **2015**, *53*, 309–329, doi:10.1111/1745-9125.12071.

CHAPTER 4

A New Sampling Method for Discrete Crime Location Choice Modelling

Having introduced the data used in this thesis and identified its potential utility for answering questions about the links between offenders' mental maps and their crime locations, this chapter focuses on the method used to answer these questions. We employed discrete spatial choice modelling (DSCM) to test hypotheses derived from the theoretical model, following previous studies that used this method to examine offenders' crime location choices. But using DSCM with our 'big data' created a computational challenge. The following paper, published in the *Journal of Quantitative Criminology*, describes this challenge and a novel solution to it that facilitated the analyses presented in subsequent chapters.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5. Reproduced with permission from Springer Nature.

Journal of Quantitative Criminology
https://doi.org/10.1007/s10940-021-09526-5

ORIGINAL PAPER

The Importance of Importance Sampling: Exploring Methods of Sampling from Alternatives in Discrete Choice Models of Crime Location Choice

Sophie Curtis-Ham[1] · Wim Bernasco[2,3] · Oleg N. Medvedev[1] · Devon L. L. Polaschek[1]

Accepted: 6 July 2021

Abstract

Objectives The burgeoning field of individual level crime location choice research has required increasingly large datasets to model complex relationships between the attributes of potential crime locations and offenders' choices. This study tests methods of sampling aiming to overcome computational challenges involved in the use of such large datasets.

Methods Using police data on 38,120 residential and non-residential burglary, commercial and personal robbery and extra-familial sex offense locations and the offenders' pre-offense activity locations (e.g., home, family members' homes and prior crime locations), and in the context of the conditional logit formulation of the discrete spatial choice model, we tested a novel method for importance sampling of alternatives. The method over-samples potential crime locations near to offenders' activity locations that are more likely to be chosen for crime. We compared variants of this method with simple random sampling.

Results Importance sampling produced results more consistent with those produced without sampling compared with simple random sampling, and provided considerable computational savings. There were strong relationships between the locations of offenders' prior criminal and non-criminal activities and their crime locations.

Conclusions Importance sampling from alternatives is a relatively simple and effective method that enables future studies to use larger datasets (e.g., with more variables, wider study areas, or more granular spatial or spatio-temporal units) to yield greater insights into crime location choice. By examining non-residential burglary and sexual offenses, in New Zealand, the substantive results represent a novel contribution to the growing literature on offenders' spatial decision making.

Keywords Crime location choice · Discrete choice modelling · Police data · Routine activity nodes · Sampling from alternatives

---

✉ Sophie Curtis-Ham
SC398@students.waikato.ac.nz

Extended author information available on the last page of the article

Published online: 31 July 2021

Introduction

Understanding crime location choice at an individual level is a growing research undertaking with potential to inform criminal investigation and prevention activities beyond improvement in the explanation and prediction of where and when crime concentrates (Bernasco 2017; Ruiter 2017). There is an opportunity to deepen this understanding by increasing its geographic granularity. Whereas most prior studies of crime location choice focused on neighborhoods, recent research has demonstrated the relevance of smaller and more ecologically valid spatial units of analysis, such as streets segments (Frith et al. 2017), census blocks (Bernasco et al. 2013), and even individual properties (Vandeviver et al. 2015). Such smaller units can better account for heterogeneity in relevant variables that would be diluted within larger spatial units, so that the effects of these variables would become undetectable. In addition, because criminal opportunities are subject to cyclical temporal variations over days, weeks, months and seasons (Bernasco et al. 2017; van Sleeuwen et al. 2018), we need not only ask questions about optimal *places* for crime but also about optimal *times and places*, in combination, for crime. Further, there is a need to extend the scope or range of variables included in models of crime location choice. In particular, the measurement of individual activity spaces has been limited to residential addresses and prior offending locations only. Increasing the number of variables in crime location choice models can mean more offenses are needed in the dataset to enable sufficient statistical power.

However, because crime location choice studies combine data at the level of individual crimes with data aggregated to the spatial (or spatio-temporal) unit of analysis (e.g., neighborhoods, street segments or census blocks in four six-hour time blocks), the models that they utilize require estimation along three rather than two dimensions. Their estimation involves optimization along the dimensions of crimes × locations × variables, instead of crimes × variables or locations × variables. For example, a study with 1,000 crimes and 1,000 potential alternative locations for each crime, yields a dataset of 1 million rows. A study with more offenses, smaller units of analysis or a wider study area could involve upwards of 10,000 choices and 10,000 alternatives, yielding an unwieldy dataset of 100 million rows. If a temporal dimension were added, even as few as 1,000 spatial alternatives and 10 time periods would produce 10,000 space–time alternatives. These extensions require computer storage and processing capabilities that exceed the limits of available computing equipment outside of specialist computing labs (Vandeviver et al. 2015). Therefore, for model estimation to remain tractable outside of high-performance computer labs, there is a need to explore ways of reducing the computational burden. In fact, with the quick proliferation of model extensions and the advent of big data, high-performance computing environments may only offer a very short-term solution.

The issue of computational burden due to many choice alternatives has been addressed in other fields and on other types of decisions. Based on the results of McFadden (1977), researchers have estimated choice models by sampling from the decision makers' available choice sets when studying decisions such as residential choice: where to find/buy a house (Duncombe et al. 2001) and transportation route choice: how to travel from A to B (Frejinger et al. 2009). Within the discrete spatial choice modelling (DSCM) paradigm for studying crime location choice (see Ruiter 2017, for a review), only four studies (detailed below) have 'sampled from alternatives' instead of using the full set of locations that could have been chosen by any given offender. Specifically, these studies have used simple random sampling (McFadden 1977). However, research in other location choice domains suggests that simple random sampling may not necessarily be an optimal strategy if a priori knowledge on choice probabilities is available from previous research. For example, importance sampling strategies, that over-sample from the a priori most likely alternatives to be chosen (McFadden 1977), can lead to estimates that are closer to those produced by including all alternatives (Lemp and Kockelman 2012; Hassan et al. 2019). We therefore examined the effects of different methods of sampling from alternatives on the results of discrete spatial choice using a real-world crime dataset. To our knowledge this is the first study to directly compare different sampling methods in the crime location choice context. It contributes to a burgeoning literature on crime location choice (Ruiter 2017) and provides initial indications of the effects of different sampling strategies which could help guide future studies in this paradigm. Although not the primary purpose of this paper, we also provide new insights into crime location choice from a country that has not yet featured in the DSCM literature.

We begin with a review of DSCM crime location choice studies with a focus on sample sizes and methods. We then discuss the literature on sampling from alternatives, which informed the selection of sampling strategies to compare in the present study. The experimental design and data used in this study are then described, along with the discrete choice model method used. Our results present the effects of sampling strategies on model coefficients and measures of fit, and we conclude by discussing their implications for sampling in future crime location choice studies.

Studies of crime location choice

A growing body of criminological research has sought to model offenders' decisions[1] about where to commit crime using a discrete choice approach. Discrete choice models (McFadden 1984) are common in other domains where decision makers are choosing from a range of alternatives—consumer products (Nevo 2001), transport modes (Nguyen et al. 2017), travel destinations (Huybers 2005)—and where researchers are interested in the attributes of the alternatives, and of the choosers, that influence the outcome of the decision. Studying crime location choice using discrete choice methods enables researchers to isolate attributes of offenders and potential crime locations that are associated with a location being chosen for crime (Townsley 2016; Bernasco 2017; Ruiter 2017). Attributes of locations (such as the number of potential crime targets present), of offenders (such as their age or level of criminal expertise) or of location-offender combinations (such as the location's distance from the offender's home), are input as predictor variables and the outcome variable is categorical: which of the possible locations was chosen for crime commission (e.g., Bernasco and Nieuwbeerta 2005; Long et al. 2018; Frith 2019)? To date, over thirty studies have applied discrete spatial choice modelling (DSCM) to crime data (see Ruiter, 2017 and subsequent studies e.g., Bernasco 2019; Hanayama et al. 2018; Long et al. 2018; Song et al. 2019). Their results have provided significant insights with implications for theory and practice in terms of crime investigation, prediction, and prevention (Bernasco 2019; Curtis-Ham et al. 2020) but there is potential to expand the research agenda further

[1] We use the terms 'decision' and 'choice' to refer to location choice as revealed in behaviour. The choice may not feel like a choice to the offender and may not even take place consciously. It can also reflect a decision to visit a place for a non-criminal purpose, whereupon a crime opportunity is identified and acted on (Ruiter 2017).

(Ruiter 2017; Curtis-Ham et al. 2020). Finding ways to overcome barriers to this expansion is therefore important.

In DSCM studies of crime location choice, the number of alternative locations that could be chosen for crime can be very large, as can the number of crime choices. Further, estimating discrete choice models requires attributes of all alternatives to be linked to the attributes of every choice. The datasets of all alternatives x all choices used (without sampling) so far have ranged from 5,502 rows (262 street segments × 21 drug deals; Bernasco and Jacques 2015) to 93,919,959 rows (138,321 houses × 679 burglaries; Vandeviver and Bernasco 2020). The more typical range is between approximately 500,000 to 5 million rows, involving 1000 to 5000 alternatives and 500 to 5000 offenses (Townsley et al. 2015; e.g., Bernasco et al. 2015; Frith et al. 2017). When these datasets outstrip the capacity of the typical research computer,[2] two options to reduce the dataset exist: reduce the number of choices, or the number of alternatives for each choice. However, the former may not be desirable. As with any form of regression, a smaller sample (here, of choices) means less statistical power, reducing the ability to detect associations between the attributes of alternatives and choices. Further, the more attributes being examined, the larger the sample (the more choices) needed to ensure sufficient power. We therefore focus on the latter option: sampling the alternatives.

Computational limitations have prompted the use of sampling from alternatives in four crime location choice DSCM studies to date. For example, Vandeviver et al's. (2015) dataset included 503,589 alternative houses that could be chosen by each of 650 residential burglars. Including all alternatives for each offender would yield 327,332,850 rows, beyond the processing capability of even the specialist computer lab used by the researchers. They therefore randomly sampled one of every 8 alternative addresses for each offender, yielding a manageable 40,916,200 rows (given the lab's computing capacity). Likewise, Bernasco et al. (2013) randomly sampled 5,999 from the 24,593 alternative census blocks that could have been chosen.[3] In addition, the authors randomly sampled 6,000 street robberies from 12,938 cases,[4]resulting in 36 million rows, processed on a consumer level computer with 12 GB of RAM. Promisingly, the estimates and standard errors were very close to those produced by models using the full choice set for a smaller sample of 2,000 offenders. In a further study with the same dataset Bernasco et al. (2017) randomly sampled 7,999 and 11,999 of the census blocks for models estimated using only the offenses occurring on a given day of the week and 2-h period of the day, respectively. Simple random sampling from alternatives was also used by Bernasco (2010) to reduce a potential dataset of almost 45 m rows to 2.8 m (sampling 1,499 of 23,984 alternative postcodes for 1871 burglaries).

To increase the robustness of the estimates by reducing the influence of random error introduced by sampling, an additional bootstrapping process was used by Vandeviver et al. (2015) and Bernasco et al. (2013). In these studies, model outputs were combined across 20 and 25 sampling iterations respectively. Although bootstrapping multiple samples iteratively can solve an issue of insufficient RAM to hold the full dataset, dividing a single long

[2] A dataset might be too large to hold in RAM, or it might take days, weeks or even months to run the model, depending on the speed of the processor.

[3] Note that the chosen alternative is always included in the sample; the random sample is taken from the remaining alternatives (McFadden 1977; Ben-Akiva and Lerman 1985).

[4] The fact that both sample sizes (of alternatives and of robberies) totalled 6,000 is coincidental. There is no reason why this should be the case.

processing time into many shorter processing times may not solve a problem of insufficient CPU power (processing speed).

However, simple random sampling may not adequately capture the alternatives considered by individual decision makers that contain the most information about variables relevant to their choice. In crime location choice it is very unlikely that offenders consider every possible alternative (at house, street or neighborhood level) equally carefully when deciding where to commit crime (Ruiter 2017). They are most likely to consider places of which they have existing knowledge (Brantingham and Brantingham 1991; Ruiter 2017; Menting 2018). To the analyst, locations known to offenders through their routine activities thus contain the most 'signal' about the variables influencing offenders' crime location choices, such as how well they know the location and its crime potential (Curtis-Ham et al. 2020). But the larger the number of alternatives, and the smaller the proportion of alternatives sampled, the less likely a random sample is to include these more informative alternatives. Thus, for example, the consistency between estimates from simple random sampling of 24% of alternatives and from the full choice set achieved by Bernasco et al. (2013) might not generalize to smaller samples. We next consider other means of sampling from alternatives that could potentially provide more robust estimates if applied to conditional logit models of crime location choice through prioritizing the sampling of the most informative alternatives. We focus on conditional logit because it is by far the most used model in in DSCM studies to date, and also because the development of sampling from alternatives in other discrete choice models is not yet completely developed (Guevara and Ben-Akiva 2013a, b; Guevara et al. 2016).

Alternative methods of sampling from alternatives

An alternative to simple random sampling for DSCMs is importance sampling, where alternatives that are more likely—based on a priori beliefs—to be chosen are preferentially sampled (McFadden 1977; Ben-Akiva and Lerman 1985). Importance samples are typically stratified: alternatives most likely to be chosen are sampled at a higher rate, followed by alternatives with lower (a priori) choice probabilities, for a number of strata defined by the researchers (Li et al. 2005). Methods of importance sampling range in complexity. Most simply, one could have a single stratum sampled randomly at a higher rate than the remainder. For example, Bhat et al. (1998) defined a single 'most feasible choice set' stratum as potential travel destinations that were within the maximum distance travelled in any of the trips in the dataset (see similarly, Shiftan 1998). More complex methods have defined multiple strata using a combination of their spatial location (often with reference to journey start points) and other theoretically relevant attributes. For example, in a study of residential location choice, zones located within a central city area (more desirable) and in the same income bracket as decision makers' current home zones were preferentially sampled (Bowman and Ben-Akiva 2001; see similarly Jonnalagadda et al. 2001). Several studies have used Moran's I to identify, statistically, strata made up of zones that are both spatially proximal and similar on a relevant variable such as the number of employees, when modelling work trip destinations (Li et al. 2005; Park et al. 2013; Kim and Lee 2017).

More sophisticated methods to establish the prior probability of each alternative being chosen, to inform its sampling probability, have been proposed. These include: using fuzzy logic to identify the routes (in a route choice scenario) most likely to be considered by individuals (Hassan et al. 2019); and using the choice probabilities output by an initial random sample (Lemp and Kockelman 2012). Moreover, when compared with simple random sampling, such importance sampling methods lead to more robust results, producing smaller standard errors and higher predictive accuracy (Lemp and Kockelman 2012; Hassan et al. 2019).

Present study

We therefore propose a method for importance sampling in the crime location choice context and compare variants of this method with simple random sampling from alternatives. We also explore the impact of sample size, given previous demonstrations of the effects of sample size on model performance with both random (Nerella and Bhat 2004; von Haefen and Domanski 2013) and importance sampling (Park et al. 2013; Hassan et al. 2019). We employ a simple method for determining importance sampling strata akin to the use of distance from origin point in studies of trip destination choice (Bhat et al. 1998; Shiftan 1998). However, in crime location choice, the focus is increasingly on the relationship between multiple routine activity locations—the various locations frequented in daily life such as home, work, school, shops and family and friends' homes—and crime locations, rather than the origin and destination point of a specific journey (Ruiter 2017; Bernasco 2019; Menting et al. 2020). We also have theoretical and empirical grounds for believing that alternatives close to these routine activity 'nodes' are more likely to be chosen for crime commission than other alternatives (Brantingham and Brantingham 1991; Ruiter 2017; Bernasco 2019; Menting et al. 2020). The sampling method thus prioritizes the inclusion of alternatives closer to *any* of the routine activity nodes in the dataset as more likely to be in a given individual's choice set. To determine which sampling procedures best approximate the true estimates, we also run models using the full set of alternatives as a baseline. We separately study 5 different crime types, to account for the potential for different spatial relationships to exist for different crime types (Curtis-Ham et al. 2020) and to enable assessment of whether the results from different sampling methods generalize across crime types.

Method

Offense and offender data

The data used in this study included solved residential and non-residential burglary, commercial and personal robbery and extra-familial sex offenses occurring between 2009 and 2018 from a national dataset obtained from the New Zealand Police. For each of the offenders recorded as having committed these offenses, the location of their most recent offense was the location choice of interest. The dataset also included the locations of a range of pre-offense activity nodes. These included: past and present homes of the offender and their family members, school and other educational institutions attended, workplace, prior offenses they had committed or experienced as victims or witnesses, non-crime incidents in which they were involved, and places they were arrested, stopped or otherwise noted for intelligence purposes. Curtis-Ham et al. (2021) describe the dataset in detail; it includes approximately 4.5 million activity locations for 60,607 offenders. In this study we used random samples of 50% of the offenders within each offense type.[5]

Because these activity node locations are only recorded where needed for operational purposes during policing activities, they are not a complete or systematic set of pre-offense activity locations for each offender. However, they constitute a wider array of activity nodes than used in previous crime location studies based on administrative data (e.g., Lammers et al. 2015; Menting et al. 2016; van Sleeuwen et al. 2018). Further, Curtis-Ham et al. (2021) analyze the extent—number and geographic range-of offenders' pre-offense activity nodes in this dataset, concluding that the data hold potential for use in crime location choice research. Indeed, the results of the present study confirm that the activity nodes included in the dataset provide considerable 'signal' in explaining offenders' crime locations.

Unit of analysis

In this study we used the NZ Census Statistical Area 2 (SA2) as the set of locations from which an offense location could be chosen. SA2s roughly equate to neighborhoods and typically contain 2000–4000 residents in metropolitan areas (1000–5000 in rural areas). There were 2153 SA2s, with land areas of 0.063km^2 to 12,042.36km^2, reflecting the relative population density in urban and rural areas (median 1.962km^2).[6] SA2s are comparable to the units used in other crime location choice studies (e.g., Clare et al. 2009; Townsley et al. 2015).

Outcome variable

The outcome variable was the choice of SA2: in which of the 2153 SA2 areas of New Zealand did the offender commit the index offense? Whereas in the *theoretical model* we assume that all SA2s appear in the offender's choice set, in the *estimation* of the parameters of this model for each offense only a subset of the SA2s is used.

Predictor variables

As our focus was on testing different sampling methods, rather than testing detailed explanatory factors, we constructed a simple model using six predictors reflecting the proximity of offenders' routine activity nodes to each SA2 in their choice set, and an additional seventh predictor reflecting the level of crime opportunity in each SA2. Previous crime location choice studies have demonstrated that the odds of a neighborhood being chosen for crime are greatest when there is an activity node in the same neighborhood and lower for neighborhoods that are further from any of the offender's activity nodes (Bernasco 2019;

[5] This study forms part of a wider programme of research for which the data were divided into 'training' and 'test' samples (50% each). The training data were used for all analyses where models were trained (such as the present study). The test data were reserved for later studies testing model accuracy when applied to new data.

[6] The 2018 SA2 shapefile and metadata were downloaded from https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/. We excluded 83 SA2s which cover large bodies of water along coastlines and over lakes.

Menting et al. 2020). We therefore included dichotomous variables reflecting the presence or absence of offenders' activity nodes in increasing distance bands in relation to each SA2 in their choice set. The six distance bands were: within the same SA2, or within 0-200 m, 200-500 m, 500 m-1 km, 1-2 km or 2-5 km outside of the SA2 boundary. Following previous crime location choice research (e.g., Menting et al. 2020), the activity node variables reflected whether the *nearest* activity node to the SA2 fell into a given distance band. For example, if the closest node was in the same neighborhood, we coded 'Same SA2' as true and all other node distance variables false; if the closest node was 4 km outside the SA2 boundary, we coded 'Node within 2-5 km' as true and all other node distance variables false.

Since crime location choice is not only a product of the locations of which offenders are aware from their routine activities but of the opportunities available at those locations (Brantingham and Brantingham 1991; Menting 2018), we also included a measure of opportunity relevant to each crime type. The opportunity measures were sourced from Statistics NZ Census and Business Demography data (http://nzdotstat.stats.govt.nz/) and are comparable to opportunity measures used in other crime location choice studies (Townsley et al. 2015; Lammers 2018; Long et al. 2018; e.g., Frith 2019). For residential burglary, opportunity was the number of households in the SA2.[7] For non-residential burglary it was the number of business units in any industry.[8] For commercial robbery, it was the number of business units for industries of the types targeted in commercial robbery.[9] For personal robbery and extra-familial sex offenses it was the number of commercial or public business units, as an indicator of ambient population and thus the number of potential crime targets.[10]

Sampling strategies

We compared nine strategies for sampling from alternatives to the results from the full set of alternatives. The sampling strategies are described in Table 1 and form three groups, within which we varied the sample size. The first group involved 'distance importance sampling' (DIS) strategies where we included all SA2s within 5 km of the offender's activity nodes and added more strata at increasing distances from which alternatives were randomly sampled with decreasing probability. Similar distance-based importance strata were used in previous non-crime location choices studies (Ben-Akiva and Bowman 1998; Shiftan 1998; Li et al. 2005). In the absence of evidence as to how many strata to include over what distance in studying crime location choice, we included three additional distance strata

[7] There were large changes in residential population in many SA2s over the data period due to the Christchurch earthquakes and housing developments in response to increasing urban populations. Census 2013 data was used for offenses occurring between 2009 and 2015, and census 2018 data were used for offenses occurring between 2016 and 2018.

[8] Business demography statistics remained consistent over the data period so 2018 was used for simplicity.

[9] Industry categories included: G Retail Trade, H Accommodation and Food Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services. M Professional, Scientific and Technical Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services, M Professional, Scientific and Technical Services, N Administrative and Support Services, R Arts and Recreation Services, S Other Services. See Curtis-Ham et al. (2021) for details of how 'commercial' robberies were identified.

[10] All industries as for commercial robbery plus: I Transport, Postal and Warehousing, J Information Media and Telecommunications, O Public Administration and Safety, P Education and Training, Q Health Care and Social Assistance.

Table 1 Strategies used for sampling from alternatives

Label	Sample includes chosen SA2 plus:	Sample size equivalence[a]
DIS1	Stratum: 1: SA2s that contained or had activity nodes within 5 km of the SA2 boundary	*
	Remainder stratum: 10 randomly sampled from the remaining SA2s outside of other strata	
DIS2	As above plus:	**
	Stratum 2: 20 sampled from remaining SA2s between 5 and 10 km of any activity node.[b]	
DIS3	As above plus:	
	Stratum 3: 15 sampled from remaining SA2s between 10 and 50 km of any activity node.[b]	
DIS4	As above plus:	***
	Stratum 4: 10 sampled from remaining SA2s between 50 and 100 km of any activity node.[b]	
SIS1	Stratum 1 plus 30 sampled from remaining SA2s	**
SIS2	Stratum 1 plus 55 sampled from remaining SA2s	***
SIS3	Stratum 1 plus 100 sampled from remaining SA2s	****
SRS1	Random sample of 137 to 261 (depending on crime type)	*
SRS2	Random sample of 227 to 345 (depending on crime type)	****

[a]Asterisks indicate strategies with comparable sample sizes

[b]Identified as SA2s whose centroids were within the specified distance range from the centroid of any SA2 containing an activity node. If there were fewer than the specified number of SA2s available to sample, 100% were sampled

beyond the initial 0-5 km (strategy DIS1): 5-10 km (DIS2), 10-50 km (DIS3), 50-100 km (DIS4). These enabled us to investigate the incremental benefit (if any) of importance sampling SA2 alternatives at a series of increasing distances.

The second group, 'simple importance sampling' (SIS), included all SA2s within 5 km of the offender's activity nodes and increased the number of additional SA2s sampled from the remainder. The simple importance strategies were included to test whether any advantage of the distance importance sampling was attributable to the inclusion of SA2s in the distance strata or merely to the inclusion of additional SA2s regardless of their location.

The third group included two simple random sampling (SRS) strategies based on the smallest and largest sample sizes achieved with the previous strategies (to enable comparison of like for like in terms of sample size). For the first SRS strategy (SRS1) we randomly sampled the median number of SA2s included in the choice sets when using the first distance importance sampling strategy (DIS1). This strategy resulted in the *smallest* choice sets of all the strategies (see Table 2 below). For the second SRS strategy (SRS2) we randomly sampled the median number of SA2s included in the choice sets when using the sampling strategy that resulted in the *largest* choice sets (SIS3: all SA2s within 5 km of activity nodes plus 100 from the remainder, see Table 2).

The number of additional SA2s to sample were determined with reference to previous research suggesting that robust estimates could be achieved by randomly sampling 12.5% of the full set of alternatives (Nerella and Bhat 2004) and studies using stratified importance samples as small as 1–7% (Bowman and Ben-Akiva 2001; Jonnalagadda et al. 2001;

Table 2 Choice set size, total dataset size and run time per sampling strategy and offense

Offense	Sampling strategy & size equivalence (*)[a]		Total N in dataset	% of All alts N[b]	n alternatives per choice set			Run time (s)	
					Min	Median	Max	Mean	SD
Res. Burg.(n = 17,054)	DIS1	*	4,213,998	11.5%	11	210	1252	42.4	1.3
	DIS2	**	4,501,322	12.3%	11	229	1272	45.8	0.6
	DIS3		4,754,811	12.9%	11	244	1287	48.9	1.1
	DIS4	***	4,925,329	13.4%	20	254	1297	50.7	1.1
	SIS1	**	4,555,078	12.4%	31	230	1272	46.2	0.9
	SIS2	***	4,981,428	13.6%	56	255	1297	51.1	0.9
	SIS3	****	5,748,858	15.7%	101	300	1342	58.8	0.9
	SRS1	*	3,598,394	9.8%	210	210	210	35.0	0.8
	SRS2	****	5,133,254	14.0%	300	300	300	50.5	0.8
	All alts		36,717,262	100.0%	2153	2153	2153	649.1	386.7
Non. Res. Burg.(n = 10,353)	DIS1	*	2,417,735	10.8%	11	190	1114	25.3	1.0
	DIS2	**	2,587,458	11.6%	11	209	1134	26.1	0.8
	DIS3		2,740,966	12.3%	12	224	1149	28.2	0.8
	DIS4	***	2,844,471	12.8%	20	234	1159	29.4	0.8
	SIS1	**	2,624,795	11.8%	31	210	1134	27.0	0.6
	SIS2	***	2,883,620	12.9%	56	235	1159	29.8	0.9
	SIS3	****	3,349,505	15.0%	101	280	1204	34.1	0.7
	SRS1	*	1,977,423	8.9%	190	190	190	18.9	0.5
	SRS2	****	2,909,193	13.1%	280	280	280	28.5	0.9
	All alts		22,290,009	100.0%	2153	2153	2153	244.8	5.9

Table 2 (continued)

Com. Rob.(n = 1,977)	DIS1	*	563,161	13.2%	13	261	1028	5.7	0.6
	DIS2	**	598,905	14.1%	13	281	1048	5.8	0.2
	DIS3		628,470	14.8%	28	296	1063	6.3	0.4
	DIS4	***	648,240	15.2%	38	306	1073	6.5	0.3
	SIS1	**	602,701	14.2%	33	281	1048	5.9	0.2
	SIS2	***	652,126	15.3%	58	306	1073	6.7	0.6
	SIS3	****	741,091	17.4%	103	351	1118	7.4	0.3
	SRS1	*	517,974	12.2%	261	261	261	4.8	0.2
	SRS2	****	695,904	16.3%	351	351	351	6.8	0.3
	All alts		4,256,481	100.0%	2153	2153	2153	40.5	1.9
Pers. Rob.(n = 4,315)	DIS1	*	1,243,204	13.4%	13	255	1067	13.0	0.6
	DIS2	**	1,321,381	14.2%	13	275	1087	13.6	0.6
	DIS3		1,385,822	14.9%	13	290	1102	14.4	0.6
	DIS4	***	1,428,972	15.4%	23	300	1112	15.9	2.6
	SIS1	**	1,329,504	14.3%	33	275	1087	13.6	0.8
	SIS2	***	1,437,379	15.5%	58	300	1112	14.5	0.5
	SIS3	****	1,631,554	17.6%	103	345	1157	16.6	0.8
	SRS1	*	1,104,640	11.9%	255	255	255	10.8	0.3
	SRS2	****	1,492,990	16.1%	345	345	345	15.0	0.4
	All alts		9,290,195	100.0%	2153	2153	2153	91.4	1.3

Table 2 (continued)

Sex offense (n = 4,421)	DIS1	*	784,537	8.2%	11	137	1113	8.8	2.3
	DIS2	**	854,011	9.0%	11	156	1133	8.5	1.1
	DIS3		919,350	9.7%	14	171	1148	8.5	0.4
	DIS4	***	963,554	10.1%	24	181	1158	9.3	0.6
	SIS1	**	872,957	9.2%	31	157	1133	8.0	0.1
	SIS2	***	983,482	10.3%	56	182	1158	9.1	0.2
	SIS3	****	1,182,427	12.4%	101	227	1203	12.2	0.5
	SRS1	*	610,098	6.4%	137	137	137	5.9	0.3
	SRS2	****	1,007,988	10.6%	227	227	227	10.2	0.3
	All alt		9,518,413	100.0%	2153	2153	2153	92.2	1.7

[a] Asterisks indicate strategies with comparable sample sizes

[b] Total N of the sampled dataset as a percentage of the Total N of the unsampled dataset using all alternatives

Li et al. 2005; Kim and Lee 2017). While the size of the choice set varies across individuals in the present study, the sampling strategies aimed to generate choice sets averaging between 5 and 15% of the 2153 SA2s (about 100–320 SA2s).

Model specification and estimation

The *conditional logit model* (McFadden 1974) is a statistical model for the probability that a decision maker n, who must choose from a set of alternatives C, chooses alternative i, and can be expressed as:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j \int C} e^{\beta' x_{ni}}} \tag{1}$$

where x_{ni} is a list of attributes that vary across alternatives and may also vary across decision makers, and ß is a vector of the parameters that represent the effects of these attributes on the outcome of the decision. The ß parameters can be estimated by maximum likelihood estimation, based on the actual choices observed. From the size, direction and statistical significance of the estimated β parameters, conclusions can be drawn about the relevant criteria that decision makers use. Typically, exponentiated ß estimates (e^{β}) are reported. They are called odds ratios and represent the effect of a one-unit increase in the x_{ni} variables on the odds of an alternative to be chosen.

The log-likelihood of the conditional logit model is:

$$l = \sum_{n=1}^{N} \sum_{i \int C, i \neq j} \left(y_{ni} \ln \left(P_{ni} \right) \right) = \sum_{n=1}^{N} \sum_{i \int C, i \neq j} \left(y_{ni} \ln \left(\frac{e^{\beta' x_{ni}}}{\sum_{j \int C} e^{\beta' x_{nj}}} \right) \right) \tag{2}$$

where y_{in} is the observed choice, such that $y_{ni} = 1$ if decision maker n chooses alternative i and $y_{ni} = 0$ if another alternative is chosen.

Sampling from alternatives is an estimation technique in which we use a subset D of the full choice set C to estimate the ß parameters. McFadden (1977) proved that under the *positive conditioning property* (whereby each alternative in the choice set C has a positive probability of being included in the estimation set D), unbiased parameter estimates are consistently estimated by maximizing a modified likelihood function with an added correction term $-ln(\pi(D|i))$ in the utility function:

$$\ell = \sum_{n=1}^{N} \sum_{i \int C, i \neq j} \left(y_{ni} \ln \left(\frac{e^{\beta' x_{ni} - \ln(\pi(D|i))}}{\sum_{j \int C} e^{\beta' x_{nj} - \ln(\pi(D|j))}} \right) \right) \tag{3}$$

where $\pi(D|i)$ is the probability of alternative i to be included in the estimation sample D. This modified estimation procedure is quite general, as it applies to *any* sample that conforms to the positive conditional property. All *importance sampling* strategies that we use in the present paper conform to the positive conditioning property, and were therefore estimated with the additional offset term. In all cases, the probability of the alternative being sampled depended on the number of alternatives remaining to be sampled after including those with activity nodes within 5 km.

In the case of a *uniform conditional probability*, as in the case of *simple random sampling*, each alternative from the full choice set C has the same positive probability of being included in D. The $\pi(D|i)$ thus cancels out in Eq. (3), and the model parameters can be estimated by the regular log-likelihood Eq. (2).

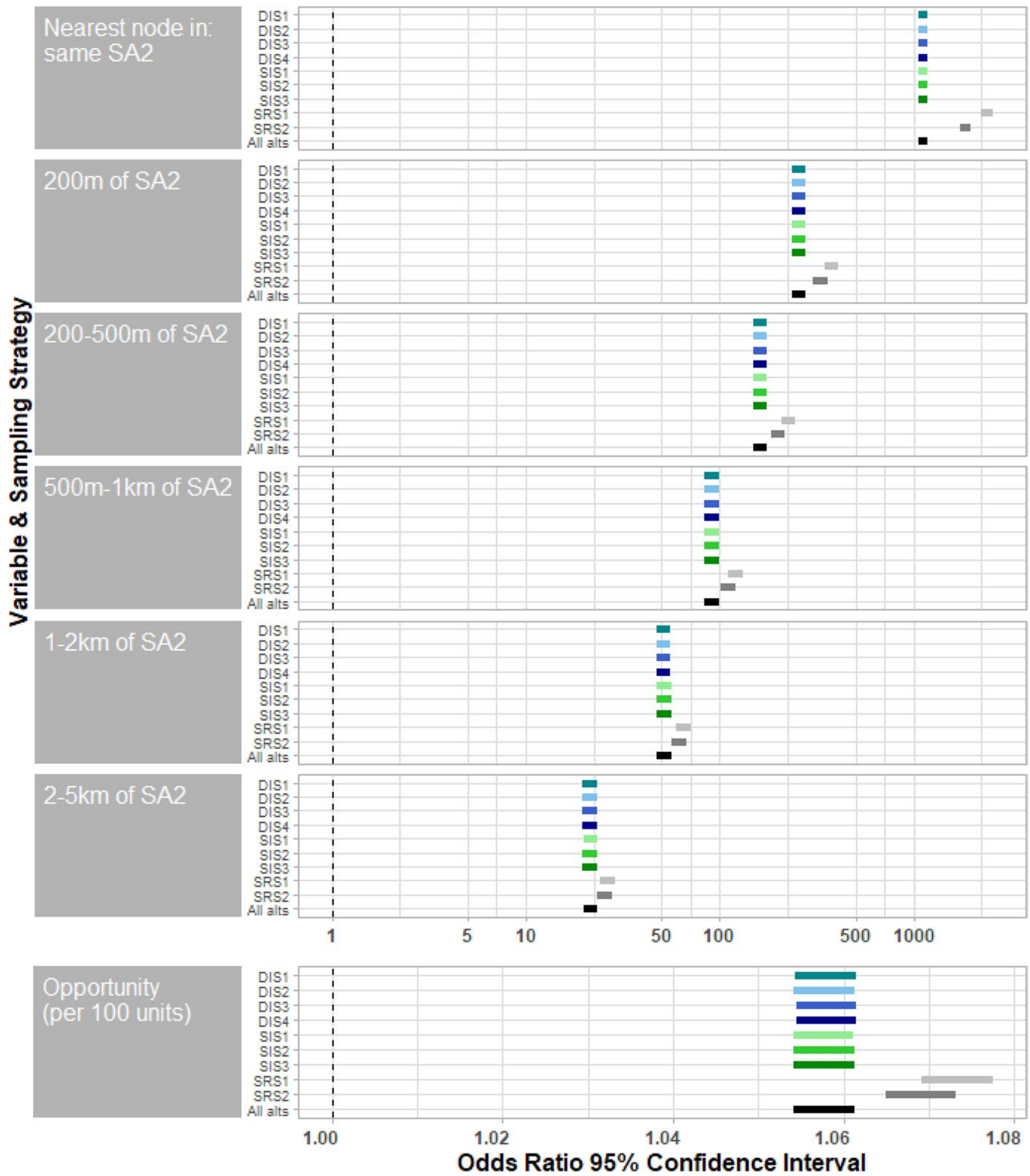


Fig. 1 Parameter estimates for residential burglary location choice by sampling strategy

All models were run on an HP Elitebook with 16gb of RAM and an Intel i7-7600U 2.8 GHz CPU using the clogit function from the survival package (Therneau 2020) in R (R Core Team 2013).

Comparison metrics

We assessed the quality of the sampling strategies with a range of criteria. First, we consider the size of the dataset produced by each strategy, the aim being to minimize the number of rows and thus the computational burden while producing robust results. We also compare the processing times for the models based on each strategy as an indicator of computational efficiency (over 10 iterations, using the microbenchmark package: Mersman 2019). The times are of course specific to the computer used to run the analyses, but provide a general indication of the relative time saving of the sampling methods by comparison to using the full choice set. Second, we examine whether the strategies produce the same results as those from baseline model including the full choice set. Following previous studies comparing alternative-sampling methods (Park et al. 2013; Hassan et al. 2019), we consider the parameter estimates (Odds Ratios and their 95% confidence intervals) and model fit (McFadden's pseudo-R^2). In discussing the results of these measures, we also consider the complexity of the strategy, with simpler strategies that produce robust results being preferred. The R script used to sample the data, run models, and analyze the results is appended as Supplementary Material.

Results

We first present the results of the sampling strategies in terms of the overall dataset sizes and the distributions of the size of the choice set per offender, before comparing model estimates and fit.

Sample sizes and sampling probabilities

Table 2 shows for each offense and sampling strategy: the number of offenders (choices), the total number of rows in the dataset, the proportion of the 'all alternatives' dataset included in the sampled dataset, the minimum, median and maximum number of alternatives included in the choice set per offender, and the model processing times (average over 10 iterations). Within the importance sampling strategies, the size of choice sets varied widely, reflecting the variability in the number of activity nodes per offender in the dataset. The importance sampling strategies resulted in as few as 11 alternatives in the choice set (0.5% of the 2153 SA2s). But on average, offenders' choice sets contained at least 150–350 alternatives, representing roughly 7–16% of the 2153 SA2s, depending on crime type and sampling strategy. All of the sampling methods considerably reduced the total size of the datasets and therefore the computational burden, with the 'random minimum' strategy producing the smallest datasets. The proportions of SA2s sampled per *stratum* are provided in Table S1 in the Supplementary Materials. Run times reflected the size of the datasets, with sampling from alternatives models running in 5–18% of the time taken to run the all-alternatives models. Further, the computational savings were greatest for the datasets with more offenses (e.g., 5–9% for residential burglary), suggesting that the bigger the offense sample (number of choices) the greater the gains from sampling from alternatives.

Parameter estimates

Figure 1 displays the parameter estimates, by sampling strategy, for the residential burglary offenses (see Figs. 2–5 in Appendices 1 to 4 for other offenses). For the activity node variables, the odds ratios (ORs) represent the increase in probability of an SA2 being chosen for crime given the presence of an activity node in the given distance band. For example, offenders were over 1000 times more likely to commit a residential burglary in an SA2 in which they also had an activity node. These odds decreased over increasing distances to the nearest activity node but remained statistically significant; the presence of an activity node up to 5 km from an SA2 was associated with an over 20-fold increase in the likelihood that

Table 3 McFadden Pseudo R^2 values by sampling strategy and crime type

Sampling strategy & size[a]		Res. Burg	Non-res. Burg	Com. Rob	Pers. Rob	Sex Offense
DIS1	*	0.386	0.395	0.316	0.346	0.374
DIS2	**	0.386	0.395	0.316	0.346	0.374
DIS3		0.386	0.395	0.316	0.346	0.374
DIS4	***	0.386	0.395	0.316	0.346	0.374
SIS1	**	0.386	0.395	0.316	0.346	0.374
SIS2	***	0.386	0.395	0.316	0.346	0.374
SIS3	****	0.386	0.395	0.316	0.346	0.374
SRS1	*	0.404	0.421	0.323	0.358	0.418
SRS2	****	0.398	0.411	0.320	0.354	0.400
All alternatives		0.386	0.395	0.316	0.346	0.374

[a]Asterisks indicate strategies with comparable sample sizes

the SA2 would be chosen. The odds of an SA2 being chosen for a residential burglary also increased by 1.06 times for every 100 households in the SA2.

When comparing the sampling strategies with the ‘all alternatives’ model that includes all 2153 SA2s in each offender’s choice set, all importance sampling strategies resulted in parameter estimates and standard errors that did not differ significantly from the full model. None of the incremental additions of the three distance strata (DIS2, DIS3, DIS4) beyond the initial 5 km (DIS1) provided additional benefits in terms of correspondence to the full model. Nor did increasing the sample size (SIS1, SIS2, SIS3) beyond the minimum achieved by including all SA2s with activity nodes within 5 km and 10 SA2s from the remainder (DIS1). However, the two simple random sampling strategies (SRS1 and SRS2) tended to produce coefficients that deviated widely from the full model coefficients (and larger standard errors). Deviation from the full model was larger for the variables reflecting the presence of activity nodes in closer proximity to the SA2, and for the smaller of the two random samples (SRS1). Conversely, the ORs for variables reflecting the presence of activity nodes farther from the SA2 were closer to those produced by the full model. The same broad pattern was found for the other offense types, though by comparison with burglary, for personal robbery and sex offenses the simple random sampling strategies (SRS1 and SRS2) produced ORs closer to those of the full model for variables reflecting activity nodes at longer distances from the SA2.[11] For commercial robbery, the confidence intervals produced by simple random sampling (SRS1 and SRS2) overlapped with those of the full model for all variables.

[11] We also compared bootstrapped versions of the single stratum importance sampling strategy (DIS1) and the smallest simple random sampling (SRS1), since the ‘strategy to beat’ to produce robust results with the smallest dataset was the single stratum importance sample, to which the smaller simple random sampling strategy was closest in sample size. The estimates and standard errors for 20 bootstrap iterations were combined using Rubin’s rule (Rubin 1987) implemented in the Amelia package in R (King et al. 2000). The bootstrapped strategies produced the same pattern as the single iterations, as shown in Fig. 6 in Appendix 5. Of note, bootstrapping the simple random sampling did not produce estimates any closer to those from the full model.

Model fit

Table 3 presents the McFadden's pseudo r-squared values per sampling strategy and the full model, for each offense type. As with the parameter estimates, importance sampling strategies (DIS1-4 and SIS1-3) led to pseudo r-squared values that matched the values from the full model using all alternatives in each choice set. The simple random strategies led to slightly higher values than the full model, with the smaller random samples (SRS1) deviating the most. That smaller random samples led to higher R^2 values is consistent with previous studies (Park et al. 2013; Hassan et al. 2019).

Discussion

This study examined the effects of different methods of sampling from alternatives on the results of discrete spatial choice models of offenders' choice of crime locations, for burglary, robbery and extra-familial sex offenses. Our results suggest that overall, importance sampling that ensures the inclusion of choice alternatives near to offenders' activity nodes can lead to coefficients and model fit on par with the results of the full model using all alternatives in each choice set, while reducing the computational burden considerably. Simple random sampling, however, tends to risk overestimating both parameter estimates and model fit. Since all importance sampling strategies produced comparable results, considering both the size of the dataset and the complexity of the strategy, the single stratum based strategy (all SA2s with activity nodes within 5 km plus 10 SA2s from the remainder, DIS1) was the optimal strategy to produce robust results with the smallest dataset, fastest run time and simplest method. That preferentially sampling choice alternatives with higher choice probability outperforms simple random sampling is consistent with previous studies in other discrete spatial choice domains (Lemp and Kockelman 2012; Hassan et al. 2019).

In the only other study to compare sampling from alternatives to the full choice set in a crime location choice context, Bernasco et al. (2013) examined street robberies in the city of Chicago. They found that the coefficients and standard errors from a model using a simple random sample of 24% of 24,593 census block alternatives for 6000 robberies were very close to those from a model using the full choice set for 2000 robberies. There are several possible explanations for simple random sampling producing robust results in that context by comparison to the present study. First, and likely foremost, the proportion of alternatives sampled was considerably larger. Second, offenders may be familiar with a higher proportion of locations (alternatives) within a city than within an entire country; thus there would be a higher chance of sampling alternatives relevant to offenders' decisions. Third, by including only offenders with residential addresses in the city, who were probably familiar with more parts of the city than outsiders, their study likely includes more offenders with greater familiarity with more alternatives than the present study. In contrast, when using a national dataset, simple random sampling may not capture enough alternatives containing or near to activity nodes, to adequately capture the 'signal' from those alternatives that fall within offenders' awareness space (Brantingham and Brantingham 1991).

A further explanation relates to the size of the units of analysis. Bernasco et al. (2013) used small spatial units—census blocks with an average of 118 residents—by comparison to thousands in our neighborhood sized SA2s. It may be that simple random sampling performs poorly when sampling neighborhoods because features relevant to offenders' location choices (e.g., activity nodes, targets) concentrate in few neighbourhoods. Thus random sampling may have a higher risk of excluding all of them. Conversely, these features may be present in more census blocks and thus the risk of excluding them all from the analysis by selecting areas randomly may be smaller.[12]

Overall, our results suggest that (a) simple random sampling does not necessarily lead to robust results and (b) ensuring that the sampling from alternatives strategy captures enough alternatives within individual offenders' awareness space is an important consideration when designing discrete crime location choice research. Particularly in countries or regions with high levels of inter-city mobility among offending populations, such as New Zealand (Curtis-Ham et al. 2021) and some European countries (Menting et al. 2020; Polišenská, 2008, as cited in Vandeviver et al. 2015; van Daele et al. 2012; van Daele and Vander Beken 2010) or parts of the USA (Bichler et al. 2012), studies of crime location choice may benefit from a wider focus. More widely focused studies employing neighborhood level units would also likely benefit from importance sampling.

The relative benefits of importance sampling also depended on crime type. Simple random sampling for commercial robbery produced results more consistent with the full model than for other offenses. A range of factors could explain this finding, likely in combination. First, commercial robbery has more specific targets so offenders may need to seek opportunities that are outside their activity space (at least as revealed by the present data). The prior choice probabilities for SA2s may thus be more evenly distributed within and outside the activity space limit (5 km) such that importance sampling and simple random sampling achieve more similar results. Second, robberies are more likely to involve co-offenders (Bright et al. 2020). With group offending, the awareness space of a single offender has less influence on crime location choice (Bernasco 2006; Lammers 2018), which would similarly lessen the difference between importance and simple random sampling. Third, commercial robbery offenders had more activity nodes on average than other offenders, meaning larger proportions of alternatives were sampled in both the importance and simple random sampling strategies. Lastly, commercial robbery had the smallest sample of offenders. The CIs are thus wider than for other offenses and wider CIs mean more potential for overlap between the random sample CIs and the full model CIs. If the relatively better performance of simple random sampling for this small sample of offenders were solely attributable to the CIs, it would be preferable to use importance sampling with future small offender samples, given it yielded results in line with the full model across the range of offender sample sizes covered by our different crime types.

The results for the different crime types suggest that importance sampling would be more important for crime types not included in this study that have more in common with burglary, personal robbery or sex offenses than commercial robbery. For example, property crimes where the targets are relatively ubiquitous, such as shoplifting, thefts of and from cars, and thefts from the person are more comparable to non-residential burglary or personal robbery and thus would likely benefit from importance sampling. However, predatory offenses targeting specific victim populations that require offenders to seek opportunities outside, or bearing less relation to their personal activity locations (e.g., sexual or other violent offenses targeting prostitutes in red light districts: Rossmo 2000) may be more akin to commercial robbery, with less need to over-sample alternatives in offenders' awareness space.

Our substantive findings as to the strong association between prior activity locations and crime location choice are also of significance. Even a simple model based on the presence (or not) of a range of activity nodes within a range of distances to a potential crime location, and the level of opportunity in that location, explained a substantial amount of

[12] We are grateful to an anonymous reviewer for contributing this explanation.

variance in crime location choice. These results are consistent with criminological theory (Brantingham and Brantingham 1991) and prior crime location choice studies that also used a range of activity nodes and found higher odds of crimes near offenders' activity nodes, declining with distance (Bernasco 2019; Menting et al. 2020). Our results are, however, novel in several respects. To our knowledge no prior crime location choice study has examined non-residential burglary or extra-familial sex offenses separately from other crimes. Our results confirm that activity space proximity is strongly related to crime location choice for these offenses.

Further, existing studies that included a comparably wide array of activity nodes have only measured their relationship to crime in general (Bernasco 2019; Menting et al. 2020), which may mask crime-type specific patterns. By disaggregating crime types, the present study revealed that while the overall trend of decreasing choice probability over increasing distance from activity nodes applies to each crime type studied, some notable variation exists. For example, commercial robbery displayed the smallest odds of crime location choice in close proximity to activity nodes. Commercial robbery tends to involve specific types of premises (e.g., convenience stores and petrol/gas stations) and offenders may need to search further afield to find targets that are not just available but suitable, considering for example their level of security, layout, and ease of escape (Taylor 2002; Altizio and York 2007). The ORs for sex offenses were closer to those of burglars than robbery offenders, with particularly high odds (~1000x) of crime in SA2s in which they had an activity node. These high odds may be partly explained by the inclusion of sex offenses that took place at the offender's home address. We note that sex offenders are a heterogenous group, with the present cohort including offenses against both adults and children, and known and stranger victims. These subgroups may have stronger or weaker associations between their home or other activity locations and crime locations, but victim information was not in the data to enable further disaggregation.

Some caveats apply to the present findings on the advantages of importance sampling from alternatives in the crime location choice context. The findings are based on a single study in one country, from one data source, requiring replication with datasets from other jurisdictions. Future crime DSCM studies where sampling of alternatives would be needed to overcome computational limits could benefit from conducting initial tests with a small subset of choices comparing activity node based importance sampling and simple random sampling to the full model, before opting for one or other sampling strategy. We also encourage further research exploring the circumstances in which importance sampling outperforms simple random sampling to guide crime location choice studies. For example, such research might systematically vary the study area size, number and size of the spatial units, crime type, types of variables (activity node and opportunity related) and data sources. The present results suggest that importance sampling may be particularly important when estimating variables that are idiosyncratic (i.e., that vary simultaneously across alternatives and across offenders, such as awareness space) or have skewed distributions, such as (again) awareness space but also opportunity variables that are highly skewed. Future research could also investigate the effect of decreasing the proportion of alternatives that are importance sampled, to establish the point at which the estimates become unreliable, by comparison to including 100% in the first (5 km) stratum as was done here.

Several limitations of the present data source also warrant acknowledgement. First, the results may only generalize to location choices of offenders who have been identified and proceeded against. If the predictor variables impact the likelihood of the offender being caught, data from solved cases may not be representative of crime location choices of all offenders (Bernasco et al. 2013; Ruiter 2017). Selection bias will exist if, for example, offending near home or prior crime locations makes it more likely that the offender is caught, or more likely that the offense is reported to the police in the first place. It was not possible with the present dataset to test for these two types of selection bias. However, prior research has found a lack of association between spatial variables and clearance rates (Bernasco et al. 2013; Lammers 2014; Chiu and Leclerc 2020) and that similar sources of bias in police data (in particular reporting rates) have less effect on analysis using larger spatial units like the neighborhoods used in this study (Buil-Gil et al. 2021).

Second, the data do not include all activity nodes of all offenders. The extent to which any offender's pre-offense activity locations are recorded depends on the extent of their prior contact with police, so many activity nodes naturally remain unknown to police. However, given that peoples' current activity nodes tend to cluster together (Golledge 1999; Schönfelder and Axhausen 2002), it is highly likely that the recorded activity locations are indicative of other, latent, activity nodes. The fact that the odds of crime location choice remained significant and large (ORs 16.1 to 21.5) even 2-5 km from activity nodes suggests that the data may indeed capture additional nodes or awareness space. It also suggests that distance bands beyond 5 km from activity nodes may explain additional variance in location choice and should thus be included in future research.

Lastly, the present results are confined to the use of conditional logit rather than other discrete choice models. But other models can be more appropriate when modelling crime location choice. For example nested logit models could better account for decisions made at tiers of spatial units such as neighborhoods and specific houses (Vandeviver and Bernasco 2020) and mixed logit models are useful for accounting for variation in preferences between different offenders (Townsley et al. 2016; Frith 2019). Further, both models also relax the assumption of independence of irrelevant alternatives (IIA) which applies to conditional logit. The IIA assumption requires that the probability of a given alternative being chosen be independent of the characteristics of other alternatives (Ben-Akiva and Lerman 1985). In spatial choice scenarios, it is likely that alternatives are not independent; the choice may be influenced by the presence, or characteristics, of nearby alternatives (Bernasco 2010). In choosing where to commit a burglary, for example, an offender may be more likely to choose a neighborhood with attractive burglary targets that is surrounded by other neighborhoods with attractive targets, than a neighborhood that has the same level of attractive targets surrounded by less attractive neighborhoods. However, there is no proof that sampling from alternatives, randomly or otherwise, produces robust estimates for these models (von Haefen and Domanski 2013). Future crime location choice research might therefore explore means of sampling for these models, following recent developments in sampling methods for them in other domains (Guevara and Ben-Akiva 2013a, b; von Haefen and Domanski 2013).

Conclusion

The findings of this paper have important implications for future crime location choice studies, and make a novel contribution to the growing literature on offenders' spatial decision making. We presented a relatively simple and effective method for importance sampling from alternatives which if adopted in future crime DSCM studies could enable the use of larger datasets (e.g., with more variables, wider study areas, or more granular spatial or spatio-temporal units) to yield greater insights into crime location choice. Our results suggest that future DSCM crime location choice studies with such large datasets should sample from alternatives (rather than sampling from offenders/offenses, which reduces statistical power), and should consider conducting initial tests to determine whether simple random or importance sampling is optimal. Further, this is the first New Zealand based study in the DSCM paradigm, and the first to specifically examine non-residential burglary and sexual offenses. In finding a strong relationship between the locations of offenders' prior criminal and non-criminal activities and their crime locations, the results support the generalizability of Crime Pattern Theory (Brantingham and Brantingham 1991, 1993) and previous DSCM studies across jurisdictions and crime types.

Appendix 1

See Fig. 2

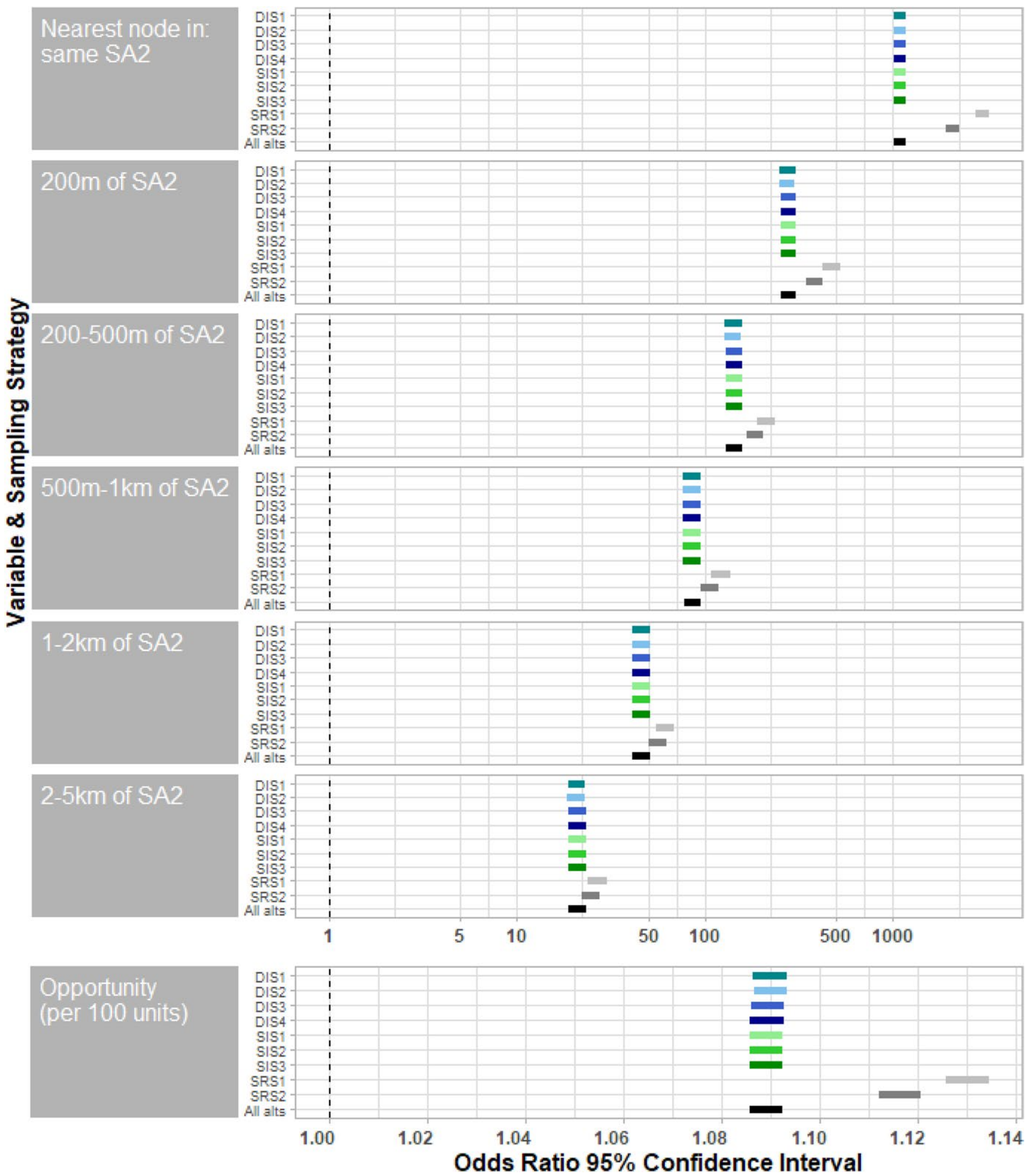


Fig. 2 Parameter estimates for non-residential burglary location choice by sampling strategy

Appendix 2

See Fig. 3

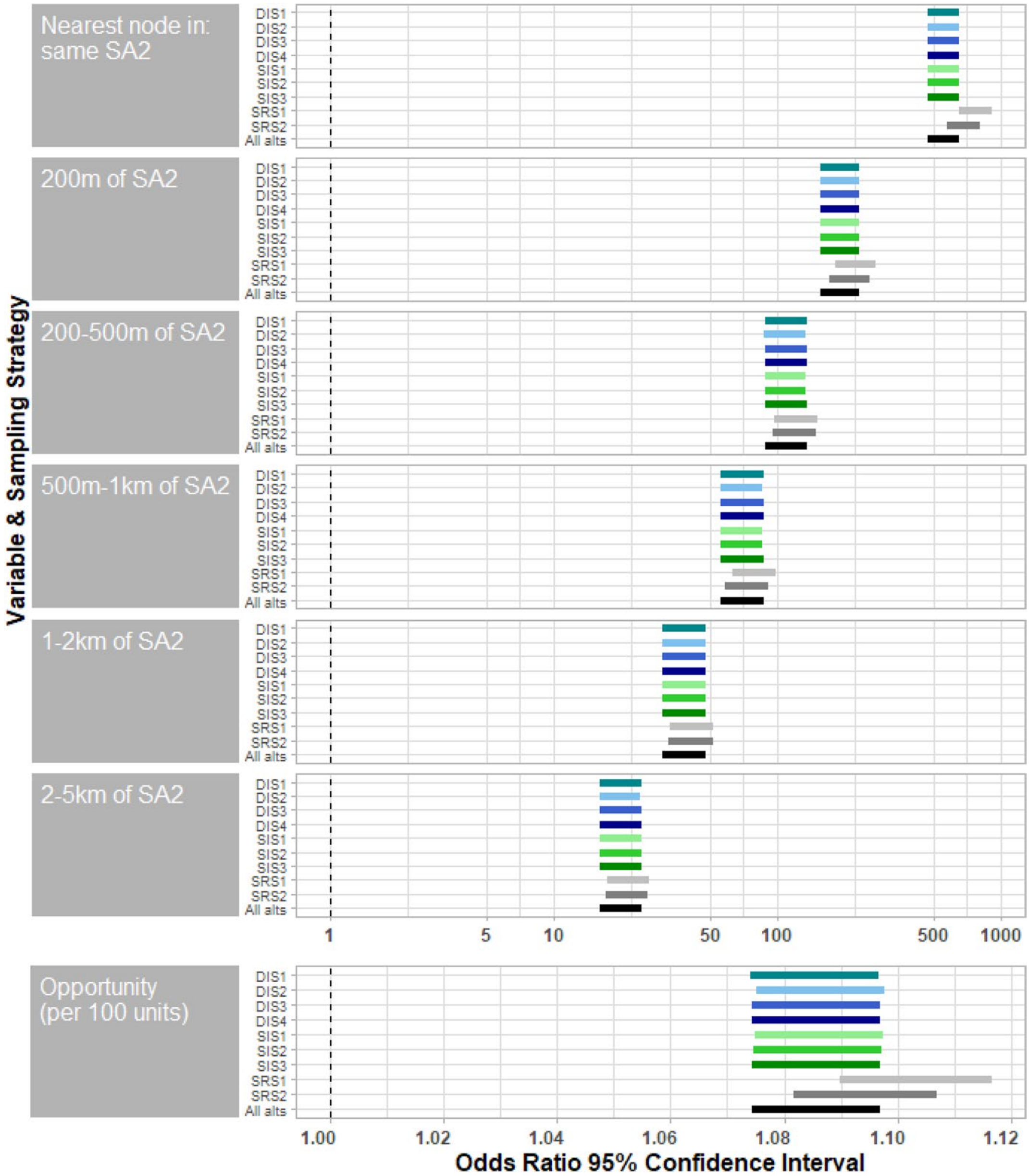


Fig. 3 Parameter estimates for commercial robbery location choice by sampling strategy

Appendix 3

See Fig. 4

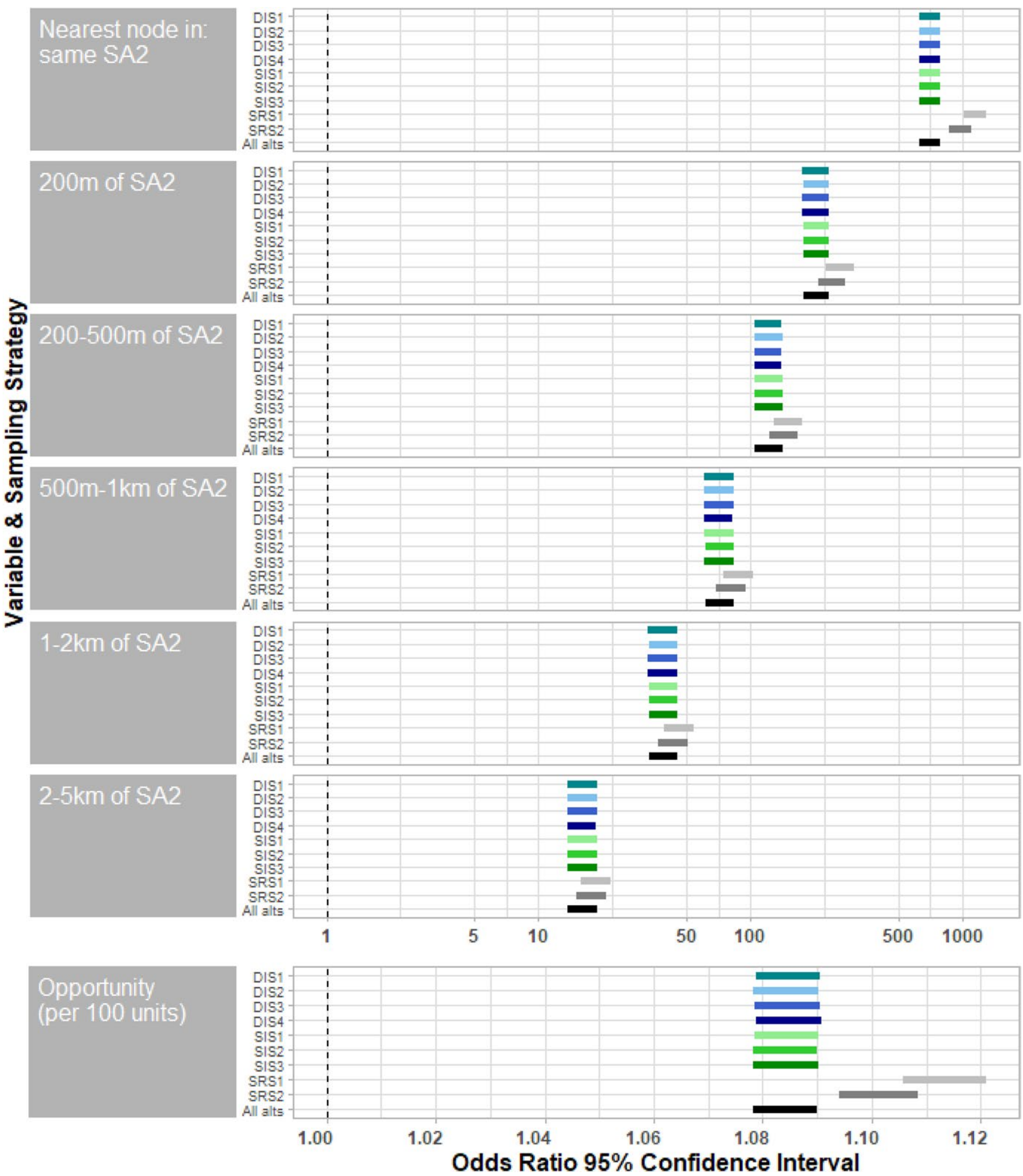


Fig. 4 Parameter estimates for personal robbery location choice by sampling strategy

Appendix 4

See Fig. 5

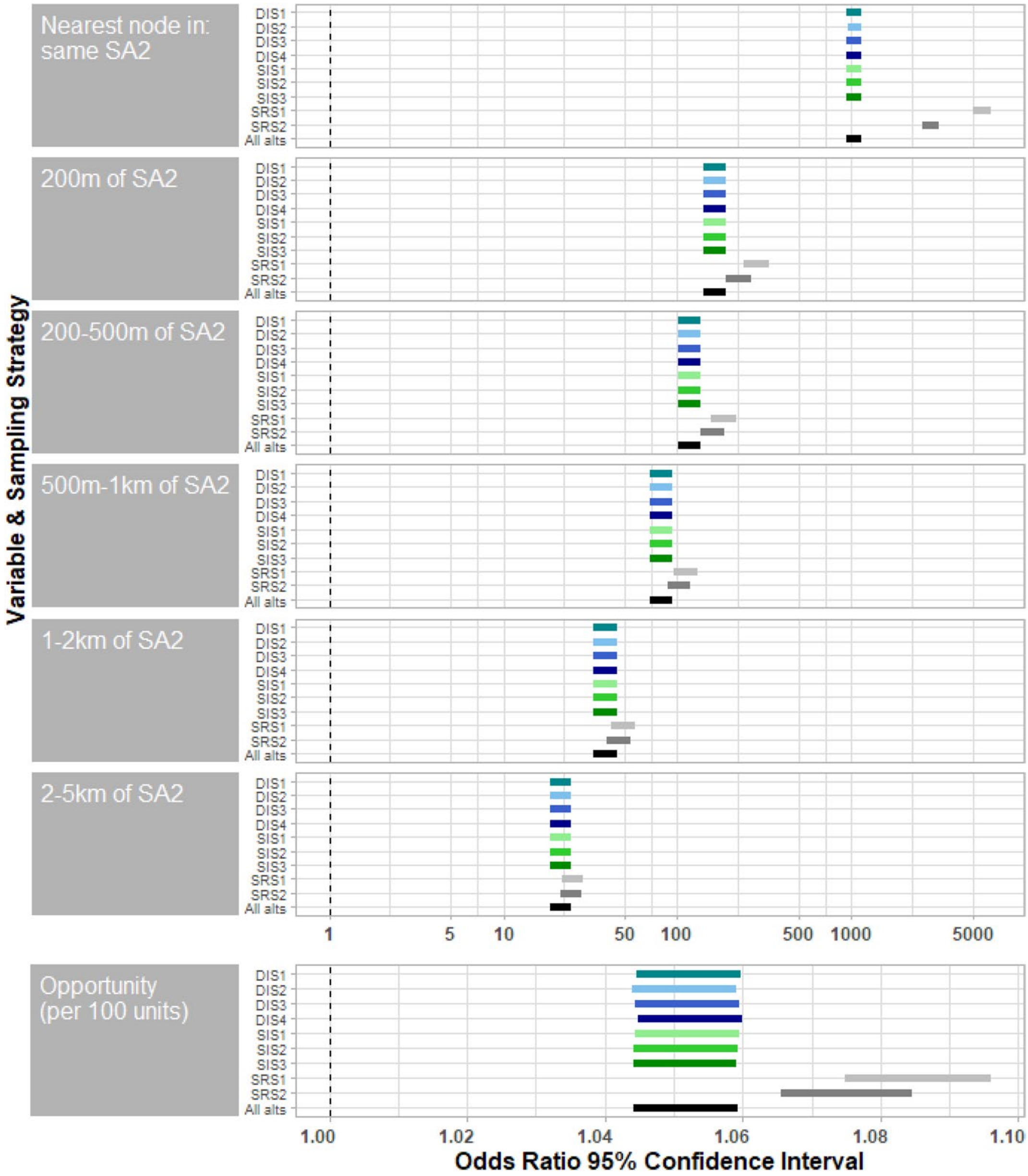


Fig. 5 Parameter estimates for extra-familial sex offense location choice by sampling strategy

Appendix 5

See Fig. 6

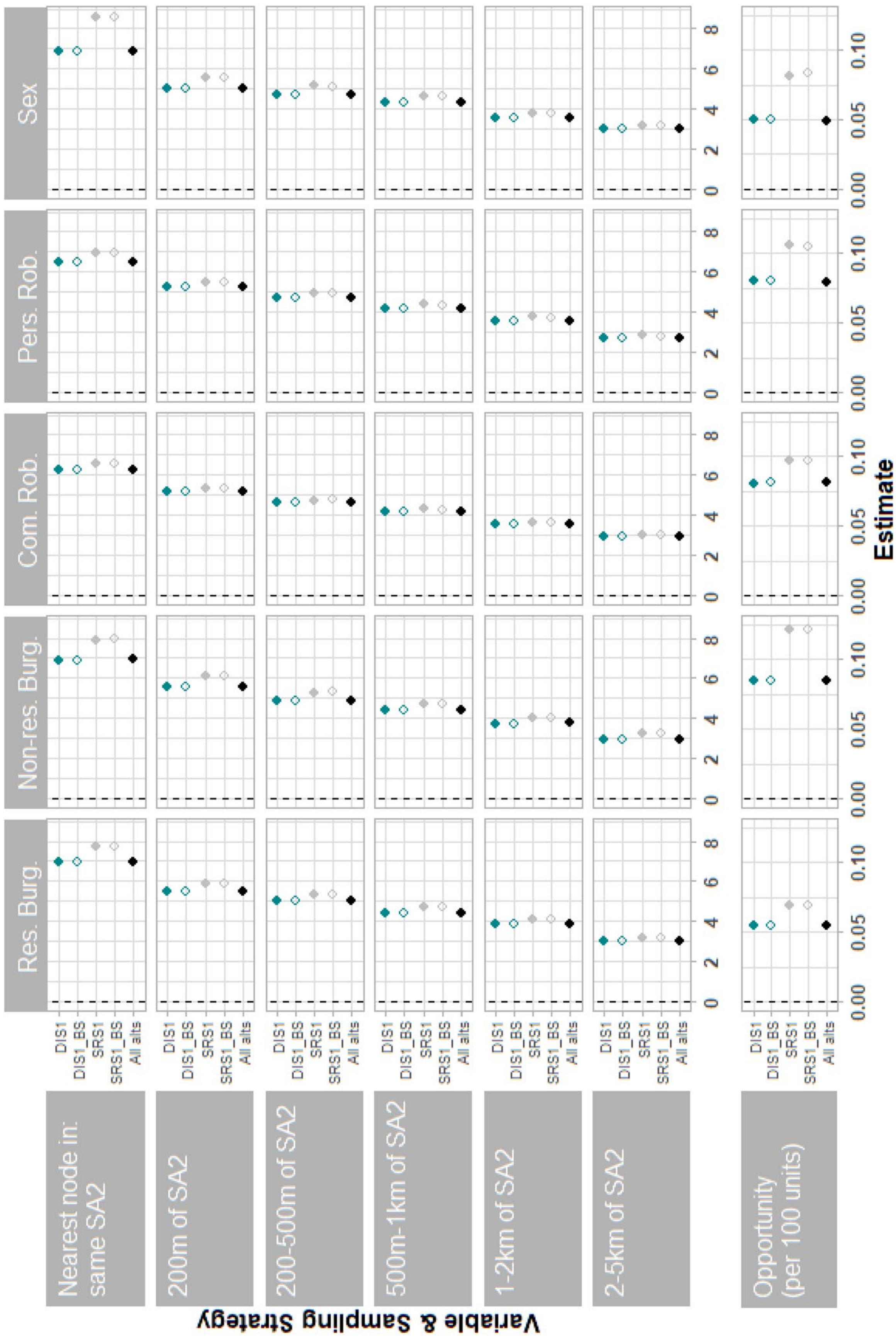


Fig. 6 Parameter estimates for bootstrapped (BS) and non-bootstrapped sampling strategies

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10940-021-09526-5.

Acknowledgements We gratefully acknowledge the assistance of the NZ Police staff who provided access to and advice on the data used in this research and who reviewed the manuscript prior to submission.

Author contributions Conceptualization: Sophie Curtis-Ham; Methodology: Sophie Curtis-Ham, Wim Bernasco; Formal analysis and investigation: Sophie Curtis-Ham; Writing—original draft preparation: Sophie Curtis-Ham; Writing—equations and accompanying text: Wim Bernasco; Writing—review and editing: Sophie Curtis-Ham, Wim Bernasco, Oleg Medvedev, Devon Polaschek; Funding acquisition: Sophie Curtis-Ham; Resources: Sophie Curtis-Ham; Supervision: Devon Polaschek, Oleg Medvedev. All authors read and approved the final manuscript.

Funding This research forms part of SCH's PhD thesis, which is funded by a University of Waikato doctoral scholarship.

Declaration

Conflicts of interest SCH is employed as a researcher at New Zealand Police. This study was not conducted as a part of that employment.

Ethics approval This research study was conducted retrospectively from data obtained for operational purposes. Ethics approval was obtained from the Psychology Research and Ethics Committee of the University of Waikato (reference #19:13). Approval of access to data for this study was obtained from the NZ Police Research Panel (reference EV-12–462). The results presented in this paper are the work of the authors and do not represent the views of New Zealand Police.

References

Altizio A, York D (2007) Robbery of convenience stores. U.S. Department of Justice, Office of Community Oriented Policing Services, Washington, DC
Ben-Akiva ME, Bowman JL (1998) Integration of an activity-based model system and a residential location model. Urban Stud 35:1131–1153. https://doi.org/10.1080/0042098984529
Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: Theory and application to travel demand. MIT Press, Cambridge, MA
Bernasco W (2006) Co-offending and the choice of target areas in burglary. J Investig Psych Offender Profil 3:139–155. https://doi.org/10.1002/jip.49
Bernasco W (2010) Modeling micro-level crime location choice: application of the discrete choice framework to crime at places. J Quant Criminol 26:113–138. https://doi.org/10.1007/s10940-009-9086-6
Bernasco W (2017) Modeling offender decision making with secondary data. In: Bernasco W, Van Gelder J-L, Elffers H (eds) The Oxford handbook on offender decision making. Oxford University Press, Oxford, England, pp 569–586
Bernasco W (2019) Adolescent offenders' current whereabouts predict locations of their future crimes. PLoS ONE 14:e0210733. https://doi.org/10.1371/journal.pone.0210733
Bernasco W, Jacques S (2015) Where do dealers solicit customers and sell them drugs? a micro-level multiple method study. J Contemp Crim Justice 31:376–408. https://doi.org/10.1177/1043986215608535
Bernasco W, Nieuwbeerta P (2005) How do residential burglars select target areas? a new approach to the analysis of criminal location choice. Br J Criminol 45:296–315. https://doi.org/10.1093/bjc/azh070
Bernasco W, Block R, Ruiter S (2013) Go where the money is: modeling street robbers' location choices. J Econ Geogr 13:119–143. https://doi.org/10.1093/jeg/lbs005
Bernasco W, Johnson SD, Ruiter S (2015) Learning where to offend: effects of past on future burglary locations. Appl Geogr 60:120–129. https://doi.org/10.1016/j.apgeog.2015.03.014
Bernasco W, Ruiter S, Block R (2017) Do street robbery location choices vary over time of day or day of week? a test in Chicago. J Res Crime Delinq 54:244–275. https://doi.org/10.1177/0022427816680681
Bhat C, Govindarajan A, Pulugurta V (1998) Disaggregate attraction-end choice modeling formulation and empirical analysis. Transp Res Rec 1645:60–68. https://doi.org/10.3141/1645-08

Bichler G, Malm A, Christie-Merrall J (2012) Urban backcloth and regional mobility patterns as indicators of juvenile crime. In: Andresen MA, Kinney JB (eds) Patterns, prevention, and geometry of crime. Routledge, London, England, pp 118–136
Bowman JL, Ben-Akiva ME (2001) Activity-based disaggregate travel demand model system with activity schedules. Transp Res Part A 35:1–28. https://doi.org/10.1016/S0965-8564(99)00043-9
Brantingham PL, Brantingham PJ (1991) Notes on the geometry of crime. In: Brantingham PJ, Brantingham PL (eds) Environmental criminology, 2nd edn. Waveland Press, Prospect Heights, IL, pp 27–54
Brantingham PL, Brantingham PJ (1993) Environment, routine, and situation: Toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction Publishers, Piscataway, NJ, pp 259–294
Bright D, Whelan C, Morselli C (2020) Understanding the structure and composition of co-offending networks in Australia. Australian Institute of Criminology Australia
Buil-Gil D, Moretti A, Langton SH (2021) The accuracy of crime statistics: assessing the impact of police data bias on geographic crime analysis. J Exp Criminol. https://doi.org/10.1007/s11292-021-09457-y
Chiu Y-N, Leclerc B (2020) Predictors and Contexts of Unsolved and Solved Sexual Offenses. Crime Delinq 66:1268–1295. https://doi.org/10.1177/0011128719879027
Clare J, Fernandez J, Morgan F (2009) Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. Aust N Z J Criminol 42:139–158. https://doi.org/10.1375/acri.42.2.139
Curtis-Ham S, Bernasco W, Medvedev ON, Polaschek DLL (2020) A framework for estimating crime location choice based on awareness space. Crime Sci 9:1–14. https://doi.org/10.1186/s40163-020-00132-7
Curtis-Ham S, Bernasco W, Medvedev ON, Polaschek DLL (2021) A national examination of the spatial extent and similarity of offenders' activity spaces using police data. ISPRS Int J Geo-Inf 10(2):47. https://doi.org/10.3390/ijgi10020047
Duncombe W, Robbins M, Wolf DA (2001) Retire to where? a discrete choice model of residential location. Int J Popul Geogr 7:281–293. https://doi.org/10.1002/ijpg.227
Frejinger E, Bierlaire M, Ben-Akiva M (2009) Sampling of alternatives for route choice modeling. Transportation Research Part B: Methodological 43:984–994. https://doi.org/10.1016/j.trb.2009.03.001
Frith MJ (2019) Modelling taste heterogeneity regarding offence location choices. J Choice Modell 33:100187. https://doi.org/10.1016/j.jocm.2019.100187
Frith MJ, Johnson SD, Fry HM (2017) Role of the street network in burglars' spatial decision-making. Criminology 55:344–376. https://doi.org/10.1111/1745-9125.12133
Golledge R (1999) Human wayfinding and cognitive maps. In: Golledge R (ed) Wayfinding behavior: Cognitive mapping and other spatial processes. Johns Hopkins University Press, Baltimore, MD, pp 5–45
Guevara CA, Ben-Akiva ME (2013a) Sampling of alternatives in multivariate extreme value (MEV) models. Transportation Research Part B: Methodological 48:31–52. https://doi.org/10.1016/j.trb.2012.11.001
Guevara CA, Ben-Akiva ME (2013b) Sampling of alternatives in logit mixture models. Transportation Research Part B: Methodological 58:185–198. https://doi.org/10.1016/j.trb.2013.08.011
Guevara CA, Chorus CG, Ben-Akiva ME (2016) Sampling of alternatives in random regret minimization models. Transp Sci 50:306–321. https://doi.org/10.1287/trsc.2014.0573
Hanayama A, Haginoya S, Kuraishi H, Kobayashi M (2018) The usefulness of past crime data as an attractiveness index for residential burglars. J Investigative Psychology and Offender Profiling 15:257–270. https://doi.org/10.1002/jip.1507
Hassan MN, Rashidi TH, Nassir N (2019) Consideration of different travel strategies and choice set sizes in transit path choice modelling. Transportation (dordrecht). https://doi.org/10.1007/s11116-019-10075-x
Huybers T (2005) Destination choice modelling: what's in a name? Tour Econ 11:329–350. https://doi.org/10.5367/000000005774352999
Jonnalagadda N, Freedman J, Davidson WA, Hunt JD (2001) Development of microsimulation activity-based model for San Francisco: destination and mode choice models. Transp Res Rec 1777:25–35. https://doi.org/10.3141/1777-03
Kim J, Lee S (2017) Comparative analysis of traveler destination choice models by method of sampling alternatives. Transp Plan Technol 40:465–478. https://doi.org/10.1080/03081060.2017.1300242

King G, Honaker J, Joseph A, Scheve K (2000) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. American Political Science Review 95:49–69
Lammers M (2014) Are arrested and non-arrested serial offenders different? a test of spatial offending patterns using DNA found at crime scenes. J Res Crime Delinq 51:143–167. https://doi.org/10.1177/0022427813504097
Lammers M (2018) Co-offenders' crime location choice: do co-offending groups commit crimes in their shared awareness space? Br J Criminol 58:1193–1211. https://doi.org/10.1093/bjc/azx069
Lammers M, Menting B, Ruiter S, Bernasco W (2015) Biting once, twice: the influence of prior on subsequent crime location choice. Criminology 53:309–329. https://doi.org/10.1111/1745-9125.12071
Lemp JD, Kockelman KM (2012) Strategic sampling for large choice sets in estimation and application. Transp Res Part A 46:602–613. https://doi.org/10.1016/j.tra.2011.11.004
Li M-T, Chow L-F, Zhao F, Li S-C (2005) Geographically stratified importance sampling for the calibration of aggregated destination choice models for trip distribution. Transp Res Rec 1935:85–92. https://doi.org/10.3141/1935-10
Long D, Liu L, Feng J, Zhou S (2018) Assessing the influence of prior on subsequent street robbery location choices: A case study in ZG city. China Sustain 10:1818. https://doi.org/10.3390/su10061818
McFadden D (1977) Modelling the choice of residential location. Yale University, Cowles Foundation for Research in Economics
McFadden D (1984) Econometric analysis of qualitative response models. In: Griliches P, Intriligator MD (eds) Handbook of econometrics. Elsevier, Amsterdam, The Netherlands, pp 105–142
Menting B (2018) Awareness x opportunity: testing interactions between activity nodes and criminal opportunity in predicting crime location choice. Br J Criminol 58:1171–1192. https://doi.org/10.1093/bjc/azx049
Menting B, Lammers M, Ruiter S, Bernasco W (2016) Family matters: effects of family members' residential areas on crime location choice. Criminology 54:413–433. https://doi.org/10.1111/1745-9125.12109
Menting B, Lammers M, Ruiter S, Bernasco W (2020) The influence of activity space and visiting frequency on crime location choice: findings from an online self-report survey. Br J Criminol 60:303–322. https://doi.org/10.1093/bjc/azz044
Mersman O (2019) microbenchmark: Accurate timing functions. Version 1.4–7URL https://CRAN.R-project.org/package=microbenchmark
Nerella S, Bhat CR (2004) Numerical analysis of effect of sampling of alternatives in discrete choice models. Transp Res Rec 1894:11–19. https://doi.org/10.3141/1894-02
Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. Econometrica 69:307–342
Nguyen HTA, Chikaraishi M, Fujiwara A, Zhang J (2017) Mediation effects of income on travel mode choice: Analysis of short-distance trips based on path analysis with multiple discrete outcomes. Transp Res Rec 2664:23–30. https://doi.org/10.3141/2664-03
Park H, Park D, Kim C et al (2013) A comparative study on sampling strategies for truck destination choice model: case of Seoul metropolitan area. Can J Civ Eng 40:19–26. https://doi.org/10.1139/cjce-2012-0433
R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton, FL
Rubin D (1987) Multiple Imputation for Nonresponse in Surveys, 1st edn. Wiley, NY
Ruiter S (2017) Crime location choice. In: Bernasco W, Van Gelder J-L, Elffers H (eds) The Oxford handbook of offender decision making. Oxford University Press, Oxford, pp 398–420
Schönfelder S, Axhausen KW (2002) Measuring the size and structure of human activity spaces: The longitudinal perspective. ETH, Zurich
Shiftan Y (1998) Practical approach to model trip chaining. Transp Res Rec 1645:17–23. https://doi.org/10.3141/1645-03
Song G, Bernasco W, Liu L et al (2019) Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. J Quant Criminol. https://doi.org/10.1007/s10940-019-09406-z
Taylor N (2002) Robbery against service stations and pharmacies: recent trends. Australian Institute of Criminology, Canberra, Australia
Therneau T (2020) A Package for Survival Analysis in R. Version 3.1–12URL https://CRAN.R-project.org/package=survival
Townsley M (2016) Offender mobility. In: Wortley R, Townsley M (eds) Environmental criminology and crime analysis. Routledge, London, England, pp 142–161

Townsley M, Birks D, Bernasco W et al (2015) Burglar target selection: a cross-national comparison. J Res Crime Delinq 52:3–31. https://doi.org/10.1177/0022427814541447
Townsley M, Birks D, Ruiter S et al (2016) Target selection models with preference variation between offenders. J Quant Criminol 32:283–304. https://doi.org/10.1007/s10940-015-9264-7
van Daele S, Vander Beken T (2010) Journey to crime of "itinerant crime groups." Policing Int J. 33:339–353. https://doi.org/10.1108/13639511011044920
van Daele S, Vander Beken T, Bruinsma GJN (2012) Does the mobility of foreign offenders fit the general pattern of mobility? Eur J Criminol 9:290–308. https://doi.org/10.1177/1477370812440065
van Sleeuwen SEM, Ruiter S, Menting B (2018) A time for a crime: temporal aspects of repeat offenders' crime location choices. J Res Crime Delinq 55:538–568. https://doi.org/10.1177/0022427818766395
Vandeviver C, Bernasco W (2020) "Location, location, location": effects of neighborhood and house attributes on burglars' target selection. J Quant Criminol 36:779–821. https://doi.org/10.1007/s10940-019-09431-y
Vandeviver C, Neutens T, van Daele S et al (2015) A discrete spatial choice model of burglary target selection at the house-level. Appl Geogr 64:24–34. https://doi.org/10.1016/j.apgeog.2015.08.004
von Haefen RH, Domanski A (2013) Estimating mixed logit models with large choice sets. In: International Choice Modelling Conference. Sydney

Authors and Affiliations

Sophie Curtis-Ham[1] · Wim Bernasco[2,3] · Oleg N. Medvedev[1] · Devon L. L. Polaschek[1]

[1] Te Puna Haumaru NZ Institute of Security and Crime Science & Te Kura Whatu Oho Mauri School of Psychology, Te Whare Wānanga o Waikato University of Waikato, Hamilton 3240, New Zealand

[2] Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), 1081 HV Amsterdam, The Netherlands

[3] Department of Spatial Economics, School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

CHAPTER 5

Different Types of Activity Location and Crime Location Choice

The research presented in this chapter used the data explored in Chapter 3 and the method developed in Chapter 4, to test hypotheses derived from the theoretical model described Chapter 2, about the links between offenders' mental maps and their crime locations. In the theoretical model, offenders' likelihood of committing crime near a given activity location in their mental map depends on attributes of that location, such as how frequently they visit and how similar their activities there are to the future crime. But we cannot always measure these attributes directly, so one approach to testing the theoretical model is to compare broad types of activity location that would typically differ on one attribute, but not others. The following manuscript, currently under review, exemplifies this approach. It examines, for example, whether people are more likely to commit crime near their home addresses versus the homes of family members—which would typically be visited less frequently, and near locations where they previously offended versus locations where they had experienced crime as a victim or witness—which would typically bear less similarity to a future crime.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Relationships between offenders' crime locations and different prior activity locations as recorded in police data.

Relationships Between Offenders' Crime Locations and Different Prior Activity Locations as Recorded in Police Data

Abstract

Understanding the relationships between individual offenders' crime locations and their prior activity locations is important to enable individual level predictions to support crime prevention and investigation strategies. This study examined a wider range of crimes and activity locations than included in previous studies, to determine whether offenders are more likely to commit crime near some types of activity locations than others. Using discrete spatial choice models, we identified relationships between proximity to pre-crime activity locations recorded in a police database (e.g., offenders' homes, family members' homes, schools, prior crimes, and other police interactions) and the locations of 17,054 residential burglaries, 10,353 non-residential burglaries, 1,977 commercial robberies, 4,315 personal robberies and 4,421 extra-familial sex offences, in New Zealand. Offenders were generally more likely to commit crime closer to their activity locations than farther away, and closer to those visited more frequently (e.g., home versus family homes) or more likely to impart relevant knowledge about crime opportunities (e.g., prior crimes versus prior victim or witness locations). The patterns for different activity locations and crime types broadly support a recently proposed extension to crime pattern theory and highlight the benefits of differentiating activity location and crime types.

Keywords

crime location choice, crime pattern theory, discrete spatial choice, police data, routine activity locations

Introduction

We know from routine activity theory (Cohen & Felson, 1979) and crime pattern theory (Brantingham & Brantingham, 1991, 1993a), that crimes occur where opportunity (i.e., the presence of a suitable and available target) overlaps with offenders' awareness spaces: the locations known to offenders through their routine non-criminal activities such as where they live, work, or socialise with family or friends. Further, recent theoretical development suggests that some types of activity locations are more salient to offenders' crime location choices than others (Curtis-Ham et al., 2020). But little research has explored empirically the extent to which various types of activity locations differ in their effects on crime location choices, despite the importance to policing of being able to predict—at an individual level—where a person will commit crime, given knowledge of the person's different activity locations. Studies to date have only compared a limited subset of activity locations (e.g., offenders' homes, homes of family members and locations of previous offences; Ruiter 2017). Leveraging a large national dataset of offenders' pre-crime activity locations recorded in a police database, the present study compares the associations between a wider range of activity locations and offenders' crime location choices than included in prior studies. Secondarily, it adds to the literature by examining a wider range of crime types separately, in a previously unresearched context (New Zealand). In combination, the data allow us to test hypotheses derived from the theoretical model proposed by Curtis-Ham et al. (2020) and prior theoretical and empirical work, as described next.

Activity Locations and Distance Decay

In principle, offenders could identify crime opportunities near any of their activity locations, or 'nodes' (Brantingham & Brantingham, 1991, 1993a). Qualitative studies highlight the importance of home, work and other non-criminal activity locations, as having potential to generate awareness of crime opportunities (e.g., Alston, 1994; Costello & Wiles, 2001; Davies & Dale, 1996; van Daele, 2009). Importantly for prediction purposes, recent studies have begun to quantify how much more likely people are to commit crime near to their activity locations than farther away. Some studies have considered a small subset of activity locations, showing that offenders tend to commit crime near their homes, the homes of close relatives (parents, children and siblings), and the locations of previous offences (Menting et al., 2016; Ruiter, 2017; van Sleeuwen et al., 2018). This tendency also applies to former, rather than current, homes of offenders and of their relatives (Bernasco, 2010b; Bernasco & Kooistra, 2010; Menting et al., 2016).

Two studies have included a wider range of activity locations and demonstrated that offenders are more likely to commit crime near any activity location than farther away. Bernasco (2019) included all locations frequented by 70 Dutch adolescents during a four-day period up to four years before their offences; Menting et al. (2020) included friends' and partners' homes, victimisation locations, school, work and leisure locations visited by 78 Dutch offenders (aged 18-26) in the month before their offences. However, the small sample sizes in both studies precluded comparisons between types of activity locations.

Common to the above studies—and many past analyses of the frequency of offending at different distances to offenders' current homes (see Ackerman & Rossmo, 2015; Townsley, 2016)—is a 'distance decay' pattern. The probability of crime is highest close to the activity locations under study and declines with distance. However, how steeply the odds of crime decay with distance from these different nodes when examined separately, remains an open empirical question.

In light of these considerations, we hypothesised that each type of activity location in the present data would be positively associated with crime location choice, and a distance decay pattern would appear with respect to each activity location type.[1] However, we set no a priori hypotheses regarding differences in decay rates between different activity nodes. We considered this aspect to be exploratory analysis.

H1 Offenders are more likely to commit crime in closer proximity to each activity location than farther away.

Relative Influence of Different Activity Locations

To guide expectations about the relative influence of different activity locations, we drew on the theoretical framework of Curtis-Ham et al. (2020). Systematising existing theory and empirical evidence, the framework proposes that offenders' crime location choices are a product of both the *reliability* and the *relevance* of their knowledge of the areas around their

---

[1] Buffer zones of reduced crime likelihood *immediately* surrounding home or other activity nodes (Rossmo 2000) would not have been detectable given the spatial unit of analysis used in this study, as with other neighbourhood level studies.

activity nodes. People are more likely to commit crime near nodes that have produced reliable knowledge (i.e., greater familiarity) that is relevant to the crime (i.e., suggestive of a good crime opportunity). Specific attributes of offenders' activities at these locations affect how much reliable and crime-relevant knowledge they obtain. The frequency, recency and duration of offenders' activities at their activity nodes determines how reliable is their knowledge of the area. The similarity of their activities to the crime at hand determines the extent to which they have obtained relevant knowledge of opportunities for that crime. The more similar the activities in terms of the behaviour, timing or type of location involved, the more conducive they are to identifying the crime opportunities in the vicinity.

Our hypotheses comparing different activity nodes' associations with crime location choice reflect their differences on single factors set out in the theoretical framework where other factors can be assumed constant. The second hypothesis compares types of node that vary in terms of how frequently they are visited: offenders' own homes and homes of their family members. These nodes are likely comparable in recency (i.e., whether they are a current or former address) and duration (i.e., how long the offender or family member lived there). They are also comparable in terms of type of location (residential), timing (one might visit or stay with family at any time of day or day of week) and general behaviour ("hanging out", sleeping). Given that—on average—offenders will visit their own home more frequently than their family members' homes, we hypothesised:

> H2 Offenders are more likely to commit crime near their own homes than the homes of family members.

We also compared between homes of different categories of family members. Given that offenders' associations with intimate partners and their addresses are likely to have been of shorter duration (on average) than for immediate family, and given the likelihood that immediate family members are visited more frequently than other relatives, we hypothesised:

> H3 Offenders are more likely to commit crime near to the homes of immediate family members (parents, children and siblings), than those of (a) current or former intimate

partners (ex or current spouses and other intimate relationships) and (b) other relatives (grandparents, grandchildren and other relatives).[2]

Curtis-Ham et al. (2020) suggest that one dimension of behavioural similarity would be exposure to criminal behaviour in different roles. *Committing* a crime is a similar activity to committing a future crime and likely to provide relevant knowledge of the location's crime opportunities. Experiencing crime as a *victim or witness* is less similar, and so would provide less relevant knowledge of the location's crime opportunities. We therefore tested:

H4 Offenders are more likely to commit crime near to places they have previously offended than to places where they have previously been a crime victim or witness.

Activity Locations and Different Crime Types

Although the same mechanisms are theorised for all crime types, because different types of activity node generate knowledge relevant to different types of crime (Curtis-Ham et al., 2020), we also considered the interaction of node type and crime type. The theoretical framework includes a 'location type similarity' factor that reflects that activities that involve the same type of location (e.g., residential or commercial) as targeted by a certain crime are more likely to generate knowledge of nearby targets for that crime. Thus, residential nodes such as offenders' homes or their family's homes would yield relevant knowledge of nearby residential burglary targets and less relevant knowledge with respect to commercial targets. In line with this suggestion, home-crime distances tend to be longer for commercially than residentially focused crimes (Ackerman & Rossmo, 2015; Townsley, 2016). Accordingly, we tested:

H5 Residential burglars are more likely to commit crime near to home and family homes than non-residential burglars and commercial robbers.

H5 excludes personal robberies and extra-familial sex offences, which involve moving targets—people—rather than specific types of premises from which to gauge location

[2] This hypothesis assumes that the shorter duration of intimate relationships offsets the likely higher frequency of visiting intimate partners compared with other family members.

type similarity. Although personal robberies tend to concentrate in places with high numbers of potential victims (e.g., commercial, nightlife and transit hubs), many occur in residential areas (Bernasco & Block, 2009; Block & Davis, 1996; Hart & Miethe, 2015). Similarly, victims of extrafamilial sex offences could be identified and first contacted in a target rich environment such as a bar or club, or through the offenders' social network, or in their residential neighbourhood, but the offence itself could occur away from these initial 'encounter' locations, even at the offender's home (Chopin & Caneppele, 2018, 2019; A. Hewitt et al., 2012; A. N. Hewitt et al., 2020).

Opportunity and Crime Location Choice

Crime occurs where offenders' awareness space—the locations they're aware of around their activity nodes—converges with crime opportunities (Brantingham & Brantingham, 1991). Opportunities exist where suitable targets (for a given crime) are present in the absence of capable guardians (Cohen & Felson, 1979). Crime location choice is typically modelled as the product of activity node proximity while controlling for the locations of crime opportunities using covariates such as the number of potential targets, or presence of crime generators (Ruiter, 2017). We therefore expected hypotheses 1 to 5 to hold while controlling for the presence of opportunities.[3]

Additionally, including opportunity variables in models of crime location choice explains variance beyond that accounted for by proximity to activity nodes alone (Bernasco, 2019). DSCM studies have frequently found that opportunity measures are associated with crime location choice while controlling for proximity to activity nodes (Ruiter, 2017). This finding could reflect unmeasured activity locations that concentrate where crime opportunities—and people in general—concentrate (Boivin & Felson, 2018; Frank et al., 2011) or crimes committed outside of awareness space where opportunities that have been either stumbled upon or sought out (Curtis-Ham et al., 2020). We therefore expected

---

[3] Note that crime locations are a product of an *interaction* between offenders' awareness space and opportunities (Menting, 2018), an interaction that is reflected in the statistical model we use to test the hypotheses.

opportunity to be associated with crime location choice while controlling for the presence of activity nodes.

Data and Method

Offender and Activity Node Data

Data on offences and offenders' activity nodes were extracted from the New Zealand Police National Intelligence Application (NIA). Offences included all residential and non-residential burglaries, commercial and personal robberies and extra-familial sex offences committed between 2009 and 2018 for which an offender had been identified with sufficient evidence to proceed against. Including these offences balanced considerations of scope; volume; seriousness and priority to police; and inclusion of both crimes covered in previous DSCM studies and other crimes considered by research on offender mobility: non-residential burglary and extrafamilial sex offences (Chopin & Caneppele, 2018, 2019).[4] Including crimes with different target types—residential versus commercial and buildings versus people—and motivations—property acquisition versus sexual violence—enabled us to verify whether the crime-independent hypotheses (H1-H4) were robust across crime types; important given the theories from which these hypotheses derived expressly apply for all crime. Using a national dataset overcame limitations of previous studies that have typically excluded offenders with no activity locations in the study area, likely inflating associations between activity locations and crime locations (the exceptions are limited to small European countries: Bernasco & Kooistra, 2010; Menting et al., 2020; van Sleeuwen et al., 2021).

Activity node data derived from a range of records including address history, linked persons and their address histories, education, offences, incidents/call-outs, arrests, stops and intelligence. These were grouped into ten node types as described below. Not all activity

---

[4] Sex offences ranged from indecent exposure to rape. Some would have occurred at offenders' homes (regardless of how and where the victim was identified) reflecting the ability of offenders to make this choice. The dataset did not enable separation of known and stranger victims, nor adult and child victims to explore how this choice depended on victimology.

locations are on record for each offender as this information is only collected where required for investigative or operational purposes. Nonetheless, the raw data included almost 5.5 million activity node records, for approximately 66,000 offenders. After removing offence and activity node records that did not reliably establish the presence of the offender (e.g., some specific frauds and offences involving publication or remote communication), or the time or location with sufficient specificity, approximately 60,000 offenders and 4.5 million nodes remained. Moreover, there were enough nodes of each type recorded of each type for enough offenders to enable the present analyses.

To yield as many prior activity locations as possible per crime location choice and to prevent nesting, for each of the five offence types—modelled separately—each offender's most recent offence was identified as their 'reference offence'(following Bernasco, 2010a); its location was the variable of interest. An offender could appear in multiple models if they committed more than one type of offence (e.g., both a residential burglary and a personal robbery). An offence could recur in the reference offences if it was also a co-offender's reference offence (following Menting et al., 2020; and Chamberlain & Boggess, 2016 who found that statistically adjusting for nesting of offenders in offences made no difference to the results). We included only offenders with at least one pre-offence activity node and only their nodes that pre-dated the reference offence.[5] We analysed 17,054 residential burglaries, 10,353 non-residential burglaries, 1,977 commercial robberies, 4,315 personal robberies and 4,421 extra-familial sex offences.[6] Demographic characteristics of the offenders are provided in Table S2.1 in the online supplementary materials.

Unit of Analysis

The question that the discrete spatial choice model sets out to answer is: which attributes of locations make offenders decide to choose them (or avoid them) for committing

---

[5] Resulting in the removal of 1.2% of residential burglars, 1.6% of non-residential burglars, 0.5% of commercial robbers, 0.9% of personal robbers and 8.9% of sex offenders.

[6] We randomly sampled 50% of reference offences to train each model, reserving the rest for future studies testing models' predictions.

crime? The 'locations' are the spatial units of analysis of the model, an advantage of DSCM being that it considers both units that were chosen, and those that could have been but were not. The units can range from specific addresses (Vandeviver & Bernasco, 2020), to administrative units such as census tracts or postcodes (e.g., Townsley et al., 2015). Selecting the spatial unit involves balancing a range of considerations by the analyst (Bernasco, 2010a). These include theoretical relevance (how big is one unit of 'activity space'?), spatial spill-over (if the unit is too small the choice is influenced by the attributes of surrounding units), spatial heterogeneity (if the unit is too big then variation within the unit that could affect the choice is not captured), and computational processing (if there are too many units the capacity of available computing equipment may be exceeded).

We used the New Zealand Census Unit 'Statistical Area 2' (SA2).[7] SA2s approximate neighbourhoods, typically containing 2000-4000 residents in urban areas and 1000-3000 residents in rural areas. Outliers with smaller populations represent industrial and commercial areas, remote regions and bodies of water with no residential population. SA2s are comparable to the units used in other neighbourhood level DSCM studies (e.g., Clare et al., 2009; Townsley et al., 2015).

In DSCM analysis, the dataset needs to include not only the attributes of chosen locations but those of alternative locations that could have been chosen but were not. To overcome computational challenges involved in including every SA2 in each offender's 'choice set' of possible alternatives—given the number of alternatives, offenders, variables and available computing capacity—we sampled from all potential SA2 alternatives (n=2153) to construct a manageable set for each offender. We followed a stratified importance sampling approach (Ben-Akiva & Lerman, 1985; McFadden, 1977) shown to yield estimates consistent with those produced by including all alternatives when modelling crime location choice (Curtis-Ham et al., 2021). Each offender's choice set included: the chosen SA2; all SA2s that contained an activity node or had an activity node within 5km (from the SA2

[7] The 2018 SA2 shapefile and metadata were downloaded from https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/. We excluded 83 SA2s made up of bodies of water from the analysis.

boundary); and 10 SA2s randomly selected from the remaining SA2s (i.e., those that were more than 5km away from any activity nodes). The median land area of the sampled SA2s was 1.2km^2 (quartiles 0.84, 2.2 km^2) on average across the five offences modelled.

Outcome Variable

The outcome variable reflects the location choice for each offender: from their set of potential SA2s, it is the one in which their reference offence was committed.

Activity Node Variables

To assess our hypotheses about the relationships between different types of activity nodes and crime location choice, each SA2 alternative in each offender's choice set was dummy coded depending on the presence (1) or absence (0) of each type of node in the SA2. *Home* included any residential address of the offender. Family homes included the residential addresses of persons linked to the offender in NIA as *immediate family* (parents, children and siblings, including step relationships), past or present *intimate partners* (spouses, partners and boy/girlfriends) and *other relatives* (grandparents/children, other relatives, foster and other care/guardianship relationships).[8,9] *School* included school and tertiary education locations. *Work* addresses were included where available, as a distinct type of activity node from education. *Prior offences* included crimes committed by the offender with sufficient evidence to prosecute. *Prior victim/witness* included offences involving the offender in a non-offending role such as a victim or witness, but not as an offender or suspect. *Prior incidents* included non-crime events reported to or detected by police such as domestic disputes, drunk and disorderly behaviour, mental distress, suspicious behaviour and truancy. Finally, the category *Other location* captured addresses recorded in the offender's address history with codes such as 'arrested at', 'seen at', 'frequents', 'trespassed from', 'spoken to at', 'stopped

---

[8] Family nodes were coded mutually exclusively, prioritising immediate family then intimate partners over other relatives.

[9] A limitation is that home addresses of some family members, especially intimate partners, may pre- or post-date the relationship and therefore may not have been visited by the offender. The data did not include relationship dates to enable restriction of activity locations to relationship periods.

at', 'other' and 'unknown'. Only activity nodes dated within 5 years of the reference offence were coded.[10] See Tables S1.1 and S2.2 in the online supplementary materials for detailed definitions and the distribution of nodes per offender, respectively.

We included activity locations that were less common in the data, less routinely visited, and less recent, erring towards inclusion of data. Activity locations generate awareness of crime opportunities that can result in the offender returning to commit crime years after their activities, at least as captured in available data (Bernasco, 2010b, 2019; Bernasco et al., 2015; Bernasco & Kooistra, 2010; Lammers et al., 2015; Menting et al., 2016). A single prior crime in a neighbourhood is associated with an increased likelihood of future offending there (Bernasco et al., 2015; Long et al., 2018), as are activities with relatively low frequency in short activity sampling periods (Bernasco, 2019: four days; Menting et al., 2020: one month). Thus, activities do not have be particularly recent or 'routine' to impart knowledge of crime opportunities that operates on future crime location choices (though the more recent or routine, the higher the likelihood of future crime). Further, single data points such as prior offences and incidents can be indicative of other activities carried out more routinely in those locations that are not captured in the data.

Each SA2 alternative was also coded 1 or 0 based on the presence or absence of the nearest of each type of node within five distance bands extending to 5km from the SA2 (with thresholds at 200m, 500m, 1km, 2km, and 5km from the SA2 boundary). Prior studies have shown significant associations between crime location choice and presence of activity nodes at an average of 3km away (Menting et al., 2020). In choosing 5km we balanced exploring the extent to which activity nodes farther away were associated with crime location choice with minimising the number of distance bands included, given that each distance band added 10 variables indicating the presence or absence of each of the 10 node types. Using these dichotomous variables per distance band rather than the linear distance to each type of node

---

[10] This cut-off ensured comparability in recency between different node types. The data initially included crime and incident records dating back to 2004—reflecting the start of the database and limited capture of historical records— and address records dating back to the offender's date of birth.

enabled us to visualise and compare the pattern of distance decay across node and crime types. Additionally, using distance bands provided a more easily interpretable scale than using contiguous spatial units at increasing lag orders (e.g., Menting et al., 2020), including distance bands both shorter and longer than the typical distance between neighbouring SA2s.[11]

As an example, if a given SA2 alternative contained a home node it was coded 1 for the variable 'home in SA2' and 0 for the variables 'home at 0-200m', 'home at 200-500m' and so on for the remaining distance bands. If a given SA2 alternative did *not* contain a home node, but the nearest home node was between 200m and 500m of the SA2, the SA2 was coded 0 for the variable 'home in SA2', 1 for the variable 'home node at 200-500m' and 0 for all other home node x distance band variables. If there were no nodes of any kind within 5km of a given SA2 alternative, the SA2 was coded 0 for all node type x distance band variables. Consistent with previous DSCM studies (e.g., Menting et al., 2020), this method prioritised the most proximal activity nodes while controlling for the presence of activity nodes farther away. The reference category for each type of node at each distance band was no node of that type within 5km of the SA2, which could either mean that the node did not exist in the first place (e.g., no police-recorded prior offences), or that it existed but the nearest one was farther than 5 km away. A stable model for commercial robbery was only achievable using three distance bands for work nodes (applied for all crime types for comparability): 'in SA2 to 200m', '200m-1km' and '1-5km'.[12] See Table S2.3 in the online supplementary materials for descriptive statistics for the node type x distance band variables.

Opportunity Variables

In line with previous DSCM studies, we operationalised opportunity as the number of potential targets, or an indicator thereof, in each spatial unit. Because the focus of this study

---

[11] Median nearest neighbour distance between centroids of sampled SA2s = 983m (average across models).

[12] This method enabled the model to converge and produce Variation Inflation Factors (VIFs) demonstrating no problematic multicollinearity. Across all crime types, the median VIF was 1.56 and the maximum was 5.69, well under the commonly applied threshold of 10 (Bernasco and Block 2011; O'Brien 2007).

was on offenders' activity locations rather than differentiating criminogenic features of the environment, we identified one opportunity variable for each crime type indicative of the number of potential targets per SA2. These variables came from New Zealand Statistics Census and Business Demography data (http://nzdotstat.stats.govt.nz/) and are further detailed in Table S1.2 in the online supplementary materials. For residential burglary we used the number of dwellings (see similarly Bernasco & Nieuwbeerta, 2005; Frith, 2019; Townsley et al., 2015). For non-residential burglary we used the total number of business units, which directly measures the number of non-residential burglary targets in the SA2. For commercial robbery we used the number of business units in industry categories that mapped to the types of crime locations used to identify commercial robberies. This was a more direct measure of potential targets than previous studies which used indirect measures such as retail footprint (Bernasco & Kooistra, 2010) and number of retail employees (Bernasco & Block, 2009). For personal robbery we used the number of business units in commercial or public industries, combining the types of facilities included in prior DSCM studies as separate covariates (Bernasco et al., 2013; Long et al., 2018; Song et al., 2019). This measure is a proxy indicator of the ambient population in a given spatial unit: how many potential targets are there, on average, for work, education, shopping or recreation purposes (Andresen, 2011). Ambient population predicts personal robbery and other offences against mobile targets (such as sexual assaults) better than residential population, thus forming the better single indicator of target distribution (Andresen, 2011; Andresen & Jenion, 2010; Rummens et al., 2021). We thus adopted the same measure of opportunity for sex offences.

Modelling Method

Consistent with most previous DSCM studies, we used conditional logit models to test our hypotheses (McFadden, 1984). The conditional logit model evaluates the likelihood of a choice alternative (SA2) being selected given its attributes (e.g., presence of a home in the SA2, presence of nearest prior crime 200-500m away). In each model (one per crime type) there were 61 variables: 10 node types x 6 distances (SA2 plus 5 bands) + 1 opportunity variable. We report model coefficients as Odds Ratios (ORs). For node types the ORs reflect the increase in odds of an SA2 being chosen given the presence of an activity node of that type at that distance band, compared to no node of that type within 5km of the SA2. For opportunity, the ORs reflect the increase in odds for every additional 100 opportunity units (e.g., households, businesses) in the SA2. When comparing ORs within models, if the ORs' 95% confidence intervals (CIs) overlapped we used Wald's Chi-Square tests of difference.

Results

The models provided a good fit to the data with Pseudo R^2 values between 0.35 and 0.42 (McFadden, 1973), which are at the upper end of the range of values found in other DSCM studies (Table 1). Figure 1 displays the 95% CIs of the ORs for each model (crime type), simultaneously enabling visual comparison variable-wise, distance-wise and crime-wise. Tables S3.1 and S4.1 in the online supplementary materials provide the ORs and CIs, and the Wald test results, respectively.

Table 1. Model summary statistics

Statistic	Res. Burg.	Non-Res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
N offence choices	17,054	10,353	1,977	4,315	4,421
N SA2 choice alternatives	4,213,998	2,417,735	563,161	1,243,204	784,537
Pseudo R^2	0.42	0.42	0.35	0.38	0.37

Note. SA2 = Statistical Area 2

As predicted by hypothesis 1, offenders were generally more likely to commit crime closer to activity nodes than farther away. All associations between crime location choice and the presence of a node in the same SA2 were significant except for work for residential burglary, prior victim/witness events for commercial robbery and school for sex offences. In several instances, contrary to expectation, offenders were less likely to commit crimes in SA2s with activity nodes present at the longest distance bands, than if those nodes were not present in any distance band (prior victim/witness events for burglary; school for commercial robbery and sex offences).

A distance decay pattern is clear in most of the boxes in Figure 1. For example, in the top leftmost box (home nodes and residential burglary), the ORs monotonically decrease with increasing distance from the focal SA2. ORs were not always significantly different from one distance band to the next, but ORs for the shortest distance band (in SA2) were significantly greater than ORs for the longest distance band except for other family homes and incidents for commercial robbery and work for personal robbery. The only other statistically significant exception to monotonic decay was intimate partner homes for non-residential burglary with slightly higher odds at 0-200m than in the SA2.

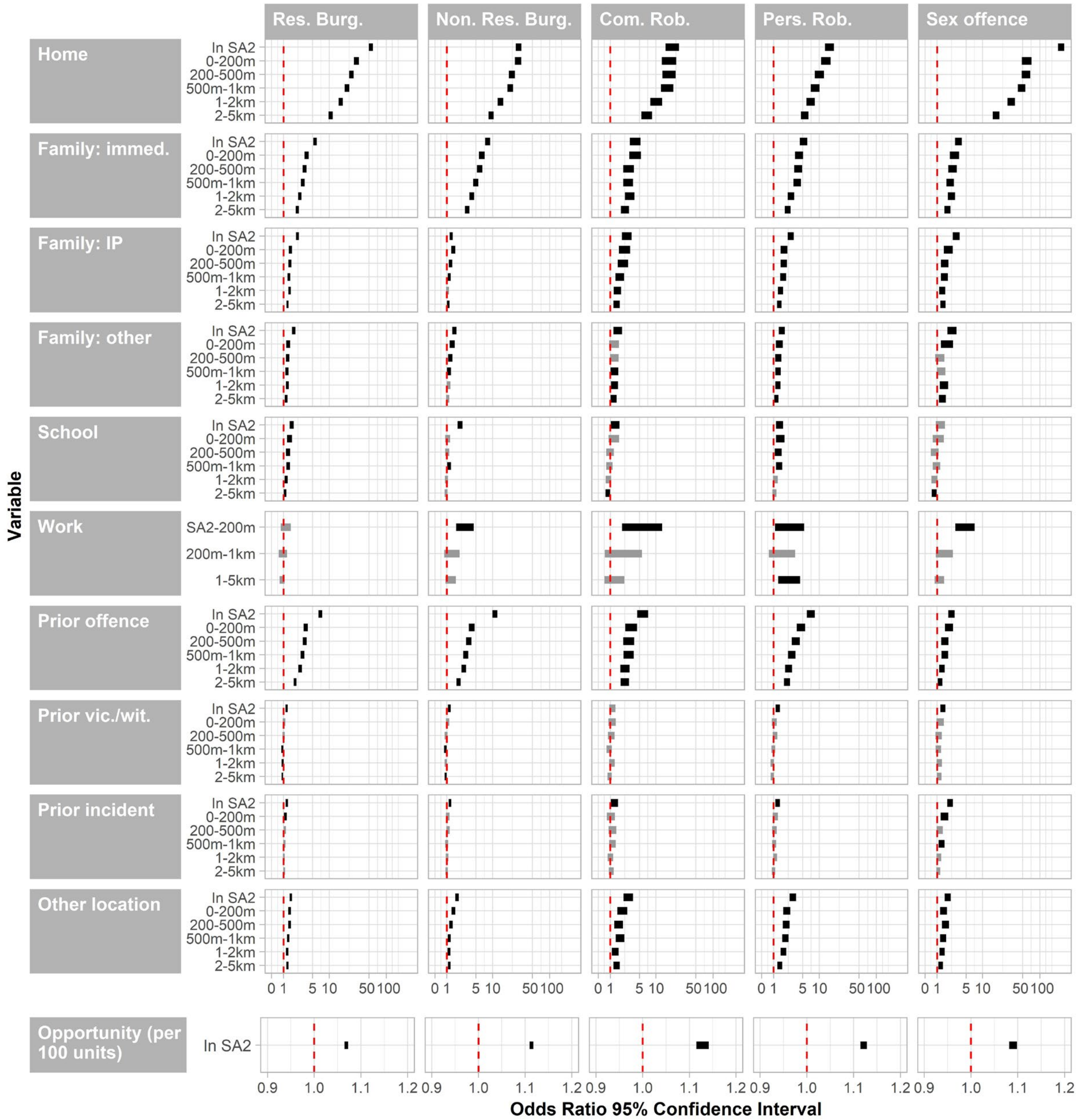


Figure 1. Odds ratio (OR) 95% confidence intervals (CIs) for node type and opportunity variables per distance band for each model (crime type)

Note: The shorter the bar, the smaller the CI. Grey bars indicate non-significant associations with CIs that cross the vertical line at 1. Bars that do not overlap indicate significant differences between ORs. Bars that overlap indicate possible non-significant differences between ORs (see Wald test results).

Confirming hypothesis 2, offenders were more likely to commit crime near their own homes than their family members' homes: all differences between 'home in SA2' and each type of 'family home in SA2' were significant (Figure 1). Regarding hypotheses 3(a) and 3(b), offenders were significantly more likely to commit crime in the same SA2 as immediate family members' homes than in the same SA2 as homes of current or former intimate partners (except sex offenders and commercial robbers) and of other relatives.

Hypothesis 4 was supported. Offenders were more likely to commit crime near to places they previously offended than places where they had previously been a crime victim or witness (no prior offence CIs overlapped with prior victim/witness CIs for 'in SA2'). We found partial support for hypothesis 5. Residential burglars were roughly twice as likely to commit crime near home than were non-residential burglars and commercial robbers. Residential burglars were—as hypothesised—more likely to offend near intimate partner homes than non-residential burglars, but they were less likely to offend near immediate family homes than non-residential burglars, and CIs overlapped for the remaining H5 comparisons (other family homes; commercial robbery).

Lastly, crime opportunity was positively associated with crime location choice, when controlling for the presence of activity nodes. For every increase of 100 households or relevant businesses (depending on crime type) in the SA2, the odds of an offence in that SA2 increased by 7% for residential burglary, 12% for non-residential burglary, 14% for commercial robbery, 13% for personal robbery and 10% for sex offences.

Discussion

This study aimed to identify near which of a range of activity locations (nodes) recorded in police data offenders are more likely to commit crime. Almost all nodes—including those not considered separately in previous studies: intimate partner and other relatives' homes, school, work, police involved incidents and other police contacts—were significantly and positively associated with crime location choice, at least at short distances. The lack of association between work nodes and residential burglary could reflect a lack of residential burglary opportunities around work nodes likely concentrated in non-residential areas; the lack of association between prior victim/witness events as commercial robbery indicates that these events were not conducive to the acquisition of knowledge of commercial robbery opportunities. There are several possible explanations for the lack of association between offenders' school/education nodes and sex offences. First, the aggregation of offences against children and adults in the present dataset could have masked any victim age specific association. Second, past research suggests that relatively few extrafamilial child sex offenders access their victims at or near schools (14%: Leclerc & Felson, 2016) and that those who do tend to carry out the offence elsewhere (Mogavero & Hsu, 2017). The few negative associations—at longer distances—were marginally significant and small. We treat these results with caution, but they could reflect the unobserved underlying environmental backcloth that shapes different activities (Brantingham & Brantingham, 1993a, 1993b) and the distribution of crime opportunity, or idiosyncrasies of the present dataset. Replicating this study elsewhere to see if similar patterns emerge would assist in resolving whether these exceptions were mere anomalies or not (and if not, the implications for theory).

Consistent with our expectations based on crime pattern theory (Brantingham & Brantingham, 1991, 1993a) crime was almost always most likely in the immediate vicinity of activity nodes, declining monotonically with distance. We place more weight on the general pattern and treat the few marginal exceptions with caution given the number of comparisons and limitations of the data (discussed below). The patterns of distance decay for different activity nodes evidence the extent of offenders' awareness space around different activity nodes and the likelihood that they will identify criminal opportunities within that space. Home and prior crimes showed the strongest associations with crime location choice, over longer distances, followed by homes of immediate family. These results are consistent with prior studies of residential burglary, robbery and crime in general (Menting et al., 2016), but provide novel confirmation for non-residential burglaries and extra-familial sex offences specifically. The apparent wider awareness space around home nodes makes sense: home nodes tend to anchor other routine activity locations that might not appear in the data (Golledge, 1999; Wang et al., 2013) and more 'paths' exist between home and other nodes, generating familiarity with a wider area around home (Schönfelder & Axhausen, 2002; Wang et al., 2013). In contrast, associations between prior victim/witness events and incidents and offenders' crime locations were largely limited to the same SA2, suggesting more constrained awareness space around these nodes. That the sex offenders were much more likely to offend in the same SA2 as they lived (OR 235) than other offenders likely reflects that some of these offences occurred at home.

That home, and to a lesser extent, immediate family homes and prior offence sites, were strongly associated with crime locations even at the 2-5km distance range could be explained by the relatively low urban population density of New Zealand (Demographia, 2019). When targets are more dispersed there are fewer crime opportunities within a given distance range, resulting in longer distances between activity nodes and crime and an extended decay curve. Accordingly, New Zealand burglary and sex offenders' home-crime distances are longer on average than those of offenders from other countries (Hammond, 2014; Lundrigan et al., 2010; Lundrigan & Czarnomski, 2006; Scott, 2012). Likewise, Townsley et al. (2015), found a smaller association between distance to home and residential burglary location choice in Australia than in the UK and the Netherlands, reflecting differences in urban population density. Our results are likely to generalise to other low population density jurisdictions and imply that studies in such jurisdictions could benefit from using an even longer distance range.

Differences in relative associations between crime locations and different nodes were largely consistent with the hypotheses derived from Curtis-Ham et al.'s (2020) theoretical model, thus lending further empirical support to this extension of crime pattern theory. Offenders were generally more likely to commit crime closer to activity nodes that were higher frequency (home versus family homes; immediate family versus other relatives); likely to impart more relevant knowledge about crime opportunities (prior crimes versus prior victim or witness locations); and of the same location type as the crime target (home for residential burglary versus non-residential burglary and commercial robbery). They were also more likely to offend near immediate family homes—presumed to be more enduring nodes—than near intimate partners' homes, though this difference was less pronounced for the commercial robbery and sex offenders. It is unclear why residential burglars were much more likely to offend near home than offenders targeting non-residential properties but not always more likely near family homes. Offenders' homes are just as likely as their family members' homes have more residential and fewer non-residential crime opportunities nearby. The difference must therefore lie in how knowledge of those opportunities is acquired through their activities at different residential nodes, a question that could be explored via survey or interview research in the future.

Several limitations of this study affect how our results can be interpreted. First, although the results demonstrate the ability of police data to indicate crime-salient elements of offenders' activity spaces, we have missed or understated theoretically important relationships for under-represented nodes, such as school and work. Future research could supplement police data with other administrative datasets that comprehensively capture these locations, akin to the use of government address registry data by Bernasco and colleagues (Bernasco, 2006; Bernasco & Kooistra, 2010; Menting et al., 2016).

Second, because we studied solved cases—albeit common practice in discrete crime location choice research—the results may not generalise to all offenders (Bernasco et al., 2013; Ruiter, 2017). For example, if those who offend close to home are more likely to be identified, relationships between home and offence locations would be stronger for solved than unsolved crimes. Similarly, if police are more likely to link an offender to a crime committed in a place that the police *know* the offender has visited, than to a crime committed in a place with no recorded link to the offender, we would also find stronger node-crime associations for solved than unsolved crimes. Conversely, if offenders are more likely to return to locations of prior crimes they were not caught for (not captured in the present data), relationships between prior crimes and offence locations would be weaker for solved than unsolved prior crimes. Correspondingly, Long et al. (2018) found that offenders were less likely to return to the location of a crime for which they were immediately caught than of a crime for which they initially avoided capture. However, they still were more likely to return to the location of a prior crime with immediate capture, than somewhere they had not previously offended. Further, the few studies to examine the spatial decision-making of offenders in solved versus unsolved crimes, suggest that differences between these groups are minimal (Bernasco et al., 2013; Lammers, 2014).

Lastly, variation may exist between offenders within offence types. For example, the sex offence data includes a range of offences, from indecent exposure to rape, against both adults and children, and known and stranger victims, and sex offenders employ a range of methods for identifying, selecting and approaching victims (Balemba & Beauregard, 2013; Beauregard & Busina, 2013; Mogavero & Hsu, 2017). Similarly, several prior studies have highlighted differences between offenders in the association between home or prior offence locations and future crime locations: for example, based on age and criminal history (Frith, 2019; Townsley et al., 2016). Future research could consider separately examining relationships between different activity nodes and crime location choice for specific subtypes of offences and groups of offenders.

Future research might also benefit from a more granular approach when examining the relationship between offence locations and offenders' prior victim/witness, police incident and other police-recorded activity locations. Within these categories there is likely variation in factors that influence whether those activities generate relevant knowledge of nearby crime opportunities, such as the type of offence or incident and its timing (Curtis-Ham et al., 2020).[13] Exploring this variation by measuring, more directly, the factors that mediate the influence of prior activity locations on crime location choice would help to identify why, for example, commercial robbers were no more likely to offend near prior victim/witness locations than farther away.

We conclude by highlighting the practical implications of quantifying the links between offenders' different activity locations and their crime locations, as we did in this study. This quantification enables prediction of most likely locations at which an individual will offend, given the locations, and nature, of their activity nodes. Such predictions could be used to identify offenders' high-risk locations for offending, to inform risk management strategies. In

[13] This disaggregation could also explore whether prior crimes are more likely to be of the same type as the reference offence than victim/witness experiences: an additional potential explanation for the present findings.

crime investigations, geographic profiling involves using offence locations to infer the likely location of an offender's home or other node (Rossmo, 2000). Our results could help to identify the kind of node that is more likely to be (e.g., home, school, prior crime), thus enabling prioritisation of suspects with higher likelihood nodes in the area predicted by the geographic profile (see examples of the geographic profiling process in Rossmo & Rombouts, 2008; Rossmo & Summers, 2015). Alternatively, given a list of suspects, our results could help identify who is more likely to have committed a crime at the location of an unsolved crime, given the nature and proximity of their activity nodes.

References

Ackerman, J. M., & Rossmo, D. K. (2015). How far to travel? A multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology*, *31*(2), 237–262. https://doi.org/10.1007/s10940-014-9232-7

Alston, J. D. (1994). *The serial rapist's spatial pattern of target selection* [Masters thesis, Simon Fraser University]. http://summit.sfu.ca/item/5080

Andresen, M. A. (2011). The ambient population and crime analysis. *The Professional Geographer*, *63*(2), 193–212. https://doi.org/10.1080/00330124.2010.547151

Andresen, M. A., & Jenion, G. W. (2010). Ambient populations and the calculation of crime rates and risk. *Security Journal*, *23*(2), 114–133. https://doi.org/10.1057/sj.2008.1

Balemba, S., & Beauregard, E. (2013). Where and when? Examining spatiotemporal aspects of sexual assault events. *Journal of Sexual Aggression*, *19*(2), 171–190. https://doi.org/10.1080/13552600.2012.703702

Beauregard, E., & Busina, I. (2013). Journey 'during' crime: Predicting criminal mobility patterns in sexual assaults. *Journal of Interpersonal Violence*, *28*(10), 2052–2067. https://doi.org/10.1177/0886260512471084

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.

Bernasco, W. (2006). Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling*, *3*(3), 139–155. https://doi.org/10.1002/jip.49

Bernasco, W. (2010a). Modeling micro-level crime location choice: Application of the discrete choice framework to crime at places. *Journal of Quantitative Criminology*, *26*(1), 113–138. https://doi.org/10.1007/s10940-009-9086-6

Bernasco, W. (2010b). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology*, *48*(2), 389–416. https://doi.org/10.1111/j.1745-9125.2010.00190.x

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLOS ONE*, *14*(1), e0210733. https://doi.org/10.1371/journal.pone.0210733

Bernasco, W., & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, *47*(1), 93–130. https://doi.org/10.1111/j.1745-9125.2009.00140.x

Bernasco, W., Block, R., & Ruiter, S. (2013). Go where the money is: Modeling street robbers' location choices. *Journal of Economic Geography*, *13*(1), 119–143. https://doi.org/10.1093/jeg/lbs005

Bernasco, W., Johnson, S. D., & Ruiter, S. (2015). Learning where to offend: Effects of past on future burglary locations. *Applied Geography*, *60*(Supplement C), 120–129. https://doi.org/10.1016/j.apgeog.2015.03.014

Bernasco, W., & Kooistra, T. (2010). Effects of residential history on commercial robbers' crime location choices. *European Journal of Criminology*, *7*(4), 251–265. https://doi.org/10.1177/1477370810363372

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *The British Journal of Criminology*, *45*(3), 296–315. https://doi.org/10.1093/bjc/azh070

Block, R., & Davis, S. (1996). The environs of rapid transit stations: A focus for street crime or just another risky place? In R. V. Clarke (Ed.), *Preventing mass transit crime* (Vol.

6, pp. 237–257). Criminal Justice Press. https://popcenter.asu.edu/sites/default/files/problems/street_robbery/PDFs/BlockDavis1996.pdf

Boivin, R., & Felson, M. (2018). Crimes by visitors versus crimes by residents: The influence of visitor inflows. *Journal of Quantitative Criminology*, *34*(2), 465–480. https://doi.org/10.1007/s10940-017-9341-1

Brantingham, P. L., & Brantingham, P. J. (1991). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (2nd ed., pp. 27–54). Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993a). Environment, routine, and situation: Toward a pattern theory of crime. In R. V. Clarke & M. Felson (Eds.), *Routine activity and rational choice* (pp. 259–294). Transaction Publishers.

Brantingham, P. L., & Brantingham, P. J. (1993b). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Chamberlain, A. W., & Boggess, L. N. (2016). Relative difference and burglary location: Can ecological characteristics of a burglar's home neighborhood predict offense location? *Journal of Research in Crime and Delinquency*, *53*(6), 872–906. https://doi.org/10.1177/0022427816647993

Chopin, J., & Caneppele, S. (2018). The mobility crime triangle for sexual offenders and the role of individual and environmental factors. *Sexual Abuse*, *31*(7), 812–836. https://doi.org/10.1177/1079063218784558

Chopin, J., & Caneppele, S. (2019). Geocoding child sexual abuse: An explorative analysis on journey to crime and to victimization from French police data. *Child Abuse & Neglect*, *91*, 116–130. https://doi.org/10.1016/j.chiabu.2019.03.001

Clare, J., Fernandez, J., & Morgan, F. (2009). Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. *Australian & New Zealand Journal of Criminology*, *42*(2), 139–158. https://doi.org/10.1375/acri.42.2.139

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608. https://doi.org/10.2307/2094589

Costello, A., & Wiles, P. (2001). GIS and the journey to crime: An analysis of patterns in South Yorkshire. In K. J. Bowers & A. Hirschfield (Eds.), *Mapping and analysing crime data: Lessons from research and practice* (pp. 27–60). Taylor & Francis.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5

Davies, A., & Dale, A. (1996). Locating the stranger rapist. *Medicine, Science and the Law*, *36*(2), 146–156. https://doi.org/10.1177/002580249603600210

Demographia. (2019). *Demographia world urban areas: 15th annual edition 201904*. Demographia. http://www.demographia.com/db-worldua.pdf

Frank, R., Andresen, M. A., Cheng, C., & Brantingham, P. L. (2011). Finding criminal attractors based on offenders' directionality of crimes. In N. Memon & D. Zeng (Eds.), *2011 European Intelligence and Security Informatics Conference* (pp. 86–93). CPS. https://doi.org/10.1109/EISIC.2011.34

Frith, M. J. (2019). Modelling taste heterogeneity regarding offence location choices. *Journal of Choice Modelling*, *33*, 100187. https://doi.org/doi.org/10.1016/j.jocm.2019.100187

Golledge, R. (1999). Human wayfinding and cognitive maps. In R. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 5–45). Johns Hopkins University Press.

Hammond, L. (2014). Geographical profiling in a novel context: Prioritising the search for New Zealand sex offenders. *Psychology, Crime & Law*, *20*(4), 358–371. https://doi.org/10.1080/1068316X.2013.793331

Hart, T. C., & Miethe, T. D. (2015). Configural behavior settings of crime event locations: Toward an alternative conceptualization of criminogenic microenvironments. *Journal of Research in Crime and Delinquency*, *52*(3), 373–402. https://doi.org/10.1177/0022427814566639

Hewitt, A., Beauregard, E., & Davies, G. (2012). "Catch and release": Predicting encounter and victim release location choice in serial rape events. *Policing: An International Journal*, *35*(4), 835–856. https://doi.org/10.1108/13639511211275814

Hewitt, A. N., Chopin, J., & Beauregard, E. (2020). Offender and victim 'journey-to-crime': Motivational differences among stranger rapists. *Journal of Criminal Justice*, *69*, 101707. https://doi.org/10.1016/j.jcrimjus.2020.101707

Lammers, M. (2014). Are arrested and non-arrested serial offenders different? A test of spatial offending patterns using DNA found at crime scenes. *Journal of Research in Crime and Delinquency*, *51*(2), 143–167. https://doi.org/10.1177/0022427813504097

Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology*, *53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071

Leclerc, B., & Felson, M. (2016). Routine activities preceding adolescent sexual abuse of younger children. *Sexual Abuse*, *28*(2), 116-131. https://doi.org/10.1177/1079063214544331

Long, D., Liu, L., Feng, J., & Zhou, S. (2018). Assessing the influence of prior on subsequent street robbery location choices: A case study in ZG city, China. *Sustainability*, *10*(6), 1818. https://doi.org/10.3390/su10061818

Lundrigan, S., & Czarnomski, S. (2006). Spatial characteristics of serial sexual assault in New Zealand. *Australian & New Zealand Journal of Criminology*, *39*(2), 218–231. https://doi.org/10.1375/acri.39.2.218

Lundrigan, S., Czarnomski, S., & Wilson, M. (2010). Spatial and environmental consistency in serial sexual assault. *Journal of Investigative Psychology and Offender Profiling*, *7*(1), 15–30. https://doi.org/10.1002/jip.100

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1977). *Modelling the choice of residential location* (No. 477; Cowles Foundation Discussion Papers). Yale University. https://EconPapers.repec.org/RePEc:cwl:cwldpp:477

McFadden, D. (1984). Econometric analysis of qualitative response models. In P. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 105–142). Elsevier. https://doi.org/10.1016/S1573-4412(84)02016-X

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2016). Family matters: Effects of family members' residential areas on crime location choice. *Criminology*, *54*(3), 413–433. https://doi.org/10.1111/1745-9125.12109

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: Findings from an online self-report survey. *The British Journal of Criminology*, *60*(2), 303–322. https://doi.org/10.1093/bjc/azz044

Mogavero, M. C., & Hsu, K.-H. (2017). Sex offender mobility: An application of crime pattern theory among child sex offenders. *Sexual Abuse*, *30*(8), 908–931. https://doi.org/10.1177/1079063217712219

Rossmo, D. K. (2000). *Geographic profiling*. CRC Press.

Rossmo, D. K., & Rombouts, S. (2008). Geographic profiling. In R. Wortley & L. Mazerolle (Eds.), *Environmental criminology and crime analysis* (pp. 136–149). Willan. https://www.taylorfrancis.com/books/e/9781136308451

Rossmo, D. K., & Summers, L. (2015). Routine Activity Theory in crime investigation. In *The Criminal Act* (pp. 19–32). Palgrave Macmillan, London. https://doi.org/10.1057/9781137391322_3

Ruiter, S. (2017). Crime location choice. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook of offender decision making* (pp. 398–420). Oxford University Press.

Rummens, A., Snaphaan, T., Van de Weghe, N., Van den Poel, D., Pauwels, L. J. R., & Hardyns, W. (2021). Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? *ISPRS International Journal of Geo-Information*, *10*(6), 369. https://doi.org/10.3390/ijgi10060369

Schönfelder, S., & Axhausen, K. W. (2002). *Measuring the size and structure of human activity spaces: The longitudinal perspective* [Working Paper]. ETH; DOI: 10.3929/ethz-a-004444846. https://www.research-collection.ethz.ch/handle/20.500.11850/36482

Scott, D. (2012). *The travelling distances of stranger intruder sex offenders* [Research Report]. New Zealand Police.

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-019-09406-z

Townsley, M. (2016). Offender mobility. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (pp. 142–161). Routledge.

Townsley, M., Birks, D., Bernasco, W., Ruiter, S., Johnson, S. D., White, G., & Baum, S. (2015). Burglar target selection: A cross-national comparison. *Journal of Research in Crime and Delinquency*, *52*(1), 3–31. https://doi.org/10.1177/0022427814541447

Townsley, M., Birks, D., Ruiter, S., Bernasco, W., & White, G. (2016). Target selection models with preference variation between offenders. *Journal of Quantitative Criminology*, *32*(2), 283–304. https://doi.org/10.1007/s10940-015-9264-7

van Daele, S. (2009). Itinerant crime groups: Mobility attributed to anchor points? In L. Pauwels, P. Ponsaers, G. Vande Walle, T. Vander Beken, F. Vander Laenen, G. Vermeulen, M. Cools, S. De Kimpe, B. De Ruyver, & M. Easton (Eds.), *Contemporary issues in the empirical study of crime* (Vol. 1, pp. 211–225). Maklu.

van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A time for a crime: Temporal aspects of repeat offenders' crime location choices. *Journal of Research in Crime and Delinquency*, *55*(4), 538–568. https://doi.org/10.1177/0022427818766395

van Sleeuwen, S. E. M., Ruiter, S., & Steenbeek, W. (2021). Right place, right time? Making crime pattern theory time-specific. *Crime Science*, *10*(1), 1–10. https://doi.org/10.1186/s40163-021-00139-8

Vandeviver, C., & Bernasco, W. (2020). "Location, location, location": Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*, *36*(4), 779–821. https://doi.org/10.1007/s10940-019-09431-y

Wang, X., Grengs, J., & Kostyniuk, L. (2013). Visualizing travel patterns with a GPS dataset: How commuting routes influence non-work travel behavior. *Journal of Urban Technology*, *20*(3), 105–125. https://doi.org/10.1080/10630732.2013.811986

CHAPTER 6

Familiarity and Activity Similarity in Crime Location Choice

The research presented in this chapter continued the empirical tests of the proposed theoretical model systematising the links between the activity locations in offenders' mental maps and their crime locations (Chapter 2). In contrast to the approach described in Chapter 5 comparing broad types of activity locations, it more directly measured the attributes of activity locations proposed to affect offenders' crime locations, such as how frequently or recently they were visited. Many of these attributes had been examined individually in the literature from which the theoretical model was derived (see Chapter 2). But the model suggests that they contribute to two interacting constructs—reflecting offenders' familiarity with locations and the similarity of their past activities in those locations to the future crime.[1] The following manuscript, currently under review, focuses on the interaction of composite familiarity and activity similarity variables compiled from the individual attributes. It examines whether people are more likely to commit crime near activity locations with which they are more familiar and where their activities were more similar to the future crime, than near activity locations with lower levels of familiarity and activity similarity.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Familiar locations and similar activities: Examining the interaction of reliable and relevant knowledge in offenders' crime location choices.

---

[1] In the theoretical model these constructs are labelled 'knowledge reliability' and 'knowledge relevance' respectively.

Familiar Locations and Similar Activities: Examining the Interaction of Reliable and Relevant Knowledge in Offenders' Crime Location Choices

Abstract

Objectives: This study tested a recently theorized interaction between the reliability and relevance of offenders' knowledge of locations in their crime location choices.

Methods: We analysed associations between offenders' pre-offence activity locations recorded in police data and the locations of 17,054 residential burglaries, 10,353 non-residential burglaries, 1,977 commercial robberies, 4,315 personal robberies and 4,421 extra-familial sex offenses, in New Zealand. The activity locations included home addresses, family members' home addresses, work, school and locations of prior offences, victimizations, non-crime incidents and other police contacts. These locations were coded on theorised reliability and relevance attributes (e.g., how frequently and recently they were visited, and how similar the activities were to the location choice offence). We used discrete spatial crime location choice models to test the interaction of reliability and relevance attributes.

Results: Offenders were most likely to commit crime where their prior activity locations indicated they had highly reliable and highly relevant knowledge—where they were both highly familiar with the area and had conducted similar activities in the past. Crime was less likely where offenders had less familiarity or had conducted less similar activities in the past.

Conclusions: The results highlight the importance of accounting for both the reliability and relevance of offenders' activity locations when modelling or predicting their crime location choices. The study contributes to the extension of crime pattern theory and the ability to predict individuals' future crime locations based on their past activity locations

Keywords

crime location choice, crime pattern theory, discrete spatial choice, police data, routine activity locations

Introduction

Understanding why people commit crime where they do has important implications for policing and criminal justice. Being able to predict—at an individual level—where someone will offend can help criminal justice agencies to tailor strategies to manage offenders' risk of re-offending by identifying their high-risk locations and helping them to avoid these locations or otherwise change their routine activities. The same knowledge can also be utilised in the aftermath of crime. It can help police to prioritise suspects in crime investigations by identifying suspects most likely to have committed the crime given its location (Curtis-Ham et al., 2020).

At a more fundamental level, an understanding of the spatial decisions of offenders contributes to the development of scientific theory. According to crime pattern theory, an influential theoretical framework in environmental criminology, people commit crime where their awareness space—the areas they know around activity locations where they live, work and socialise—overlaps with crime opportunities (P. L. Brantingham & Brantingham, 1991). A lot of research has investigated the kinds of places where these overlaps occur in the aggregate: the places where many people are likely to commit crime (see for example, Bruinsma & Johnson, 2018; Weisburd et al., 2016). However, making individualised explanations and predictions requires understanding how people's individual activity locations influence their crime locations: where in relation to these locations are they most likely to identify and exploit crime opportunities? The present study adds to the literature that addresses this question.

As we elaborate below, empirical research on this question has to date focused on the relationships between specific attributes of people's activity locations associated with their crime location choices, such as how frequently they have visited the location (e.g., Bernasco, 2019) or whether they committed a similar crime there before (e.g., van Sleeuwen et al., 2018). Curtis-Ham et al. (2020) recently formalised these relationships in a theoretical model, which proposes that attributes of individuals' activity locations operate on crime location choice via two psychological mechanisms reflecting how their activities generate knowledge of crime opportunities. In the model, offending is more likely where, and when, offenders' have *reliable* knowledge (affected by the frequency, recency and duration of their past activities there) that is *relevant* to the future crime (affected by the similarity of those past activities to the future crime). In other words, the more *familiar* the location is to the offender and the more *similar* the offender's prior activities in the location are to the crime at hand, the more likely is the offender to choose the location for committing the crime.

The model theorises that knowledge reliability and relevance interact: both are necessary, but insufficient in isolation, for crime commission (Curtis-Ham et al., 2020). But since prior research has focused on specific attributes of activity locations in isolation, this interaction has not been empirically validated. Therefore, to add to the theory's empirical support, and to our ability to explain and predict where individuals will commit crime, we examined this interaction using data on offences and pre-offence activity locations for a large sample of burglary, robbery and extra-familial sex offenders in New Zealand. In the first section of this paper, we elaborate on the existing theoretical and empirical literature considering relationships between individuals' prior activity locations and their crime location choices.

Prior Activities and Crime Location Choice

Crime Pattern Theory postulates that people offend where crime opportunities overlap with places they are familiar with from their routine activities (P. L. Brantingham & Brantingham, 1991, 1993a, 1993b). The theoretical model proposed by Curtis-Ham et al. (2020) extended Crime Pattern Theory to consider what kind of activities are likely to generate the requisite level of familiarity to create awareness of and act on crime opportunities. As mentioned, the model distinguishes between attributes of offenders' prior activities that affect the *reliability* of knowledge and attributes that affect the *relevance* of knowledge. The reliability of knowledge reflects how familiar offenders become with potential crime locations, and is a function of the frequency, recency and duration of prior activities in these locations. The relevance of knowledge is the extent to which they become aware of crime opportunities and is a function of the similarity between prior activities and future offences. Specifically, the more similar are the prior activities to a future offence, the more likely they are to generate relevant knowledge of the location's potential for such offending. The model further suggests several dimensions along which similarity can be considered, by comparing the behaviour (e.g., same or different prior crime), timing (e.g., time of day or day of week), and location (e.g., residential versus commercial) involved in the prior activities and future offence.

A range of studies provide support for the association between individual prior activity factors and crime location choice. For example, the odds of an offender committing crime are higher in proximity to activity locations they have visited more frequently, regardless of the broad type of activity location, such as home, school, workplace or venue for socialising (Bernasco, 2019; Menting et al., 2020). Each of the three familiarity factors have also been evidenced with reference to specific types of activity node. People are more likely to commit crime near places where they have *frequently* committed prior crimes than near places where they have committed few or no prior crimes (Lammers et al., 2015). The odds are also higher near more *recent* home and family home addresses (Bernasco, 2010b; Bernasco & Kooistra, 2010; Lammers et al., 2015; Menting et al., 2016) and near places where they committed more recent prior crimes (Bernasco et al., 2015; Lammers et al., 2015; Long et al., 2018). With respect to *duration*, the longer that an offender has resided at a home address, the higher the odds of crime in its vicinity (Bernasco & Kooistra, 2010; Lammers et al., 2015).

Similarity of the behaviour, timing, and type of location involved in offenders' prior activities have also been explored. People are more likely to offend where they have previously committed the *same kind of crime* than at the site of a different crime type (Lammers et al., 2015; van Sleeuwen et al., 2018). Examining whether people were more likely to offend at certain broad types of activity location, Curtis-Ham et al. (under review) found that crime was more likely where people had previously committed a crime than where they had been a victim or witness to a crime. This finding further supports the model's proposition that people offend in places where their past activities were more similar, behaviourally, to the future crime. With regard to *timing*, offences are also more likely at places where offenders have previously committed crimes that are closer in terms of times of day and week to the time of day and week of the present offence than at previous crime sites that are less closely matched temporally to the current offence (van Sleeuwen et al., 2018) and at places they have routinely visited during the day for daytime offences and during the night for night-time offences (van Sleeuwen et al., 2021).

Considering *location type*, Bernasco (2010) found stronger associations between residential activity locations (current and prior home addresses) and crime locations for residential burglary than for robbery and car theft. Curtis-Ham et al.'s (under review) comparison of broad types of activity location likewise found that residential burglars were more likely to offend near a past or present home address than were commercial robbers and non-residential burglars. These studies controlled for the presence of opportunities (e.g., residential population, number of residences), so these associations do not simply reflect that residential burglaries—for example—are more likely in residential areas. Rather, the location similarity factor captures that people motivated to commit a residential burglary are more likely to acquire relevant knowledge of residential burglary opportunities near activity locations that are residential in nature than near non-residential activity locations (see Curtis-Ham et al, 2020).

These previous studies have been confined to considering a single variable across a range of activity locations (Bernasco, 2019; Menting et al., 2020)[1] or one or two variables within one kind of activity location (e.g., recency + duration of home addresses: Bernasco, 2010b; Bernasco & Kooistra, 2010; frequency of prior crimes: Lammers et al., 2015). While

---

[1] Both these studies used surveys to capture adolescents' activity locations and subsequent offending, which enabled a wide range of activity locations to be captured but precluded analysis of multiple variables due to the small samples.

these studies have shed light on individual elements of the theoretical model, no research to date has included both a wide range of activity locations and a wide range of activity location variables, which would help to validate the theoretical model as a whole.

Moreover, the theoretical model proposed that offenders are more likely to commit crime where they have knowledge of crime opportunities that is *both* reliable and relevant (Curtis-Ham et al., 2020). In other words, neither familiarity with a location nor awareness that it presents good crime opportunities is sufficient in isolation; both are necessary. Reliability and relevance thus interact, such that places with highly reliable and relevant knowledge would be most likely to be chosen, places that are high on one dimension but not the other are less likely choices, and places low on both reliability and relevance are the least likely. But the interaction of reliability and relevance of prior activities has not yet been investigated empirically.[2]

We therefore sought to test the theoretical model by incorporating all the variables specified in the model across a wide array of activity locations, and by examining their association with crime location choice via the interaction of reliability and relevance, prompting the following hypothesis:

> H1 Offenders are more likely to commit crime near activity nodes that are high on both reliability and relevance, less likely to commit crime near activity nodes that are high on only reliability or relevance and least likely to commit crime near activity nodes that are low on both reliability and relevance.

---

[2] Very indirect evidence for this interaction comes from Menting's (2018) finding that the odds of crime are reduced when there are fewer crime opportunities near offenders' homes (high reliability, less potential to gain relevant knowledge). However, this study was concerned with the interaction of the presence of activity locations and the presence of opportunities. It did not consider the interaction of attributes of activity locations reflecting the acquisition of reliable and relevant knowledge about those opportunities.

Because neither Crime Pattern Theory nor the extended framework developed by Curtis-Ham et al. (2020) hypothesises that the functional role of an activity location for the offender (e.g., whether it is their home, school, workplace or other activity node) affects its likelihood of being selected for crime, we do not distinguish activity locations by their functional role. We test H1 by considering all types of activity locations available in the data combined, rather than within each functional type of location. This approach is consistent with the aforementioned studies of activity frequency and crime location choice (Bernasco, 2019; Menting et al., 2020).

The hypothesis (H1) predicts that crime is more likely 'near' to activity nodes with certain attributes, but what is 'near'? In the crime location choice literature, near is often operationalised as meaning 'within the same spatial unit of analysis'. This could mean, for example, in the same 200m x 200m grid cell (Bernasco, 2019), the same census block (e.g., Bernasco et al., 2017), or the same neighbourhood as identified by postcodes (e.g., Lammers, 2018). Further, spatial units at a series of contiguous lags can be included, to estimate associations at decreasing 'nearness'. Doing so has consistently produced a 'distance decay' pattern, whereby offenders are more like to commit crime in spatial units closer to their activity nodes than in those farther away (Bernasco, 2019; Bernasco & Block, 2009; Kuralarasan & Bernasco, 2021; Menting et al., 2020). We therefore included a range of 'near' distances and expected the hypothesised associations to be stronger at shorter distances than longer distances.

Lastly, we consider the role of opportunity. Opportunity means the presence of a suitable target in the absence of capable guardians (Cohen & Felson, 1979). Crime occurs where opportunities exist and offenders have awareness of them through their prior activities or other indirect information sources (P. L. Brantingham & Brantingham, 1991). In this study our interest was in isolating the relationship between the nature of prior *activities* and future crime locations, rather than the mere presence of opportunities and future crime locations. We therefore control for the presence of crime opportunities, using indicators of the number of potential targets akin to those used in previous crime location choice research (Ruiter, 2017).

In sum, empirical evidence shows that offenders tend to commit crime near locations where their past activities have been more frequent, recent, enduring *or* similar to the present crime. But the proposed interaction between these factors has not been empirically explored. This study therefore tested whether offenders are likely to offend near places of which they have knowledge that is both reliable and relevant.

Method

To test our hypothesis, we used discrete spatial choice models (DSCMs). DSCMs are discrete choice models applied to spatial choices, such as where to buy a house (McFadden, 1977), where to shop for groceries (Hillier et al., 2015), or, in the present case, where to commit a crime. DSCMs have been applied in many studies of crime location choice given their ability to model the association between the attributes of potentially chosen locations (choice 'alternatives') and the choice of location (Ruiter, 2017). Attributes of possible alternatives can include location-specific variables such as the number of potential crime targets in the location. They can also include person-specific variables such as how frequently the individuals have visited the location previously, or how far away it is from their home address.

The locations of possible alternatives are the unit of analysis, which in the present study are the neighbourhoods ('Statistical Area 2' Census Units, or 'SA2s') from which offenders are selecting when deciding where to commit crime.[3] Consistent with previous

---

[3] SA2s usually contain 2000-4000 residents (1000-3000 in rural areas). The SA2s included in this study following sampling from alternatives as described in this section were a median of 1.2km^2 land area, with 50% sized between 0.84 and 2.2 km^2. The 2018 SA2 shapefile and metadata were downloaded from

studies (e.g., Clare et al., 2009; Townsley et al., 2015), we use SA2s to examine meso-level spatial choices. In addition to the theoretical suitability of neighbourhoods as our chosen level of spatial aggregation, using SA2s balances the risk of ignoring spatial spill-over mechanisms, where choices between small units are affected by nearby units, and the risk of ignoring spatial heterogeneity, where the use of large units may fail to capture important variation within those units. The inclusion in our models of the attributes of surrounding areas within a series of distance bands around each SA2 alternative (described in detail below) further reduces the risk of ignoring spatial spill-over mechanism on our results by explicitly accounting for the presence of activity locations in nearby units.

Estimating DSCMs requires a dataset that includes one row for each potentially chosen alternative for each decision-maker (here, offenders). But including all 2153 SA2s for each offender, would, for example, yield a dataset of almost 37 million rows for the 17,054 residential burglaries, presenting a computational barrier given available computing equipment. We therefore sampled from alternatives to create a smaller set of alternatives for each offender using stratified importance sampling (Ben-Akiva & Lerman, 1985; McFadden, 1977). For each offender we included the chosen SA2, all SA2s with any activity nodes within 5km of the SA2 boundary and 10 SA2s randomly selected from the remaining SA2s. Curtis-Ham et al. (2021) demonstrated that this strategy is optimal when compared with simple random sampling, yielding estimates close to those produced without sampling while reducing computational burden. The outcome variable and variables of interest were then coded for each SA2 in each offender's resulting set of alternatives, as described in the following sections.

---

https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/. Eighty-three SA2s made up of bodies of water were excluded from the analysis.

Offence and Activity Node Data

Offence and activity node data were sourced from the New Zealand Police National Intelligence Application (NIA). Initially, we collected all burglary, robbery and extra-familial sex offences committed between 2009 and 2018 where an offender had been identified with sufficient evidence for police against them.[4] These data included the location (address and geographic coordinates), date and time at which the offence occurred; other offence attributes indicating the type of location and type of crime; and demographic details of the offender. Such data on detected or 'cleared' offences are commonly used in research on offenders' decisions where or whether to offend (Bernasco, 2017), including the majority of DSCM studies of crime location choice (Ruiter, 2017).

Following data cleaning, [5] we identified each offender's most recent offence ('reference offence') for each of the five crime types included in this study, and took a random sample of 50% per crime category,[6] yielding 17,054 residential burglaries, 10,353 non-residential burglaries, 1,977 commercial robberies, 4,315 personal robberies and 4,421 extra-familial sex offences. Analysing these offence types enabled us to verify whether the hypothesis held for a range of offences that involved different types of target (premises versus people) and motivations (acquiring property versus sexual violence), while keeping the number of separate analyses to be run manageable. We note that the theoretical model we tested is universal to all crime so our hypothesis would not change if different crime types were included. Each crime type was analysed separately, though if an offender had committed

---

[4] These offenders may then have been arrested, charged, or dealt with by way of a formal warning or other out-of-court diversionary action.

[5] Data cleaning of *offences* involved removal of approximately 3% of offence records due to missing or imprecise location or timing information. Of the remaining offenders, 1.2% of residential burglars, 1.6% of non-residential burglars, 0.5% of commercial robbers, 0.9% of personal robbers and 8.9% of sex offenders were removed as having no pre-offence activity locations to include in the analysis (following cleaning of the activity location data).

[6] We sampled in order to reduce computing effort and to reserve data for other analyses planned in a wider research program.

more than one type of crime they would appear in multiple analyses (e.g., both commercial robbery and residential burglary).[7] But each offender only appeared in each analysis once (i.e., one reference offence per offender per model). The dataset included offenders of any age (median ages 18-21 for the property offences and 28 for sex offences) and gender (ranging from 80% male for personal robbery to 97% male for sex offences).

Pre-offence activity locations were also sourced from NIA and included the locations of: offenders' and their family members' home addresses (about 15% and 46% of activity locations, respectively); school and other educational institutions attended by the offenders (1%); their workplaces (<1%); previous offences they committed (15%); offences they were involved in as a victim or witness (4%); non-crime incidents in which they were involved (4%); and miscellaneous encounters with or sightings by police labelled, for example, as 'spoken to at', 'seen at', 'frequents' or 'arrested at' (14%). Activity locations could date back to the offender's date of birth, except for prior offences/incidents which dated from 2004 (due to limited transfer of these records when NIA was created). Within this timeframe we included all activity locations regardless of how recent they were because recency was one of the variables that made up our overall reliability measure as described below.

Previous DSCM studies have used similar police data on offenders' prior crimes and home addresses (e.g., Frith, 2019; Long et al., 2018), sometimes supplemented with public registry data on offenders' and their family members' home addresses (e.g., Menting, 2018; Menting et al., 2016). The present dataset enabled us to examine reliability and relevance factors across a more representative array of activity locations. Information about these activity locations is recorded where needed by police for operational or investigative purposes. Our data thus do not capture all pre-offence activity locations for all offenders, but

---

[7] 87% of offenders appeared in one analysis (i.e., for one reference offence), 12% in two analyses, 1% in three, 0.1% in four and <0.01% in all five analyses. Although these statistics suggest specialisation in offending, note that some offenders committed prior crimes outside the reference offence types—displaying versatility—or committed only the reference offence—indicating neither specialisation nor versatility (see DeLisi & Piquero, 2011 for further detail on these concepts).

most offenders had at least 10 pre-offence activity locations, following data cleaning and the removal of locations that did not pre-date the offenders' most recent offence.[8] The activity location data included a range of attributes that were used to code the reliability and relevance variables as described below.

Outcome Variable

The outcome variable reflects each offender's choice of crime location: in which of the 2153 SA2s the reference offence was committed.

Prior Activity Location Variables

Figure 1 outlines the steps taken to code attributes of prior activity locations and convert these into attributes of the spatial unit of analysis (SA2s), which were as follows. First, reliability and relevance variables (e.g., recency, behaviour similarity) were calculated for each prior activity node (e.g., home address, family member's home address, location of one or more prior crimes). Table 1 summarises how the reliability and relevance variables were operationalised; Supplementary Materials S1 gives further detail. Not all variables were present in the data for all activity locations, requiring a range of assumptions to be made to 'fill in the gaps'. These gaps partly arose due to the collection of two types of activity location records: 'time-span nodes' that spanned a time period with start and end dates (offender and family home addresses, school, and work), and 'events' that took place at a specified time and date (prior crimes, victim/witness events, non-crime incidents and the miscellaneous police contacts). It was necessary to code some variables based on assumptions about the nature and timing of activities at the time-span nodes. For example, all home nodes were assumed to have maximal time of day and day of week similarity to the reference offence as one routinely visits home at any time of day. Home nodes were also

---

[8] Data cleaning of *activity locations* involved removal of approximately 12% of activity location records due to missing or imprecise location or timing information, excluding certain categories of prior offences and incidents that preliminary checks showed did not reliably indicate that the person was present at the location of the offence/incident, and excluding offenders who no longer appeared in the *offence* data following data cleaning (see note 7).

assumed to have low behavioural similarity to the reference offence. All variables were then re-scaled so that a value of 1 represented maximal frequency, recency, duration and similarity (i.e., 100th percentile out of all offenders' activity locations in the dataset for each crime type model).

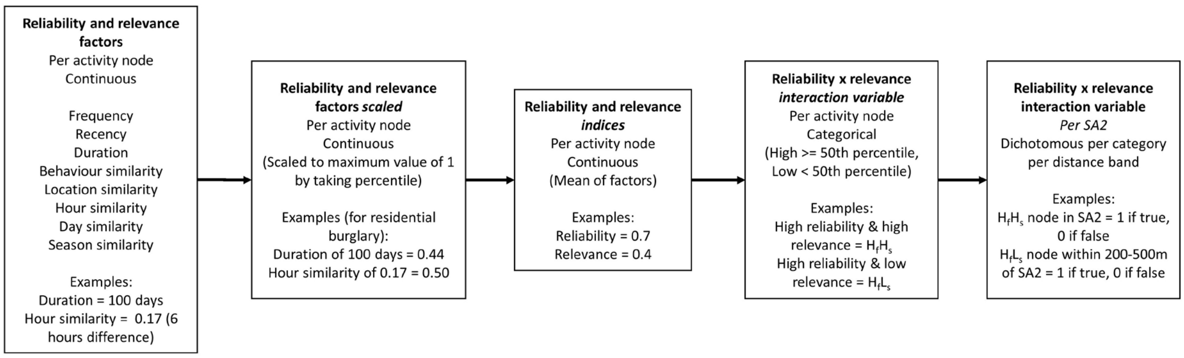


Figure 1. Overview of the process for constructing the prior activity location variables

Table 1. Reliability and relevance variable construction summary

Factor	**Definition**	**Coding rules**
Frequency (0 to 1)	Proportion of days over the offender's association with a given node that it was visited	Home: daily = 1 Family members' homes: fortnightly = 0.07 School/education: 200 days per year = 0.55 Work: 240 days per year = 0.66 Events: number of events at location/duration.
Recency (0 to 1)	Reciprocal of the number of days between reference offence and most recent activity at the node.	Time-span nodes: used latest end date of all records for that node (e.g., home address). Events: used date of latest event.
Duration (0 to max duration)	Number of days between earliest and latest date at the node.	Time-span nodes: earliest start date to latest end date of all records for that node (e.g., home address). Events: date of first event to date of latest event.
Behaviour similarity (1 to 7)	Ordinal scale of similarity based on Kuang et al.'s (2017) study of the similarity of different crimes (see Supplementary Materials S1).	7 Prior crime of same crime type 6 Prior crime of similar crime type 5 Prior crime of not similar crime type 4 Prior crime experienced as a victim/witness, of same crime type

Factor	Definition	Coding rules
		3 Prior crime or incident experienced as a victim/witness, of similar type 2 Prior crime or incident experienced as a victim/witness, of not similar type 1 Other activity node (all other nodes)
Location similarity (1 to 3)	Ordinal scale of similarity between activity node and reference offence[a]	3 Same location type 2 Unknown similarity 1 Different location type
Hour similarity (0 to 1)	Reciprocal of the difference in hours between reference offence hour of the day and activity hour of the day.	Home & family members' homes: any time of day. School/education: 08:00-16:00. Work: 08:00-18:00. Events: time of the event.
Day similarity (0 to 1)	Reciprocal of the difference in days between reference offence day of the week and activity day of the week.	Home & family members' homes: any day. School/education & work: Monday to Friday. Events: day of the event.
Season similarity (0 to 1)	Reciprocal of the difference in seasons between reference offence season and activity season.	Time span nodes: if time span covered the same season as reference offence, 1, else take the minimum difference in seasons to season of time span start date or end date. Events: minimum difference to event season.

[a] Location types: residential, commercial premises, public premises, or street/open spaces/transit

Second, we averaged the reliability variables (frequency, recency and duration) and relevance variables (behaviour, location type, hour, day and season similarity) to create indices representing the latent constructs of reliable and relevant knowledge. Preliminary analyses added each variable stepwise to determine whether each explained additional variance in crime location choice, resulting in the decision to retain all variables to create the reliability and relevance indices (see Supplementary Materials S1). We weighted each variable equally in the absence of a priori guidance as to the relative importance of the individual variables to reliability and relevance (a topic we note that future research could explore). In accordance with the theoretical model, activity locations that were visited frequently, recently and over a longer period of time had high reliability scores; activity locations with high similarity on multiple similarity dimensions had high relevance scores.

Third, we combined the indices into a categorical variable reflecting the interaction to be tested, based on whether the activity node was 'high' (greater or equal to the 50^{th} percentile) or 'low' (below the 50^{th} percentile) on each index. Activity locations with high reliability (familiarity) and high relevance (similarity) were coded H_fH_s,[9] those with high reliability and low relevance H_fL_s, low reliability and high relevance L_fH_s, and low reliability and low relevance L_fL_s.[10]

Last, each SA2 (in each offender's set) was coded based on the presence (1) or absence (0) of a given activity node (e.g., home, school, prior offence) of a given category (H_fH_s / H_fL_s / L_fH_s / L_fL_s). Each SA2 was then also coded based on the presence (1) or absence (0) of activity nodes outside the SA2 at a series of distance bands, being: between 0 and 200m outside the SA2 boundary, then 200-500m, 500m-1km, 1km-2km and 2km-5km. Only the activity nodes in the distance band containing the nearest activity node(s) were included. For example, if there were no activity nodes in the SA2, but there was an H_fH_s node and an L_fL_s node within 0-200m, the variables 'H_fH_s node 0-200m' and 'L_fL_s node within 0-200m' were coded 1 and all remaining variables (e.g., 'H_fH_s node in SA2', 'L_fH_s node within 200-500m' and so on) were coded 0. This approach is consistent with previous studies (e.g., Menting et al., 2020) and enables examination of patterns of distance decay over increasing distances-to-nearest-nodes, by comparison to the use of continuous distance variables. The reference category for each variable (e.g., H_fH_s node within SA2, L_fH_s node 200-500m from the SA2) is no node within 5km of the SA2.

Opportunity (Control) Variables

Opportunity variables were derived from NZ Statistics Census and Business Demography data (http://nzdotstat.stats.govt.nz/). Consistent with past DSCM research, these

---

[9] Where subscript f = familiarity/reliability and s = similarity/relevance for ease of reference.

[10] We also explored using three levels (high, medium and low) to create more granular categories (H_fH_s/H_fM_s… L_fL_s): see Supplementary Materials S1.

variables represent, directly or indirectly, the number of targets (or an indicator thereof) per SA2. For each crime type (i.e., each model), we included a single variable indicative of the target or victim distribution for that crime type.

For crimes targeting particular types of premises—residential burglary, non-residential burglary and commercial robbery---the opportunity measures directly captured the number of those types of premises per SA2 (as per Bernasco & Nieuwbeerta, 2005; Frith, 2019; Townsley et al., 2015). These were: number of households for residential burglary; number of business units in any industry for non-residential burglary; and the number of business units in commercial industries (e.g., retail, accommodation and food services) that matched the types of premises used to identify commercial robbery (the 'commercial' location types described in relation to the location similarity variable above).

For personal robbery and sex offences, we used the number of business units in commercial or public industries (as for commercial robbery plus industries such as transport, education and health care). This variable served as an indirect indicator of ambient population, which is a better measure of the target distribution than residential population, for personal robbery and other offences targeting people rather than premises (Andresen, 2011; Andresen & Jenion, 2010; Rummens et al., 2021). See Supplementary Materials S1 for further detail.

Analytic Approach

In line with most previous DSCM studies, we employed conditional logit models (McFadden, 1984). The mechanics of the conditional logit model have been described in detail elsewhere (Bernasco, 2010a), but to summarise, the model estimates the expected utility of each choice alternative (SA2), given the attributes of that alternative, and selects the alternative that maximises utility. It yields the estimated probability that an offender n will choose alternative i from the set of alternatives C, given its attributes x_{ni} and parameters β representing the effects of these attributes on the decision:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j \epsilon C} e^{\beta' x_{ni}}}$$

(1)

The β parameters are estimated by maximum likelihood estimation. Because we sampled from alternatives, using a subset *D* of the full set of alternatives *C*, the log-likelihood function is:

$$\ell = \sum_{n=1}^{N} \sum_{i \epsilon C, i \neq j} \left(y_{ni} \ln \left(\frac{e^{\beta' x_{ni} - ln(\pi(D|i))}}{\sum_{j \epsilon C} e^{\beta' x_{nj} - ln(\pi(D|j))}} \right) \right)$$

(2)

where $\pi(D|i)$ is the probability of alternative *i* to be included in *D*, and y_{in} is the observed choice: either 1 if *n* chooses *i* or 0 if they do not (see Curtis-Ham et al., 2021 for further detail).

We exponentiate the estimated coefficients (β) for the activity location and control variables to report them as odds ratios (ORs). The ORs indicate the increase/decrease in the odds of an offender choosing an SA2 given the presence of an activity node with a given attribute (e.g., activity node with high reliability and high relevance within 0-200m of the SA2) by comparison with there being no activity node within 5km of the SA2. For the opportunity covariates, the ORs represent the change in odds for each 100 unit increase in households or businesses. Odds ratios within models were compared with Wald's Chi-Square difference tests. For each crime type we tested H1 with a conditional logit model involving four reliability-relevance interaction categories (H_fH_s to L_fL_s) x 6 distance bands (within SA2 plus 5 outside) + 1 opportunity variable = 25 estimates. Descriptive statistics showing the proportion of all sampled SA2s with nodes of each category (H_fH_s to L_fL_s) in each distance band, and for the opportunity variables, are provided in Supplementary Materials S2.

Results

Overall, the results confirmed our hypothesis and are summarised in figure 2, which displays the 95% confidence intervals around the odds ratios (ORs). The exact statistics and Wald tests statistics are reported in Supplementary Materials S3 and S4, respectively. The general trend for all crime types was that offenders were more likely to commit crime near to activity nodes that were high on both reliability and relevance, lower in proximity to activity nodes that were high on only reliability or relevance and lowest in proximity to activity nodes that were low on both reliability and relevance. This trend was most pronounced in the 'within SA2' distance band. In other words, the closer to activity nodes, the greater the differentiation between nodes of different levels of reliability and relevance.

Specifically, the odds of crime location choice near high reliability high relevance (H_fH_s) nodes were statistically significantly greater than for low reliability high relevance (L_fH_s) nodes at all six distance bands for all five offences except residential burglary and the 2-2km distance band. The odds were significantly greater for H_fH_s than H_fL_s for sixteen of thirty H_fH_s vs H_fL_s comparisons (5 offences x 6 distance bands), including all comparisons for the 'within SA2' band. The odds of crime location choice near L_fH_s nodes were statistically significantly greater than near L_fL_s nodes for only nine of thirty L_fH_s vs L_fL_s comparisons.[11] In one instance the results were significant in the opposite direction to the hypothesis (for residential burglary and the 1-2km distance band, $H_fL_s > H_fH_s$). We approach exceptions to the hypothesised trend at longer distance bands with caution: the data contain a measurement error due to how activity locations and their attributes were recorded and to the assumptions involved in constructing the variables of interest. This measurement error is likely to be most consequential at greater distances, where we expect (and find) weaker associations between the presence of activity nodes and crime location choice.

[11] The ORs for L_fH_s were significantly greater than for L_fL_s for: all offences except commercial robbery in the 'within SA2' band; residential burglary in the 200-500m and 500-1000m bands; non-residential burglary in the 0-200m and 200-500m bands; and commercial robbery in the 500-1000m band.

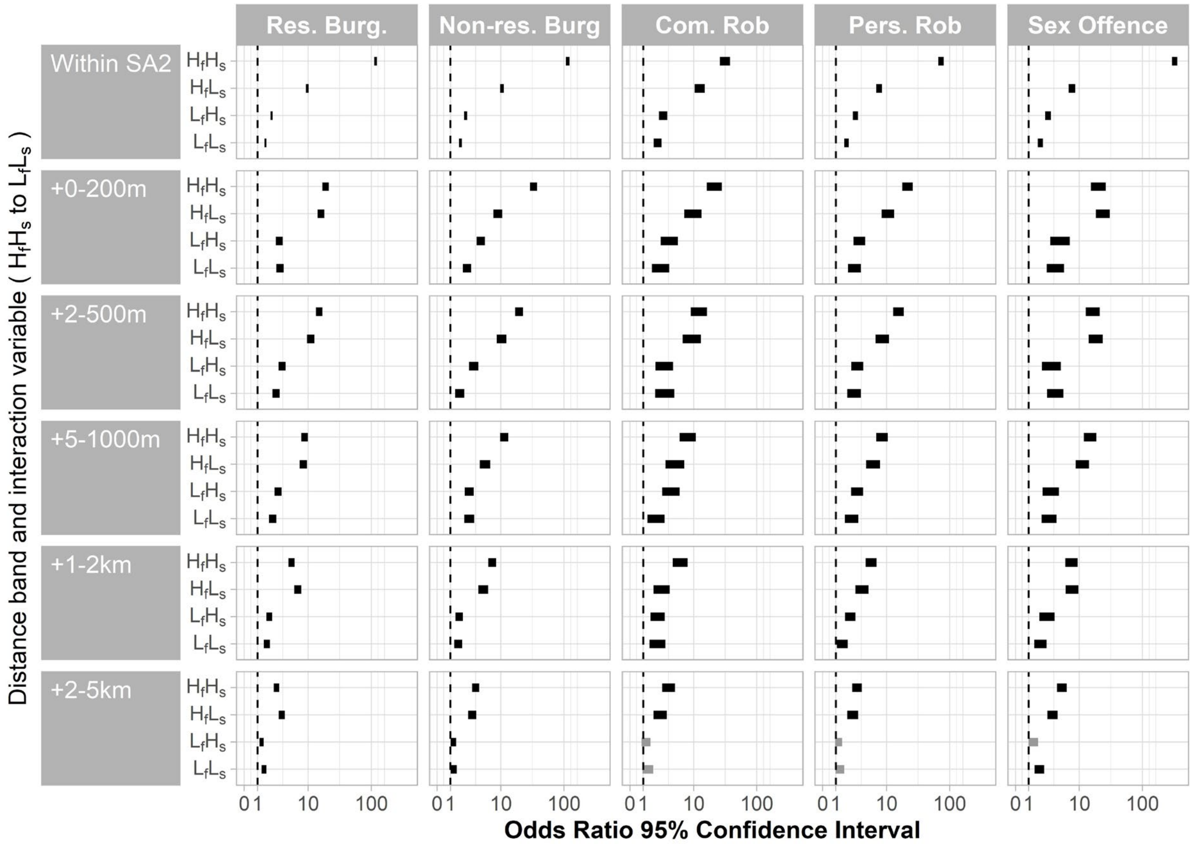


Figure 2. Odds ratio 95% confidence intervals for reliability x relevance interaction variables per distance band for each model (crime type).

Note: The shorter the bar, the smaller the CI. Non-significant associations have CIs that cross the vertical line at 1 and are grey. ORs with bars that do not overlap are significantly different. ORs with bars that overlap may be significantly different: see Wald tests results in Supplementary Materials S4. In "H_fL_s", H = high, L = low, f = familiarity/reliability and s = similarity/relevance.

Also as expected, a general pattern of distance decay, or a decline in the odds of crime location choice with increasing distance bands, is exhibited in figure 2. Compare, for example, the ORs for H_fH_s nodes (the top-most point in each box) over the distance band boxes down the residential burglary column. We likewise found, as expected, significant and positive associations between crime location choice and the opportunity covariates. For every 100 units (e.g., households, commercial premises) in the SA2, the odds of crime increased by a factor of 1.06 for residential burglary (95% CI: 1.05-1.06), 1.11 for non-residential burglary (1.10-1.11), 1.11 for commercial robbery (1.10-1.12), 1.11 for personal robbery (1.10-1.11) and 1.08 for sex offences (1.07-1.09).

Table 2 displays the model summary statistics for each crime type. The Pseudo R^2 values, ranging from 0.27 to 0.36, indicate a good fit to the data (McFadden, 1973) and are comparable with previous DSCM studies.

Table 2. Model summary statistics

Statistic	Res. Burg.	Non-Res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
N offence choices	17,054	10,353	1,977	4,315	4,421
N SA2 choice alternatives	4,213,998	2,417,735	563,161	1,243,204	784,537
Pseudo R^2	0.35	0.36	0.27	0.31	0.32

Note: SA2 = Statistical Area 2

Discussion

This study aimed to advance our ability to explain and predict where individual offenders will commit crime, by testing whether offenders are more likely to commit crime in locations where their prior activities have generated *both* reliable and relevant knowledge of crime opportunities, for a range of crime types. Our results confirmed that in general, crime was most likely at locations with which offenders were more familiar *and* near which they had conducted similar activities in the past, and least likely at locations with which offenders were not familiar *and* had not conducted similar activities. These results support the hitherto untested interaction between reliability and relevance constructs theorised in the model proposed by Curtis-Ham et al. (2020).

An alternative explanation for the findings is that people are simply more likely to be caught for crimes committed near high reliability/high relevance activity nodes than for crimes committed near low reliability or low relevance nodes. In other words, since we only studied solved crimes, the results are confounded by an unmeasured variable reflecting the rate at which police solve crimes committed near activity nodes with different levels of reliability and relevance. However, studying offences linked by DNA to both identified (caught) and unidentified offenders, Lammers (2014) found that people who offended near past crimes—high relevance nodes— were not more likely to get caught, or caught sooner, than people whose offences were farther apart. Similarly, Bernasco et al (2013) found no or weak correlations between robbery detection rates and a range of environmental features that represent common activity locations (e.g., bars, restaurants, retail outlets). Further, several studies using survey data on self-reported crimes—both detected by police and undetected—have found relationships between the frequency (Menting et al., 2020) and timing (van Sleeuwen et al., 2021) of past visits to locations and future crime locations that are consistent with studies of frequency and timing using solved crime data (Bernasco, 2019; van Sleeuwen et al., 2018). It is therefore unlikely that offenders' probability of detection covaries with reliability and relevance factors to an extent that would produce our pattern of results.

Another alternative explanation is that offenders are more likely to have high reliability/high relevance activity nodes in places with more crime opportunities and thus the results reflect the underlying opportunity structure, rather than individual offenders' activities. We controlled for the distribution of opportunities using measures that reflected the distribution of potential crime targets for each crime type. But it remains possible that other unmeasured aspects of the environment that are conducive to both crime and activities that generate reliable and relevant knowledge could explain some of the variance we attribute to those activities. Particularly with recent prior crimes of the same type—which are high reliability and high relevance—it may be difficult to disentangle the effect of existing opportunities from the effect of knowledge gained from previously exploiting those opportunities. However, research on repeat and near repeat offending supports our interpretation of the results: a 'boost' effect—from the highly reliable and relevant knowledge gained from a recent crime—explains offenders' tendency return to prior crime locations beyond the effect of opportunity structure alone (Bowers & Johnson, 2004; Lantz & Ruback, 2017).

The finding of an interaction between reliability and relevance builds on prior crime location choice studies that employed discrete spatial choice models to study individual reliability and relevance variables in isolation (e.g., Bernasco, 2019; Ruiter, 2017; van Sleeuwen et al., 2021), by demonstrating how these variables operate collectively. This interaction had only been hinted at by previous findings that offenders are much more likely to commit crime near their current homes—a high reliability, high timing similarity activity node—(Bernasco, 2010b; Lammers et al., 2015) and places they have committed multiple or recent offences of the same type and/or at the same period of the week (Lammers et al., 2015; van Sleeuwen et al., 2018), than other activity nodes.

The pattern of declining likelihood of crime commission with increasing distance—clearest for the high reliability/high relevance activity locations—is also consistent with previous studies (Bernasco, 2019; Bernasco & Block, 2009; Kuralarasan & Bernasco, 2021; Menting et al., 2020). However, a novel finding is that this distance decay pattern was flatter and not always monotonic in relation to activity locations lower on either or both dimensions, implying that people are just as unlikely to commit crime close to low reliability or low relevance activity locations as farther away from them (but still within 5km).

That the opportunity variables were also associated with crime location choice when controlling for the presence and attributes of offenders' activity locations is also consistent with previous literature. Although past DSCM studies have often found associations with opportunity variables, they have typically only included offenders' homes and one or two other nodes (e.g., family members' homes, prior crimes), enabling only partial control for offenders' activity locations (e.g., Lammers et al., 2015; Menting et al., 2016; van Sleeuwen et al., 2018). However, similar opportunity effects were also observed by Bernasco (2019) and Menting et al. (2020) when controlling for a wide range of activity locations. That people offend where there are more opportunities regardless of whether they have an activity location nearby could reflect activity locations missing from the data, awareness of opportunities generated through processes other than direct experience, or offending outside of awareness space (Curtis-Ham et al., 2020).

The above findings broaden the empirical evidence base for crime pattern theory (P. L. Brantingham & Brantingham, 1991, 1993a) and its recent extension in the model proposed by Curtis-Ham et al. (2020) in a number of ways. First and foremost, by including all attributes of activity locations from the model—previously only studied in isolation— our findings provide the first direct evidence of the theorised interaction between reliable and relevant location knowledge in offenders' crime location choices. This finding highlights the importance of considering both the quantitative (reliability) and qualitative (relevance) attributes of offenders' prior activity locations when examining their crime location choices in research or practice.

Second, the findings demonstrate the salience of a broad range of activity locations to offenders' crime location choices. We included a variety of different activity locations and observed strong associations observed between proximity to activity locations—particularly those with high reliability and relevance—and offenders' crime locations. Much of the individual-level research on crime pattern theory has focused on offenders' current homes as a proxy for their awareness space (Ackerman & Rossmo, 2015; Ruiter, 2017; Townsley, 2016). However, the present study corroborates—on a much larger scale—previous small-scale and qualitative studies that suggested that many activity locations other than home can play an important role generating awareness of crime opportunities (e.g., Bernasco, 2019; Costello & Wiles, 2001; Davies & Dale, 1996; Menting et al., 2020; Wright & Decker, 1994), consistent with the original formulation of crime pattern theory by the Brantinghams (1991, 1993a).

Third, by including a range of crimes reflecting different types of targets and motivations and analysing them separately, this study lends support to the generalizability of both crime pattern theory and Curtis-Ham et al.'s extension to different crimes. Previous studies of the connection between offenders' activity locations and their crime locations have tended to study a wide range of crimes in combination (e.g., Lammers et al., 2015; Menting et al., 2016) or to focus on a single crime type (e.g., Frith et al., 2017; Kuralarasan & Bernasco, 2021; Long et al., 2018; c.f. Bernasco, 2010b; Frith, 2019). The former studies are likely to have masked crime specific patterns, with the results simply reflecting high volume crime types; the latter studies leave questions about the generalizability of their findings across crime types. The present study overcame both these limitations. Importantly, the consistency of the observed interaction of reliability and relevance across crime types indicates that it likely generalises beyond the crimes included in this study. That said, we recommend future studies expand the range of crimes included, to test this suggestion empirically. The consistency across crimes also supports the generalizability of the results beyond the analysed samples, which represented different proportions of burglary, robbery and extra-familial sex offences due to their different detection rates.[12] The use of a dataset from a hitherto understudied part of the world also adds support to the generalizability of crime pattern theory, and its recent extension, across crime behaviours and contexts.

An unexpected finding of potential theoretical significance was the smaller difference between L_fH_s and L_fL_s activity nodes than between H_fL_s and L_fL_s activity nodes. The lack of distinction between low reliability/high relevance and low reliability/low relevance activity locations indicates that if offenders are less familiar with a location, the nature of their previous activities there matters less. In other words, there appears a notional minimum amount of reliable location knowledge below which relevance does not matter. Conversely, once a threshold level of familiarity is reached, the behaviour, type of location and timing involved in offenders' prior activities in a location make a greater difference to whether they ultimately choose to offend nearby. This interpretation is speculative, however, because this result could be an artefact of the data and methods employed in this study.

---

[12] We had no data on unsolved crimes to calculate detection rates specific to the crimes we analysed, but official statistics (at policedata.nz) for 2009-2018 show that at 180 days after being reported police, approximately 9% of burglaries, 30% of robberies and 22% of sex offences had resulted in either court or non-court action against an identified offender.

Our results should be interpreted in light of several key limitations of the study. First, our measures of reliability and relevance—constructed by combining attributes of offenders' activity locations—were proxies for offenders' knowledge; we did not assess their knowledge directly. Although it is common in discrete choice research to analyse people's preferences through behavioural proxies for their internal decision-making processes (Bernasco, 2017), the words 'knowledge' and 'familiarity' as used in this paper should be interpreted as short-hand for 'knowledge/familiarity as inferred from observed behaviour'. Second, the nature of the present dataset required us to apply some broad-brush assumptions during coding of reliability and relevance variables, rather than measure the variables directly. For example, we used the average frequency and timing of visits to family members (from a time-use survey), because the data did not measure individual visits, and applied a simple scale for behavioural similarity, because the data lacked detail about the activities conducted at some activity locations (e.g., the 'other police contacts'). Considering these sources of error, in drawing conclusions we put more emphasis on the general pattern of interaction, seen consistently across the crime types, than the specific effect sizes observed.

Further research could supplement the present findings using complementary methods that overcome the above limitations. For example, in-depth interviews could be used to investigate offenders' perceived familiarity with locations and similarity of prior activities to a given crime, and the relative roles of these constructs in the decision where to commit that crime. Surveys could be used to capture the timing, location and behaviour involved in offenders' pre-offence activities with greater precision and detail, and to advance on recent survey based DSCM studies by explicitly including the reliability and relevance variables, though sample sizes would need to be increased considerably (Bernasco, 2019; Menting et al., 2020; van Sleeuwen et al., 2021). Given more detailed descriptions of the behaviour involved in offenders' reference offences and prior activities, similarity could be measured on a range of dimensions (see Farrell, 2015), or through text analysis (e.g., Birks et al., 2020; Kuang et al., 2017). Our interpretation of the observed interaction of reliability and relevance discussed above could also be corroborated by future research including a wider range of opportunity variables to better isolate the effects of opportunities versus activities, examining other crime types, and seeing if the threshold effect we observed appears in other jurisdictions.

Another avenue for future research in continuing the empirical investigation of Curtis-Ham et al.'s (2020) model, would be to consider different subgroups of offenders. For example, separating sex offenders by age of or relationship to victim might reveal differences masked in the present dataset (which did not include such victim information). Given individual differences in risk tolerance (Palminteri & Chevallier, 2018), and differences in criminal expertise and offending styles (P. J. Brantingham et al., 2020; Nee, 2015; Sanders et al., 2016), it is plausible that some offenders would prefer to commit crime in very familiar places where their prior activities have generated readily transferrable knowledge, while others would be more likely to seize crime opportunities despite a relative lack of reliable and relevant knowledge of the location.

We conclude by highlighting the practical implications of the present study and future research on this subject. Understanding the interaction of reliability and relevance is important for informing predictions about where individual offenders are likely to offend, which in turn can inform crime prevention and investigation strategies. Our results suggest that both reliability and relevance factors should be taken into account when making such predictions about where individuals will offend. People are most likely to offend in places that they know well—as indicated, collectively, by the frequency, recency and duration of their activities in those places—*and* that their activities have been most conducive to acquiring relevant knowledge of crime opportunities—as indicated by the similarity in behaviour, type of location and timing involved. Practitioners working with offenders in the community could apply these factors to identify offenders' highest risk locations for re-offending and devise plans to manage the risk such locations present. Practitioners conducting geographic profiling (Rossmo, 2000) in police investigations could apply these factors to assess suspects' likelihood of having committed a crime given the location and attributes their nearby activity locations, and prioritise the suspects accordingly.

In sum, by simultaneously accounting for the various activity location attributes theorised to affect crime location choice and using a large administrative dataset with a wide array of activity locations, we provide the first empirical evidence of the joint importance of both reliable and relevant knowledge to offenders' crime location choices. Our findings add to the empirical evidence base for crime pattern theory and its recent extension by Curtis-Ham et al. (2020). By advancing our understanding of the kinds of places that people are more likely to commit crime, this research enhances our ability to predict where individuals will offend, with important implications for crime prevention and investigation.

References

Ackerman, J. M., & Rossmo, D. K. (2015). How far to travel? A multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology*, *31*(2), 237–262. https://doi.org/10.1007/s10940-014-9232-7

Andresen, M. A. (2011). The ambient population and crime analysis. *The Professional Geographer*, *63*(2), 193–212. https://doi.org/10.1080/00330124.2010.547151

Andresen, M. A., & Jenion, G. W. (2010). Ambient populations and the calculation of crime rates and risk. *Security Journal*, *23*(2), 114–133. https://doi.org/10.1057/sj.2008.1

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.

Bernasco, W. (2010a). Modeling micro-level crime location choice: Application of the discrete choice framework to crime at places. *Journal of Quantitative Criminology*, *26*(1), 113–138. https://doi.org/10.1007/s10940-009-9086-6

Bernasco, W. (2010b). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology*, *48*(2), 389–416. https://doi.org/10.1111/j.1745-9125.2010.00190.x

Bernasco, W. (2017). Modeling offender decision making with secondary data. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook on offender decision making* (pp. 569–586). Oxford University Press. http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199338801.001.0001/oxfordhb-9780199338801-e-28

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLOS ONE*, *14*(1), e0210733. https://doi.org/10.1371/journal.pone.0210733

Bernasco, W., & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, *47*(1), 93–130. https://doi.org/10.1111/j.1745-9125.2009.00140.x

Bernasco, W., Johnson, S. D., & Ruiter, S. (2015). Learning where to offend: Effects of past on future burglary locations. *Applied Geography*, *60*(Supplement C), 120–129. https://doi.org/10.1016/j.apgeog.2015.03.014

Bernasco, W., & Kooistra, T. (2010). Effects of residential history on commercial robbers' crime location choices. *European Journal of Criminology*, *7*(4), 251–265. https://doi.org/10.1177/1477370810363372

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *The British Journal of Criminology*, *45*(3), 296–315. https://doi.org/10.1093/bjc/azh070

Bernasco, W., Ruiter, S., & Block, R. (2017). Do street robbery location choices vary over time of day or day of week? A test in Chicago. *Journal of Research in Crime and Delinquency*, *54*(2), 244–275. https://doi.org/10.1177/0022427816680681

Birks, D., Coleman, A., & Jackson, D. (2020). Unsupervised identification of crime problems from police free-text data. *Crime Science*, *9*(1), 1–19. https://doi.org/10.1186/s40163-020-00127-4

Bowers, K., & Johnson, S. (2004). Who commits near repeats? A test of the boost explanation. *Western Criminology Review*, *5*(3), 12–24.

Brantingham, P. J., Brantingham, P. L., Song, J., & Spicer, V. (2020). Crime hot spots, crime corridors and the journey to crime: An expanded theoretical model of the generation of crime concentrations. In K. M. Lersch & J. Chakraborty (Eds.), *Geographies of Behavioural Health, Crime, and Disorder* (Vol. 126, pp. 61–86). Springer. https://doi.org/10.1007/978-3-030-33467-3_4

Brantingham, P. L., & Brantingham, P. J. (1991). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (2nd ed., pp. 27–54). Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993a). Environment, routine, and situation: Toward a pattern theory of crime. In R. V. Clarke & M. Felson (Eds.), *Routine activity and rational choice* (pp. 259–294). Transaction Publishers.

Brantingham, P. L., & Brantingham, P. J. (1993b). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Bruinsma, G. J. N., & Johnson, S. D. (Eds.). (2018). *The Oxford handbook of environmental criminology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.38

Clare, J., Fernandez, J., & Morgan, F. (2009). Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. *Australian & New Zealand Journal of Criminology*, *42*(2), 139–158. https://doi.org/10.1375/acri.42.2.139

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608. https://doi.org/10.2307/2094589

Costello, A., & Wiles, P. (2001). GIS and the journey to crime: An analysis of patterns in South Yorkshire. In K. J. Bowers & A. Hirschfield (Eds.), *Mapping and analysing crime data: Lessons from research and practice* (pp. 27–60). Taylor & Francis.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). *Relationships between offenders' crime locations and different prior activity locations as recorded in police data*.

Davies, A., & Dale, A. (1996). Locating the stranger rapist. *Medicine, Science and the Law*, *36*(2), 146–156. https://doi.org/10.1177/002580249603600210

DeLisi, M., & Piquero, A. R. (2011). New frontiers in criminal careers research, 2000–2011: A state-of-the-art review. *Journal of Criminal Justice*, *39*(4), 289–301. https://doi.org/10.1016/j.jcrimjus.2011.05.001

Farrell, G. (2015). Crime concentration theory. *Crime Prevention & Community Safety*, *17*(4), 233–248. i3h. http://search.ebscohost.com/login.aspx?direct=true&db=i3h&AN=110338577&site=ehost-live

Frith, M. J. (2019). Modelling taste heterogeneity regarding offence location choices. *Journal of Choice Modelling*, *33*, 100187. https://doi.org/doi.org/10.1016/j.jocm.2019.100187

Frith, M. J., Johnson, S. D., & Fry, H. M. (2017). Role of the street network in burglars' spatial decision-making. *Criminology*, *55*(2), 344–376. https://doi.org/10.1111/1745-9125.12133

Hillier, A., Smith, T., Cannuscio, C. C., Karpyn, A., & Glanz, K. (2015). A discrete choice approach to modeling food store access. *Environment and Planning. B, Planning & Design.*, *42*(2), 263–278. https://doi.org/10.1068/b39136

Kuang, D., Brantingham, P. J., & Bertozzi, A. (2017). Crime topic modeling. *Crime Science*, *6*(1), 1–20. https://doi.org/10.1186/s40163-017-0074-0

Kuralarasan, K., & Bernasco, W. (2021). Location choice of snatching offenders in Chennai city. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09514-9

Lammers, M. (2014). Are arrested and non-arrested serial offenders different? A test of spatial offending patterns using DNA found at crime scenes. *Journal of Research in Crime and Delinquency*, *51*(2), 143–167. https://doi.org/10.1177/0022427813504097

Lammers, M. (2018). Co-offenders' crime location choice: Do co-offending groups commit crimes in their shared awareness space? *The British Journal of Criminology*, *58*, 1193–1211. https://doi.org/10.1093/bjc/azx069

Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology*, *53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071

Lantz, B., & Ruback, R. B. (2017). A networked boost: Burglary co-offending and repeat victimization using a network approach. *Crime & Delinquency*, *63*(9), 1066–1090. https://doi.org/10.1177/0011128715597695

Long, D., Liu, L., Feng, J., & Zhou, S. (2018). Assessing the influence of prior on subsequent street robbery location choices: A case study in ZG city, China. *Sustainability*, *10*(6), 1818. https://doi.org/10.3390/su10061818

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1977). *Modelling the choice of residential location* (No. 477; Cowles Foundation Discussion Papers). Yale University. https://EconPapers.repec.org/RePEc:cwl:cwldpp:477

McFadden, D. (1984). Econometric analysis of qualitative response models. In P. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 105–142). Elsevier. https://doi.org/10.1016/S1573-4412(84)02016-X

Menting, B. (2018). Awareness × opportunity: Testing interactions between activity nodes and criminal opportunity in predicting crime location choice. *The British Journal of Criminology*, *58*, 1171–1192. https://doi.org/10.1093/bjc/azx049

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2016). Family matters: Effects of family members' residential areas on crime location choice. *Criminology*, *54*(3), 413–433. https://doi.org/10.1111/1745-9125.12109

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: Findings from an online self-report survey. *The British Journal of Criminology*, *60*(2), 303–322. https://doi.org/10.1093/bjc/azz044

Nee, C. (2015). Understanding expertise in burglars: From pre-conscious scanning to action and beyond. *Aggression and Violent Behavior*, *20*(Supplement C), 53–61. https://doi.org/10.1016/j.avb.2014.12.006

Palminteri, S., & Chevallier, C. (2018). Can we infer inter-individual differences in risk-taking from behavioral tasks? *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02307

Rossmo, D. K. (2000). *Geographic profiling*. CRC Press.

Ruiter, S. (2017). Crime location choice. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook of offender decision making* (pp. 398–420). Oxford University Press.

Rummens, A., Snaphaan, T., Van de Weghe, N., Van den Poel, D., Pauwels, L. J. R., & Hardyns, W. (2021). Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? *ISPRS International Journal of Geo-Information*, *10*(6), 369. https://doi.org/10.3390/ijgi10060369

Sanders, A. N., Kuhns, J. B., & Blevins, K. R. (2016). Exploring and understanding differences between deliberate and impulsive male and female burglars. *Crime & Delinquency*, *63*(12), 1547–1571. https://doi.org/10.1177/0011128716660519

Townsley, M. (2016). Offender mobility. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (pp. 142–161). Routledge.

Townsley, M., Birks, D., Bernasco, W., Ruiter, S., Johnson, S. D., White, G., & Baum, S. (2015). Burglar target selection: A cross-national comparison. *Journal of Research in Crime and Delinquency*, *52*(1), 3–31. https://doi.org/10.1177/0022427814541447

van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A time for a crime: Temporal aspects of repeat offenders' crime location choices. *Journal of Research in Crime and Delinquency*, *55*(4), 538–568. https://doi.org/10.1177/0022427818766395

van Sleeuwen, S. E. M., Ruiter, S., & Steenbeek, W. (2021). Right place, right time? Making crime pattern theory time-specific. *Crime Science*, *10*(1), 1–10. https://doi.org/10.1186/s40163-021-00139-8

Weisburd, D., Eck, J. E., Braga, A. A., Telep, C. W., Cave, B., Bowers, K., Bruinsma, G. J. N., Gill, C., Groff, E. R., Hibdon, J., Hinkle, J. C., Johnson, S. D., Lawton, B., Lum, C., Ratcliffe, J. H., Rengert, G., Taniguchi, T., & Yang, S.-M. (2016). *Place matters: Criminology for the twenty-first century*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342087

Wright, R., & Decker, S. H. (1994). *Burglars on the job: Streetlife and residential break-ins*. Northeastern University Press.

CHAPTER 7

A New Geographic Profiling Method for Mapping and Ranking Suspects

Equipped with an empirically supported theoretical model systematising the links between the activity locations in offenders' mental maps and their crime locations, the research presented in this chapter returned to the practical problem formulated at the start of this thesis: how to infer who committed a crime from where it was committed. More specifically, how information about the mental maps of potential suspects—the location and nature of their known activity locations—can be used to identify which of those suspects is more likely to have offended at the location of the crime. The final study of this thesis applied the theoretical model in an algorithmic solution to this problem—the Geographic Profiling Suspect Mapping and Ranking Technique (GP-SMART). The following manuscript, subsequently published in the *Journal of Investigative Psychology and Offender Profiling*, describes GP-SMART and the results of tests of its accuracy in placing the actual offender among the top ranked suspects.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART. *Journal of Investigative Psychology and Offender Profiling*. https://doi.org/10.1002/jip.1585

A New Geographic Profiling Suspect Mapping And Ranking Technique for Crime Investigations: GP-SMART

Abstract

This study developed and tested a new geographic profiling method for automating suspect prioritisation in crime investigations. The Geographic Profiling Suspect Mapping And Ranking Technique (GP-SMART) maps suspects' activity locations available in police records—such as home addresses, family members' home addresses, prior offence locations, locations of non-crime incidents, and other contacts with police—and ranks suspects based on both the proximity and nature of these locations, relative to an input crime. In accuracy tests using solved burglary, robbery and extra-familial sex offence cases (n=4,511), GP-SMART ranked the offender at or near the top of the suspect list at rates greatly exceeding chance. Highlighting the benefit of its novel inclusion and differentiation of many different types of activity location, GP-SMART also outperformed baseline methods—approximating existing algorithms—that ranked suspects using only the proximity of their activity locations, or home addresses, to the input crime.

Keywords

Crime investigation, crime location choice, geographic profiling, police data, suspect prioritisation

Introduction

A common problem in crime investigations is to prioritise among the many possible suspects who could have committed the crime. Using a collection of analytic methods called 'geographic profiling' (Rossmo, 2000)—or in the UK, geographical profiling (Canter, 2003)—specialist police analysts can examine information in the geography of crimes and suspects to inform investigative activities, including prioritising suspects. This information is particularly telling because people tend to commit crime near places they know from their everyday activities such as where they live, work or socialise—their 'activity nodes' (P. L. Brantingham & Brantingham, 1991, 1993).

Often, geographic profiling involves analysing the locations of a series of crimes believed to have been committed (e.g., linked by forensic or behavioural evidence) by the same unknown offender, to predict where the offender might have an activity node (Knabe-Nicol & Alison, 2011; Rossmo, 2000). Following this crime-based prediction, suspects can be ranked by the proximity of their known activity nodes to the area predicted as most likely to contain an activity node (e.g., Canter & Hammond, 2007) or the prediction score at their known activity nodes (Rossmo & Velarde, 2008). Algorithms automating this prediction process exist, with Rigel being the most widely used by analysts (Emeno et al., 2016).

A complementary approach to the crime-based one focuses on the suspects, examining their known activity nodes to determine who among them is more likely to have committed a crime in the location of any given crime being investigated. Beneficially, this suspect-based approach can be applied to single crimes; it does not require a series of offences believed to have been committed by the same offender.[1] However, in the absence of a pre-filtered list of specific suspects for a given crime, the potential suspect pool is vast: theoretically, every individual recorded in a police service's crime and intelligence systems. In this situation, manual searches and comparisons between suspects are impossible; automated solutions are necessary. Yet there have been few attempts to automate this suspect-based approach in algorithms that could be used by analysts: we identified only five in the published literature (Bache et al., 2008; Frank, 2012; Gore et al., 2005; Snook et al., 2006; Tayebi et al., 2017) and failed to identify evidence of their use in practice from surveys and interviews of analysts (Emeno et al., 2016; Knabe, 2008; Knabe-Nicol & Alison, 2011; Öhrn, 2019).[2]

---

[1] For crime series, the suspect-based approach could be applied to each crime individually and the output predictions aggregated such that suspects more likely to have committed multiple crimes in the series are prioritised; we also discuss this point briefly in the Discussion section.

[2] We focus on algorithms whose accuracy had been assessed and published, given that establishing accuracy is a critical step prior to any implementation in practice (Rich & Shively, 2004). Studies describing algorithms without providing evidence of their accuracy were excluded (e.g., Ding et al., 2009).

Into this gap we introduce a new method, GP-SMART (Geographic Profiling: Suspect Mapping And Ranking Technique), which constitutes, to our knowledge, the most comprehensive attempt yet to map and compare suspects' activity nodes in a suspect-based geographic profiling algorithm. In this paper, we describe GP-SMART and test its predictive accuracy in ranking suspects (i.e., where do we find the actual offender in the ranked list of suspects). We first review existing suspect-based algorithms and highlight limitations in their methods and in the way they have been tested both of which we address in the present study.

Suspect-Based Geographic Profiling Algorithms

Existing suspect-based geographic profiling algorithms vary in sophistication, but they all calculate the distance between suspects' activity nodes (e.g., known home address, prior offence locations) and an input crime, then rank suspects such that those with nodes closer to the crime rank higher than those with nodes farther from the crime. The logic is grounded in theory—and a common empirical finding: people are more likely to commit crime closer to their activity nodes than farther away (Bernasco, 2019; P. L. Brantingham & Brantingham, 1991; Menting et al., 2020; Ruiter, 2017). People with nodes close to the crime are therefore more likely to have committed it than people with nodes farther away. The most basic algorithm (Snook et al., 2006) merely ranks suspects by the distance between a single node—their home address—and the input crime. Two others also use only home nodes but apply a distance decay function to estimate a probability of offending at the distance between the suspect's home and the input crime, which informs the suspect rankings (Bache et al., 2008; Gore et al., 2005).[3] Gore et al. applied distance decay curves derived empirically for each input crime from the other crimes in the dataset but they acknowledge that because the curve (for all crimes in the dataset) is highest in the shortest distance interval (0-250m), the outcome is "much the same" (p134) as ranking the suspects by distance (as per Snook et al.,

[3] We refer here to Gore et al.'s 'Incident Based Distance Decay Curve filter'. The other two suspect prioritisation methods they report are not suspect-based but instead first predict the offender's home location from the location of a single crime, then rank suspects by the prediction score at their home's location—a process equivalent to the that described above for crime series'.

2006). Bache et al. tested negative exponential and power decay functions, which produced similar accuracy results. Their algorithm additionally adjusts suspects' predicted probability based on their number of prior offences before ranking the suspects, on the grounds that suspects with more prior offences are more likely to have committed the crime. Both Frank (2012) and Tayebi et al. (2017) first estimate each suspect's activity space—the area around and between their nodes, which included home, prior offence locations, co-offenders' homes, and places likely to be commonly frequented by many offenders, such as malls and transit hubs—then rank suspects by the shortest distance from the input crime to this activity space. Like Bache et al., Tayebi et al. additionally adjust the ranks based on suspects' prior offences, in this case ranking higher suspects who have previously committed the same type of offence as the input crime.[4]

We address two limitations of these algorithms. First, they include only a small subset of suspects' activity nodes, but studies have shown that people are more likely to commit crime near any activity node, than farther away (providing crime opportunities are present). These activity nodes include both present and past homes and those of their family members or co-offenders, places of work, where they go or went to school, prior crime locations, and places they purchase drugs or fence stolen goods (Bernasco, 2019; Lammers, 2018; Menting et al., 2016; Pettiway, 1995; Rengert & Wasilchick, 1985). Logically, including more activity nodes in the algorithm should increase the likelihood of identifying one or more of the offender's nodes in close proximity to the input crime and ranking them highly accordingly. Further, a wide range of activity nodes that could be fed to the algorithm are likely to be recorded in police databases. Curtis-Ham et al. (2021a) found that burglary, robbery and extra-familial sex offenders typically had at least 10 activity nodes recorded in a national police database, which included home addresses, family members' home addresses, prior crime locations, places where they had been victims of or witnesses to crime, locations of

[4] Other algorithms exist that use suspects' activity node information to predict the location of their next offence, but do not include the step of ranking those suspects by comparing the prediction to an unsolved crime (e.g., Duan et al., 2017). However, algorithms that lack the suspect prioritisation step are not considered herein.

non-crime incidents police had been called to, and places of arrests, stop and search and other interactions with police.

Second, the algorithms that include different activity nodes treat them equally, and thus ignore relevant variability between activity nodes. People are more likely to commit crime near some activity nodes, than others (holding distance equal). Specifically, people are more likely to offend near activity locations they have visited more frequently, more recently, over a longer period of time, and where their activities were more similar to the present offence (Bernasco, 2019; Curtis-Ham et al., under review; Lammers et al., 2015; Menting et al., 2016, 2020; van Sleeuwen et al., 2018). These are places that are more familiar to the offender, and where their prior activities, being similar to the present offence, are more likely to have generated knowledge of good opportunities for that offence (Curtis-Ham et al., 2020). Logically, differentiating between activity nodes in the algorithm—by ranking suspects with high familiarity, high similarity activity nodes near the input crime higher than suspects with less crime-conducive nodes near the input crime—should increase the likelihood of ranking the offender highly.

The accuracy of the existing algorithms, in identifying and ranking the offender highly, has typically been evaluated by running the algorithm for a sample of solved input crimes, and a sample of potential suspects, and seeing where the actual offender ranked (Frank, 2012; Snook et al., 2006; Tayebi et al., 2017). However, it is difficult to interpret and compare these studies' results due to differences and limitations in their selection of suspect samples. First, Gore et al., Bache et al., and Frank sampled a relatively small number of suspects including the offender (20[5], 83[6] and 322 suspects, respectively). Their results thus do not reliably indicate how highly the offender would rank in actual investigations where the algorithm would have to prioritise among many thousands of suspects.

Second, all studies except Tayebi et al. appear to have included suspect activity nodes using records from any time in their data period, including activity nodes that were recorded after the input crime.[7] For example, Frank (2012) included suspects' home address and the locations of all their offences during the data period except one offence randomly held back for each suspect to be used as an input crime to test, thus offences committed after the input crime would have been used for prioritising suspects. Further, offenders' home addresses are often not known at the point of investigation and recorded after they are caught (in New Zealand, at least), but they could appear in research data as 'address at offence date' despite having been recorded retrospectively. These studies' results therefore do not reliably indicate how high the offender would rank in actual investigations where the algorithm could only use suspect activity nodes on record at the time of the input offence, as advocated for studying geographic profiling algorithms' accuracy (Goodwill et al., 2014; Rich & Shively, 2004; Rossmo, 2015).

---

[5] For each of 101 input offences tested, 19 suspects were randomly sampled in addition to the actual offender to create a list of 20 suspects to rank, and 25 samples were repeated for each input offence. This method enabled the authors to identify whether the offender ranked higher on the lists on average than if ranked randomly, according to the aim of the study, but their average ranks are not comparable with studies that employ much larger—and more realistic—suspect samples.

[6] For the dataset of burglary offences tested using offenders' known addresses. Other crime datasets the authors tested did not include the offenders' home addresses, which were instead estimated using a geographic profiling algorithm to predict likely home location, introducing considerable error.

[7] We base this conclusion on the lack of evidence in these studies that suspect nodes post-dating the input the offence were excluded. If they had limited the suspect nodes for each input offence to only those that pre-date that offence, there would have been a different number of suspects ranked for each input crime tested. However, these studies report a single number of suspects, rather than a distribution over the input crimes. In contrast, Tayebi et al. specify that they constructed suspects' activity spaces using data from 54 months preceding the 6-month period in which the tested input crimes were committed.

The Present Study

Building on previous suspect-based geographic profiling algorithms, we developed a new method (GP-SMART) that uses a wider range of suspect activity nodes, from police data, to identify and rank suspects for a given crime. We also investigated whether differentiating between suspects' activity nodes, using theoretically salient attributes, would lead to more accurate suspect rankings than simply ranking suspects by the distance from their nearest activity node to the input crime. Further, we assessed accuracy via more ecologically valid methods than previous studies by using (a) a pool of 16,000 potential suspects who had committed a range of offences and (b) only their activity nodes known to police at the time of the given input crime.

Method

In this section we first describe the data used in this study, then the GP-SMART process and the method used to test its accuracy. Both the GP-SMART process and accuracy tests were programmed in the open-source software R (R Core Team, 2013). An R package ('gpsmartr') implementing GP-SMART is online: https://github.com/Sophie-c-h/gpsmartr.

Data

The data used in this study were extracted from the New Zealand Police National Intelligence Application (NIA). They comprise offenders who committed[8] a residential burglary, non-residential burglary, commercial robbery, personal robbery, or extra-familial sex offence between 2009 and 2018 inclusive; the location and details of their latest offence (of each type); and details of other activity locations. Data cleaning of offences led to removal of approximately 3% of offence records due to missing or imprecise location or timing information, leaving n=60,229 offenders.[9]

[8] For whom police had sufficient evidence to support a decision to prosecute, regardless of how the case was proceeded with (e.g., prosecution or out of court proceeding such as a warning).

[9] Of these, 11,459 appeared in two of the offence categories studied separately in this study (e.g., both residential burglary and commercial robbery), 2540 appeared in three, 424 appeared in four, and 27 appeared in all five offence categories.

The activity locations included: home addresses; home addresses of family members (immediate family—parents, children or siblings, current or former intimate partners and other relatives); schools and other educational institutions attended (rarely on record); workplaces (more rarely on record); offence locations; locations where they had experienced crime as a victim or witness; locations of non-crime incidents attended by police (e.g., disorderly behaviour, mental distress, civil disputes); and other police contact locations (e.g., stop and search, intelligence notings, arrests). These locations could reach as far back as the offender's birth date, except for offences, victim/witness events and incidents, which dated from 2004 (due to limited back capture of these records in NIA). Home addresses that are only discovered and recorded after an offender is identified for an offence are typically not backdated in the NIA database, which minimises the risk of including addresses that were not known at the time of investigation, as discussed above. Data cleaning of activity locations involved removal of approximately 12% of activity location records where: location or timing information was missing or imprecise; preliminary checks showed the prior offence or incident category did not reliably indicate that the person was present at the location of the offence/incident; and the offender was no longer in the offence data following data cleaning of offences. The number of remaining activity locations recorded for each offender as at the date of their latest offence ranged from none to many hundreds, with most offenders having multiple pre-offence activity locations in the dataset.

For each crime type we randomly split the offenders into two samples. The first—'calibration'—sample was used to identify values representing the relative likelihood of people offending near activity nodes with different attributes, to use in GP-SMART. The second—'test'—sample was used to test the accuracy of the algorithm in ranking the offender among the top suspects.[10] Separating these analyses prevented the risk of 'overfitting' the likelihood values to the data, which would reduce the algorithm's

[10] To ensure complete independence of the two samples, we removed from the test sample (a) offenders who appeared in any calibration sample for another crime type, and (b) offenders whose co-offenders (people with whom they committed their latest offence) appeared in the calibration sample.

generalisability to new data and produce inflated estimates of its accuracy (James et al., 2013).

The GP-SMART Method

Figure 1 depicts a high-level summary of the GP-SMART process that maps and ranks suspects, for a given input offence and list of possible suspects. Here the initial input set of *possible* suspects was all offenders in the test data, which means people who had committed at least one of the offence types included in this research. Although it would have been more ecologically valid to include a wider set of suspects, we were limited to using the data obtained for the programme of research of which this study is one part. However, by including as possible suspects people who had committed a range of offence types (burglary, robbery and sex offences) and not necessarily the input offence type, we recreated some of the noise that would be present with a wider set of possible suspects.

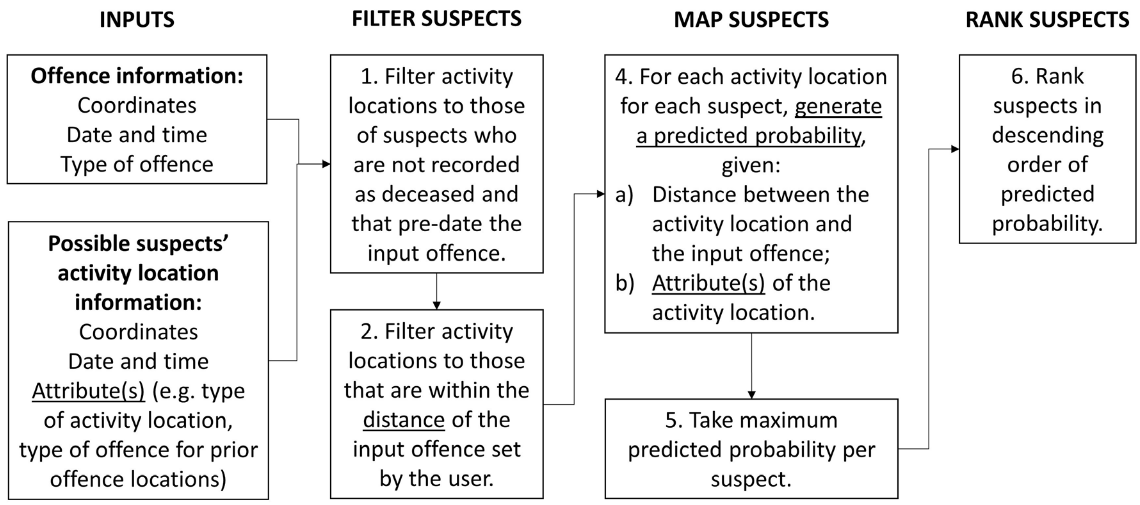


Figure 1. GP-SMART process summary

The three underlined parameters in Figure 1 represent configurable elements of GP-SMART dependent on decisions about: which activity nodes attributes to use (and how to measure them), how far to the search for activity nodes ‘near’ the input offence, and the method for predicting the probability of each suspect offending at the input offence location. In this study we used the activity node attributes theorised by Curtis-Ham et al. (2020) to influence which activity nodes offenders are more or less likely to commit crime nearby.

Figure 2 describes the attributes, which were calculated relative to the input crime, such as how recently the activity node was visited prior to the input crime, or whether the activity involved the same offence as the input crime.

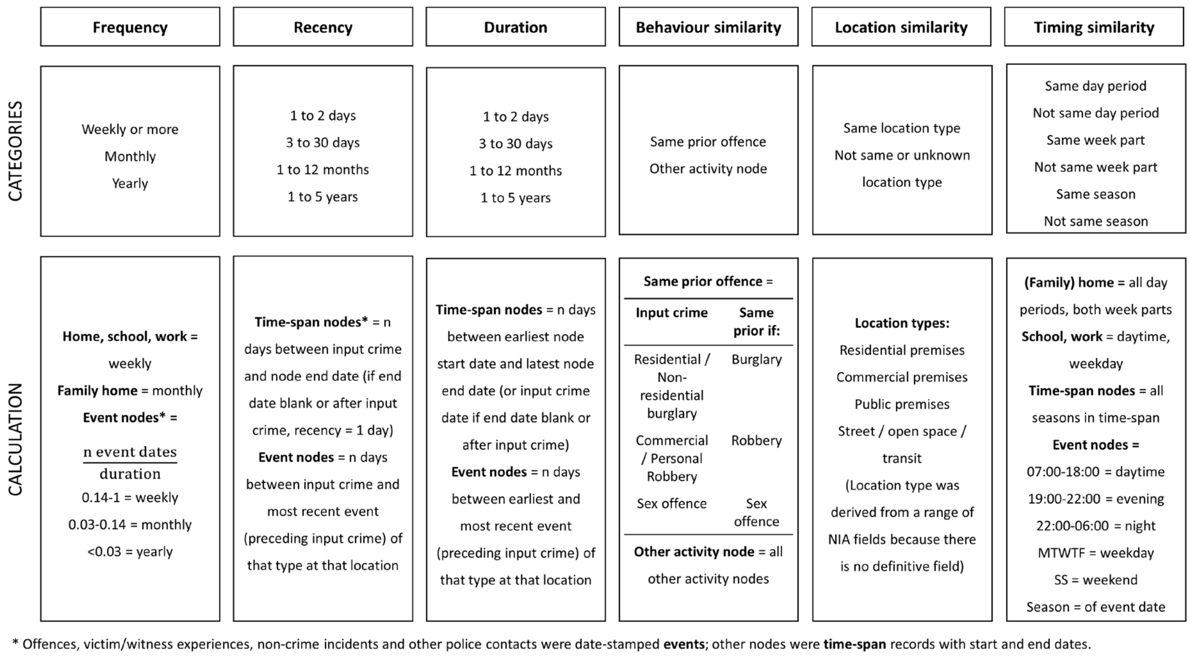


Figure 2. Attributes used in GP-SMART to differentiate activity nodes

After filtering the suspects' activity nodes to only those that pre-date the input crime,[11] and of suspects who were alive as at the input crime, GP-SMART filters to nodes within the user's specified distance: here 10km. Preliminary analyses showed that 10km

[11] For time-span nodes (home, family home, work, school), 'pre-date' meant those with start dates preceding the input crime regardless of when the record was end-dated. For some input crimes and event nodes (offences, incidents, etcetera) the exact timing was unknown and the event was recorded with start and end date-times. For event nodes, to ensure that the event pre-dated the input crime and had likely been reported by the time of the input crime, 'pre-date' meant those with end dates preceding the start date of the input crime. However, for calculating event node attributes relative to (e.g., recency, daypart similarity) the input crime, random date-times between the node and input crime start and end date-times were used, following research establishing that random date-times are a more accurate estimation of the actual timing of unknown timing offences than are the start or end date-time (Ashby & Bowers, 2013; Boldt & Borg, 2016). Time of day was not recorded for the other police contacts, so for these, time-of-day similarity was imputed based on the median over all nodes (being 'same day period').

increased the proportion of test cases where the offender was among the filtered suspects from 80-93% using 5km to 86-96%.[12]

Next, for each activity node for each suspect GP-SMART predicts the probability that the suspect would commit a crime at the input crime's location given its distance to, and the attributes of, the activity node. The prediction method we implemented in this study follows a Bayesian updating process (Figure 3).[13] The first step is conceptually similar to the application of a distance decay function (e.g., Bache et al., 2008), to estimate the likelihood the suspect would commit the input crime given its distance *D* from activity node *Ni*: $P(D_{Ni})$. However, we simply used the reciprocal of the distance between the input crime, which equates to a negative exponential distance decay function $P(D) = D^{-1}$.[14] We convert $P(D_{Ni})$ to odds (the odds of probability P is P / (1 - P)), then update the odds by values representing our beliefs about the increase or decrease in odds of the suspect committing crime near to a node given its attributes A_1 to A_j (frequency, recency and so on): $OR(A_{1\text{-}j,Ni})$ before converting the resulting posterior odds back to a probability: $P(D_{Ni}, A_{1\text{-}j,Ni})$. The distance

[12] In developing GP-SMART we also experimented with an additional filter to narrow to suspects with a prior offence of the same type as the input crime, following previous studies that applied a similar filter (Canter & Hammond, 2007; Snook et al., 2006). However, we found that the proportion of test cases where the offender appeared among the filtered suspect reduced to 13- 51%, so we removed the filter.

[13] We also trialled a machine learning approach that trained random forest models with the calibration data to predict the probability of the suspect being the offender, given the distance between a given activity node and the input crime and the attributes of the activity node. GP-SMART applied these models (one for each crime type) to the filtered suspects and their nodes to predict this probability for each suspect and each node for each test input crime (see accuracy testing method below). The random forest model prediction method yielded very similar results to the Bayesian updating method, except for personal robbery where it was less accurate in ranking suspects.

[14] We also trialled the use of distance decay and kernel density estimation functions fitted to the calibration data for each of the different types of node (home, family home and so on), following the methods set out by Levine (2013) that are implemented in the software CrimeStat, as well as the actual proportion of node-crime distances per 0.2km wide distance band observed in the calibration data. None of these methods performed as well as using the reciprocal of the distance regardless of the type of input crime or type of node.

decay function prioritises offenders who have an activity node near the crime location, and the Bayesian updating gives more weight to activity nodes that are important (such as homes) than to nodes that have proven to be less important (such as homes of relatives). In principle, the updating values could reflect any a priori belief about the relative likelihood of offending near an activity node, given its attribute(s). In practice, these values will depend on what attributes are available in the dataset, and what is known, or can be ascertained, about the associations between those attributes and crime probability.

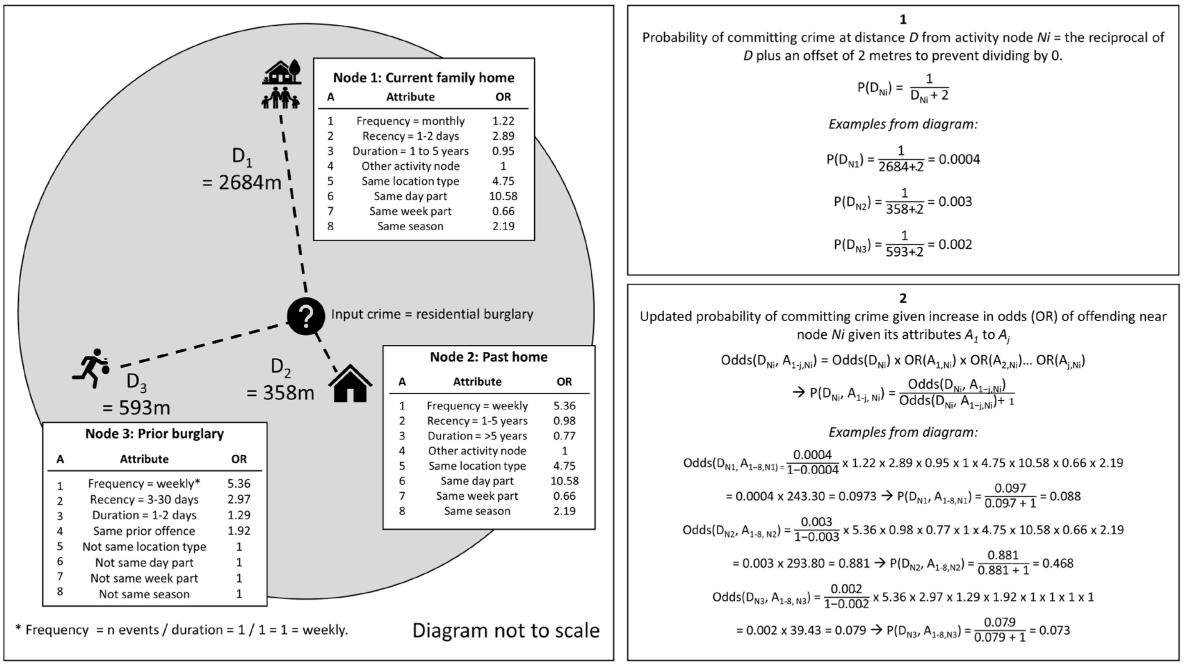


Figure 3. GP-SMART prediction method based on distance and attributes of suspects' activity nodes in relation to an input crime

For the present study we derived these values by analysing the calibration data to quantify the relative likelihood of offending near an activity node, given its attributes (for the attributes outlined above). Specifically, we applied discrete spatial choice models as described in the [Supplementary Online Material S1](). In short, these models estimated the increase in odds of an offender committing crime near an activity node with a given attribute over a location where there is no activity node nearby. In our analyses, 'near an activity node' meant in the same neighbourhood, and 'no activity node nearby' meant no node within 5km of the neighbourhood. For example, the residential burglary offenders were 5.36 times more likely to offend in a neighbourhood[15] containing an activity node visited on a weekly basis and 1.92 times more likely to offend in a neighbourhood where they had committed the same type of crime before (than in neighbourhoods with no activity nodes nearby). These estimates, expressed as odds ratios (ORs), also indicate the *relative* impact of the presence of activity nodes with *different* attributes on the likelihood of committing crime nearby. We therefore adopted these ORs as the updating values for GP-SMART; their application is exemplified in Figure 3. Supplementary Online Material S2 lists the ORs for each node attribute and crime type used in GP-SMART.[16]

The final two steps of GP-SMART (Figure 1) aggregate the predicted probabilities for each activity location into a single prediction for each suspect, then rank the suspects by their predicted probabilities. Initial tests showed that aggregating by taking the maximum yielded more accurate suspect predictions than summing or taking the mean. The maximum appears to sufficiently reflect the presence of higher probability nodes, despite not capturing the cumulative impact of having multiple nodes within the specified distance on likelihood of offending.

Accuracy Testing Method

For each crime type except commercial robbery, we randomly sampled 1000 offences from the test sample for each crime type and ran these through the GP-SMART algorithm to see how often it identified the actual offender in the top ranked suspects, from among all 16,388 possible suspects in the entire test sample. For commercial robbery we used all 511 offences in the test sample. For each sample of input crimes tested, we compared GP-SMART's accuracy with two baseline methods roughly comparable to existing suspect-based

---

[15] Specifically, the same Statistical Area 2 (SA2); see Supplementary Material S1 for detail.

[16] We also trialled a version where we assigned Odds Ratios a priori such that for each attribute the lowest category (e.g., frequency yearly, recency > 5 years, not same daypart) was 1, the next lowest category (e.g., frequency monthly, recency 1-5 years, same daypart) was 2 and so on up to the highest category (e.g., frequency weekly = 3, recency 1-2 days = 5, same daypart = 2). This a priori version produced less accurate suspect rankings than the version we report on which used Odds Ratios empirically estimated from the calibration data.

algorithms that used only home nodes or included other nodes but weighted them equally. The first baseline method filtered to home addresses pre-dating the input crime (but including both current and past homes), then filtered to home addresses within 10km of the input crime, then ranked the suspects by the distance between their nearest home address and the crime. The second baseline method was the same as the first but included all activity nodes at each filtering and ranking step.

We used a range of accuracy metrics consistent with previous studies of the accuracy of geographic profiling algorithms in ranking suspects. These were: the proportion of test cases where the offender appeared in the filtered suspects; the proportion of test cases where the offender appeared in the top 1, 5, 10 and 50 and 100 suspects; the median percentile rank of the offender; and the Gini coefficient, a global measure of the concentration of offenders in the top ranked suspects, with coefficients closer to 1 indicating higher numbers of test cases with high-ranked offenders (as per Snook et al., 2006). Suspects with nodes pre-dating the input crime who did not appear in the distance-filtered list ranked by GP-SMART were all assigned a rank at the midpoint of remaining possible ranks. For offences involving multiple offenders, we counted only the highest-ranking offender in the accuracy measures (i.e., was at least one of the offenders in the top k?). However, most offences involved a single offender (residential burglary: 92% of all offences in the test data, non-residential burglary 89%, commercial robbery 87%, personal robbery 86%, sex offences 99%).

Results

We first consider how many suspects appeared in the filtered list of suspects with prior activity nodes before and after the 10km filter. Of the 16,388 potential suspects, filtering to only those with at least one activity node pre-dating the input crime resulted in still sizeable suspect pools of at least 15,442 using all nodes and 12422 using home only. Filtering to nodes within 10km of the input crime produced median reductions in the number of suspects of between 81% and 89% using all nodes and between 91% and 95% using home only, depending on crime type. Figure 4 shows the distribution of the number of suspects in the distance-filtered lists for each input crime. The bold horizontal lines in the overlaid boxplots indicate the median numbers of suspects, which were in the thousands when including all nodes, but lower when using only home nodes; the underlaid plots provide a more granular representation of the number of cases (points) with different numbers of suspects. In only two test cases (one residential burglary and one personal robbery) were there no suspects with any activity nodes within 10km of the input crime.

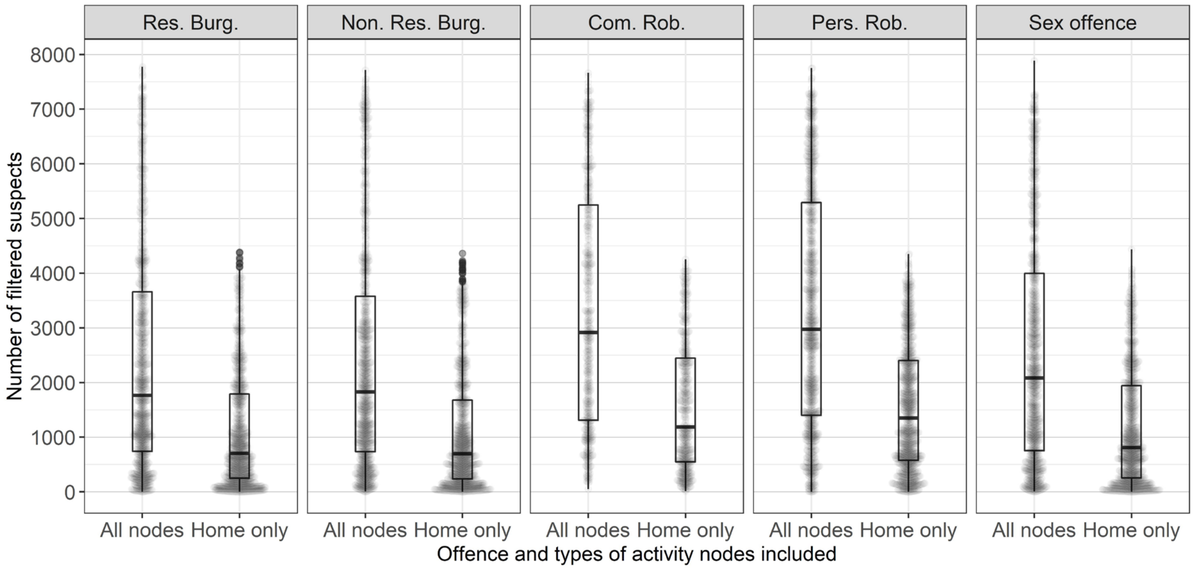


Figure 4. Distribution of the number of filtered suspects per test case

Table 1. Percent of test cases where the offender appeared in the suspect list.

Suspect list	**Res. Burg.**	**Non-res. Burg.**	**Com. Rob.**	**Pers. Rob**	**Sex offences**
Any node	100.0%	100.0%	100.0%	99.9%	99.8%
Home node	98.6%	97.9%	99.4%	99.3%	95.3%
Any node within 10km	94.9%	93.0%	93.9%	95.7%	85.8%
Home node within 10km	85.2%	81.9%	86.5%	88.4%	74.4%

Table 1 shows the proportions of test cases where the offender appeared anywhere in the suspect list, before and after the distance filter, based on any node or home only. In all except 3 cases the offender had a pre-crime activity node on record; the offender appeared in the distance-filtered ‘any node’ list for 86% to 96% of test cases, indicating the utility of applying even a simple distance-based search for suspects with *any* activity nodes within 10km of a crime. Using only home addresses reduced the chances of the offender being in the distance-filtered suspects by 8-12%.

Considering the typical number of suspects to differentiate, GP-SMART performed very well. The top panel in Figure 5 shows the proportion of test cases in which the offender appeared in the top k suspects for different values of k, comparing GP-SMART with the baseline methods ranking suspect by distance to their nearest node or home node. To put these statistics in perspective, the chances of picking the offender as the top suspect if picking randomly—the number of offenders (usually 1) / number of suspects per case—was 0.006% on average for all crime types. GP-SMART ranked the offender at the top in 20% of residential burglary cases: 3240 times the 0.006% expected by chance. Commercial robbery had the lowest accuracy, but even then the offender was ranked first in 4% of cases: 625 times the proportion expected by chance. The proportion of cases with the offender in the top 100 suspects ranged in improvement over chance from a factor of 68 for commercial robbery (43% of cases), to 100 for residential burglary (63% of cases). The Gini coefficients for GP-SMART, ranging from 0.80 to 0.87, indicate that the offenders were highly concentrated in the top ranked suspects (Figure 5 middle panel). On average, GP-SMART placed the offender in the top 0.2% to 0.9% of suspects (Figure 5 bottom panel).

GP-SMART's predictions incorporating attributes of the suspects' activity nodes improved on using distance to either any node or home only, across all accuracy measures (in Figure 5, in the top and middle panels higher numbers indicate superior accuracy but in the bottom panel lower numbers indicate superior accuracy). Considering the median percent ranks, for all crime types the offender was about half as far down the ranked list on average using GP-SMART than distance to nearest node (e.g., in the top 0.2% versus the top 0.4% for residential burglary: see Figure 5, bottom panel). Paired sample Wilcoxon signed-rank tests of difference in offender percentile ranks confirmed that GP-SMART ranked the offender statistically significantly higher up the suspect list than both baseline methods (statistics reported in Supplementary Online Material S3).[17] With robbery offences, the baseline methods yielded similar accuracy as measured by top k percentages and the Gini coefficient,

[17] We compared the percentile ranks rather than absolute ranks to control for the difference in the number of suspects in the list for each case using GP-SMART and the 'any node' baseline versus home only.

suggesting that if the offender appeared in the distance-filtered suspects they were just as likely to appear near the top of the list. But recall that the offender was less much likely to appear in the distance filtered list when using only home nodes than when using any node.

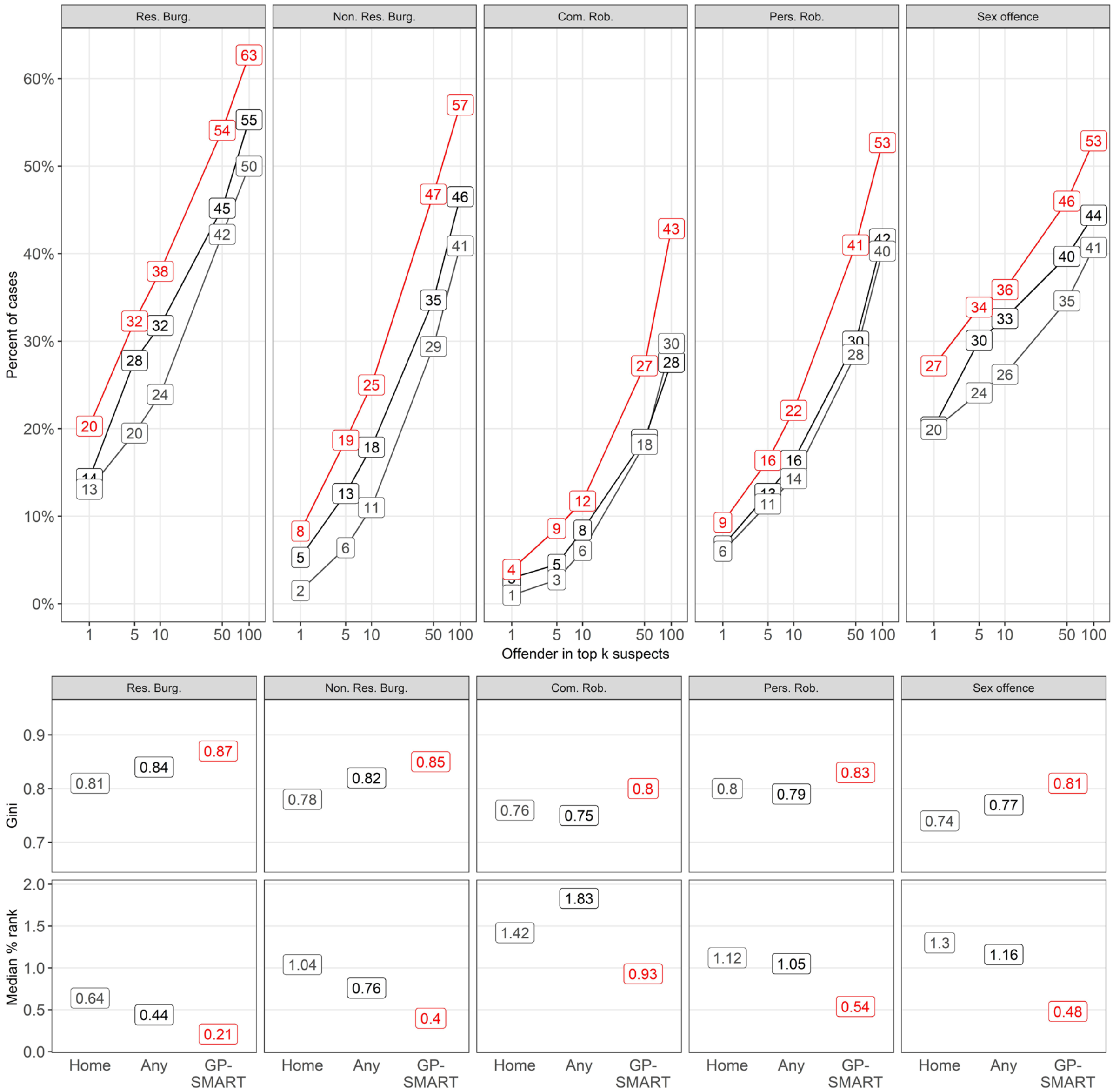


Figure 5. Accuracy statistics for ranking suspects using GP-SMART (red), distance to nearest activity node (black) and distance to nearest home node (grey)

Discussion

This study developed and tested a new suspect-based geographic profiling method, GP-SMART, which prioritises suspects based on both the nature of their activity nodes and their proximity to the crime locations, all derived from police records, under more realistic conditions than previous studies of similar geographic profiling methods. When tasked with prioritising among a large pool of potential suspects and limited to suspect activity nodes pre-dating the input crime, GP-SMART achieved promisingly high accuracy. The probability of finding the offender in the top k suspects in our study was 27-54% and 43-63% for k = 50 and k = 100 (depending on crime type: see Figure 5 top panel), which was higher than in the most comparably robust study with 5% and 9% reported by Tayebi et al. (2017). We can also compare the factors by which GP-SMART improves on chance in picking the offender in the top k suspects, with those reported for Frank's (2012) suspect-based algorithm, for k = 1, 5, 10 and 50. GP-SMART placed the offender in the top k ranks between 63 and 4348 times more often than expected by chance (depending on k value and crime type); Frank's picked the offender in the top k between 3.2 and 7.4 times more often than expected by chance.

The fact that GP-SMART yielded greater accuracy—even under more stringent test conditions—than similar algorithms is consistent with our direct comparisons of GP-SMART and two baseline methods designed to approximate existing algorithms. These comparisons confirmed that including theoretically salient information about suspects' activity nodes—reflecting their familiarity with and likelihood of identifying crime opportunities around their nodes—led to consistently greater likelihoods of finding the offender in the top ranked suspects than using only the distance to their nearest activity node or nearest home address.

GP-SMART's accuracy varied between the crime types studied. Residential burglary and sex offence locations are more strongly associated with offenders' home locations than other offences (Curtis-Ham et al., under review). Home locations score highly on the attributes used in GP-SMART to adjust the predicted probability of crime near the suspects' activity nodes—especially if recent; GP-SMART thus places high weight on home nodes. So it is unsurprising that the predictions led to greater suspect ranking accuracy for these crime types. Conversely, commercial robbery locations are less strongly associated with offenders' activity node locations, being more constrained by the locations of suitable targets (Curtis-Ham et al., 2021b, under review). GP-SMART is less accurate for commercial robbery accordingly. Should GP-SMART be extended to other crime types, its accuracy is likely to reflect the strength of associations between people's crime and activity node locations for those crime types.

Our analysis estimated how likely it is that GP-SMART's sifting process will place the offender among the top suspects, but some hefty caveats apply to these estimates due to the limitations of the data and methods used in this study—the percentages listed here can probably not be realised in practice. For example, the suspects were limited to people known to have committed one of the crime types included, providing a relatively high baseline chance of selecting the offender in the top k suspects, and likely higher accuracy estimates, by comparison to searching for this needle in the haystack of all possible suspects recorded in NIA.

Additionally, our efforts to limit the suspects' activity nodes to only those on record prior to the input offence were not infallible. For example, offences, victim/witness events or incidents recent to the input crime but only reported to police after the input crime would have counted as 'prior' activity nodes despite not having been in police records yet. Likewise, offences committed shortly before the input crime by a suspect whom the police had not yet caught at the date of the input crime would have erroneously counted as prior activity nodes. Our estimates of the likelihood of finding the offender in the top k suspects are therefore also over-estimates of the true likelihood in practice to the extent that recent offence, victim/witness event and incident nodes are only reported to police or attributed to the suspect after a delay, and that suspects' rankings are driven by recent prior offences.

However, we also omitted information that if included in GP-SMART might *increase* its accuracy. We only included a subset of activity nodes that most reliably indicated offenders' activity spaces, with high levels of geocoding accuracy and precision (Curtis-Ham et al., 2021a). It is possible that including even more activity nodes would lead to higher levels of ranking accuracy, even if they were less reliable indicators of offenders' activity spaces (e.g., home addresses of past co-offenders or other associates), or less geographically precise (e.g., traffic offences, which are recorded against stretches of road rather than specific addresses). It is also possible that including these nodes would introduce more noise than signal, and thus reduce accuracy. Additionally, in focusing on only the location and attributes of activity nodes, we did not account for the suspects' likelihood of committing the input crime given their criminal history (c.f. Bache et al., 2008; Tayebi et al., 2017). Higher accuracy could potentially be achieved by adding this information, which could be incorporated into the Bayesian updating step of GP-SMART.

Our operationalisation of the node attributes required many assumptions to fill in gaps in the data, inevitably introducing error. Results will therefore vary, depending on the nature and amount of the data available, and decisions about how to operationalise node attributes using that data, should the GP-SMART method be used in other jurisdictions. Replication is needed to see if the process works as well with data from other jurisdictions, which will vary in the types of activity nodes and attributes available as inputs.

We also recommend some additional tests prior to practical implementation of GP-SMART (or future GP-SMART type algorithms). First, how to apply GP-SMART to a series of offences linked to the same offender warrants further investigation. For example, one could average the suspects' ranks across the series, take their highest or lowest rank, or factor in the proportion of crimes in the series for which they are ranked above a certain threshold. Alternatively, the predicted probabilities for each suspect for each input crime could be combined prior to the ranking step. These methods require testing to see which leads to more accurate rankings. Future research might also provide more specific estimates as to how likely it is that the offender is among the top k suspects given the circumstances of an offence, such as the time of day or week, or where it occurred (e.g., a particular city or urban versus rural area). Lastly, as with any geographic profiling algorithm, field testing is needed to evaluate GP-SMART's utility in operational settings when faced with the true population of potential suspects and information available at the time of a live investigation (Goodwill et al., 2014; Rich & Shively, 2004).

Conclusion

The purpose of geographic profiling decision-support tools is not to identify the offender but to help sift through large numbers of suspects (Öhrn, 2019; Rossmo, 2000, 2021). The present study demonstrates the potential for the wide array of suspect activity node information recorded by police to be harnessed to support this suspect prioritisation task, and the benefit of differentiating between activity nodes. Our results suggest that GP-SMART could be a useful 'tool in the toolbox' to support geographic profiling analysis in police investigations, though we also advocate for further research to fully understand its accuracy and utility as next steps towards any practical implementation.

References

Ashby, M. P., & Bowers, K. J. (2013). A comparison of methods for temporal analysis of aoristic crime. *Crime Science*, *2*(1), 1. https://doi.org/10.1186/2193-7680-2-1

Bache, R., Crestani, F., Canter, D., & Youngs, D. (2008, August). *A Bayesian decay model for suspect prioritisation based on geographical profiling* [Paper]. 14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Las Vegas, NV. http://eprints.hud.ac.uk/id/eprint/8054/

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLOS ONE*, *14*(1), e0210733. https://doi.org/10.1371/journal.pone.0210733

Boldt, M., & Borg, A. (2016). Evaluating temporal analysis methods using residential burglary data. *ISPRS International Journal of Geo-Information*, *5*(9), 148. https://doi.org/10.3390/ijgi5090148

Brantingham, P. L., & Brantingham, P. J. (1991). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (2nd ed., pp. 27–54). Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Canter, D. (2003). *Mapping murder: The secrets of geographical profiling*. Virgin Books.

Canter, D., & Hammond, L. (2007). Prioritizing burglars: Comparing the effectiveness of geographical profiling methods. *Police Practice and Research*, *8*(4), 371–384. https://doi.org/10.1080/15614260701615086

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021a). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information*, *10*(2), 47. https://doi.org/10.3390/ijgi10020047

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021b). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). *Relationships between offenders' crime locations and different prior activity locations as recorded in police data*.

Ding, L., Steil, D., Hudnall, M., Dixon, B., Smith, R., Brown, D., & Parrish, A. (2009). PerpSearch: An integrated crime detection system. *2009 IEEE International Conference on Intelligence and Security Informatics*, 161–163. https://doi.org/10.1109/ISI.2009.5137289

Duan, L., Ye, X., Hu, T., & Zhu, X. (2017). Prediction of suspect location based on spatiotemporal semantics. *ISPRS International Journal of Geo-Information*, *6*(7), 185. https://doi.org/10.3390/ijgi6070185

Emeno, K., Bennell, C., Snook, B., & Taylor, P. J. (2016). Geographic profiling survey: A preliminary examination of geographic profilers' views and experiences. *International Journal of Police Science & Management*, *18*(1), 3–12. https://doi.org/10.1177/1461355715621070

Frank, R. (2012). SPORS: A suspect recommendation system based on offenders' reconstructed spatial profile. In N. Memon & D. Zeng (Eds.), *2012 European Intelligence and Security Informatics Conference* (pp. 38–45). CPS. https://doi.org/10.1109/EISIC.2012.26

Goodwill, A. M., van der Kemp, J. J., & Winter, J. M. (2014). Applied geographical profiling. In G. J. N. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 86–99). Springer. https://doi.org/10.1007/978-1-4614-5690-2_207

Gore, R. Z., Tofiluk, N., & Griffiths, K. (2005). Single incident geographical profiling. In F. Wang (Ed.), *Geographic information systems and crime analysis* (pp. 118–136). IGI Global. https://doi.org/10.4018/978-1-59140-453-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.

Knabe, S. (2008). *Geographic Profiling under the microscope – a critical examination of the utility of Geographic Profiling and expert Geographic Profilers* [Masters thesis]. University of Liverpool.

Knabe-Nicol, S., & Alison, L. (2011). The cognitive expertise of Geographic Profilers. In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 126–159). Routledge.

Lammers, M. (2018). Co-offenders' crime location choice: Do co-offending groups commit crimes in their shared awareness space? *The British Journal of Criminology*, *58*, 1193–1211. https://doi.org/10.1093/bjc/azx069

Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology*, *53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071

Levine, N. (2013). *Crimestat IV: A spatial statistics program for the analysis of crime incident locations, version 4.0*. National Institute of Justice. https://nij.ojp.gov/library/publications/crimestat-iv-spatial-statistics-program-analysis-crime-incident-locations

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2016). Family matters: Effects of family members' residential areas on crime location choice. *Criminology*, *54*(3), 413–433. https://doi.org/10.1111/1745-9125.12109

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: Findings from an online self-report survey. *The British Journal of Criminology*, *60*(2), 303–322. https://doi.org/10.1093/bjc/azz044

Öhrn, M. (2019). *We look at crime through the lens of geographic behaviour: Geographic profiling in operational settings* [Masters thesis, University of Gothenburg]. https://gupea.ub.gu.se/handle/2077/59807

Pettiway, L. E. (1995). Copping crack: The travel behavior of crack users. *Justice Quarterly*, *12*(3), 499–524. https://doi.org/10.1080/07418829500096111

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Rengert, G., & Wasilchick, J. (1985). *Suburban burglary: A time and a place for everything*. C.C. Thomas.

Rich, T., & Shively, M. (2004). *A methodology for evaluating geographic profiling software*. National Institute of Justice. https://www.ncjrs.gov/pdffiles1/nij/grants/208993.pdf

Rossmo, D. K. (2000). *Geographic profiling*. CRC Press.

Rossmo, D. K. (2015). Rounding up twice the usual number of suspects. In M. Maltz & S. Rice (Eds.), *Envisioning criminology* (pp. 253–260). Springer. https://doi.org/10.1007/978-3-319-15868-6_27

Rossmo, D. K. (2021). Dissecting a criminal investigation. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-021-09434-1

Rossmo, D. K., & Velarde, L. (2008). Geographic profiling analysis: Principles, methods and applications. In S. Chainey & L. Tompson (Eds.), *Crime mapping case studies* (pp. 33–43). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470987193.ch5

Ruiter, S. (2017). Crime location choice. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook of offender decision making* (pp. 398–420). Oxford University Press.

Snook, B., Wright, M., House, J. C., & Alison, L. J. (2006). Searching for a needle in a needle stack: Combining criminal careers and journey-to-crime research for criminal suspect prioritization. *Police Practice and Research*, *7*(3), 217–230. https://doi.org/10.1080/15614260500432972

Tayebi, M. A., Glässer, U., Brantingham, P. L., & Shahir, H. Y. (2017). SINAS: Suspect investigation using offenders' activity space. *Machine Learning and Knowledge Discovery in Databases*, 253–265. https://doi.org/10.1007/978-3-319-71273-4_21

van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A time for a crime: Temporal aspects of repeat offenders' crime location choices. *Journal of Research in Crime and Delinquency*, *55*(4), 538–568. https://doi.org/10.1177/0022427818766395

CHAPTER 8

Discussion

> There is a need to move towards a broader notion of geographical profiling that goes beyond the focus on the home or base. By taking on board the social psychology of crime as well as the dynamic, changing nature of criminal spatial behaviour a much richer set of theories, and in turn, investigative tools will emerge. (Canter & Youngs, 2008a, p. 20)

This thesis aimed to develop a means of enhancing geographic profiling by 'going beyond the focus on the home or base' through deepening our understanding of how the various locations within people's mental maps influence the spatial signatures of their crimes. This chapter reviews the findings of the research presented in Chapters 2 through 7 that collectively achieved this goal. It next discusses the theoretical, methodological and practical implications of the findings then considers limitations of the research impacting how the findings can be interpreted and used. The chapter concludes by suggesting avenues for future research.

Review of Findings

In crime investigations, specialist police analysts employ a collection of analysis methods known as geographic profiling, to interpret geographic clues to devise strategies to help the investigative search for the offender (Emeno et al., 2016; Rossmo, 2000, 2012, 2014; Rossmo & Rombouts, 2008). This thesis focused on one aspect of geographic profiling: using the location and nature of suspects' activity locations—such as their home address, workplace or past crime locations—for inferring who is more likely to have committed a crime or series of crimes. One process for this inference first predicts the offender's activity location area from the locations of a series of crimes believed to have been committed by the same offender, then prioritises suspects with activity locations nearer the predicted area: 'crime-based' geographic profiling (e.g., Canter & Hammond, 2007). Another process predicts suspects' likelihoods of offending at a crime's location from their activity locations, then prioritises suspects with higher predicted likelihoods: 'suspect-based' geographic profiling (e.g., Bache et al., 2008). In either scenario, knowing which of their activity locations people are more likely to commit crime nearby could help in inferring which suspects to prioritise: suspects with more crime-conducive activity locations near to the predicted activity location area or to the crime. However, prior to the present research, we lacked a systematic theory of the relative influence of different activity locations on crime locations to guide this inference. This thesis developed one, tested it empirically, and demonstrated its practical application in a suspect-based geographic profiling algorithm.

The key findings at each of these steps were as follows. Reviewing the existing theoretical and empirical literature, in Chapter 2 (Curtis-Ham et al., 2020) we proposed a theoretical model that links people's prior activities to their crime locations through two pathways, reflecting the development of *reliable* knowledge of potential crime locations and *relevant* knowledge of those locations' crime opportunities. In short, the model states that people are more likely to commit crime where they have reliable knowledge—from more frequent, recent or enduring activities—*and* where they have acquired knowledge relevant to the present crime—from activities similar to the present crime in the behaviour, type of location or timing involved.

In Chapter 3 (Curtis-Ham et al., 2021a), we established that the police dataset we intended to use to test and apply this theory included a large number and range of pre-offence activity locations, supporting their use for examining their relationships with the offenders' crime locations, and for identifying and prioritising suspects in geographic profiling. The 60,229 people in the dataset—who had committed burglary, robbery or extra-familial sex offences in New Zealand between 2009 and 2018—usually had at least 10 pre-offence activity locations on record. These included: home addresses; home addresses of family members (including parents, children or siblings, current or former intimate partners and other relatives); more rarely, schools and other educational institutions attended; even more rarely, workplaces; past offence locations; locations where they had experienced crime as a victim or witness; locations of non-crime incidents attended by police (e.g., disorderly behaviour, mental distress, civil disputes); and other police contact locations (e.g., stop and search, intelligence notings, arrests). Initial explorations of the dataset also established that those who offended in nearby locations tended to share more activity space—the areas around their activity locations—than those who offended farther apart: an important finding supporting the potential to differentiate between suspects in a geographic profiling context. Analyses presented in Chapter 4 (Curtis-Ham et al., 2021b) then developed a sampling method that overcomes the computational challenge involved in using discrete spatial choice modelling (DSCM) to test the theoretical model with this large dataset.

In Chapters 5 (Curtis-Ham et al., under review-b) and 6 (Curtis-Ham et al., under review-a) we used DSCM to test hypotheses derived from the theoretical model about the relative associations between different activity locations and offenders' crime location choices. Consistent with the model, offenders tended to commit crime closer to the types of activity locations that are typically visited more frequently (e.g., home versus family homes), and likely to impart more relevant knowledge about crime opportunities (e.g., prior crimes versus prior victim or witness locations). Analysis of the interaction of reliable and relevant knowledge proposed in the model confirmed that offenders were more likely to commit crime near activity locations that are likely to have generated both reliable and relevant knowledge, than near activity locations lower on either or both dimensions. This result demonstrates the importance of taking account of both reliability and relevance when analysing suspects' activity locations in geographic profiling.

Importantly, these findings are likely to generalise beyond the crime types and jurisdiction studied. The consistency of the results—demonstrating the interaction of reliability and relevance—across the different crime types supports the generalisability of the theoretical model. It is therefore likely that other crime types would present the same interaction pattern (see Chapter 6). Where differences between crime types emerged, these were consistent with the theoretical model and reflected differences in the distribution of targets for different types of crime. It is therefore likely that the crime type specific results would generalise other crimes with similar target distributions (see Chapters 4 and 5). Similarly, the consistency of the pattern of declining likelihood of crime with distance from activity locations—for offences both previously studied (residential burglary and robbery) and not (non-residential burglary and extra-familial sex offences)—lends support for the universality of the tendency to offend closer to activity locations than further away, regardless of the crime (see Chapters 4, 5 and 6).

A disadvantage of most previous discrete crime location choice studies is that they examined single cities and excluded local offences with no local offender and local offenders with no local offences (e.g., Clare et al., 2009; Frith, 2019; Long et al., 2021; Menting, 2018; c.f. Menting et al., 2020; see also Bichler et al., 2012, on this point). This exclusion of people committing crimes at longer distance from their activity nodes could have led to overestimation of the associations between activity nodes and crimes. By including a more representative sample of offenders, the present research provides evidence of the applicability of these associations to offenders in general, not just those with an existing connection to the locality under study.

As for the generalisability of the present findings beyond this national sample, it is implausible that there is something unique about New Zealand offenders such that the overall findings apply only here. The findings fit with the crime location choice literature from outside New Zealand that informed the model (see Chapter 2). It is also implausible to think that New Zealand offenders are unique because that would require New Zealanders in general to make location choices differently to other people, given the model is extendable to any location choice, not just crime.

Instead, differences between cities or countries are likely to manifest the *interaction* between the proposed psychological processes and different distributions of opportunities in the environment. As discussed in Chapter 5, for example, New Zealand has lower urban density than other nations, resulting in longer distances (on average) between people's activity locations and crime opportunities. These longer distances would affect what is considered 'near' but would not affect the overall tendency to offend nearer to rather than farther away from high reliability-high relevance activity locations. The one previous cross-national crime location choice study illustrates this point. Townsley et al. (2015) found that people in Dutch, British and Australian cities were more likely to offend near home than farther away, but how likely they were to offend at the same distance (e.g., 500m, 1km) from home varied between jurisdictions, reflecting differences in urban density.

In the final step from theory to practice, the thesis developed a suspect-based geographic profiling algorithm: GP-SMART, the Geographic Profiling Suspect Mapping and Ranking Technique. For any input crime, and set of potential suspects, GP-SMART identifies suspects with activity locations within a user-specified distance of the crime, calculates the probability of each suspect committing the crime given the proximity *and* attributes of their activity locations, and ranks suspects accordingly. It calculates this probability by applying the outputs of DSCM analyses—of the same police dataset as used in the previous chapters—that specified the relative likelihood of crime near activity locations with different attributes (frequency, recency, duration and activity similarity). Analysis of the accuracy of GP-SMART, presented in Chapter 7 (Curtis-Ham et al., 2022), revealed that it placed the actual offender in the top few ranked suspects at rates greatly exceeding chance and—we concluded—with sufficient frequency to warrant investigation of its utility when applied in practice. Importantly given the aim of this thesis, accounting for the theoretically salient attributes of suspects' activity locations led to more accurate suspect rankings than using distance alone without differentiating between different activity locations. However, GP-SMART represents just one means of operationalising this thesis' findings about the links between different activity locations and people's crime locations; other means of applying the findings to enhance geographic profiling are elaborated later in this chapter.

Theoretical Implications

This thesis contributes to both the environmental criminology and investigative psychology theoretical approaches to geographic profiling. As described in Chapter 1, these approaches emerged simultaneously in North America and the United Kingdom and both approaches underpin geographic profiling inferences by emphasising the connection between people's past activity locations and their future crime locations (P. L. Brantingham & Brantingham, 1991, 1993a; Canter & Youngs, 2008b, 2009). But neither of these approaches enable predictions about which of their various activity nodes a given individual will offend nearby. Therefore, the theoretical model proposed—and empirically evidenced—in this thesis contributes to both environmental criminology and investigative psychology by systematising the connections between different activity locations and future crime to enable such predictions. This section expands on the thesis' connections and contributions to each of these theoretical domains.

A central component of environmental criminology's crime pattern theory is the concept of awareness space: the places known to people through their routine activities (P. L. Brantingham & Brantingham, 1991, 1993a). In principle, people could identify and act on crime opportunities anywhere that opportunities exist within their awareness space. This aspect of crime pattern theory has arguably been overlooked by decades of research focused on the 'journey to crime' between offenders' home and crime locations (as summarised in Ackerman & Rossmo, 2015; Townsley, 2016). Although smaller scale or qualitative studies have often identified cases of offending closer to other activity nodes than a current home (e.g., Alston, 1994; Costello & Wiles, 2001; Davies & Dale, 1996; Rengert, 1996; van Daele, 2009), home-based journey-to-crime studies dominate the quantitative literature on individuals' activity locations and their crime locations. This trend has not gone unnoticed, as exemplified by David Canter's observation quoted at the start of this chapter.

The studies described in Chapters 5 and 6 reinforce the role of awareness space: various activity locations can generate awareness of crime opportunities and thus crime doesn't only occur near home. These studies demonstrated that robbery, burglary and extra-familial sex offenders in New Zealand are more likely to offend closer to their *various* activity locations—as recorded by police—than farther away. This finding is consistent with recent smaller scale studies that have expanded the range of activity locations examined in crime location choice research (Bernasco, 2019; Menting et al., 2020) and highlights the importance of considering the entirety of people's activity spaces when examining their crime location choices. The GP-SMART results underscore the practical benefits of considering nodes other than home (Chapter 7). The present results should spur further large-scale research incorporating a wider array of activity locations to enable further quantitative exploration of the relationships between people's past activity locations and their crime locations.

While reinforcing the role of awareness space, the present research also extends crime pattern theory by explaining how and why some activity locations within that space have stronger connections to people's crime locations than others. Not only does the proposed theoretical model set out the attributes of people's activity locations that predict the likelihood of them committing crime nearby, but it also suggests the psychological mechanisms through which those attributes operate: through their impact on the acquisition of reliable and relevant knowledge. As detailed in Chapter 2, identifying these mechanisms involved synthesising the criminological literature with evidence from investigative, environmental and cognitive psychology, and behavioural and cognitive geography—subfields focused on the cognitive processes involved in the acquisition and application of spatial knowledge. The model therefore builds psychologically on the criminological foundations of crime pattern theory, going some way to bridging the gap between the environmental criminology and investigative psychology approaches to geographic profiling.

Through elaborating these psychological mechanisms, the model also helps to explain why home is strongly connected with people's crime locations, as emphasised in the investigative psychology approach. Central to this approach is 'circle theory' and the principle of 'domocentricity' (Canter & Youngs, 2008b). Circle theory states that offenders tend to live within a circle whose diameter is the line between their two most far apart crimes (Canter & Gregory, 1994; Canter & Larkin, 2008); domocentricity similarly states that offenders' crimes tend to be centred on their domicile (Canter & Youngs, 2008b, 2009). Further, circle theory creates a dichotomy into which offenders can be categorised: marauders, whose homes fall within the circle as defined above, and commuters, whose homes fall outside the circle (Canter & Gregory, 1994; Canter & Youngs, 2008b) and whose crimes are likely centred on other activity nodes. This simple geometric classification has been criticised for failing to account for offenders who live outside the circle but still close to their crimes and transient offenders who neither maraud nor commute from a stable home base (Alston, 1994; Campbell, 2019). Geometry aside, the theory is clear: home is central in people's mental maps and many activities—criminal and non-criminal—therefore centre around home (Canter & Youngs, 2009). And consistent with this theory, the present research and other DSCM studies have found that people are—on average—more likely to offend near home than other non-criminal activity nodes (Chapter 5; Menting, 2018; Menting et al., 2016).

But why should home play such a dominant role in crime location choice? As discussed in Chapters 2 and 5, home is a high frequency and—usually—long duration activity node. Homes thus generate highly *reliable* knowledge of the surrounding area; current (high recency) homes even more so. People's comings and goings to and from home also facilitate the acquisition of *relevant* knowledge of crime opportunities in this area by exposing them to opportunities at different times of day and along the various paths that connect home to other nodes; even more so for crimes with residential targets or scenes. Domocentricity is therefore explained by the role home plays in generating reliable and relevant knowledge of crime opportunities.

The thesis' theoretical model provides an explanation of crime location choice that accommodates both circle theory and domocentricity, and the many exceptions, as discussed above, where people offend near other locations in their awareness space. Rather than considering the functional role of activity nodes—home, work, shopping and so on—the model prompts consideration of the attributes of these nodes that affect whether they generate reliable and relevant knowledge that would lead an offender to choose their location for crime. Considering these attributes enables predictions about the exceptions when people are likely to commit crime near a non-home activity node: when their activities at that node have been recent, frequent or enduring, and similar to a future crime in the type of behaviour, location or timing involved (i.e., conducive to identifying nearby opportunities for that crime). That people are more likely to commit crime near locations of recent or numerous past crimes—especially of the same type—than near home (Lammers, 2018; Lammers et al., 2015; van Sleeuwen et al., 2018; Vandeviver & Bernasco, 2020; c.f. Chapter 5)[1] is a case in point, such past crimes providing reliable knowledge highly relevant to future crimes.

A final theoretical implication bears on investigative psychology's concepts of 'homology' and 'differentiation'. Investigative psychology studies the links between offenders' actions in their crimes and their characteristics, to support behavioural profiling inferences of offenders' characteristics from their crime actions (Canter, 2011; Canter & Youngs, 2008b). Much investigative psychology research is focused on empirically evidencing two theoretical assumptions that underpin action-characteristic links: homology—people with similar crime actions will be similar on some characteristic dimension—and differentiation—people with different crime actions will be different on that characteristic dimension (Canter, 2000; Doan & Snook, 2008; Woodhams & Toye, 2007). Geographic profiling inferences rely on the link between offenders' spatial action of committing crime in a particular location and their spatial characteristics: their home and other activity nodes (Canter, 2011; Canter & Youngs, 2008b). The theoretical model developed in this thesis provides a causal explanation linking offenders' crime locations and their activity nodes; it also suggests that the assumptions of homology and differentiation should be met when applied to spatial actions and characteristics: people who commit crimes in similar locations should have similarly located activity nodes and people who commit crimes in different locations should not. The analysis presented in Chapter 3 confirmed this pattern, providing the first direct evidence of homology and differentiation in relation to geographic profiling. That GP-SMART could frequently differentiate the offender of a given crime from other suspects based on the location and attributes of their activity nodes (Chapter 7), is additional, indirect evidence of spatial differentiation.

[1] The present research found higher odds of crime near home than past crimes, likely because the past crimes included a wide range of offences, of which a smaller proportion would have been the same type as the reference location choice offence, than in the cited studies.

Methodological Implications

The main methodological contributions of the thesis are threefold, relating to the data used throughout, the sampling method developed in Chapter 4, and the method for testing GP-SMART's accuracy employed in Chapter 7. Turning first to the data, the dataset used in this research included an unprecedented range of offender activity locations compared with previous studies of individual offenders' activity spaces and crime locations using administrative data (e.g., Lammers et al., 2015; Menting et al., 2016; van Sleeuwen et al., 2018). However, these data are collected for myriad purposes other than relating offenders' activity locations to their crime locations and form neither a complete nor systematic record of any person's pre-offence activity locations. A secondary aim of the thesis was therefore to determine how much 'signal' could be found amid the 'noise' of police data, that could further our understanding of people's crime location choices and enhance geographic profiling.[2]

The use of this dataset contributes to the literature from a methodological perspective by establishing (1) what signals about suspects' activity locations that are salient to their choice of crime location, are present in police data readily available in crime investigations; and (2) whether those signals can inform sufficiently accurate predictions of who is most likely to have committed a given crime to help prioritise suspects in investigations. Specifically, Chapter 3 explored how much information about the location of offenders' pre-crime activities is available from the main crime and intelligence database of a national police service, establishing its prima facie potential for use in research on offenders' activity spaces and crime location choices. Subsequent analyses revealed that the type of activity location alone contains a considerable amount of signal, explaining a large amount of the variance in offenders' crime location choices (Chapter 5). However, the incompleteness of the data made

---

[2] These concepts are adopted from Nate Silver's book 'The Signal and the Noise: The art and science of prediction', which, despite having ostensibly nothing to do with crime, had a significant influence on the thinking that went into this thesis. In his words, "The signal is the truth. The noise is what distracts us from the truth" (Silver, 2012, p. 21).

it difficult to fully examine the relationships between specific attributes of people's activity locations proposed in our theoretical model and offenders' crime locations (see Chapter 6, Appendix D, Appendix E and the limitations section below). Finally, in testing GP-SMART, we found that, despite the challenges for hypothesis testing presented by the data, they nonetheless contained sufficient signal to identify and prioritise suspects with promising accuracy (Chapter 7).

What do these findings imply for future research? Certainly that future DSCM studies could use police data to account for a wider array of activity nodes when modelling people's crime location choices—thus more accurately capturing their awareness space and its role in their decision-making. And likewise, that future geographic profiling research could use police data on offenders' activity nodes to help the research domain to move beyond examining home-crime journeys (e.g., Chen & Lu, 2020; Trinidad et al., 2020) and testing algorithms using only offenders' home locations (e.g., Barreda, 2020; Canter & Hammond, 2006; Paulsen, 2006a; Svobodová, 2018). But our findings also imply that such data may not suit research questions about the association between people's crime locations and attributes of their activity nodes that are poorly captured in police data (see the method sections of Chapters 6 and 7 for the assumptions we made to fill in these gaps when coding activity node attributes). For those questions, studies should use data that enable more refined measurement of theorised activity node attributes (i.e., that capture the time, date, and behaviour involved in people's activities at each activity node: see Chapter 2 and the limitations section).

The second methodological contribution of this thesis is the development of a novel method for sampling from alternatives in discrete spatial choice modelling (DSCM) for crime location choice (Chapter 4). Because DSCM requires datasets with n rows = n choices x n alternative locations for each choice, applying it to big datasets such as ours—with thousands of choices (crimes) x thousands of alternatives (neighbourhoods in New Zealand)—creates a computational challenge. To overcome this challenge, previous crime location choice studies with similarly large datasets used high performance computer labs (Vandeviver et al., 2015; Vandeviver & Bernasco, 2020) or sampled a random set of alternatives for each crime choice (e.g., Bernasco et al., 2013; Vandeviver et al., 2015). But DSCM studies in other domains suggested that over-sampling alternatives that are more important to each decision-maker's choice (McFadden, 1977) can produce more robust parameter estimates—closer to those produced without sampling—than random sampling (Hassan et al., 2019; Lemp & Kockelman, 2012). Therefore, we developed a sampling method for crime location choice that prioritises selection of the most likely signal-containing alternatives for each offender: those near their activity nodes. Although the concept of 'importance sampling' the alternatives more likely to be chosen by each decision-maker was not new, this study was its first application to crime location choice. The demonstration that our importance sampling method leads to more robust parameter estimates than simple random sampling (for most of the crime types studied) implies that this method could be used in future DSCM crime location choice studies to save processing time, and even to enable research that would not be feasible outside of high-performance computer labs (if at all).

The method used in Chapter 7 to test the accuracy of GP-SMART is the third contribution of this thesis to the development of research methods. As detailed in that chapter, previous studies testing the accuracy of algorithms that rank suspects using data on their activity locations have tended to use small suspect pools that included the offender (e.g., n=83: Bache et al., 2008; n=322: Frank, 2012) and activity location records that post-date the 'input' crime and were thus unknown to police when investigating it (e.g., Gore et al., 2005; Snook et al., 2006). These methods likely led to inflated accuracy estimates with limited ecological validity to investigations. In investigations, one is searching for a needle in a much larger haystack of suspects—the thousands, or even millions, of people in the police database(s) containing activity location information—even if some initial filtering criteria were applied (e.g., only considering people who have committed a crime, or certain types of crime). One is also limited to activity locations that are in the database(s) at the time of the investigation.

In testing GP-SMART, to provide a more accurate picture of accuracy we constructed tests that more closely mimicked these investigative realities. For each input crime tested, the suspect pool included people who had committed any of the various offence types examined in this thesis, but only those with activity locations pre-dating the input crime. The suspect pool thus did not necessarily include the actual offender. This method created a different number of suspects for each input crime we tested, though there were over 15,000 suspects for every input crime. Although it did not replicate a real investigation exactly (see the limitations discussed in Chapter 7), the method improved on previous studies and provides an exemplar for future studies of suspect prioritisation accuracy.

Practical Implications

This thesis was primarily motivated by an opportunity to improve geographic profiling practice; its geographic profiling implications are thus the focus of this section. Secondarily, the findings have implications for enabling geographic profiling through data recording practice and preventing crime through offender management practice, which are also addressed in turn, in this section.

Geographic Profiling

As described in Chapter 1, geographic profiling involves the structured application of criminological and psychological theory to case materials, to make inferences about the offender or how to identify them (Knabe-Nicol & Alison, 2011; Rossmo & Rombouts, 2008; Rossmo & Summers, 2015). The inferences to which this thesis is relevant involve working from either suspect locations or crime locations to infer which suspects to prioritise, using quantitative—statistical or algorithmic—or qualitative—clinical—processes. In teasing out how the theoretical model can be applied to support these inferences in practice, this section considers statistical and clinical processes in relation to suspect-based and crime-based inferences in turn. It concludes by placing these inferences in their wider investigative context.

Statistical Suspect-Based Inference. This thesis developed a suspect-based statistical process—GP-SMART—that applies the theoretical model to infer from suspects' activity locations, who among them is more likely to have committed an input crime. Tests of GP-SMART's accuracy in ranking the offender among the top suspects demonstrated that it outperforms methods that do not take differences between suspects' activity locations into account (Chapter 7) and that it appears accurate enough to warrant trialling its use in practice to evaluate its utility in the field. There are several ways in which GP-SMART could be used in practice, depending on whether the investigation has already identified a short list of suspects.

First, if suspect(s) have already been identified, an analyst could use GP-SMART to examine which of these suspects are more likely to have committed the input crime, given their activity locations. In this scenario, suspect activity location information from sources other than police data could conceivably be added, provided it includes information enabling the attributes from the theoretical model (frequency, recency and so on) to be calculated. For example, the investigation might have collected information about the suspects' activity locations when undertaking surveillance, obtaining cell phone records, or scraping social media data. Ideally, further research should be conducted to identify the 'adjustment values' to use in GP-SMART that differentiate between suspects' activity locations based on their attributes as captured in these data sources.[3] Chapter 7 provides a blueprint for this process of calibrating the adjustment values, that could be applied to data from other sources, just as it can be applied to local police data to calibrate GP-SMART for use in other jurisdictions.[4]

Second, as exemplified in Chapter 7, an analyst could use GP-SMART to filter and prioritise from a much larger list of potential suspects extracted using some initial broad criteria (e.g., having committed a crime before). The value of this approach lies in identifying a short-list of suspects from among a number that would be much too large to search and prioritise manually. Just how short that list is in any given case will depend on how many suspects have activity nodes within the distance threshold the analyst sets in GP-SMART (we used 10km).

---

[3] Although environmental criminological research has increasingly used anonymised 'big data' from such sources (Snaphaan & Hardyns, 2019), connecting data on individuals' activity locations from these sources with their crime locations in police records presents technical and ethical challenges.

[4] GP-SMART has been implemented in an R package—'gpsmartr'—that was released on the open source code repository GitHub alongside the publication of Chapter 7. The user can input the adjustment values identified from this calibration process. See https://github.com/Sophie-c-h/gpsmartr.

In both scenarios, using GP-SMART—or any GP-SMART style algorithm—to automate the calculation of suspect ranks increases the speed and reliability of the process of inferring suspects' priority from their activity locations. GP-SMART took only a few seconds to process hundreds of thousands of activity locations, for over 15,000 potential suspects in each test case. Using an algorithm also guarantees consistency: each case is subjected to the same process, eliminating inaccuracy arising from the noise inevitably introduced by humans performing the same task (Kahneman et al., 2021).

Clinical Suspect-Based Inference. But what can an analyst do if they have no GP-SMART style decision-support tool to do the heavy statistical lifting? The next best option is a qualitative—but structured—approach. Clearly it is not possible to process many thousands of suspects manually, but given a short list of suspects an analyst can map their activity locations to enable a comparison between the suspects' activity locations and the location of an input crime. The theoretical model could then be applied to guide consideration of each activity location: how frequently and recently did the suspect visit the location? How long have they been going there? What activities did they undertake there and how similar are those activities—in behaviour, location type and timing—to the input crime? Do the answers to these questions suggest the suspect has highly reliable knowledge of the location that is relevant to the input crime? Applying the same logic implemented in GP-SMART, suspects with high reliability-high relevance activity locations close to the crime can be ranked as a higher priority than suspects with activity locations farther from the crime, or lower in reliability and relevance.

An advantage of this qualitative approach is that activity locations from many different sources can be mapped and considered, without the need for the statistical calibration process described in Chapter 7. For example, activity locations might come from evidence collected about specific suspects such as undercover surveillance logs, witness or CCTV sightings, phone records, bank card transactions, geo-tagged social media posts and police and non-police administrative databases. But as with clinical—human—judgement in other domains (Kahneman et al., 2021), the less quantified and structured are analysts' judgments of suspects' priority, the more noisy and inconsistent the judgments will be.

Minimising this noise by approximating the statistical process operationalised in GP-SMART as closely as possible is therefore desirable.

Statistical and Clinical Crime-Based Inference. Crime-based suspect prioritisation inferences first work from the locations of a series of crimes believed to have been committed by the same offender to predict the area most likely to contain their home or another activity location that is 'anchoring' their crime locations (Rossmo 2000). This prediction can be made statistically using tools such as Rigel (Environmental Criminology Research Inc., n.d.; Rossmo, 2000), CrimeStat (Levine, 2013), or Dragnet (Canter et al., 2003), and their recent adaptations in R (Hauge et al., 2016; Spaulding, 2020; Verity et al., 2014). Or it can be made clinically, applying simple heuristics such as 'choose the middle' of the area covered by the crimes (Bennell, Snook, et al., 2007), or applying expert intuition built over years of experience and use of the statistical tools (Knabe-Nicol & Alison, 2011). But regardless of the prediction method, the next step is to compare suspects' activity locations to the area predicted as most likely to include an activity location, to determine who to prioritise.

The findings of this thesis provide a means to differentiate between different activity locations in this suspect ranking step. Activity locations near which suspects are more likely to have committed crime—that are more likely to have anchored the crime locations—could be weighted in much the same way as in GP-SMART, so that suspects with more crime-conducive activity locations in or near the predicted area are ranked higher on the list. Further research is needed to explore how this weighting step might be implemented as an add-on to existing crime-based geographic profiling algorithms (e.g., Environmental Criminology Research Inc., n.d.; Rossmo et al., 2005), and to test whether weighting improves suspect ranking accuracy (over treating all activity locations equally or including only home nodes), as we did for GP-SMART in Chapter 7. Meantime, in the absence of an algorithmic add-on, analysts could apply the theoretical model to structure their comparisons of short-listed suspects' activity locations with the crime-based activity location predictions (or to adjust an algorithm's unweighted rankings), to prioritise suspects with high-reliability high-relevance activity locations at or near the predicted area. Naturally, the same disadvantages of clinical inference processes apply to any clinical steps in the crime-based process, as discussed for the suspect-based process above.

The Context of Suspect- and Crime-Based Suspect Prioritisation Inferences. The inferences discussed in this section, whether made algorithmically or clinically, do not operate in a vacuum. They will always need to be interpreted in light of evidence, intelligence and other spatial clues. For example, physical descriptions of the offender, forensic evidence or alibi evidence might quickly eliminate otherwise high-priority suspects from the list (see Rossmo et al., 2005 for a case study illustrating this process). In interpreting spatial clues, a key question for geographic profiling analysts is whether the crime location is more likely to reflect the availability of suitable targets than the offender's activity locations (Fumagilli & Johnson, 2020; Rossmo, 2000; van der Kemp, 2021). If targets are sparsely distributed—only found in particular locations—the offender may need to travel farther from their activity locations to find a good target. In such cases, inferences based on the proximity of suspects' activity locations to the crime or predicted activity location area will be less accurate because there is a weaker relationship between the offender's activity space and the crime. Accordingly, less weight should be put on these suspect prioritisation inferences when advising in an investigation. The commercial robbery cases we tested in Chapter 7 exemplify this issue: GP-SMART placed commercial robbery offenders in the top ranked suspects less often than other offences, reflecting the sparser availability of suitable targets (as discussed in Chapters 4 and 5). Other spatial clues suggesting particular activity locations as likely 'anchor points' for the input crime(s) that are not necessarily captured in the inference processes described above may also serve to update analysts' suspect prioritisation judgments (e.g., Daniell, 2008; as detailed in Chapter 1).

Data Recording

The geographic profiling enhancements described above cannot be achieved without accessible, complete, precise and accurate records of potential suspects' activity locations, especially the records most readily available to police in an investigation: those in police administrative databases. Although police cannot collect information about people's activity locations beyond that necessary for operational purposes, there is an argument to be made for storing this information so that it is readily retrievable and processable to aid the identification and prioritisation of suspects in future investigations. Despite New Zealand's advantage in having a national crime and intelligence database (the National Intelligence Application: NIA), different data extraction systems—with access to different parts of the database—were needed to extract and piece together the present dataset from NIA records. Additional activity location information sits inside NIA in unstructured formats (e.g., free text fields and attached documents) and outside NIA in a separate database for serious crime investigations and in case files on networked drives. In addition, in the extracted data, incompleteness, imprecision and inaccuracy led us to exclude some crime or activity location records or to impute their attributes. Details of how these issues were managed can be found in Chapters 3 to 7 and their supplementary materials (Appendices A to E).

These data challenges highlight the need to improve police data systems and data quality. The more structured, unified and streamlined a police service's systems and processes for storing and extracting activity location data, the easier it will be to harness that information for the benefit of crime investigations through the use of methods such as GP-SMART. With more complete, precise and accurate data, GP-SMART would achieve better accuracy. Issues with data quality are not unique to New Zealand Police and can impede crime and intelligence analysis beyond the present geographic profiling example (Burcher & Whelan, 2019; Burrell & Bull, 2011; Cope, 2004; Gerell, 2018; O'Connor et al., 2021; Tonkin & Weeks, 2021). The present research thus contributes further support to existing arguments for police investment in information recording and management policies, processes and technological solutions that facilitate access to and analysis of high quality data to support police decision-making (Garicano & Heaton, 2010; Home Office, 2018; O'Connor et al., 2021).

Offender Management

A final practical implication of the thesis is that the theoretical model set out in Chapter 2 could be integrated into offender risk assessment and management practices. Community-based risk assessment involves identifying offenders' acute dynamic risk factors, including previous proximal triggers for criminal behaviour (Hanson & Harris, 2000; Serin et al., 2019). These assessments can then inform risk management through formal strategies, such as community-based sentence conditions or electronic monitoring protocols, or informal strategies, such as personalised safety plans to help offenders avoid triggering situations (Serin et al., 2016). Existing psychological risk assessment tools include opportunity or victim access as an acute risk factor that can trigger offending (Serin et al., 2019; Yesberg & Polaschek, 2015). Although the present research was focused on the relationship between offenders' awareness of opportunities (from past activities) and their decision *where* to commit crime, that awareness can also influence the decision *whether* to commit crime (P. J. Brantingham & Brantingham, 2012; P. L. Brantingham & Brantingham, 1993a).

From a risk assessment perspective, therefore, it can be helpful to identify places that offenders have awareness of crime opportunities, and that therefore pose a risk of triggering offending. This suggestion is operationalised in the 'environmental corrections' approach to offender management, which focuses on identifying triggering opportunities and helping offenders to avoid them (Cullen et al., 2002; Schaefer et al., 2016). There is some evidence that this opportunity-focused approach is effective in reducing reoffending (Schaefer & Little, 2019), and it has been described as simply reframing probation practice that is standard in some jurisdictions (Johnston, 2016). But regardless of the extent to which this approach is business as usual, it could be enhanced by applying the theoretical model proposed in this thesis to identify places that are high-risk for a given offender. These are places of which they have reliable and relevant knowledge of crime opportunities, given their past activities and the type(s) of crime they are motivated to commit. Risk management strategies can then be tailored to help offenders avoid these locations or to mitigate the risk of taking advantage of offending opportunities near high-risk but unavoidable locations (friends' and family members' homes, for example).

Limitations

This section considers several general limitations of the data and methods used in this thesis that impact how the results can be interpreted and used. First, the thesis relied on data about people's offence and other activity locations recorded by police far from comprehensively or systematically. Police services collect information about the people with whom they come into contact, as needed for operational purposes such as enforcing the law, investigating crime, or maintaining public safety. These data are inherently limited in the extent to which they capture people's activity spaces and the theoretical constructs to be measured, such as the frequency of people's visits to their activity locations, and the similarity of their activities in those location to the crime they later committed. We filled in gaps in the data using well-founded but broad assumptions about the timing and nature of people's routine activities. But these assumptions will have introduced error. In the theory testing chapters (5 and 6), we therefore place more emphasis on the general patterns of results than the exact estimates when drawing conclusions. For example, it seems reasonable to conclude that—as theorised—factors such as how often a person visits a location and the similarity of their activities there to a future crime matter to their choice of location for that crime, but this thesis draws no conclusions about which factors matter more, or by how much. These questions remain for future research to explore.

Second, the research used solved cases, which might limit the generalisability of the findings to offenders who evade detection. This issue is common to all crime location choice studies that rely on police data on offenders and their offences (Ruiter, 2017). As discussed in Chapters 4, 5 and 6, analysing only solved crimes may lead to over- or under-estimation of associations between the attributes of offenders' activity locations and their crime locations if these attributes are also associated with their likelihood of detection. But the general findings in relation to the proposed theoretical model would likely be the same were unsolved crimes included. The findings are consistent with previous crime location choice studies that used data on self-reported crimes—whether those crimes were detected by police or not—to examine associations between the frequency (Menting et al., 2020) and timing (van Sleeuwen et al., 2021) of past visits to locations and future crime locations. They are also consistent with the literature on people's location choices for non-criminal activities that informed the development of the theoretical model, to which considerations of 'solved' and 'unsolved' do not apply (see Chapter 2).

Third, this thesis did not distinguish between—for example—stranger and known offenders; offences against specific types of victims (e.g., children or adults) or targets (e.g., retail stores, convenience stores or petrol stations); or offences with different modus operandi (MO; e.g., the method of entry for burglary, or the method of approaching the victim for sex offences). But these factors affect which of the offender's activity locations are likely to converge with their preferred victims or targets. For example, in a burglary involving forced entry to a recent home address to assault an ex-partner—typical of a subtype of residential burglary relating to family or interpersonal conflicts (Coupe, 2017; Fox & Farrington, 2012)—the crime is committed at the offender's (recently vacated) home.[5] Similar interpersonal conflict driven personal robberies could occur at the offender's home, by comparison with 'muggings' that occur in public places (Haberman et al., 2021). Commercial robbers travel different distances from home, on average, depending on the type of premises targeted (Van Koppen & Jansen, 1998). Extrafamilial sex offences against children are more likely to occur at or near the offender's home address (Chopin & Caneppele, 2019) or familiar location (Goodwill et al., 2016) than are those against adult victims. They also frequently occur near the kind of activity locations that facilitate contact with children (Leclerc & Felson, 2016; Mogavero & Hsu, 2017; Smallbone & Wortley, 2000), which are not the same as those that facilitate contact with adult victims (Ceccato, 2014; Hewitt, 2021). Additionally, whether the offender is a stranger and their method of approaching the victim—such as engaging them in conversation or 'blitz' attacking them by surprise—are associated with how far from home extrafamilial sex offenders offend (Chopin & Beauregard, 2020; Chopin & Caneppele, 2018; Duwe et al., 2008; Hewitt et al., 2020; LeBeau, 1987) and how likely they are to offend in a familiar location (Beauregard et al., 2007; Chopin & Caneppele, 2018). The associations between different routine activity locations and crime locations will thus vary between subtypes of offences defined by targets and MO.

[5] As I have observed in reading burglary case files for unpublished research undertaken as an employee of New Zealand Police.

By studying disaggregated crime types this thesis improved on previous crime location choice research that aggregated crime types into broader categories (e.g., Lammers et al., 2015; Menting et al., 2016). However, we traded off precision in estimating crime location-activity location associations for more specific crime *subtypes*—such as described above—for simplicity and convenience (due to the complexity and time involved in extracting the requisite data). This imprecision limits the extent to which a) the observed associations generalise across the offences from which the associations derived, and b) the estimates of GP-SMART's accuracy generalise across the cases to which the associations were applied—it will be more accurate for some subtypes than others. However, this trade-off is typical of the field: none of the DSCM studies identified during this research considered subtypes within crime types either.

Future Research

The limitations discussed above raise opportunities for future research, described in this section. This section also suggests directions for expanding the theoretical model and its empirical investigation, and next steps to further develop and evaluate GP-SMART.

In studying people's criminal decision-making, no dataset is perfect; theory testing requires triangulating between the results of studies using different imperfect datasets (Bernasco, 2017). The present findings should thus be corroborated using data sources that provide alternative means of measuring the elements of the proposed theoretical model—activity location attributes and the composite constructs of reliability and relevance—and that can capture both detected and undetected crimes. Time use surveys provide a means of capturing the timing, location and type of behaviour involved in pre-crime activities with great detail and precision. But the small sample sizes achieved by survey based crime location choice studies to date (Bernasco, 2019; Menting et al., 2020; van Sleeuwen et al., 2021) suggest that future research should direct effort to increasing sample sizes to enable inclusion of a larger number of variables and separate analyses of different crime types. The rise of private sector and administrative 'big data' capturing people's everyday activity locations also presents an opportunity for further triangulation (Solymosi & Bowers, 2018).

Harnessing this data requires overcoming practical and ethical challenges in matching individuals' activity locations to self-disclosed or officially recorded crime locations. Like the present research, such quantitative methods rely on measuring latent constructs—reliability and relevance—through observable characteristics of peoples' activity locations. In validating these constructs, future research should also explore offenders' experience of acquiring and applying reliable and relevance knowledge in their crime location choices, using qualitative interviews.

The limited granularity of this and previous studies with respect to crime types suggests that a potentially fruitful avenue for future DSCM research would be to study more granular crime subtypes. Studies could disaggregate residential burglaries and personal robberies by interpersonal conflict and other subtypes; non-residential burglaries and commercial robberies by specific types of target premises; and sex offences by age, relationship to the victim, and MO. Such disaggregation would align with previous research highlighting individual differences in crime location choice and calls for these to be accounted for in crime location choice research (Frith, 2019; Townsley et al., 2015, 2016).[6] Likewise, future research on GP-SMART style geographic profiling algorithms could calibrate and test their predictions with these specific subtypes of offences; testing with only stranger or 'whodunit' offences would also maximise ecological validity to the types of cases where geographic profiling is used.

In a similar vein, individual differences are one direction in which to expand the theoretical model. As detailed in Chapter 2, a range of environmental and individual factors that moderate the acquisition of reliable location knowledge likely also apply in the criminal context. Factors that affect offenders' ability to recognise good crime opportunities and generalise from previous exposure to such opportunities to a future crime (i.e., that affect the

---

[6] However, these studies focused on subgroups of offenders defined by their characteristics such as age and prior criminal history, not subtypes of offences defined by features such as the type of target or age of the victim. In the types of whodunit crimes to which geographic profiling is applied, investigators have information about such features of the offence, but not about the offender.

acquisition of relevant knowledge) could also be investigated. Understanding these more individualised aspects of crime locations choices would help inform more precise predictions about where individuals will commit crime in the future.

Another direction for further theoretical development would be to explore the relative importance of different activity location attributes—frequency, recency and so on—to crime location choice. As noted above, the present research leaves questions about which of these attributes matter more, or by how much, for future research to explore. For example, is it more important to have visited a location recently, or many times, regardless of when? Is there a threshold for acquiring sufficiently reliable knowledge of a location to feel confident in offending there? The Chapter 6 results hint at such a threshold but require corroboration.

An additional avenue for expanding the empirical evidence for the theoretical model is the incorporation of paths: the routes people take between activity nodes that also generate awareness of crime opportunities (P. L. Brantingham & Brantingham, 1991, 1993b). Whereas the theoretical model acknowledges the role of paths in generating reliable and relevant knowledge (see Chapter 2), paths were not accounted for in the empirical studies in this thesis for technical reasons. Studies of human mobility show that people do not travel all potential paths between their various activity nodes with the same frequency, if at all (Hart et al., 2020; Pappalardo et al., 2015; Yan et al., 2017). With the present data—which included activity locations 'active' at different time points before the crime—inferring paths between all nodes was unjustifiable (c.f., Ruiter & Davies, 2018); creating paths between only certain active nodes was unfeasible. Future research could leverage GPS or other location tracking data to directly capture paths (see e.g., Rossmo et al., 2012), avoiding the need to infer them from activity nodes and enabling examination of the association between path attributes such as frequency and recency and the likelihood of committing crime along or near the path.

Lastly, further research is needed to refine GP-SMART and test the utility of applying the theoretical model in the field, be it statistically through GP-SMART or clinically (as described above). Opportunities for further algorithmic development include: creating a GP-SMART style add on to crime-based geographic profiling algorithms (see above); developing GP-SMART for use with crime series (see Chapter 7); and extending GP-SMART to other typically whodunit crime types. But only through field-based evaluation will we know if the enhancements to geographic profiling practice proposed in this thesis truly improve the advice that geographic profiling analysts provide to investigations to increase the probability or speed of apprehending an offender. As various commentators have pointed out, it is relatively easy to test the accuracy of single geographic profiling inferences with solved cases—as exemplified in Chapter 7—but much harder to evaluate the impact of the total package of inferences and advice constituting a 'geographic profile' (Berezowski et al., 2021; Goodwill et al., 2014; Rich & Shively, 2004). It remains an important next step in validating the value of this thesis' findings nonetheless.

Conclusion

In crime investigations, geographic profiling involves inferring information about 'whodunit' from information about where and when the crime occurred. Such inferences are possible because the 'spatial signature' of a crime location reflects the 'mental map' of the offender: the places they know from their everyday activities. This thesis aimed to enhance geographic profiling by developing, validating, and applying a theoretical model that systematises the links between the activity locations in people's mental maps and their crime locations. The theoretical model adds to our existing understanding of these links by identifying the attributes of people's past activity locations that influence their future crime locations, and specifying the psychological mechanisms through which this influence occurs: through the development of reliable and relevant knowledge of crime opportunities. Leveraging a large, national dataset sourced from police records, this research also added to the empirical evidence base supporting the role of the theorised attributes and mechanisms in crime location choice. Importantly, the model enables predictions about the activity locations near which people are more likely to commit crime; the thesis operationalised these predictions in a novel geographic profiling suspect prioritisation method, the Geographic Profiling: Suspect Mapping and Ranking Technique (GP-SMART). GP-SMART exemplifies how the theoretical model can be applied algorithmically to enhance existing methods of geographic profiling; other examples were elaborated in this chapter. Given the ability of geographic profiling to contribute to “faster investigation, substantial cost savings and possibly fewer victims” (Rossmo, 2021, p. 115), improving the accuracy of geographic profiling inferences from ‘wheredunit’ to ‘whodunit’ could have significant practical benefits. Future research should therefore explore whether the geographic profiling enhancements developed in this thesis can improve investigative outcomes in practice.

REFERENCES FOR CHAPTERS 1 AND 8

Ackerman, J. M., & Rossmo, D. K. (2015). How far to travel? A multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology*, *31*(2), 237–262. https://doi.org/10.1007/s10940-014-9232-7

Alston, J. D. (1994). *The serial rapist's spatial pattern of target selection* [Masters thesis, Simon Fraser University]. http://summit.sfu.ca/item/5080

Bache, R., Crestani, F., Canter, D., & Youngs, D. (2008, August). *A Bayesian decay model for suspect prioritisation based on geographical profiling* [Paper]. 14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Las Vegas, NV. http://eprints.hud.ac.uk/id/eprint/8054/

Barreda, D. S. (2020). The application of Newton and Swoope's geographical profile to serial killers. *Journal of Investigative Psychology and Offender Profiling*, *n/a*(n/a). https://doi.org/10.1002/jip.1566

Beauregard, E., Proulx, J., Rossmo, D. K., Leclerc, B., & Allaire, J.-F. (2007). Script analysis of the hunting process of serial sex offenders. *Criminal Justice and Behavior*, *34*(8), 1069–1084. https://doi.org/10.1177/0093854807300851

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*(1), 119–132. https://doi.org/10.1177/0093854806290161

Bennell, C., Taylor, P. J., & Snook, B. (2007). Clinical versus actuarial geographic profiling strategies: A review of the research. *Police Practice and Research*, *8*(4), 335–345. https://doi.org/10.1080/15614260701615037

Berezowski, V., MacGregor, D., Ellis, J., Moffat, I., & Mallett, X. (2021). More than an offender location tool: Geographic profiling and body deposition sites. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-021-09475-6

Bernasco, W. (2010). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology*, *48*(2), 389–416. https://doi.org/10.1111/j.1745-9125.2010.00190.x

Bernasco, W. (2017). Modeling offender decision making with secondary data. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook on offender decision making* (pp. 569–586). Oxford University Press. http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199338801.001.0001/oxfordhb-9780199338801-e-28

Bernasco, W. (2019). Adolescent offenders' current whereabouts predict locations of their future crimes. *PLOS ONE*, *14*(1), e0210733. https://doi.org/10.1371/journal.pone.0210733

Bernasco, W., Block, R., & Ruiter, S. (2013). Go where the money is: Modeling street robbers' location choices. *Journal of Economic Geography*, *13*(1), 119–143. https://doi.org/10.1093/jeg/lbs005

Bichler, G., Orosco, C. A., & Schwartz, J. A. (2012). Take the car keys away: Metropolitan structure and the long road to delinquency. *Journal of Criminal Justice*, *40*(1), 83–93. https://doi.org/10.1016/j.jcrimjus.2011.12.002

Brantingham, P. J., & Brantingham, P. L. (2012). The theory of target search. In F. T. Cullen & P. Wilcox (Eds.), *The Oxford Handbook of Criminological Theory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199747238.013.0028

Brantingham, P. J., & Brantingham, P. L. (2016). The geometry of crime and crime pattern theory. In R. K. Wortley & M. Townsley (Eds.), *Environmental Criminology and Crime Analysis* (2nd ed.). Routledge.

Brantingham, P. L., & Brantingham, P. J. (1991). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (2nd ed., pp. 27–54). Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993a). Environment, routine, and situation: Toward a pattern theory of crime. In R. V. Clarke & M. Felson (Eds.), *Routine activity and rational choice* (pp. 259–294). Transaction Publishers.

Brantingham, P. L., & Brantingham, P. J. (1993b). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Burcher, M., & Whelan, C. (2019). Intelligence-led policing in practice: Reflections from intelligence analysts. *Police Quarterly*, *22*(2), 139–160. https://doi.org/10.1177/1098611118796890

Burrell, A., & Bull, R. (2011). A preliminary examination of crime analysts' views and experiences of comparative case analysis. *International Journal of Police Science & Management*, *13*(1), 2–15. https://doi.org/10.1350/ijps.2011.13.1.212

Campbell, J. (2019). *A spatial classification of criminal offenders: Moving beyond circle theory with an agent-based model approach* [MSc, George Mason University]. http://jbox.gmu.edu/bitstream/handle/1920/11536/Campbell_thesis_2019.pdf?sequence=1&isAllowed=y

Canter, D. (1977). *The psychology of place*. Architectural Press.

Canter, D. (1995). *Criminal shadows: Inside the mind of the serial killer*. HarperCollins.

Canter, D. (2000). Offender profiling and criminal differentiation. *Legal and Criminological Psychology*, *5*(1), 23–46. https://doi.org/10.1348/135532500167958

Canter, D. (2005). Confusing operational predicaments and cognitive explorations: Comments on Rossmo and Snook et al. *Applied Cognitive Psychology*, *19*(5), 663–668. https://doi.org/10.1002/acp.1143

Canter, D. (2008). Geographical offender profiling: Using insights from practical applications to enhance theoretical explorations. In D. Youngs (Ed.), *The behavioural analysis of crime: Studies in David Canter's investigative psychology* (pp. 249–256). Ashgate. http://eprints.hud.ac.uk/id/eprint/14188/

Canter, D. (2011). Resolving the offender "profiling equations" and the emergence of an investigative psychology. *Current Directions in Psychological Science*, *20*(1), 5–10. https://doi.org/10.1177/0963721410396825

Canter, D., Bennell, C., Alison, L. J., & Reddy, S. (2003). Differentiating sex offences: A behaviorally based thematic classification of stranger rapes. *Behavioral Sciences & the Law*, *21*(2), 157–174. https://doi.org/10.1002/bsl.526

Canter, D., Coffey, T., Huntley, M., & Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, *16*(4), 457–478. https://doi.org/10.1023/A:1007551316253

Canter, D., & Gregory, A. (1994). Identifying the residential location of rapists. *Journal of the Forensic Science Society*, *34*(3), 169–175. https://doi.org/10.1016/S0015-7368(94)72910-8

Canter, D., & Hammond, L. (2006). A comparison of the efficacy of different decay functions in geographical profiling for a sample of US serial killers. *Journal of Investigative Psychology and Offender Profiling*, *3*(2), 91–103. https://doi.org/10.1002/jip.45

Canter, D., & Hammond, L. (2007). Prioritizing burglars: Comparing the effectiveness of geographical profiling methods. *Police Practice and Research*, *8*(4), 371–384. https://doi.org/10.1080/15614260701615086

Canter, D., Hammond, L., Youngs, D., & Juszczak, P. (2013). The efficacy of ideographic models for geographical offender profiling. *Journal of Quantitative Criminology*, *29*(3), 423–446. https://doi.org/10.1007/s10940-012-9186-6

Canter, D., & Larkin, P. (2008). The environmental range of serial rapists. In D. Canter & D. Youngs (Eds.), *Applications of geographical offender profiling* (pp. 57–68). Ashgate.

Canter, D., & Youngs, D. (2008a). Geographical offender profiling: Applications and opportunities. In D. Canter & D. Youngs (Eds.), *Applications of geographical offender profiling* (pp. 1–18). Ashgate.

Canter, D., & Youngs, D. (2008b). Geographical offender profiling: Origins and principles. In D. Canter & D. Youngs (Eds.), *Principles of geographical offender profiling* (pp. 1–18). Ashgate.

Canter, D., & Youngs, D. (2008c). Interactive Offender Profiling System (IOPS). In S. Chainey & L. Tompson (Eds.), *Crime Mapping Case Studies* (pp. 153–160). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470987193.ch18

Canter, D., & Youngs, D. (2009). *Investigative psychology: Offender profiling and the analysis of criminal action* (1 edition). Wiley.

Casady, T. (2008). ‘Rolling the dice’: The arrest of Roosevelt Erving in Lincoln, Nebraska. In S. Chainey & L. Tompson (Eds.), *Crime Mapping Case Studies* (pp. 63–68). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470987193.ch8

Ceccato, V. (2014). The nature of rape places. *Journal of Environmental Psychology*, *40*, 97–107. https://doi.org/10.1016/j.jenvp.2014.05.006

Chen, P., & Lu, Y. (2020). Journey-to-crime and offender’s geographic background: A comparison between migrant and native offenders in Beijing. *SN Social Sciences*. https://doi.org/10.1007/s43545-020-00038-w

Chopin, J., & Beauregard, E. (2020). Scripting extrafamilial child sexual abuse: A latent class analysis of the entire crime-commission process. *Child Abuse & Neglect*, *106*, 104521. https://doi.org/10.1016/j.chiabu.2020.104521

Chopin, J., & Caneppele, S. (2018). The mobility crime triangle for sexual offenders and the role of individual and environmental factors. *Sexual Abuse*, *31*(7), 812–836. https://doi.org/10.1177/1079063218784558

Chopin, J., & Caneppele, S. (2019). Geocoding child sexual abuse: An explorative analysis on journey to crime and to victimization from French police data. *Child Abuse & Neglect*, *91*, 116–130. https://doi.org/10.1016/j.chiabu.2019.03.001

Clare, J., Fernandez, J., & Morgan, F. (2009). Formal evaluation of the impact of barriers and connectors on residential burglars’ macro-level offending location choices. *Australian & New Zealand Journal of Criminology*, *42*(2), 139–158. https://doi.org/10.1375/acri.42.2.139

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608. https://doi.org/10.2307/2094589

Cope, N. (2004). ‘Intelligence led policing or policing led intelligence?’: Integrating volume crime analysis into policing. *British Journal of Criminology*, *44*(2), 188–203. https://doi.org/10.1093/bjc/44.2.188

Cornish, D. B., & Clarke, R. V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. Springer-Verlag.

Costello, A., & Wiles, P. (2001). GIS and the journey to crime: An analysis of patterns in South Yorkshire. In K. J. Bowers & A. Hirschfield (Eds.), *Mapping and analysing crime data: Lessons from research and practice* (pp. 27–60). Taylor & Francis.

Coupe, T. (2017). Burglary decisions. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford Handbook on Offender Decision Making* (pp. 655–683). Oxford University Press. http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199338801.001.0001/oxfordhb-9780199338801-e-28

Cullen, F. T., Eck, J. E., & Lowenkamp, C. T. (2002). Environmental corrections—A new paradigm for effective probation and parole supervision. *Federal Probation*, *66*(2), 28–37.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science*, *9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021a). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information*, *10*(2), 47. https://doi.org/10.3390/ijgi10020047

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021b). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART. *Journal of Investigative Psychology and Offender Profiling*. https://doi.org/10.1002/jip.1585

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review-a). *Familiar locations and similar activities: Examining the interaction of reliable and relevant knowledge in offenders' crime location choices*.

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review-b). *Relationships between offenders' crime locations and different prior activity locations as recorded in police data*.

Daniell, C. (2008). Geographic profiling in an operational setting: The challenges and practical considerations, with reference to a series of sexual assaults in Bath, England. In S. Chainey & L. Tompson (Eds.), *Crime Mapping Case Studies* (pp. 45–53). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470987193.ch6

Davies, A., & Dale, A. (1996). Locating the stranger rapist. *Medicine, Science and the Law*, *36*(2), 146–156. https://doi.org/10.1177/002580249603600210

Dern, H., Dern, C., Horn, A., & Horn, U. (2009). The fire behind the smoke: A reply to Snook and colleagues. *Criminal Justice and Behavior*, *36*(10), 1085–1090. https://doi.org/10.1177/0093854809344819

Doan, B., & Snook, B. (2008). A failure to find empirical support for the homology assumption in criminal profiling. *Journal of Police and Criminal Psychology*, *23*(2), 61–70. https://doi.org/10.1007/s11896-008-9026-7

Douglas, J., & Olshaker, M. (1998). *Mindhunter: Inside the FBI's elite serial crime unit*. Scribner.

Duwe, G., Donnay, W., & Tewksbury, R. (2008). Does residential proximity matter? A geographic analysis of sex offense recidivism. *Criminal Justice and Behavior*, *35*(4), 484–504. https://doi.org/10.1177/0093854807313690

Emeno, K., Bennell, C., Snook, B., & Taylor, P. J. (2016). Geographic profiling survey: A preliminary examination of geographic profilers' views and experiences. *International Journal of Police Science & Management*, *18*(1), 3–12. https://doi.org/10.1177/1461355715621070

Environmental Criminology Research Inc. (n.d.). *Rigel geographic profiling*. http://www.ecricanada.com/Rigel%20Geographic%20Profiling/

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. https://doi.org/10.1177/0093854812457921

Frank, R. (2012). SPORS: A suspect recommendation system based on offenders' reconstructed spatial profile. In N. Memon & D. Zeng (Eds.), *2012 European Intelligence and Security Informatics Conference* (pp. 38–45). CPS. https://doi.org/10.1109/EISIC.2012.26

Frith, M. J. (2019). Modelling taste heterogeneity regarding offence location choices. *Journal of Choice Modelling*, *33*, 100187. https://doi.org/doi.org/10.1016/j.jocm.2019.100187

Fumagilli, A., & Johnson, C. (2020). The role of geography and time in property crime. In Amy Burrell & M. Tonkin (Eds.), *Property crime: Criminological and psychological perspectives* (pp. 95–114). Routledge. https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=6129502

Garicano, L., & Heaton, P. (2010). Information technology, organization, and productivity in the public sector: Evidence from police departments. *Journal of Labor Economics*, *28*(1), 167–201. https://doi.org/10.1086/649844

Gerell, M. (2018). Quantifying the geographical (un)reliability of police data. *Nordisk politiforskning*, *5*(02), 157–171. https://doi.org/10.18261/issn.1894-8693-2018-02-05

Golledge, R., & Stimson, R. (1997). *Spatial behavior: A geographic perspective*. Guilford Press.

Goodwill, A. M., Lehmann, R. J. B., Beauregard, E., & Andrei, A. (2016). An action phase approach to offender profiling. *Legal and Criminological Psychology*, *21*(2), 229–250. https://doi.org/10.1111/lcrp.12069

Goodwill, A. M., van der Kemp, J. J., & Winter, J. M. (2014). Applied geographical profiling. In G. J. N. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 86–99). Springer. https://doi.org/10.1007/978-1-4614-5690-2_207

Gore, R. Z., Tofiluk, N., & Griffiths, K. (2005). Single incident geographical profiling. In F. Wang (Ed.), *Geographic information systems and crime analysis* (pp. 118–136). IGI Global. https://doi.org/10.4018/978-1-59140-453-8

Gould, P. (1973). On mental maps. In R. M. Downs & D. Stea (Eds.), *Image and environment: Cognitive maps and spatial behavior* (pp. 182–220). Aldine.

Gould, P., & White, R. (1986). *Mental maps* (2nd ed.). Routledge.

Haberman, C. P., Clutter, J. E., & Lee, H. (2021). A robbery is a robbery is a robbery? Exploring crime specificity in official police incident data. *Police Practice & Research*. https://doi.org/10.1080/15614263.2021.2009345

Hammond, L., & Youngs, D. (2011). Decay functions and criminal spatial processes: Geographical offender profiling of volume crime. *Journal of Investigative Psychology and Offender Profiling*, *8*(1), 90–102. https://doi.org/10.1002/jip.132

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, *27*(1), 6–35. https://doi.org/10.1177/0093854800027001002

Hart, T. C., Birks, D., Townsley, M., Ruiter, S., & Bernasco, W. (2020). Activity nodes, activity spaces, and awareness spaces: Measuring geometry of crime's constructs with smartphone data. In T. C. Hart, K. M. Lersch, & M. Chataway (Eds.), *Space, time, and crime* (5th ed., pp. 156–176). Carolina Academic Press.

Hassan, M. N., Rashidi, T. H., & Nassir, N. (2019). Consideration of different travel strategies and choice set sizes in transit path choice modelling. *Transportation (Dordrecht)*. https://doi.org/10.1007/s11116-019-10075-x

Hauge, M. V., Stevenson, M. D., Rossmo, D. K., & Le Comber, S. C. (2016). Tagging Banksy: Using geographic profiling to investigate a modern art mystery. *Journal of Spatial Science*, *61*(1), 185–190. https://doi.org/10.1080/14498596.2016.1138246

Hewitt, A. N. (2021). The importance of disaggregation in the spatial patterning of sexual crimes: A local analysis of spatial concentrations and similarities across victim and offense types. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-021-09481-8

Hewitt, A. N., Chopin, J., & Beauregard, E. (2020). Offender and victim 'journey-to-crime': Motivational differences among stranger rapists. *Journal of Criminal Justice*, *69*, 101707. https://doi.org/10.1016/j.jcrimjus.2020.101707

Home Office. (2018). *National law enforcement data programme: Law Enforcement Data Service (LEDS) – privacy impact assessment report*. Home Office.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721542/NLEDP_Privacy_Impact_Assessment_Report.pdf

Johnston, P. (2016). Book Review: Environmental corrections: A new paradigm for supervising offenders in the community. *Practice: The New Zealand Corrections Journal*, *4*(2), 67–68. https://www.corrections.govt.nz/resources/research/journal/volume_4_issue_2_december_2016/book_review_environmental_corrections_a_new_paradigm_for_supervising_offenders_in_the_community

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.

Knabe, S. (2008). *Geographic profiling under the microscope – a critical examination of the utility of geographic profiling and expert geographic profilers* [Masters thesis]. University of Liverpool.

Knabe-Nicol, S., & Alison, L. (2011). The cognitive expertise of geographic profilers. In L. Alison & L. Rainbow (Eds.), *Professionalizing offender profiling: Forensic and investigative psychology in practice* (pp. 126–159). Routledge.

Lammers, M. (2018). Co-offenders' crime location choice: Do co-offending groups commit crimes in their shared awareness space? *The British Journal of Criminology*, *58*, 1193–1211. https://doi.org/10.1093/bjc/azx069

Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology*, *53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071

LeBeau, J. L. (1987). The journey to rape: Geographic distance and the rapist's method of approaching the victim. *Journal of Police Science & Administration*, *15*, 129–136.

Leclerc, B., & Felson, M. (2016). Routine activities preceding adolescent sexual abuse of younger children. *Sexual Abuse*, *28*(2), 116-131. https://doi.org/10.1177/1079063214544331

Lemp, J. D., & Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A*, *46*(3), 602–613. https://doi.org/10.1016/j.tra.2011.11.004

Levine, N. (2013). *Crimestat IV: A spatial statistics program for the analysis of crime incident locations, version 4.0*. National Institute of Justice. https://nij.ojp.gov/library/publications/crimestat-iv-spatial-statistics-program-analysis-crime-incident-locations

Long, D., Liu, L., Xu, M., Feng, J., Chen, J., & He, L. (2021). Ambient population and surveillance cameras: The guardianship role in street robbers' crime location choice. *Cities*, *115*, 103223. https://doi.org/10.1016/j.cities.2021.103223

McFadden, D. (1977). *Modelling the choice of residential location* (No. 477; Cowles Foundation Discussion Papers). Yale University. https://EconPapers.repec.org/RePEc:cwl:cwldpp:477

Menting, B. (2018). Awareness × opportunity: Testing interactions between activity nodes and criminal opportunity in predicting crime location choice. *The British Journal of Criminology*, *58*, 1171–1192. https://doi.org/10.1093/bjc/azx049

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2016). Family matters: Effects of family members' residential areas on crime location choice. *Criminology*, *54*(3), 413–433. https://doi.org/10.1111/1745-9125.12109

Menting, B., Lammers, M., Ruiter, S., & Bernasco, W. (2020). The influence of activity space and visiting frequency on crime location choice: Findings from an online self-report survey. *The British Journal of Criminology*, *60*(2), 303–322. https://doi.org/10.1093/bjc/azz044

Mogavero, M. C., & Hsu, K.-H. (2017). Sex offender mobility: An application of crime pattern theory among child sex offenders. *Sexual Abuse*, *30*(8), 908–931. https://doi.org/10.1177/1079063217712219

O'Connor, C. D., Ng, J., Hill, D., & Frederick, T. (2021). Thinking about police data: Analysts' perceptions of data quality in Canadian policing. *The Police Journal*, 0032258X211021461. https://doi.org/10.1177/0032258X211021461

Öhrn, M. (2019). *We look at crime through the lens of geographic behaviour: Geographic profiling in operational settings* [Masters thesis, University of Gothenburg]. https://gupea.ub.gu.se/handle/2077/59807

O'Leary, M. (2009). The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, *6*(3), 253–265. https://doi.org/10.1002/jip.111

Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology*, *4*(4), 1239–1262.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, *6*, 8166–8173. https://doi.org/10.1038/ncomms9166

Paulsen, D. (2006a). Connecting the dots: Assessing the accuracy of geographic profiling software. *Policing: An International Journal*, *29*(2), 306–334. https://doi.org/10.1108/13639510610667682

Paulsen, D. (2006b). Human versus machine: A comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling*, *3*(2), 77–89. https://doi.org/10.1002/jip.46

Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). *Predictive Policing* [Product Page]. https://www.rand.org/pubs/research_reports/RR233.html

Porter, M. D., & Reich, B. J. (2012). Evaluating temporally weighted kernel density methods for predicting the next event location in a series. *Annals of GIS*, *18*(3), 225–240. https://doi.org/10.1080/19475683.2012.691904

Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency*, *43*(3), 261–291. https://doi.org/10.1177/0022427806286566

Rengert, G. (1996). *The geography of illegal drugs*. Westview Press.

Rengert, G., & Wasilchick, J. (1985). *Suburban burglary: A time and a place for everything*. C.C. Thomas.

Rich, T., & Shively, M. (2004). *A methodology for evaluating geographic profiling software*. National Institute of Justice. https://www.ncjrs.gov/pdffiles1/nij/grants/208993.pdf

Rossmo, D. K. (1995). *Geographic profiling: Target patterns of serial murderers* [PhD thesis]. Simon Fraser University.

Rossmo, D. K. (2000). *Geographic profiling*. CRC Press.

Rossmo, D. K. (2008). Place, space, and police investigations: Hunting serial violent criminals. In D. Canter & D. Youngs (Eds.), *Principles of geographical offender profiling* (pp. 149–163). Ashgate.

Rossmo, D. K. (2012). Recent developments in geographic profiling. *Policing: A Journal of Policy and Practice*, *6*(2), 144–150. https://doi.org/10.1093/police/par055

Rossmo, D. K. (2014). Geographic profiling. In G. J. N. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 1934–1942). Springer.

Rossmo, D. K. (2021). Dissecting a criminal investigation. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-021-09434-1

Rossmo, D. K., Laverty, I., & Moore, B. (2005). Geographic profiling for serial crime investigation. In F. Wang (Ed.), *Geographic information systems and crime analysis* (pp. 60–83). IGI Global. https://doi.org/10.4018/978-1-59140-453-8

Rossmo, D. K., Lu, Y., & Fang, T. B. (2012). Spatial-temporal crime paths. In M. A. Andresen & J. B. Kinney (Eds.), *Patterns, prevention, and geometry of crime* (pp. 3–15). Routledge.

Rossmo, D. K., & Rombouts, S. (2008). Geographic profiling. In R. Wortley & L. Mazerolle (Eds.), *Environmental criminology and crime analysis* (pp. 136–149). Willan. https://www.taylorfrancis.com/books/e/9781136308451

Rossmo, D. K., & Summers, L. (2015). Routine Activity Theory in crime investigation. In *The Criminal Act* (pp. 19–32). Palgrave Macmillan, London. https://doi.org/10.1057/9781137391322_3

Rossmo, D. K., & Velarde, L. (2008). Geographic profiling analysis: Principles, methods and applications. In S. Chainey & L. Tompson (Eds.), *Crime mapping case studies* (pp. 33–43). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470987193.ch5

Ruiter, S. (2017). Crime location choice. In W. Bernasco, J.-L. Van Gelder, & H. Elffers (Eds.), *The Oxford handbook of offender decision making* (pp. 398–420). Oxford University Press.

Ruiter, S., & Davies, T. (2018, July). *BTW, a test of crime pattern theory*. Environmental Criminology and Crime Analysis Symposium, Spain.

Schaefer, L., Cullen, F. T., & Eck, J. E. (2016). *Environmental corrections: A new paradigm for supervising offenders in the community*. Sage.

Schaefer, L., & Little, S. (2019). A quasi-experimental evaluation of the "environmental corrections" model of probation and parole. *Journal of Experimental Criminology*, *16*(4), 535–553. https://doi.org/10.1007/s11292-019-09373-2

Serin, R. C., Chadwick, N., & Lloyd, C. D. (2019). Integrating dynamic risk assessment into community supervision practice. In *The Wiley international handbook of correctional psychology* (pp. 725–743). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119139980.ch45

Serin, R. C., Gobeil, R., Lloyd, C. D., Chadwick, N., Wardrop, K., & Hanby, L. (2016). Using dynamic risk to enhance conditional release decisions in prisoners to improve their outcomes. *Behavioral Sciences & the Law*, *34*(2–3), 321–336. https://doi.org/10.1002/bsl.2213

Silver, N. (2012). *The signal and the noise: The art and science of prediction*. Allen Lane, Penguin Books.

Smallbone, S. W., & Wortley, R. K. (2000). *Child sexual abuse in Queensland: Offender characteristics & modus operandi*. Queensland Crime Commission.

Snaphaan, T., & Hardyns, W. (2019). Environmental criminology in the big data era. *European Journal of Criminology*. https://doi.org/10.1177/1477370819877753

Snook, B., Taylor, P. J., & Bennell, C. (2005). Shortcuts to geographic profiling success: A reply to Rossmo (2005). *Applied Cognitive Psychology*, *19*(5), 655–661. https://doi.org/10.1002/acp.1142

Snook, B., Wright, M., House, J. C., & Alison, L. J. (2006). Searching for a needle in a needle stack: Combining criminal careers and journey-to-crime research for criminal suspect prioritization. *Police Practice and Research*, *7*(3), 217–230. https://doi.org/10.1080/15614260500432972

Solymosi, R., & Bowers, K. J. (2018). *The role of innovative data collection methods in advancing criminological understanding* (G. J. N. Bruinsma & S. D. Johnson, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.35

Spaulding, J. (2020). *Spatio-temporal analysis of crime incidents for forensic investigation* [West Virginia University]. https://researchrepository.wvu.edu/etd/7791

Svobodová, J. (2018). *Bayesian models in geographic profiling*. Masaryk University. http://arxiv.org/abs/1805.02993

Tayebi, M. A., Glässer, U., Brantingham, P. L., & Shahir, H. Y. (2017). SINAS: Suspect investigation using offenders' activity space. *Machine Learning and Knowledge Discovery in Databases*, 253–265. https://doi.org/10.1007/978-3-319-71273-4_21

Tonkin, M., & Weeks, M. J. (2021). Crime linkage practice in New Zealand. *Journal of Criminological Research, Policy and Practice*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JCRPP-01-2020-0013

Townsley, M. (2016). Offender mobility. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (pp. 142–161). Routledge.

Townsley, M., Birks, D., Bernasco, W., Ruiter, S., Johnson, S. D., White, G., & Baum, S. (2015). Burglar target selection: A cross-national comparison. *Journal of Research in Crime and Delinquency*, *52*(1), 3–31. https://doi.org/10.1177/0022427814541447

Townsley, M., Birks, D., Ruiter, S., Bernasco, W., & White, G. (2016). Target selection models with preference variation between offenders. *Journal of Quantitative Criminology*, *32*(2), 283–304. https://doi.org/10.1007/s10940-015-9264-7

Trinidad, A., Vozmediano, L., Ocáriz, E., & San-Juan, C. (2020). "Taking a walk on the wild side": Exploring residence-to-crime in juveniles. *Crime & Delinquency*, 0011128720916141. https://doi.org/10.1177/0011128720916141

van Daele, S. (2009). Itinerant crime groups: Mobility attributed to anchor points? In L. Pauwels, P. Ponsaers, G. Vande Walle, T. Vander Beken, F. Vander Laenen, G.

Vermeulen, M. Cools, S. De Kimpe, B. De Ruyver, & M. Easton (Eds.), *Contemporary issues in the empirical study of crime* (Vol. 1, pp. 211–225). Maklu.

van der Kemp, J. J. (2021). The modus via of sex offenders and the use of geographical offender profiling in sex crime cases. In N. Deslauriers-Varin & C. Bennell (Eds.), *Criminal Investigations of Sexual Offenses: Techniques and Challenges* (pp. 33–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-79968-7_4

Van Koppen, P. J., & Jansen, R. W. J. (1998). The road to the robbery: Travel patterns in commercial robberies. *British Journal of Criminology*, *38*, 230–246. https://doi.org/10.1093/oxfordjournals.bjc.a014233

van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A time for a crime: Temporal aspects of repeat offenders' crime location choices. *Journal of Research in Crime and Delinquency*, *55*(4), 538–568. https://doi.org/10.1177/0022427818766395

van Sleeuwen, S. E. M., Ruiter, S., & Steenbeek, W. (2021). Right place, right time? Making crime pattern theory time-specific. *Crime Science*, *10*(1), 1–10. https://doi.org/10.1186/s40163-021-00139-8

Vandeviver, C., & Bernasco, W. (2020). "Location, location, location": Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*, *36*(4), 779–821. https://doi.org/10.1007/s10940-019-09431-y

Vandeviver, C., Neutens, T., van Daele, S., Geurts, D., & Vander Beken, T. (2015). A discrete spatial choice model of burglary target selection at the house-level. *Applied Geography*, *64*(Supplement C), 24–34. https://doi.org/10.1016/j.apgeog.2015.08.004

Verity, R., Stevenson, M. D., Rossmo, D. K., Nichols, R. A., & Le Comber, S. C. (2014). Spatial targeting of infectious disease control: Identifying multiple, unknown sources. *Methods in Ecology and Evolution*, *5*(7), 647–655. https://doi.org/10.1111/2041-210X.12190

Woodhams, J., & Toye, K. (2007). An empirical test of the assumptions of case linkage and offender profiling with serial business robberies. *Psychology, Public Policy, and Law*, *13*(1), 59–85. https://doi.org/10.1037/1076-8971.13.1.59

Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., & Lai, Y.-C. (2017). Universal model of individual and population mobility on diverse spatial scales. *Nature Communications*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-01892-8

Yesberg, J. A., & Polaschek, D. L. L. (2015). Assessing dynamic risk and protective factors in the community: Examining the validity of the Dynamic Risk Assessment for Offender Re-entry. *Psychology, Crime & Law*, *21*(1), 80–99. https://doi.org/10.1080/1068316X.2014.935775

APPENDIX A

Supplementary Materials to Chapter 3

The following materials are available online as supplementary materials to:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information, 10*(2), 47. https://doi.org/10.3390/ijgi10020047

S1. DATA PARAMETER DEFINITIONS

Table S1.1 provides definitions and technical specifications for the parameters applied in extracting the datasets used in this research.

Table S1.1. Definitions of the parameters used to extract the data.

Variable	Definition	NIA fields
Residential burglary	Burglary of residential properties (not commercial or public properties).	ANZSOC Division = 07 Unlawful Entry With Intent/Burglary, Break and Enter. Location Type[a] = Residential
Non-residential burglary	Burglary of non-residential properties (i.e., commercial or public properties).	ANZSOC Division = 07 Location Type[a] = not Residential or Unknown
Commercial robbery	Robbery of commercial or properties (e.g., petrol stations, shops, banks)	ANZSOC Subdivision = Robbery. Location Type[a] = Commercial
Personal robbery	Robbery of people (e.g., on the street, at transit hubs, at schools), also known as street robbery	ANZSOC Subdivision = Robbery. Location Type[a] = not Commercial or Unknown
Extra-familial sex offenses	Indecent acts, assaults and sexual violations including rape not involving family members.	ANZSOC Division = 03 Sexual Assault and Related Offenses. Excluded: Offense codes for non-contact or family-specific sex offenses, where the Offense Code Description contained the words: OBJECTIONABLE, RECORDING, PUBLISH, PUBLICATION, TOUR, INCEST, SPOUSE, or FAMILY. Excluded: occurrences with a Family Violence flag.
Location ID	Unique reference created to identify and count unique locations.	Coordinates of a location, or if there are no coordinates, the address string, or if there is no address string, the Location ID field.
Offender	A person believed by Police to have committed a given offense on the basis of sufficient evidence to initiate a charge or other proceeding (e.g., written or verbal warning).	Person-Offense Link Type = Offender, Cleared Offender (this refers to the Police 'clearing' – solving - the case; it does not refer to the offender being cleared of guilt) or Youth Aid Offender.
Non-offender role	A person involved in an offense or non-crime incident but not as an offender, nor as a suspect against whom there is insufficient evidence to proceed.	Person-Offense Link Type = Victim, Complainant, Informant, Witness, Subject, Child or Young Person Exposed to Family Harm, Predominant Aggressor, Primary Victim, Mutual Participant. Excluded: Suspect, Offender/Cleared Offender/Youth Aid Offender, Other (a catch all category sometimes used as an alternative to Suspect).
Home address	Residential address where a person is recorded as living, or having lived, for any period of time.	Address Type = Bail Address, Boarding House Of, Boards At, Contact Address, E M Bail Address, Home Address, Leases, MOJ Address, Place Known to Sleep, Postal Address, Previous Address, Rented By, Rents, Residence Of, Resides At, Sleeps At, 86, 338, 423, 491, 1366 (numeric codes are invalid address types; these numeric codes were checked and are indicative of home addresses).
Family home address	Residential address where persons linked to the offender as family members, either as an intimate partner or relative, is recorded as living, or having lived, for any period time during the offender's lifetime. Addresses limited to the offender's childhood years before memories are formed were extremely rare (<2% of records) and thus retained as	Person-Person Link Type = 1. Immediate family (Child, Identical Twin, Next Of Kin, Parent, Sibling, Step child, Step parent, Step Sibling). 2. Intimate Partner (Boyfriend/girlfriend, De-facto, Ex Married/Partner Living Together, Ex partner, Ex Partner Not Living Together, Ex-boyfriend/girlfriend, Married, Partner, Partner - Living together, Partner - Not Living Together)

	unlikely to make any difference to the results of the present analyses.	3. Other relative (Care Giver, Caregiver - Other Relative, Foster Parent, Grandchild, Grandparent, In Custody Of, Legal Guardian, Other Relative, Relative, Under Care Of, Under care of - Other Relative, Under Foster Care Of). Excluded: extra-familial caregivers (often institutional or medical staff), friend, associate, stranger, known to each other. Address record 'End Date' >= the offender's date of birth.
Education	School, alternative education facility or tertiary institution the offender is recorded as having attended.	Education screen > Education Institution/Other Name = not Null. Or: Employment screen > Employer Name = not Null, and contains the words: INTERMEDIATE, SECONDARY, NORMAL, PRIMARY, KURA, TKKM, COLLEGE, YMCA, GRAMMAR, PROGRAMME, COLLEGE, EDUCATION, LEARNING, TRUST, COURSE, CPIT, BACHELOR, INSTITUTE, TRAINING, TURANGA, POUTAMA, UNIVERSITY, WANANGA, DIPLOMA, DEGREE, POLYTECH, UNITEC, or CAMPUS. Inspection of Employer Names and checks of age at association dates showed these key words to be a reliable indicator that these school/education details had been entered into the employment instead of education screen and were not reflective of employment as teachers.
Non-traffic offenses	Offenses that do not relate to driving behavior or other traffic violations. Traffic offenses were excluded because initial checks showed that their inclusion would introduce an unacceptable number of geographically imprecise records, considering the spatial units of analysis (exact coordinates and SA1 census units). For these codes, an estimated 40% or more of the addresses were street-only rather than address specific and tended to involve long streets, the geocoded coordinates of which could fall hundreds of meters from the actual place the offense occurred.	Offense code = 1000-9999. Excluded: Traffic Offense codes (those preceded by A through Z).
Non-crime incidents	Incidents to which the Police respond that are not crimes/offenses, and not traffic or rescue related (for the same reasons as excluding traffic offenses).	Incident code = 1A – 1Z, 6A-6Z. Excluded: Traffic and land/water rescue codes: 1I, 1Q, 1U, 1V, 6F, 6I, 1L, 1W.

[a] Location Type categories were: Residential (e.g., private dwellings and associated outbuildings/grounds, residential construction sites, farms, rest homes), Commercial (e.g., shops, restaurants, bars, entertainment facilities, businesses, factories, malls, offices, gyms, motels, hotels, hostels, campgrounds, commercial construction sites), Public (e.g., school/education, sports facilities/grounds, hospitals and medical clinics, transit stations, police stations, courts, prisons, church/religious facility, community buildings, marae), Street (e.g., streets, roads, footpaths, parks and open spaces, car parking facilities, wharf, harbor, beach, lake, river, sea), Transit (e.g., bus, car, train, boat, ship, plane in transit) and Unknown (i.e., not able to be coded as any of above, e.g., "other", "miscellaneous", "unspecified", "unknown", "online"). Location type was coded using a custom rule applied to a range of indicators because there is no single field in the NIA database that is a reliable indicator of location type (covering the entire data period). The rule was tested against a series of alternatives in order to identify the version that most accurately identified location type, by checking the narrative descriptions of a random sample of 1% of the reference offenses of each crime type. The optimal rule achieved 95%-98% accuracy across the reference offense categories (burglary, robbery and sex offenses). Reference tables mapping the NIA Scene Type and Address Type codes (e.g., farm, restaurant, hospital) to the present location type categories (residential, commercial, etcetera), and the rule incorporating these and other location type indicators, as an R Script, are available from the corresponding author.

S2. DATA FILTERS AND SAMPLE ATTRITION

Data filtering steps

The following steps were taken to filter out inaccurate or imprecise data. First, offenses/incident codes which do not reliably establish the offender's presence at the offense/incident address were excluded, because they could not be assumed to represent the offender's activity space. This step is described in greater detail in the following section. Second, offenses/incidents recorded with start and end timestamps (within which the exact time of the offense/incident could not be established) covering more than 30 days were excluded. This threshold minimized data loss (under 2% of records). Third, offenses/incidents with no coordinates, or whose coordinates were not located in a census unit ('SA1') due to data entry errors were excluded. Addresses recorded against generic locations such as suburbs or cities as indicated by keywords such as 'NFA'/'no fixed abode' and 'unknown' in the address string were also excluded, as their coordinates were insufficiently precise to link the offender to a specific location. Finally, offenses/incidents relating to offenders who no longer appeared in the reference offense dataset (having been excluded in any of the above steps) were removed from the activity node datasets. Table S2.1 shows the number and proportion of records excluded by each filter. Sample sizes are expressed as the number of unique person-locations, where a location is either a discrete event in space and time (offenses/incidents) or an address without a specific related event record (offender, family and education addresses).

Table S2.1. Sample size per dataset following exclusions.

Exclusions	Reference offenses[a]	Offenses[a]	Incidents	Addresses	Family addresses	Education
Initial N	144,839	842,582*	647,548	1,501,873	2,462,611	35,470
			New N (-%)			
Offender not present	-	802,614 (-4.7%)	499,449 (-22.9%)	-	-	-
Long timespan	142,267 (-1.8%)	793,224 (-1.2%)	493,282 (-1.2%)	-	-	-
No coordinates	140,765 (-1.1%)	773,120 (-2.5%)	484,957 (-1.7%)	1,375,571 (-8.4%)	2,162,276 (-12.5%)	33,931 (-4.3%)
No SA1	140,707 (-0.04%)	772,629 (-0.06%)	484,662 (-0.06%)	1,372,004 (-0.3%)	2,156,828 (-0.3%)	33,931 (-0%)
Imprecise address	140,670 (-0.02%)	772,474 (-0.02%)	484,516 (-0.03%)	1,352,327 (-1.4%)	2,135,972 (-1.0%)	-
Offender not in ref. offense data	-	749,428 (-3.1%)	454,462 (-6.6%)	1,282,080 (-5.2%)	2,014,328 (-5.7%)	32,831 (-3.2%)

[a] The 'reference offense' dataset included all *possible* reference offenses i.e., all burglaries, robberies and extra-familial sex offenses committed between 2009 and 2018, from which the most recent per offender and offense type were identified. The 'offenses' dataset included the 'reference offense' dataset plus all other offenses committed between 2004 and 2018. For each individual offender, there is no more than one reference offense per offense type (burglary, robbery, or extra-familial sexual offense). All prior offenses were committed before the reference offense. For offenders who committed multiple types of offenses (burglary and/or robbery and/or extra-familial sexual offense) a particular offense (say a burglary on January 16, 2016) can be included as a burglary reference offense and also as a prior offense of a reference robbery or a reference sexual crime committed subsequently (e.g., on March 2, 2017).

Exclusion of codes not reliably indicating the offender's presence

Small samples of all codes in each of the categories shown in Table S2.2 were checked against NIA records to determine whether the offender was present at the offense/incident address. These categories were then excluded because:

1. For offenses (committed by the offender), less than 75% involved the offender being present at the recorded address.
2. For incidents (offenses involving the offender in a non-offending role or non-crime incidents), less than 75% involved the offender being present at the recorded address.

Including these offenses would have meant overestimating activity space. But excluding them does not result in a large underestimation of activity space: of the offenses in these categories which do involve the person's presence, many occur at their home address which is already captured in the address dataset. Further, over 80% of the records excluded from the prior incidents dataset for this reason were breaches of bail, which are predominantly recorded against the person's home address or against the location of the offense for which they are on bail. Both locations are already in the address and prior offense datasets respectively. Table S2.2 lists the codes excluded with this filter.

Table S2.2. Codes excluded from offenses (O) or incidents (I) offender not present.

Category	Code excluded	From
Make/publish/supply objectionable material	2792, 2793, 2794, 2795, 2946, 2961, 2965, 2966, 2967, 6563	O, I
Possess/make/publish intimate visual recording	2991, 2992, 2993	O, I
Other publication offense (e.g., breach name suppression)	3811, 3816, 7444	I
Harmful digital communication	1765	O
Telephone based offense	6551, 6552, 6554, 6555, 6559	O
Post based offense	6569	O
Unauthorized computer access	4631, 4632, 4634, 4636	O, I
Harassment	1841, 1849, 6553	O
Receiving stolen goods	4413, 4417, 4418, 4419, 4422, 4423, 4429	O, I
Fraud involving loss by deception	4553, 4554, 4555, 4556, 4557, 4558, 4567	O
Forgery/altering documents	4511, 4512, 4531, 4532, 4538, 4541, 4542, 4575, 4576, 4581	I
Unspecified "other" incident	1Z	I
Complaint	6A	I
Breach of police/court bail	7193, 7194, 6B, 6D, 6E	O, I
Fail to attend court or program	3746, 3832, 3835, 3855, 3855, 7191, 7192, 7955, 7956	O, I

S3. GEOCODING PROCESS, ACCURACY AND PRECISION

For all data except the education records, geocoding of addresses takes place within the NIA database which involves addresses being matched against a regularly updated address database and allocated Easting and Northing coordinates which were extracted as part of the datasets. Of a random sample of 400 records (100 from each dataset) with coordinates checked for geocoding accuracy, 98% were geocoded to the correct address (95% CI: 96%-99%), 99% to the correct SA1 (95% CI: 98%-99.8%). 'Street records' based on street name rather than specific addresses (constituting 24%/10%/7%/2% of the address/offense/incident/family address samples respectively, 89% overall) were deemed accurate when the coordinates fell on an address on the correct street and at least part of the street fell within the same SA1/SA2 as the coordinates.

The potential imprecision of street records was measured as the maximum distance between the record's map point and any point along the street (or street segment within the postcode or suburb specified in the address string). The median potential imprecision was 805.5 meters (95% CI 309 – 1083, n=31), thus even if street record points were maximally imprecise, the actual location of the offense/incident/address would typically fall within 1km of the SA1 into which the record fell.

The school/education data was geocoded in R using a custom function that queried the name of the school/institution in Google Maps using an API and returned the top matching address and its latitude and longitude. The coordinates were then transformed into Easting and Northing using ArcGIS. Checks of a random sample of school/education records with coordinates (n=372) showed 95% were geocoded to the correct address and SA1 (95% CI: 93%-97%).

S4. USE OF DATE INFORMATION IN IDENTIFYING 'PRIOR' ACTIVITY NODES

Offenses and incidents

Offenses/incidents recorded as having occurred between a start and end time-range were deemed to be 'prior' to the reference offense when the 'end date' (the latest point in time that the prior offense/incident could have occurred) pre-dated the reference offense 'start date' (the earliest date on which the reference offense could have occurred). This approach ensured that only offenses/incidents that definitely preceded the reference offense counted as prior activity nodes.

Address, family address and education records

The address, family address and education records included dates indicating the timeframe for which the information was current. Records with 'start dates' on or after the reference offense 'start date' were excluded as not preceding the reference offense. Issues with the way these date fields are used were managed as follows.

First, the start date often reflects the date that Police became aware of the address, school, etcetera rather than the true start of the individual's association with that node. To the extent of this error, it was accepted that the duration of a person's association with that node would be underestimated, or that some nodes would be missed as the Police did not find out about them until after the reference offense.

Second, some home, family home, work and 'unspecified' node records were not updated with end dates and so would still appear current after the association with the node has ended. To the extent of this error, the recency of the node would be overestimated. On inspection, the 'unspecified' node records tended to reflect one-off events such as arresting or noting the presence of the individual at a given location. Unspecified node records with no end date were therefore allocated end dates the same as the record's start date to reflect their nature as singular events rather than enduring associations.

Third, with education records the opposite error appeared: start and end dates were frequently the same date. To the extent of this error, the recency of the node would be underestimated. This was addressed by 1) end-dating education records with the start of any subsequent education record for the same individual, then 2) end-dating the most recent school record whichever was the latter of the record's end date or the date on which the offender turned 18, the typical school leaving age in New Zealand, and tertiary education records to the date on which the offender turned 22 (based on a 3 year bachelor degree, typical of NZ tertiary institutions). School and tertiary education records were distinguished using key words in the name of the institution. To the extent that the offender cohort are more likely than the general population to leave school or tertiary study at an earlier age, education node recency and duration are likely to be overestimated, though this error is believed to be less than the underestimation error from using the original end dates.

Lastly, outlier record start dates were detected in the offender address and family address datasets. Records with start dates that were blank (n=12/169), prior to 1950 (n= 6/8) or greater than the end date (n=10/5) were allocated the record end date and vice versa (i.e., the dates of records with start and end dates apparently the wrong way around were switched).

S5. OFFENSE DISTANCE AND ACTIVITY SPACE SIMILARITY CORRELATIONS

Table S5.1. Kendall's tau-b correlations between reference offense distance and activity space similarity measures, by reference offense, Territorial Authority type and pre-offense timeframe. For each correlation, N=10,000 offender pairs and $p < 0.001$.

Reference offense	TA Type	Time	Co-offenders[a]	Percent shared activity space	Minimum node distance	Median nearest node distance
Res. Burg.	Urban	0-1y	0.01%	-0.13	0.26	0.28
		ever	0.01%	-0.19	0.23	0.25
	Rural	0-1y	0.12%	-0.20	0.29	0.25
		ever	0.12%	-0.24	0.26	0.23
Non-res. Burg.	Urban	0-1y	0.02%	-0.11	0.22	0.25
		ever	0.02%	-0.16	0.20	0.22
	Rural	0-1y	0.30%	-0.21	0.29	0.24
		ever	0.30%	-0.24	0.26	0.22
Com. Rob.	Urban	0-1y	0.11%	-0.08	0.17	0.21
		ever	0.11%	-0.12	0.14	0.18
	Rural	0-1y	2.38%	-0.17	0.21	0.17
		ever	2.38%	-0.19	0.19	0.15
Pers. Rob.	Urban	0-1y	0.03%	-0.10	0.18	0.20
		ever	0.03%	-0.13	0.15	0.18
	Rural	0-1y	0.82%	-0.16	0.20	0.19
		ever	0.82%	-0.19	0.18	0.17
Sex Offense	Urban	0-1y[b]	0.01%	-0.07	0.24	0.23
		ever	0.01%	-0.13	0.21	0.20
	Rural	0-1y[b]	0.12%	-0.12	0.24	0.18
		ever	0.12%	-0.18	0.22	0.18

[a] Percentage of pairs who were co-offenders (committed the reference offense together).

[b] Excludes 1 and 5 pairs respectively where one of the offenders had no 1-year activity nodes.

S6. CO-OFFENDER AND OTHER SAME LOCATION OFFENDER COMPARISONS

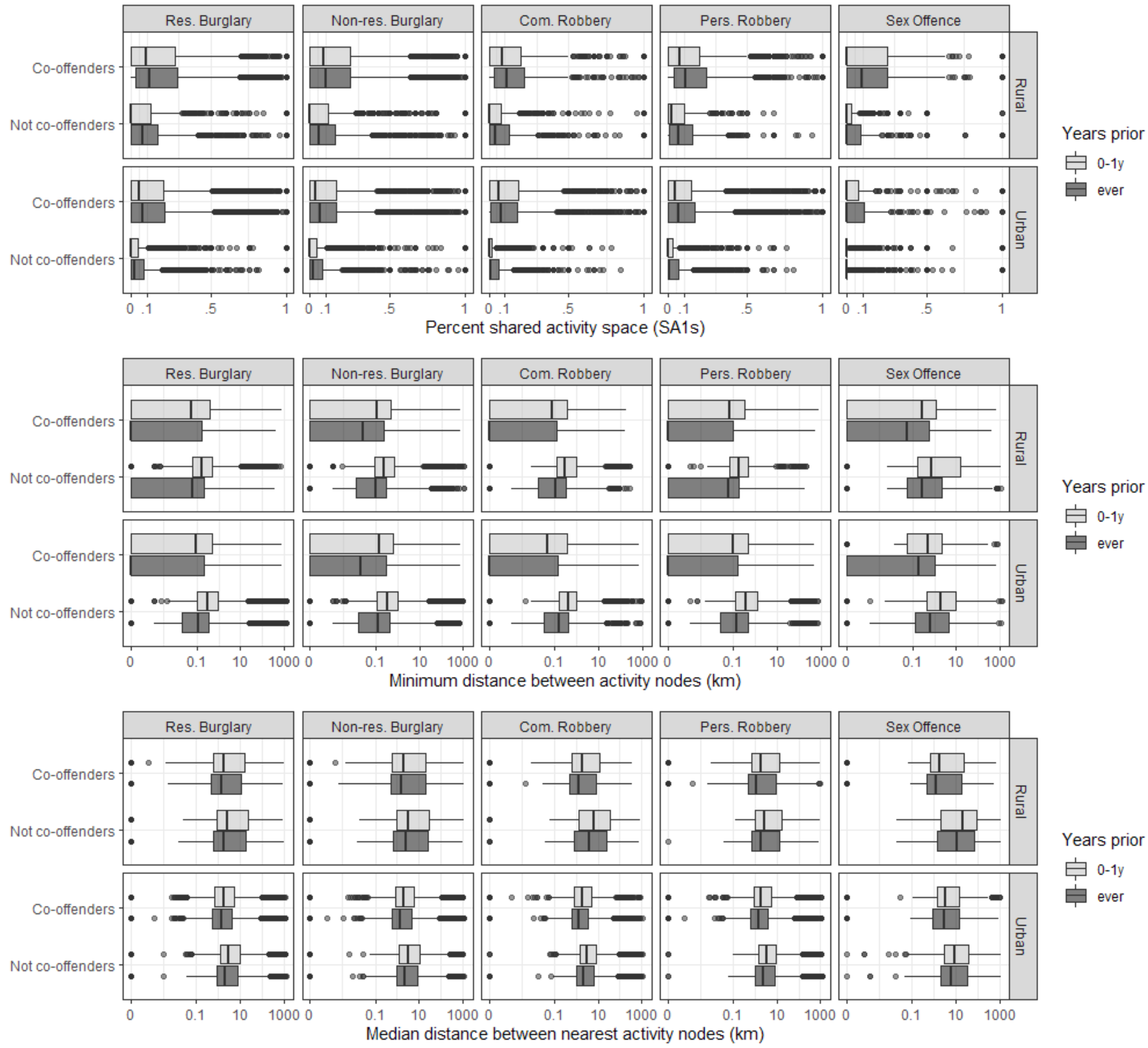


Figure S6.1. Distributions of percent shared activity space (top), minimum distance between activity nodes (middle) and median distance between nearest activity nodes (bottom) comparing co-offending pairs and pairs of other offenders whose reference offenses were within 100m, by reference offense, Territorial Authority type and pre-offense timeframe.

Table S6.1. Difference in activity space similarity between co-offending pairs and pairs whose reference offenses were within 100m, by reference offense, Territorial Authority type and pre-offense timeframe. All significant at $p < 0.001$. Effect size (r): 0.10 - 0.3 = small; 0.30 - 0.5 = moderate; >0.5 = large.

					Percent shared activity space (%)						**Minimum node distance (km)**						**Median nearest node distance (km)**					
Ref.	**TA**		**N pairs**		Median						Median						Median					
offense	**Type**	**Time**	Co-off.	not	Co.	not	U	95% CI a		r	Co.	not	U	95% CI b		r	Co.	not	U	95% CI b		r
Res. Burg.	Urban	0-1y	4096	5556	4.90	0.00	7674427	0.00	2.40	0.31	0.09	0.32	14972974	0.12	0.15	0.27	1.83	2.82	13722343	0.72	0.92	0.18
		ever	4096	5556	7.10	2.00	8092848	3.00	3.80	0.25	0.00	0.11	14429664	0.04	0.05	0.23	1.35	1.97	13317011	0.41	0.55	0.15
	Rural	0-1y	4068	4560	9.60	0.00	6927156	2.90	4.50	0.23	0.06	0.18	11541299	0.06	0.07	0.21	1.88	2.48	10252981	0.30	0.46	0.09
		ever	4068	4560	11.40	7.50	7675035	2.60	3.80	0.15	0.00	0.07	11178598	0.01	0.02	0.18	1.34	1.69	10137631	0.18	0.30	0.08
Non-res. Burg.	Urban	0-1y	3757	6127	3.40	0.00	8325850	0.00	0.00	0.27	0.14	0.35	14667767	0.13	0.15	0.23	1.89	3.06	13611936	0.67	0.89	0.15
		ever	3757	6127	6.30	2.00	8819688	1.70	2.60	0.20	0.02	0.13	14006861	0.03	0.05	0.19	1.42	2.15	13433862	0.43	0.58	0.14
	Rural	0-1y	4131	6254	8.30	0.00	10004000	0.00	1.90	0.20	0.12	0.23	15949354	0.08	0.09	0.20	1.93	3.37	14587890	0.42	0.61	0.11
		ever	4131	6254	10.00	5.70	10646645	2.00	3.20	0.15	0.03	0.10	15321334	0.02	0.03	0.16	1.63	2.53	14204968	0.24	0.38	0.08
Com. Rob.	Urban	0-1y	1755	3091	5.90	0.00	1598046	3.30	4.60	0.39	0.05	0.41	3890625	0.21	0.26	0.36	1.92	3.10	3365880	0.84	1.12	0.20
		ever	1755	3091	7.30	1.30	1659905	3.80	4.90	0.33	0.00	0.15	3777199	0.07	0.10	0.33	1.40	2.24	3330800	0.57	0.78	0.19
	Rural	0-1y	792	1009	8.10	0.00	279219	1.10	4.80	0.28	0.08	0.31	550021	0.13	0.19	0.33	1.96	6.29	497018	1.00	2.21	0.21
		ever	792	1009	11.10	4.10	285882	3.70	5.70	0.25	0.00	0.12	537557	0.05	0.07	0.30	1.34	3.78	486699	0.56	1.18	0.19
Pers. Rob.	Urban	0-1y	2866	4084	4.20	0.00	3900564	0.00	2.10	0.32	0.10	0.37	8004053	0.17	0.20	0.32	1.96	3.27	7052183	0.75	1.01	0.17
		ever	2866	4084	6.50	0.90	3796792	3.30	4.10	0.31	0.00	0.14	7883732	0.06	0.08	0.30	1.38	2.34	7109005	0.57	0.75	0.18
	Rural	0-1y	1527	786	7.70	2.30	453571	1.20	4.10	0.21	0.06	0.19	768891	0.06	0.09	0.23	1.80	2.53	666615	0.22	0.58	0.09
		ever	1527	786	10.70	6.90	475152	2.50	4.30	0.17	0.00	0.06	758138	0.01	0.03	0.23	1.11	1.79	686694	0.22	0.48	0.12
Sex Offense	Urban	0-1ya	277	1994	0.00	0.00	220641	0.00	0.00	0.19	0.54	1.86	368302	0.55	1.12	0.19	3.47	8.50	341909	1.66	3.65	0.13
		ever	277	1994	0.00	0.00	207133	0.00	0.00	0.18	0.19	0.67	362367	0.18	0.41	0.18	2.90	6.14	336305	1.05	2.43	0.12
	Rural	0-1ya	221	509	0.00	0.00	40583	0.00	0.00	0.27	0.26	0.75	73388	0.18	0.46	0.24	1.90	22.14	75747	2.98	14.86	0.28
		ever	221	509	9.60	0.00	39594	0.30	6.70	0.25	0.05	0.28	71977	0.06	0.16	0.22	1.25	11.30	73731	1.42	6.10	0.25

[a] 95% confidence interval around the median difference in percent shared activity space.

[b] 95% confidence interval around the median difference in distance.

APPENDIX B

Supplementary Materials to Chapter 4

The following materials are available online as supplementary materials to:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5. Reproduced with permission from Springer Nature.

Additionally, the R script for this paper is available at: https://static-content.springer.com/esm/art%3A10.1007%2Fs10940-021-09526-5/MediaObjects/10940_2021_9526_MOESM2_ESM.txt

Supplementary material to: 'The Importance of Importance Sampling: Exploring Methods of Sampling from Alternatives in Discrete Choice Models of Crime Location Choice'

Proportions of alternatives sampled per stratum in importance sampling

Table S1 below shows for each offense, the minimum, median and maximum sampling probabilities for each *stratum* (as opposed to strategy) in the stratified importance sampling strategies. The probability of inclusion in the 'within 5km' stratum was always 1, and the probability of selection in the 'remainder' strata is the probability of selection from the remaining alternatives. These results confirm that the sampling probabilities for the distance-based strata decrease over the increasing distance bands, as desired, to reflect our a priori belief that choice probability decreases with distance. They also show that in some cases, the importance sampling results in the inclusion of all SA2s within a given distance of activity nodes (indicated by instances with a sampling probability of 1). On average, the sampling probabilities per strata are small, but when combined with the SA2s with nodes in 5km, produce samples of 7-16% of the 2153 SA2s.

Table S1. Distribution of sampling probabilities for each importance sampling stratum

Offense	Sample stratum	Minimum	Median	Maximum
Res. Burg.	20 in 5 to 10km	0.061	0.260	1.000
	15 in 10 to 50km	0.014	0.052	1.000
	10 in 50 to 100km	0.010	0.050	1.000
	10 remainder[a]	0.005	0.008	0.104
	30 remainder	0.014	0.015	0.033
	55 remainder	0.026	0.028	0.060
	100 remainder	0.046	0.051	0.110
Non. Res. Burg.	20 in 5 to 10km	0.059	0.278	1.000
	15 in 10 to 50km	0.014	0.056	1.000
	10 in 50 to 100km	0.009	0.051	1.000
	10 remainder	0.005	0.007	0.189
	30 remainder	0.014	0.015	0.029
	55 remainder	0.026	0.028	0.052
	100 remainder	0.046	0.051	0.095
Com. Rob.	20 in 5 to 10km	0.062	0.217	1.000
	15 in 10 to 50km	0.013	0.049	1.000
	10 in 50 to 100km	0.009	0.048	0.769
	10 remainder[a]	0.005	0.008	0.064
	30 remainder	0.014	0.016	0.026
	55 remainder	0.026	0.029	0.048
	100 remainder	0.047	0.053	0.088
Pers. Rob.	20 in 5 to 10km	0.067	0.211	1.000
	15 in 10 to 50km	0.015	0.046	1.000
	10 in 50 to 100km	0.009	0.049	1.000

Offense	Sample stratum	Minimum	Median	Maximum
	10 remainder[a]	0.005	0.008	0.185
	30 remainder	0.014	0.016	0.027
	55 remainder	0.026	0.029	0.050
	100 remainder	0.047	0.052	0.091
Sex offense	20 in 5 to 10km	0.061	0.313	1.000
	15 in 10 to 50km	0.015	0.065	1.000
	10 in 50 to 100km	0.010	0.066	1.000
	10 remainder[a]	0.005	0.007	0.175
	30 remainder	0.014	0.015	0.029
	55 remainder	0.026	0.027	0.052
	100 remainder	0.046	0.049	0.095

[a] Proportion sampled out of all alternatives remaining after the first stratum (all SA2s with activity nodes within 5km) had been included. Proportions sampled out of all alternatives remaining after the second, third and fourth distance strata have been sampled are not shown.

APPENDIX C

Supplementary Materials to Chapter 5

The following materials are under review as supplementary materials to:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Relationships between offenders' crime locations and different prior activity locations as recorded in police data.

Supplementary Materials to 'Relationships Between Offenders' Crime Locations and Different Prior Activity Locations as Recorded in Police Data'

S1 Activity node type and opportunity variable definitions

Table S1.1. Node type variable definitions and data specification

Node type	Definition	NIA database specification
Home	Residential address where a person is recorded as living, or having lived, for any period of time.	Address history screen > Address Type = Bail Address, Boarding House Of, Boards At, Contact Address, EM Bail Address, Home Address, Leases, MOJ Address, Place Known to Sleep, Postal Address, Previous Address, Rented By, Rents, Residence Of, Resides At, Sleeps At, 86, 338, 423, 491, 1366 (numeric codes are invalid address types; these numeric codes were checked and are indicative of home addresses).
Family home	Residential address where persons linked to the offender as family members, either as an intimate partner or relative, is recorded as living, or having lived, for any period time during the offender's lifetime. Addresses limited to the offender's childhood years before memories are formed were extremely rare (<2% of records) and thus retained as unlikely to make any difference to the results of the present analyses.	Person-Person Link Type = 1. Immediate family (Child, Identical Twin, Next Of Kin, Parent, Sibling, Step child, Step parent, Step Sibling). 2. Intimate Partner (Boyfriend/girlfriend, De-facto, Ex Married/Partner Living Together, Ex partner, Ex Partner Not Living Together, Ex-boyfriend/girlfriend, Married, Partner, Partner - Living together, Partner - Not Living Together) 3. Other relative (Care Giver, Caregiver - Other Relative, Foster Parent, Grandchild, Grandparent, In Custody Of, Legal Guardian, Other Relative, Relative, Under Care Of, Under care of - Other Relative, Under Foster Care Of). Excluded: extra-familial caregivers (often institutional or medical staff), friend, associate, stranger, known to each other. Address history screen (for family member) > Address Type = as per Home above and address record 'End Date' >= the offender's date of birth.
School	School, alternative education facility or tertiary institution the offender is recorded as having attended.	Education screen > Education Institution/Other Name = not Null. Or: Employment screen > Employer Name = not Null, and contains the words: INTERMEDIATE, SECONDARY, NORMAL, PRIMARY, KURA, TKKM, COLLEGE, YMCA, GRAMMAR, PROGRAMME, COLLEGE, EDUCATION, LEARNING, TRUST, COURSE, CPIT, BACHELOR, INSTITUTE, TRAINING, TURANGA, POUTAMA, UNIVERSITY, WANANGA, DIPLOMA, DEGREE, POLYTECH, UNITEC, or CAMPUS. Inspection of Employer Names and checks of age at association dates showed these key words to be a reliable indicator that these school/education details had been entered into the employment

Node type	Definition	NIA database specification
		instead of education screen and were not reflective of employment as teachers.
Work	Work address	Address history screen > Address Type = Alternate Work Address, Firearms Dealer - Place Of Business, Work Address. We also attempted to identify work addresses from Employer Name in the Employment screen but these typically only included a company name with no specific details of location, branch or any other address, precluding sufficiently complete or accurate geocoding to enable their inclusion.
Prior offence	Offences that do not relate to driving behaviour or other traffic violations. Traffic offences were excluded because initial checks showed that their inclusion would introduce an unacceptable number of geographically imprecise records.	Offence code = 1000-9999. Excluded: Traffic Offence codes (those preceded by A through Z). Person-Offence Link Type = Offender, Cleared Offender (this refers to the Police 'clearing' – solving - the case; it does not refer to the offender being cleared of guilt) or Youth Aid Offender. Or Address history screen > Address Type = Obtained Drugs From, Place Drugs obtained by, Operates Illegal Activities At
Victim/ witness	Offences in which the offender was involved but not as an offender, nor as a suspect against whom there is insufficient evidence to proceed.	Offence code = as per Prior offence above. Person-Offence Link Type = Victim, Complainant, Informant, Witness, Subject, Child or Young Person Exposed to Family Harm, Primary Victim. Excluded: Suspect, Offender/Cleared Offender/Youth Aid Offender, Other (a catch all category sometimes used as an alternative to Suspect).
Incident	Incidents to which the Police respond that are not crimes/ offences, and not traffic or rescue related (for the same reasons as excluding traffic offences).	Incident code = 1A – 1Z, 6A-6Z. Excluded: Traffic and land/water rescue codes: 1I, 1Q, 1U, 1V, 6F, 6I, 1L, 1W. Person-Offence Link Type = Complainant, Informant, Witness, Subject, Child or Young Person Exposed to Family Harm, Predominant Aggressor, Primary Victim, Mutual Participant.
Other	Other police contacts or sightings.	Address history screen > Address Type = Arrested At, Decamped From, Died At, Found At, Frequented By, Frequents, Keyholder, Last Drank At, Other(Specify), Owns, Place Arrested For, Place Decamped By, Place Found, Place Seen, Place Spoken To, Place Stopped For, Place trespassed from, Recovered, Seen At, Spoken To At, Stopped At, Trespassed from, Unknown – NIS, Visited At, 45, 46, 47, 90, 91, 96, 177, 178, 180, 909, 1340, 1424 (numeric codes are invalid address types; these numeric codes were checked and are indicative of these other types of addresses).

Table S1.2. Opportunity variable definitions and data specification

Offence	Variable	Census/Business demography specification
Residential Burglary	Number of residential dwellings	NZ Census > Number of private dwellings. Census 2013 data was used for reference offences occurring from 2009-2015 and census 2018 data were used for 2016-2018 reference offences because a small number of SA2s' residential populations increased or decreased drastically between Census years. Marlborough SA2s were adjusted as per Statistics New Zealand guidance and corrections available at: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/b09ae9c0-de19-4418-952b-c9e4cbf7e1a2. For reference, Residential Burglaries were those where the Location Type of the offence was Residential*.
Non-residential Burglary	Number of business units	Business demography statistics > Industry Category = All. Business demography statistics remained consistent over the data period so 2018 was used for simplicity. For reference, Non-residential Burglaries were those where the Location Type was not Residential or Unknown*.
Commercial Robbery	Number of commercial business units	Business demography statistics > Industry Category = G Retail Trade, H Accommodation and Food Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services. M Professional, Scientific and Technical Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services, M Professional, Scientific and Technical Services, N Administrative and Support Services, R Arts and Recreation Services, S Other Services. For reference, Commercial Robberies were those where the Location Type of the offence was Commercial*.
Personal Robbery & Extrafamilial Sex Offences	Number of commercial and public business units	Business demography statistics > Industry Category = As for Commercial Robbery plus I Transport, Postal and Warehousing, J Information Media and Telecommunications, O Public Administration and Safety, P Education and Training, Q Health Care and Social Assistance. For reference, Personal Robberies were those where the Location Type was not Commercial or Unknown*.

* Location Type categories were: Residential (e.g., private dwellings and associated outbuildings/grounds, residential construction sites, farms, rest homes), Commercial (e.g., shops, restaurants, bars, entertainment facilities, businesses, factories, malls, offices, gyms, motels, hotels, hostels, campgrounds, commercial construction sites), Public (e.g., school/education, sports facilities/grounds, hospitals and medical clinics, transit stations, police stations, courts, prisons, church/religious facility, community buildings, marae), Street (e.g., streets, roads, footpaths, parks and open spaces, car parking facilities, wharf, harbour, beach, lake, river, sea), Transit (e.g., bus, car, train, boat, ship, plane in transit) and Unknown (i.e., not able to be coded as any of above, e.g., "other", "miscellaneous", "unspecified", "unknown", "online"). Location type was coded using a custom rule applied to a range of indicators because there is no single field in the NIA database that is a reliable indicator of location type (covering the entire data period). The rule was tested against a series of alternatives in order to identify the version that most accurately identified location type, by checking the narrative descriptions of a random sample of 1% of the reference offences of each crime type. The optimal rule achieved 95%-98% accuracy across the reference offence categories (burglary, robbery and sex offences). Reference tables mapping the NIA Scene Type and Address Type codes (e.g., farm, restaurant, hospital) to the present location type categories (residential, commercial, etcetera), and the rule incorporating these and other location type indicators, as an R Script, are available from the corresponding author.

S2 Descriptive statistics

Tables S2.1 to S2.4 present a range of descriptive statistics describing the offenders, the number of nodes per offender, the proportion of sampled SA2 alternatives with activity nodes in them or nearby, and the distribution of opportunity variables for the sampled SA2 alternatives. Non-parametric statistics are used because all distributions were positively skewed.

Table S2.1 presents demographic statistics for the offenders for each crime type.

Table S2.1. Offender demographic statistics

Reference offence	**Median Age (IQR)**	**% Male**	**% Female**
Residential Burglary	21 (14)	83.3%	16.6%
Non-residential Burglary	18 (12)	88.0%	12.0%
Commercial Robbery	19 (8)	87.7%	12.3%
Personal Robbery	19 (10)	80.7%	19.3%
Extrafamilial Sex Offences	28 (24)	96.8%	3.2%

Table S2.2 shows the average number of SA2s per person that contained each type of node, for each crime type. The minimum for all node types and crime types was 0.

Table S2.2. Median, IQR and Maximum number of SA2s per person containing each node type

Node type	**Res. Burg.**			**Non-res. Burg.**			**Com. Rob.**			**Pers. Rob.**			**Sex offences**		
	Med	**IQR**	**Max**	**Med**	**IQR**	**Max**	**Med**	**IQR**	**Max**	**Med**	**IQR**	**Max**	**Med**	**IQR**	**Max**
Home	4	4	54	3	4	54	4	4	28	4	4	54	3	3	44
Family home: immediate	5	6	86	5	6	77	7	6	89	7	7	73	1	4	71
Family home: intimate partner	3	7	175	1	6	151	2	5	98	3	7	117	0	4	135
Family home: other relative	0	4	114	0	3	85	0	5	72	0	6	120	0	0	57
School	0	1	6	0	1	6	0	1	5	0	1	7	0	0	5
Work	0	0	6	0	0	5	0	0	2	0	0	3	0	0	3
Prior offence	4	4	79	3	5	98	5	4	56	5	4	73	1	2	99
Prior victim/ witness	1	1	32	1	1	23	1	2	16	1	2	21	0	1	42
Prior incident	1	1	17	1	1	18	1	1	13	1	1	14	1	1	23
Other location	4	4	57	3	5	88	4	5	61	5	4	58	1	3	92

Table S2.3 provides descriptive statistics for the node type variables as a proportion of all SA2 alternatives sampled for each crime type. To illustrate, for the residential burglaries, 2.4 percent of all potentially chosen SA2 units contained a home node of the offender, for 1.6 percent the nearest home node was 0-200m away, for 2.5 percent the nearest home node was 200-500m away and for 18.3 percent the nearest home node was 2-5km away.

Table S2.3. Proportion of sampled SA2s for which the nearest activity node was within a given distance band, for each type of node and each distance band

Node type	**Distance band**	**Res. Burg.**	**Non-res. Burg.**	**Com. Rob.**	**Pers. Rob.**	**Sex offences**
Home	In the SA2	0.024	0.023	0.019	0.020	0.025
	0-200m	0.016	0.016	0.014	0.015	0.018
	200-500m	0.025	0.024	0.022	0.023	0.028
	500-1000m	0.040	0.039	0.037	0.038	0.046
	1-2km	0.074	0.073	0.073	0.072	0.088
	2-5km	0.183	0.182	0.199	0.185	0.235
Family home: immediate	In the SA2	0.032	0.033	0.033	0.033	0.023
	0-200m	0.021	0.022	0.023	0.022	0.016
	200-500m	0.031	0.032	0.034	0.033	0.023
	500-1000m	0.048	0.050	0.052	0.051	0.036
	1-2km	0.085	0.089	0.094	0.092	0.066
	2-5km	0.204	0.216	0.231	0.222	0.163
Family home: intimate partner	In the SA2	0.029	0.024	0.019	0.026	0.021
	0-200m	0.019	0.017	0.014	0.018	0.014
	200-500m	0.028	0.024	0.020	0.027	0.021
	500-1000m	0.043	0.037	0.032	0.041	0.033
	1-2km	0.076	0.066	0.059	0.072	0.060
	2-5km	0.176	0.153	0.147	0.171	0.145
Family home: other relative	In the SA2	0.013	0.013	0.015	0.015	0.009
	0-200m	0.009	0.009	0.010	0.010	0.006
	200-500m	0.013	0.014	0.016	0.016	0.010
	500-1000m	0.021	0.021	0.025	0.025	0.016
	1-2km	0.038	0.039	0.046	0.046	0.029
	2-5km	0.095	0.097	0.115	0.113	0.073
School	In the SA2	0.002	0.002	0.002	0.002	0.001
	0-200m	0.001	0.002	0.002	0.002	0.001
	200-500m	0.003	0.003	0.003	0.003	0.002
	500-1000m	0.005	0.006	0.006	0.006	0.003
	1-2km	0.011	0.012	0.015	0.014	0.007
	2-5km	0.032	0.035	0.045	0.044	0.021
Work	In the SA2	0.000	0.000	0.000	0.000	0.000
	0-200m	0.000	0.000	0.000	0.000	0.000

Node type	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
	200-500m	0.000	0.000	0.000	0.000	0.001
	500-1000m	0.000	0.000	0.000	0.000	0.001
	1-2km	0.001	0.001	0.001	0.001	0.003
	2-5km	0.003	0.003	0.003	0.002	0.009
Prior offence	In the SA2	0.023	0.024	0.024	0.023	0.016
	0-200m	0.017	0.018	0.019	0.018	0.012
	200-500m	0.023	0.024	0.026	0.024	0.017
	500-1000m	0.037	0.039	0.042	0.039	0.028
	1-2km	0.068	0.071	0.080	0.072	0.054
	2-5km	0.159	0.161	0.194	0.175	0.137
Prior victim/ witness	In the SA2	0.006	0.006	0.007	0.006	0.006
	0-200m	0.005	0.005	0.006	0.005	0.005
	200-500m	0.008	0.008	0.009	0.008	0.007
	500-1000m	0.014	0.014	0.015	0.015	0.013
	1-2km	0.028	0.027	0.032	0.030	0.026
	2-5km	0.078	0.076	0.095	0.086	0.074
Prior incident	In the SA2	0.007	0.006	0.005	0.006	0.008
	0-200m	0.005	0.005	0.004	0.005	0.006
	200-500m	0.008	0.008	0.007	0.008	0.010
	500-1000m	0.014	0.014	0.013	0.014	0.018
	1-2km	0.029	0.029	0.027	0.028	0.037
	2-5km	0.083	0.080	0.083	0.083	0.105
Other location	In the SA2	0.023	0.023	0.022	0.023	0.019
	0-200m	0.017	0.018	0.018	0.019	0.015
	200-500m	0.023	0.024	0.025	0.025	0.021
	500-1000m	0.037	0.038	0.040	0.040	0.034
	1-2km	0.070	0.071	0.078	0.075	0.066
	2-5km	0.168	0.168	0.199	0.184	0.171

Table S2.4 describes the distribution of the opportunity variable (i.e., number of dwellings or business units) for the sampled SA2s, for each crime type. The minimum for all crime types was 0 dwellings/business units.

Table S2.4. Median, IQR and maximum number of opportunity units per SA2

Offence type	**Median**	**IQR**	**Max**
Residential Burglary	1167	273	4707
Non-residential Burglary	306	123	4059
Commercial Robbery	195	90	3354
Personal Robbery	231	102	3822
Sex offences	231	102	3822

S3 Odds ratios for activity node associations

Table S3.1 shows the odds ratios (ORs) for the associations between activity node proximity and crime location choice, for each activity node type, distance band and crime type. Each column represents a separate model.

Table S3.1. Odds ratios (CIs) for activity node associations by node type, distance and crime

Node type	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
Home	In the SA2	53.97*** (49.64-58.68)	28.66*** (25.54-32.17)	19.43*** (14.89-25.34)	15.12*** (12.62-18.12)	235.28*** (207.72-266.5)
	0-200m	30.18*** (27.26-33.41)	28.06*** (24.72-31.85)	16.99*** (12.75-22.65)	13.09*** (10.76-15.92)	59.12*** (48.88-71.5)
	200-500m	24.65*** (22.38-27.15)	21.88*** (19.35-24.73)	17.07*** (13.05-22.35)	9.99*** (8.26-12.09)	57.36*** (48.26-68.18)
	500-1000m	20.65*** (18.83-22.65)	20.48*** (18.22-23.02)	15.77*** (12.27-20.27)	8.42*** (7.01-10.11)	48.19*** (41.23-56.33)
	1-2km	16*** (14.66-17.47)	13.78*** (12.3-15.44)	10.11*** (7.95-12.86)	6.98*** (5.87-8.3)	31.73*** (27.41-36.74)
	2-5km	10.62*** (9.79-11.53)	9.42*** (8.47-10.48)	6.82*** (5.46-8.51)	5.41*** (4.61-6.35)	17.3*** (15.13-19.77)
Family home: immediate	In the SA2	5.55*** (5.14-5.98)	8.17*** (7.37-9.06)	4.13*** (3.26-5.24)	5.18*** (4.4-6.09)	3.55*** (3.04-4.14)
	0-200m	3.84*** (3.48-4.23)	6.4*** (5.68-7.21)	4.12*** (3.18-5.34)	4.23*** (3.53-5.07)	2.9*** (2.32-3.64)
	200-500m	3.49*** (3.18-3.83)	5.88*** (5.23-6.6)	3.06*** (2.36-3.95)	4.08*** (3.42-4.86)	2.65*** (2.14-3.28)
	500-1000m	3.22*** (2.94-3.52)	4.95*** (4.43-5.55)	2.98*** (2.34-3.79)	3.9*** (3.3-4.61)	2.35*** (1.91-2.88)
	1-2km	2.79*** (2.56-3.04)	4.15*** (3.72-4.63)	3.22*** (2.58-4.01)	2.91*** (2.47-3.43)	2.52*** (2.1-3.04)
	2-5km	2.45*** (2.26-2.65)	3.35*** (3.02-3.72)	2.54*** (2.06-3.13)	2.47*** (2.12-2.88)	2.03*** (1.7-2.42)
Family home: intimate partner	In the SA2	2.46*** (2.29-2.65)	1.40*** (1.25-1.56)	2.78*** (2.18-3.53)	2.88*** (2.47-3.36)	3.17*** (2.68-3.74)
	0-200m	1.67*** (1.51-1.85)	1.61*** (1.41-1.83)	2.47*** (1.84-3.31)	2.02*** (1.66-2.45)	2.12*** (1.66-2.7)
	200-500m	1.6*** (1.45-1.76)	1.33*** (1.18-1.52)	2.27*** (1.73-2.99)	1.99*** (1.66-2.38)	1.7*** (1.34-2.16)
	500-1000m	1.51*** (1.38-1.65)	1.21** (1.07-1.37)	1.9*** (1.47-2.46)	1.93*** (1.63-2.28)	1.68*** (1.36-2.08)
	1-2km	1.57*** (1.45-1.71)	1.06 (0.94-1.19)	1.65*** (1.3-2.09)	1.65*** (1.41-1.93)	1.47*** (1.2-1.79)
	2-5km	1.38*** (1.28-1.49)	1.11* (1.01-1.23)	1.57*** (1.28-1.93)	1.52*** (1.32-1.76)	1.56*** (1.31-1.84)
Family home: other	In the SA2	2.02*** (1.84-2.22)	1.71*** (1.51-1.95)	1.70*** (1.29-2.23)	1.77*** (1.48-2.13)	2.59*** (2.05-3.27)
	0-200m	1.43***	1.48***	1.28	1.51***	1.92***

Node type	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
relative		(1.25-1.64)	(1.26-1.75)	(0.89-1.84)	(1.2-1.9)	(1.34-2.75)
	200-500m	1.38*** (1.22-1.57)	1.31*** (1.12-1.53)	1.34 (0.98-1.83)	1.4** (1.13-1.73)	1.18 (0.82-1.7)
	500-1000m	1.31*** (1.17-1.48)	1.18* (1.02-1.38)	1.35* (1.02-1.8)	1.39** (1.14-1.7)	1.34 (0.98-1.83)
	1-2km	1.35*** (1.21-1.5)	1.14 (1-1.31)	1.36* (1.06-1.75)	1.38*** (1.15-1.66)	1.62*** (1.24-2.12)
	2-5km	1.24*** (1.13-1.36)	1.06 (0.94-1.2)	1.28* (1.02-1.61)	1.23* (1.04-1.45)	1.47** (1.16-1.87)
School	In the SA2	1.79*** (1.59-2.02)	2.37*** (2.07-2.7)	1.41* (1.03-1.91)	1.53*** (1.22-1.93)	1.26 (0.89-1.78)
	0-200m	1.57*** (1.34-1.85)	1.06 (0.86-1.3)	1.26 (0.84-1.87)	1.59*** (1.21-2.1)	1.03 (0.63-1.67)
	200-500m	1.41*** (1.23-1.63)	1.02 (0.86-1.21)	0.92 (0.64-1.32)	1.4** (1.1-1.78)	0.73 (0.47-1.14)
	500-1000m	1.42*** (1.26-1.61)	1.19* (1.03-1.39)	0.88 (0.64-1.2)	1.49*** (1.21-1.83)	0.9 (0.62-1.3)
	1-2km	1.24*** (1.1-1.39)	0.95 (0.82-1.1)	0.81 (0.61-1.07)	1.13 (0.93-1.38)	0.74 (0.52-1.04)
	2-5km	1.14* (1.03-1.27)	0.92 (0.8-1.05)	0.76* (0.59-0.97)	1.06 (0.89-1.26)	0.71* (0.53-0.97)
Work	In the SA2 - 200m	1.13 (0.75-1.7)	2.94*** (1.92-4.53)	5.34*** (2.21-12.89)	2.43* (1.1-5.33)	4.69*** (3.07-7.18)
	200-1000m	0.88 (0.59-1.33)	1.33 (0.77-2.3)	1.74 (0.54-5.59)	1.43 (0.57-3.58)	1.56 (0.89-2.73)
	1-5km	0.84 (0.65-1.1)	1.29 (0.88-1.88)	1.1 (0.49-2.48)	2.51** (1.42-4.44)	1.14 (0.77-1.68)
Prior offence	In the SA2	6.95*** (6.44-7.5)	11.04*** (9.99-12.2)	5.77*** (4.58-7.27)	7.03*** (5.98-8.26)	2.51*** (2.14-2.95)
	0-200m	3.69*** (3.35-4.07)	4.17*** (3.67-4.73)	3.42*** (2.6-4.5)	4.62*** (3.85-5.55)	2.21*** (1.78-2.74)
	200-500m	3.53*** (3.22-3.88)	3.62*** (3.19-4.11)	3.05*** (2.34-3.99)	3.67*** (3.05-4.41)	1.74*** (1.4-2.17)
	500-1000m	3.15*** (2.89-3.45)	3.15*** (2.78-3.55)	3.05*** (2.38-3.9)	3.04*** (2.54-3.64)	1.72*** (1.41-2.11)
	1-2km	2.84*** (2.61-3.09)	2.87*** (2.56-3.22)	2.55*** (2.01-3.22)	2.58*** (2.18-3.06)	1.45*** (1.2-1.76)
	2-5km	2.17*** (2-2.35)	2.18*** (1.95-2.43)	2.52*** (2.03-3.11)	2.38*** (2.03-2.79)	1.26** (1.06-1.5)
Prior victim/ witness	In the SA2	1.28*** (1.19-1.39)	1.22*** (1.1-1.35)	1.17 (0.91-1.49)	1.36*** (1.16-1.6)	1.54*** (1.3-1.83)
	0-200m	1.04 (0.93-1.17)	1.06 (0.92-1.22)	1.13 (0.83-1.53)	1.04 (0.84-1.27)	1.26 (0.95-1.67)
	200-500m	1.01 (0.91-1.13)	0.91 (0.79-1.04)	1.05 (0.79-1.39)	1.1 (0.91-1.33)	1.1 (0.84-1.44)
	500-1000m	0.89* (0.81-0.99)	0.85* (0.74-0.96)	0.88 (0.68-1.15)	0.95 (0.8-1.14)	1.09 (0.86-1.38)
	1-2km	0.91* (0.83-0.99)	0.9 (0.8-1.01)	1.12 (0.89-1.4)	0.86 (0.73-1.01)	1.19 (0.97-1.46)
	2-5km	0.90**	0.88*	0.92	0.87	1.18

Node type	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
		(0.83-0.97)	(0.79-0.98)	(0.75-1.13)	(0.75-1.01)	(0.99-1.42)
Prior incident	In the SA2	1.31*** (1.21-1.41)	1.27*** (1.14-1.41)	1.35* (1.05-1.74)	1.36*** (1.16-1.6)	2.34*** (2.02-2.72)
	0-200m	1.16* (1.04-1.3)	1.06 (0.92-1.22)	1.01 (0.71-1.44)	1.13 (0.91-1.4)	1.69*** (1.32-2.15)
	200-500m	1.1 (1-1.23)	1.09 (0.95-1.25)	1.15 (0.84-1.57)	1.05 (0.86-1.28)	1.22 (0.96-1.56)
	500-1000m	1.08 (0.98-1.19)	0.98 (0.86-1.11)	1.17 (0.89-1.52)	1.02 (0.85-1.23)	1.40** (1.14-1.72)
	1-2km	1.03 (0.94-1.12)	1.01 (0.9-1.13)	0.99 (0.78-1.27)	1.11 (0.95-1.31)	1.15 (0.95-1.39)
	2-5km	1.06 (0.98-1.15)	0.97 (0.87-1.07)	1.06 (0.87-1.31)	0.97 (0.84-1.13)	1.09 (0.92-1.28)
Other location	In the SA2	1.7*** (1.58-1.84)	2.03*** (1.84-2.24)	3.00*** (2.39-3.77)	3.20*** (2.74-3.75)	2.07*** (1.76-2.44)
	0-200m	1.57*** (1.43-1.72)	1.63*** (1.44-1.84)	2.19*** (1.68-2.86)	2.35*** (1.96-2.82)	1.59*** (1.27-1.98)
	200-500m	1.58*** (1.44-1.72)	1.38*** (1.21-1.56)	1.78*** (1.37-2.33)	2.28*** (1.9-2.73)	1.79*** (1.45-2.21)
	500-1000m	1.44*** (1.32-1.57)	1.20** (1.07-1.35)	1.95*** (1.53-2.48)	2.17*** (1.83-2.58)	1.55*** (1.28-1.89)
	1-2km	1.34*** (1.23-1.45)	1.17** (1.05-1.31)	1.44** (1.14-1.82)	1.96*** (1.67-2.31)	1.47*** (1.22-1.76)
	2-5km	1.37*** (1.27-1.48)	1.21*** (1.09-1.35)	1.57*** (1.28-1.94)	1.58*** (1.35-1.84)	1.33*** (1.13-1.57)

*p < 0.05, ** p < 0.01, *** p < 0.001

S4 Wald test results

Table S4.1 shows the results of Wald Chi Square tests for statistical differences used to test hypotheses comparing Odds Ratios (ORs) within models where their 95% confidence intervals (CIs) overlapped. Comparisons appear in the order reported in the text. All variable 1 ORs are greater than variable 2 ORs.

Table S4.1. Results of Wald tests comparing ORs where CIs overlapped

H[a]	Offence	Variable 1	Variable 2	X^2
1	Com. Rob.	Family: other in SA2	Family: other 2-5km	2.79
1	Com. Rob.	Incident in SA2	Incident 2-5km	0.06
1	Pers. Rob.	School in SA2	School 2-5km	9.24**
1	Sex	School in SA2	School 2-5km	7.86**
1	Sex	Victim/Witness in SA2	Victim/Witness 2-5km	7.07**
1	Non. Res. Burg.	Family: IP 0-200m	Family: IP in SA2	4.42*
1	Pers. Rob.	School 0-200m	School in SA2	0.06
1	Non. Res. Burg.	Family: IP 2-5km	Family: IP 1-2km	0.76
1	Non. Res. Burg.	School 500m-1km	School 200-500m	2.87
1	Pers. Rob.	Work 1-5km	Work 200m-1km	1.57
1	Sex	Family: other 1-2km	Family: other 500m-1km	1.22
1	Sex	Incident 500m-1km	Incident 200-500m	0.95
3(a)	Com. Rob.	Family: immed. in SA2	Family: IP in SA2	0.16
3(a)	Sex	Family: immed. in SA2	Family: IP in SA2	1.06
3(b)	Sex	Family: immed. in SA2	Family: other in SA2	4.69*

*p < 0.05, ** p < 0.01
[a] Hypothesis to which the comparison is relevant.

APPENDIX D

Supplementary Materials to Chapter 6

The following materials are under review as supplementary materials to:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Familiar locations and similar activities: examining the interaction of reliable and relevant knowledge in offenders' crime location choices.

Online supporting information for: 'Familiar Locations and Similar Activities: Examining the Interaction of Reliable and Relevant Knowledge in Offenders' Crime Location Choices'

S1 Variable Construction: Methodological Detail

This appendix provides detailed descriptions of how the individual reliability and relevance variables, the reliability and relevance indices and interaction variable, and the opportunity covariates were constructed.

Individual Reliability and Relevance Variables

Reliability and relevance variables (e.g., recency, behaviour similarity) were calculated for each prior activity node (e.g., home address, family member's home address, location of one or more prior crimes). In developing coding rules, we distinguished between activity nodes that spanned a time period with start and end dates but no time-of-day information (offender and family home addresses, school, and work), and those that represented events that took place at a specified time and date (prior crimes, victim/witness events, non-crime incidents and the miscellaneous police contacts). Where crime, victim/witness and incident events were recorded as having occurred at an unknown time within a start and end timeframe, we used a random date-time within that timeframe to calculate similarity of timing to the reference offence. This method has been shown to predict actual offence timing most accurately in other studies (Ashby & Bowers, 2013; Boldt & Borg, 2016). Although the 'other police contacts' were entered with 'start' and 'end' dates in the same section of the database as the home addresses, preliminary checks showed the vast majority to relate to one-off events on the 'start' date, so they were treated as events.

Frequency was expressed as the proportion of days over the offender's association with a given node that it was visited. Offenders' home nodes were coded as 1 on the assumption that one visits home daily. Homes of their family members were coded as 0.07 representing fortnightly visits (26 visits per 365 days a year = 0.07) on the basis of research showing that most New Zealanders visit family members at least 'several times a month'

(ISSP Research Group, 2009). School/education nodes were coded as 0.55 (40 weeks at 5 days a week = 200 visits per 365 days = 0.55) based on the typical school year in New Zealand and work nodes as 0.66 (48 weeks at 5 days a week = 240 visits per 365 days = 0.66) based on the typical work year. For prior crime, incident and other police contact nodes the rate was calculated as the number of unique dates on which events were recorded at the location (i.e., frequency as a count) divided by the number of days between the first and last event at the location (i.e., duration).[1]

Recency represents how close in linear time the prior activities were to the reference offence. To create a variable representing temporal proximity, rather than difference, we took the reciprocal of the number of days between the reference offence and the most recent activity at that node. This created a continuous variable with maximum of 1 representing maximal recency with reference to the date of the reference offence. For example, for a home node that was current at the reference offence date (i.e., no end date or end date was after the reference offence) there was 1 day between the most recent activity and the reference offence (using 0 days would have resulted in dividing by zero), resulting in a reciprocal of 1/1=1. A node where a prior crime had been committed a week ago was 1/7 = 0.14. A home address recorded with an end date a year prior to the reference offence was 1/365 = 0.003.

Duration was calculated as the length of time in days between the first recorded association with a given activity node and the latest. For former home, family home, school

---

[1] For most event-based nodes, only one event was recorded at the location, resulting in a rate of 1/1=1 (maximum frequency), which is likely an overestimation of frequency for these nodes. We considered this acceptable given some overestimation was desirable to account for the fact that a single data point recorded by police likely captures only one of many visits to a location. Further, the overestimation is offset by the inclusion of the duration variable, for which these one-off event nodes score the minimum of 1 day.

and work nodes, duration was the number of days between the recorded start and end dates. For such nodes with end dates after the reference offence, indicating a current association, duration was the number of days between the recorded start date and the reference offence. For offence, incident, and other police contact nodes, duration was the length of time between the first and last offence/incident/contact at that location.

Behavioural similarity between prior activities and the reference offences was derived using the results of Kuang et al.'s (2017) study of the similarity of different crimes. Kuang et al. analysed text descriptions of offences in Los Angeles and calculated cosine similarity between different offence categories such as burglary, robbery, theft and assault. We used the similarity matrix presented in Kuang et al. (figure 6) to assign equivalent NZ offence categories similarity scores in relation to burglary and robbery, which were then grouped into three categories: same offence, similar offence, and not similar offence. Similarity categories in relation to sex offences and other offences and police callout incidents that were not reported in Kuang et al., were assigned a priori and checked for agreement among the present authors. Victim/witness nodes were also categorised as involving the same, similar or not similar type of offence/incident.

Table S1.1 below provides detail of the resulting similarity values for all prior crime and incident types. For burglary and robbery offences, similarity categories were assigned based on the similarity matrix of Kuang et al. (2017). For sex offences (which were not included in their analysis), other violent and interpersonal offences (e.g., abuse, threats and harassment) were considered similar and all other offences/incidents were considered not similar. A spreadsheet listing the specific offence and incident codes falling within each offence/incident category is available from the corresponding author. The resulting ordinal variable was: prior offending same (7), similar (6), not similar (5), prior victimization/witness offence/incident same (4), similar (3), not similar (2), and other activity node (1).

Table S1.1 Prior offence/incident behaviour similarity values relative to reference offence

Offence/ incident	**Burglary**	**Robbery**	**Sex offences**
Abuse/neglect*	Similar	Not similar	Similar
Assault	Not similar	Similar	Similar
Assault (weapon)	Not similar	Similar	Similar
Breach court order	Similar	Not similar	Not similar
Breach non-association order*	Similar	Not similar	Similar
Burglary	Same	Not similar	Not similar
Disorder*	Similar	Similar	Not similar
Drugs*	Not similar	Not similar	Not similar
Embezzlement	Not similar	Not similar	Not similar
Firearms	Similar	Similar	Similar
Forgery/documents	Not similar	Not similar	Not similar
Fraud/deception	Similar	Similar	Not similar
Harassment*	Not similar	Not similar	Similar
Homicide	Not similar	Similar	Similar
Kidnap	Not similar	Similar	Similar
Mental distress*	Not similar	Not similar	Not similar
Miscellaneous*	Similar	Similar	Not similar
Property damage	Similar	Not similar	Not similar
Resist/obstruct	Not similar	Similar	Similar
Robbery	Not similar	Same	Similar
Sex	Not similar	Similar	Same
Shoplift < $1000	Not similar	Not similar	Not similar
Shoplift > $1000	Not similar	Not similar	Not similar
Suspicious behaviour*	Not similar	Not similar	Not similar
Theft	Similar	Not similar	Not similar
Theft from person	Not similar	Similar	Not similar
Theft from vehicle	Similar	Not similar	Not similar
Theft of bicycle	Similar	Not similar	Not similar
Theft of vehicle	Not similar	Not similar	Not similar
Threats	Not similar	Not similar	Similar

Offence/ incident	Burglary	Robbery	Sex offences
Trespass	Similar	Not similar	Not similar
Weapon	Not similar	Similar	Similar

*Prior non-crime incidents falling under these headings were assigned to the victim/witness similarity categories as it is not possible to be an offender in relation to a non-crime incident.

Location similarity reflects the type of location (rather than geographical proximity) involved in a prior activity, by comparison of the prior activity location with the location of the reference offence. Locations refer to structures and places in the built environment that have a specific function, such as schools, stores, stations, or parks, and not to whole areas like neighbourhoods. Activity nodes and reference offences were categorised into 4 location types: residential, commercial premises, public premises, and street/open spaces/transit[2]. 'Residential' included private dwellings and associated outbuildings/grounds, residential construction sites, farms, rest homes. 'Commercial' included shops, restaurants, bars, entertainment facilities, businesses, factories, malls, offices, gyms, motels, hotels, hostels, campgrounds, commercial construction sites. 'Public' included school/education, sports facilities/grounds, hospitals and medical clinics, transit stations, police stations, courts, prisons, church/religious facility, community buildings, marae. 'Street/open space/transit' included streets, roads, footpaths, parks and open spaces, car parking facilities, wharf, harbour, beach, lake, river, sea, bus, car, train, boat, ship, plane. Comparing the location types of the activity nodes and reference offences resulted in an ordinal variable for location similarity: same location type (3), unknown similarity—where either the reference offence or the node was missing location type information so similarity could be either same or

[2] Location type was identified through a range of indicators because there is no single field in NIA for it. Location type was coded as missing when no indicator enabled it to be coded as any of above (e.g., where codes such as "miscellaneous", "other" or "online" were used).

different—(2), different location type (1).

Hour similarity represents how close in terms of hour of the day the prior activities were to the reference offence. To create a variable representing similarity, rather than difference, we took the reciprocal of the number of hours difference between the reference offence and the prior activity at each node. This created a continuous variable with maximum of 1 representing the same hour of the day (1/1=1, to avoid using 0 as a denominator). For example, a prior activity occurring at 13:00 and a reference offence at 01:00 are maximally different at 12 hours apart. This case would result in an hour similarity value of 1/13 = 0.08. Home and family home nodes, where the offender could have been present at any time of day or day of week, received an hour similarity value of 1. School and work nodes received an hour similarity value of 1 if the reference offence occurred during typical school (08:00-16:00) or work (08:00-18:00) hours, otherwise a similarity value based on the difference between the start or end of the typical school/work day and the time of the reference offence. For the 'other police contact' nodes, which did not include time of day information, hour similarity was imputed based on the median hour similarity over all nodes.

Day similarity represents how close in terms of day of the week the prior activities were to the reference offence. Again, to create a similarity rather than difference, we the reciprocal of the number of days of the week (from 1 indicating the same day of the week, to 4) between the prior activity and the reference offence. This created a variable with a maximum of 1 representing the same day of the week. For example, a prior activity occurring on Monday and a reference offence on Thursday are maximally different at 3 days apart. This case would result in a day similarity value of 1/4 = 0.25. Home and family home nodes, where the offender could have been present at any time of day or day of week, received a day similarity value of 1. School and work nodes received a day similarity value of 1 when the reference offence fell on a weekday, and 0.5 when it fell on a weekend.

Season similarity was likewise the reciprocal of the number of seasons (from 1 indicating the same season, to 3 for the opposite season) between the prior activity and the reference offence (with a maximum of 1 for same season). For example, if a home address spanned more than a year, similarity = 1. If an offender only lived at a temporary home address during spring, and the offence occurred in autumn, similarity = 1/2 = 0.5. If a prior offence occurred in summer and the reference offence occurred in winter, similarity = 1/3 = 0.33.

If multiple events occurred at the same node, yielding multiple values for each of the relevance (similarity) variables described above, the node was assigned the maximum of these values.

Reliability and Relevance Indices

Prior to aggregating the above variables into reliability and relevance indices, we checked whether each variable explained independent variance in offenders' crime location choices following a stepwise approach described in this section. The stepwise analysis followed the steps detailed in figure A1, which first involved converting the individual reliability and relevance variables to categorical variables so that each SA2 could be dummy coded based on the presence (1) or absence (0) of an activity node of a given category of a given variable, within a given distance band. The categories for each variable are shown in table A2. As with the main analyses, SA2s were only coded as 1 if no activity nodes were present in a shorter distance band, so that only the attributes of the nearest activity nodes are captured. Starting with a base model using the variable (all categories, all distance bands) explaining the most variance and iteratively adding whichever variable resulted in the most variance explained through its inclusion. Iterations continued until either all variables were included in the model or adding more variables did not explain significantly more variance (as measured by a likelihood ratio test). For residential and non-residential burglary, all variables added significantly more variance. For commercial robbery, all variables except duration added significantly more variance. For personal robbery and sex offences, all variables except day similarity added significantly more variance. Given the small number and lack of consistency of the exceptions, we elected to retain all variables in the main analysis, rather than introduce further complexity by varying the calculation of the reliability and relevance indices across crime types.

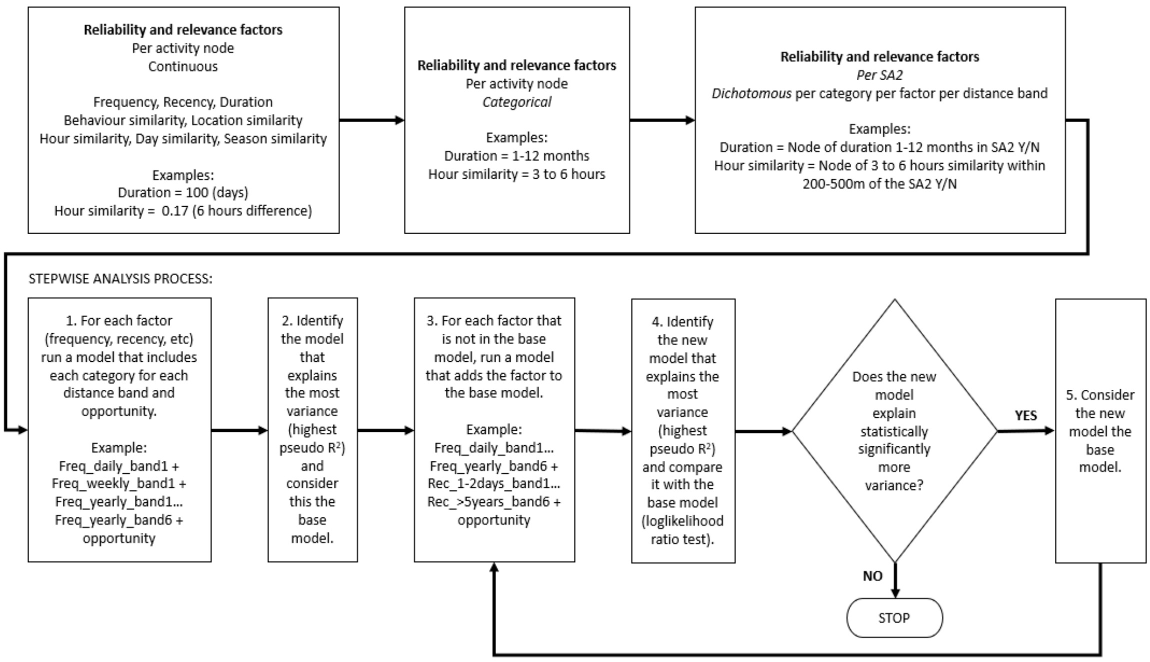


Figure S1.1 Process for stepwise analysis of individual variables

Table S1.2 Categories for each reliability and relevance variable

Variable	**Category**
Frequency	Weekly or more (0.14-1)
	Monthly (0.03-0.14)
	Yearly or less (<0.03)
Recency and duration	1-2 days
	3 to 30 days
	1 to 12 months
	1 to 5 years
	Over 5 years
Behaviour similarity	Prior crime of same crime type
	Prior crime of similar crime type
	Prior crime of not similar crime type
	Prior crime experienced as a victim/witness of same crime type

Variable	Category
	Prior crime or incident experienced as a victim/witness of similar type
	Prior crime or incident experienced as a victim/witness of not similar type
	Other activity node
	Behaviour similarity unknown[3]
Location similarity	Same location type
	Unknown similarity
	Different location type
Hour similarity	Within 0 to 2 hours
	Within 3 to 6 hours
	Within 7 to 12 hours
Day similarity	Same day
	One day difference
	Two to three days difference
Season similarity	Same season
	One season difference
	Two seasons difference

[3] This category was used for a small number of offender address records that indicated they were locations of prior crimes but the type of crime was not specified. There were so few of these records that the estimates had wide confidence intervals, or, in some cases, the model could not converge on any estimate. In the ordinal variable used in the main analysis this category was therefore merged with 'other activity node'.

Reliability and Relevance Interaction Variable

We reproduced the analyses using two versions of the reliability and relevance interaction variable using three levels (high, medium, low), to see whether a more detailed interaction variable (9 interaction categories: $H_fH_s/H_fM_s \ldots L_fL_s$), could reveal more detailed insights. The two versions varied the thresholds for identifying the high, medium and low categories, to test whether the results were robust to the threshold decision. The thresholds used were (1) < 25th percentile = low and >75th percentile = high and (2) <33rd percentile = low and >66th percentile = high. The results (not reported) of the three-category models varied by threshold and tended to produce wide confidence intervals for some interaction categories, likely because of the smaller numbers of nodes falling into each category by comparison to the simpler high/low version. Additionally, the three-category models produced worse fit to the data, as measured by Pseudo-R^2. We therefore report only the results of the more robust two-category models.

Opportunity (Control) Variables

The opportunity covariate was different for each crime type. For residential burglary the opportunity covariate was the number of households, from New Zealand Census data. Due to large changes in residential population in many SA2s over the data period, census 2013 data was used for reference offences occurring between 2009 and 2015, and census 2018 data for those occurring between 2016 and 2018. For the remaining crime types the opportunity variables were counts of businesses from Statistics New Zealand Business Demography data. For business counts, only the statistics for 2018 were used because they were highly correlated with those for 2013. Table S1.3 provides details of industry categories included in the counts of businesses used as opportunity variables for non-residential burglary, commercial robbery, personal robbery and sex offences, respectively.

Table S1.3 Industry categories included in business counts for different crime types

Crime type	**Industry categories included in business counts**
Non-residential burglary	All industries.
Commercial robbery	G Retail Trade, H Accommodation and Food Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services. M Professional, Scientific and Technical Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services, M Professional, Scientific and Technical Services, N Administrative and Support Services, R Arts and Recreation Services, S Other Services
Personal robbery and sex offences	As for Commercial robbery plus: I Transport, Postal and Warehousing, J Information Media and Telecommunications, O Public Administration and Safety, P Education and Training, Q Health Care and Social Assistance.

References

Ashby, M. P., & Bowers, K. J. (2013). A comparison of methods for temporal analysis of aoristic crime. *Crime Science*, *2*(1), 1. https://doi.org/10.1186/2193-7680-2-1

Boldt, M., & Borg, A. (2016). Evaluating temporal analysis methods using residential burglary data. *ISPRS International Journal of Geo-Information*, *5*(9), 148. https://doi.org/10.3390/ijgi5090148

ISSP Research Group. (2009). *International social survey programme 2007: Leisure time and sports—ISSP 2007* (ZA4850 Data file Version 2.0.0). GESIS Data Archive. https://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA4850

Kuang, D., Brantingham, P. J., & Bertozzi, A. (2017). Crime topic modeling. *Crime Science*, *6*(1), 1–20. https://doi.org/10.1186/s40163-017-0074-0

S2 Descriptive Statistics

Table S2.1 provides descriptive statistics for the node category x distance band variables as a proportion of all SA2 alternatives sampled for each crime type. To illustrate, for the residential burglaries, 4.4% of all potentially chosen SA2 units contained a H_fH_s activity node, 3% contained a L_fL_s activity node and 20% had a H_fH_s node within 2-5km and no closer nodes.

Table S2.1 Proportion of sampled SA2s for each node category and each distance band

Category	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
HfHs	Within SA2	0.044	0.046	0.041	0.046	0.030
	0-200m	0.023	0.025	0.023	0.026	0.017
	200-500m	0.031	0.033	0.032	0.035	0.024
	500-1000m	0.046	0.050	0.048	0.051	0.038
	1-2km	0.082	0.088	0.087	0.091	0.071
	2-5km	0.204	0.217	0.219	0.225	0.191
H_fL_s	Within SA2	0.023	0.020	0.022	0.018	0.022
	0-200m	0.013	0.011	0.013	0.011	0.014
	200-500m	0.015	0.013	0.016	0.013	0.018
	500-1000m	0.022	0.019	0.025	0.019	0.028
	1-2km	0.040	0.034	0.045	0.034	0.053
	2-5km	0.096	0.086	0.114	0.089	0.142
L_fH_s	Within SA2	0.038	0.040	0.037	0.044	0.025
	0-200m	0.020	0.021	0.021	0.024	0.014
	200-500m	0.027	0.029	0.029	0.033	0.019
	500-1000m	0.041	0.043	0.043	0.048	0.029
	1-2km	0.072	0.077	0.076	0.084	0.051
	2-5km	0.175	0.186	0.186	0.201	0.125
L_fL_s	Within SA2	0.030	0.027	0.025	0.022	0.026
	0-200m	0.016	0.015	0.014	0.012	0.015
	200-500m	0.020	0.018	0.017	0.015	0.020
	500-1000m	0.029	0.026	0.025	0.022	0.031
	1-2km	0.051	0.046	0.046	0.039	0.056
	2-5km	0.120	0.113	0.115	0.097	0.139

Table S2.2 describes the distribution of the opportunity variable (i.e., number of dwellings or business units) for the sampled SA2s, for each crime type. The minimum for all crime types was 0 dwellings/business units.

Table S2.2 Median, IQR and maximum number of opportunity units per SA2

Offence type	**Median**	**IQR**	**Max**
Residential Burglary	1167	273	4707
Non-residential Burglary	306	123	4059
Commercial Robbery	195	90	3354
Personal Robbery	231	102	3822
Sex offences	231	102	3822

S3 Odds Ratios for Activity Node Variables

Table S3.1 shows the odds ratios (ORs) and the 95% confidence intervals for the associations between activity node proximity and crime location choice, for each activity node category, distance band and crime type. Each column represents a separate model.

Table S3.1 Odds ratios (95% confidence intervals) for each activity node category, distance band and crime

Category	**Distance band**	**Res. Burg.**	**Non-res. Burg.**	**Com. Rob.**	**Pers. Rob.**	**Sex offences**
HfHs	In the SA2	117.43***	113.78***	31.37***	72.96***	321.78***
		(111.75-123.41)	(106.51-121.55)	(26.07-37.74)	(65.82-80.88)	(292.78-353.65)
	0-200m	19.00***	33.04***	21.34***	21.69***	20.17***
		(16.88-21.39)	(29.10-37.52)	(16.25-28.02)	(17.94-26.22)	(15.44-26.36)
	200-500m	15.13***	19.46***	12.14***	15.56***	16.46***
		(13.48-16.99)	(16.9-22.41)	(9.02-16.34)	(12.8-18.93)	(12.74-21.26)
	500-1000m	8.78***	11.33***	8.02***	8.49***	14.93***
		(7.77-9.93)	(9.74-13.18)	(5.92-10.87)	(6.84-10.52)	(11.89-18.73)
	1-2km	5.35***	7.24***	6.04***	5.61***	7.45***
		(4.74-6.04)	(6.24-8.39)	(4.55-8.03)	(4.54-6.93)	(5.9-9.41)
	2-5km	2.9***	3.77***	3.8***	3.18***	5.19***
		(2.58-3.26)	(3.26-4.35)	(2.93-4.93)	(2.6-3.89)	(4.28-6.28)
H_fL_s	In the SA2	9.72***	10.44***	12.48***	7.6***	7.65***
		(9.24-10.22)	(9.75-11.18)	(10.37-15.03)	(6.84-8.45)	(6.77-8.65)
	0-200m	16.02***	8.91***	9.71***	10.53***	23.75***
		(14.12-18.17)	(7.57-10.49)	(7.05-13.36)	(8.4-13.2)	(18.31-30.81)
	200-500m	11.04***	10.29***	9.33***	8.55***	18.21***
		(9.68-12.59)	(8.61-12.29)	(6.63-13.12)	(6.69-10.93)	(14.11-23.5)
	500-1000m	8.39***	5.51***	4.87***	6.02***	11.19***
		(7.32-9.63)	(4.51-6.74)	(3.38-7.03)	(4.62-7.83)	(8.74-14.32)
	1-2km	6.8***	5.13***	2.78***	3.88***	7.6***
		(5.98-7.75)	(4.25-6.18)	(1.93-3.99)	(2.97-5.08)	(5.99-9.65)
	2-5km	3.62***	3.28***	2.62***	2.61***	3.54***
		(3.19-4.11)	(2.75-3.92)	(1.94-3.54)	(2.03-3.36)	(2.87-4.36)

Category	Distance band	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
L_fH_s	In the SA2	2.33***	2.46***	3.03***	2.96***	2.96***
		(2.21-2.45)	(2.29-2.64)	(2.53-3.64)	(2.66-3.3)	(2.61-3.36)
	0-200m	3.28***	4.66***	3.9***	3.5***	4.77***
		(2.82-3.81)	(3.95-5.49)	(2.73-5.55)	(2.74-4.48)	(3.23-7.03)
	200-500m	3.66***	3.5***	3.12***	3.19***	3.36***
		(3.16-4.24)	(2.89-4.25)	(2.12-4.59)	(2.46-4.13)	(2.25-5.03)
	500-1000m	3.12***	2.87***	4.15***	3.17***	3.25***
		(2.68-3.63)	(2.35-3.5)	(2.92-5.89)	(2.45-4.09)	(2.27-4.65)
	1-2km	2.06***	1.77***	2.33***	2.32***	2.77***
		(1.77-2.39)	(1.45-2.16)	(1.65-3.28)	(1.8-2.99)	(1.97-3.89)
	2-5km	1.33***	1.24*	1.21	1.2	1.34
		(1.16-1.52)	(1.03-1.48)	(0.88-1.65)	(0.94-1.53)	(0.99-1.82)
L_fL_s	In the SA2	1.68***	1.9***	2.34***	1.94***	2.06***
		(1.59-1.77)	(1.76-2.04)	(1.94-2.81)	(1.73-2.17)	(1.8-2.35)
	0-200m	3.37***	2.59***	2.64***	2.81***	3.97***
		(2.89-3.91)	(2.14-3.13)	(1.78-3.91)	(2.11-3.73)	(2.8-5.63)
	200-500m	2.83***	1.81***	3.18***	2.75***	3.95***
		(2.41-3.33)	(1.42-2.31)	(2.09-4.83)	(2.02-3.75)	(2.83-5.52)
	500-1000m	2.42***	2.88***	2.1***	2.45***	3.05***
		(2.05-2.86)	(2.31-3.6)	(1.36-3.25)	(1.79-3.36)	(2.22-4.19)
	1-2km	1.82***	1.68***	2.29***	1.49*	2.02***
		(1.55-2.14)	(1.34-2.11)	(1.56-3.35)	(1.07-2.08)	(1.47-2.77)
	2-5km	1.53***	1.25*	1.34	1.30	1.93***
		(1.33-1.77)	(1.02-1.54)	(0.95-1.9)	(0.97-1.74)	(1.51-2.47)

*** p<.001, *p<.05.

S4 Wald Test Results

Table S4.1 shows the results of Wald Chi Square tests relevant to hypothesis H1 (H_fH_s vs H_fL_s, H_fH_s vs L_fH_s, H_fH_s vs L_fL_s, H_fL_s vs L_fL_s and L_fH_s vs L_fL_s) comparing Odds Ratios (ORs) with overlapping 95% confidence intervals (CIs). All variable 1 ORs are greater than variable 2 ORs.

Table S4.1 Results of Wald tests comparing ORs where CIs overlapped

Offence	Distance band	Variable 1	Variable 2	X^2
Residential Burglary	0-200m	H_fH_s	H_fL_s	2.54
Residential Burglary	0-200m	L_fH_s	L_fL_s	0.05
Residential Burglary	200-500m	L_fH_s	L_fL_s	4.45*
Residential Burglary	500m-1km	H_fH_s	H_fL_s	0.17
Residential Burglary	500m-1km	L_fH_s	L_fL_s	4.05*
Residential Burglary	1-2km	H_fL_s	H_fH_s	4.96*
Residential Burglary	1-2km	L_fH_s	L_fL_s	1.00
Residential Burglary	2-5km	L_fH_s	L_fL_s	1.68
Non-residential Burglary	500m-1km	L_fH_s	L_fL_s	0.00
Non-residential Burglary	1-2km	L_fH_s	L_fL_s	0.09
Non-residential Burglary	2-5km	H_fH_s	H_fL_s	0.99
Non-residential Burglary	2-5km	L_fH_s	L_fL_s	0.01
Commercial Robbery	Within SA2	L_fH_s	L_fL_s	3.01
Commercial Robbery	0-200m	L_fH_s	L_fL_s	1.67
Commercial Robbery	200-500m	H_fH_s	H_fL_s	0.93
Commercial Robbery	200-500m	L_fH_s	L_fL_s	0.00
Commercial Robbery	500m-1km	H_fH_s	H_fL_s	3.15
Commercial Robbery	500m-1km	L_fH_s	L_fL_s	4.62*
Commercial Robbery	1-2km	L_fH_s	L_fL_s	0.00
Commercial Robbery	2-5km	H_fH_s	H_fL_s	2.38
Commercial Robbery	2-5km	L_fH_s	L_fL_s	0.16
Personal Robbery	0-200m	L_fH_s	L_fL_s	1.10
Personal Robbery	200-500m	L_fH_s	L_fL_s	0.42
Personal Robbery	500m-1km	H_fH_s	H_fL_s	2.84
Personal Robbery	500m-1km	L_fH_s	L_fL_s	1.26

Offence	Distance band	Variable 1	Variable 2	X^2
Personal Robbery	1-2km	H_fH_s	H_fL_s	3.26
Personal Robbery	1-2km	L_fH_s	L_fL_s	3.61
Personal Robbery	2-5km	H_fH_s	H_fL_s	1.04
Personal Robbery	2-5km	L_fH_s	L_fL_s	0.14
Sex offences	0-200m	H_fH_s	H_fL_s	0.53
Sex offences	0-200m	L_fH_s	L_fL_s	0.43
Sex offences	200-500m	H_fH_s	H_fL_s	0.22
Sex offences	200-500m	L_fH_s	L_fL_s	0.33
Sex offences	500m-1km	H_fH_s	H_fL_s	2.11
Sex offences	500m-1km	L_fH_s	L_fL_s	0.06
Sex offences	1-2km	H_fH_s	H_fL_s	0.01
Sex offences	1-2km	L_fH_s	L_fL_s	1.57
Sex offences	2-5km	L_fH_s	L_fL_s	2.90

$^*p < 0.05$

APPENDIX E

Supplementary Materials to Chapter 7

The following materials are available online as supplementary materials to:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART. *Journal of Investigative Psychology and Offender Profiling*. https://doi.org/10.1002/jip.1585

Supplementary Materials to 'A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART'

S1 Discrete Spatial Choice Modelling Process

The discrete spatial choice modelling (DSCM) described in this section was originally conducted as part of a different study within the same programme of research as the present study. It quantified the relative likelihood of people offending in a location given the attributes of nearby activity nodes, thus providing values we could use in GP-SMART to see if including them would improve accuracy over not accounting for activity nodes' attributes. This section repeats material that appears in Curtis-Ham et al. (under review) describing that study; we summarise the DSCM method here for replicability, should others wish to adopt a similar approach to develop and calibrate their own GP-SMART style algorithm. Readers are referred to Curtis-Ham et al. (under review) for detail of the rationales for methodological choices made in that study (e.g., unit of analysis, choice of covariates) and to Bernasco (2021) for practical guidance on DSCM for crime location choice analysis.

Data

The calibration datasets used in the DSCM analysis consisted of the latest offences, and pre-offence activity locations, of 17,054 residential burglars, 10,353 non-residential burglars, 1,977 commercial robbers, 4,315 personal robbers and 4,421 extra-familial sex offenders. Each crime type was analysed with a separate model.

Unit of analysis

In DSCM, the units of analysis are the locations from which decision-makers (here, offenders) are choosing when deciding where to conduct an activity (here, crime). We used units approximating neighbourhoods: 'Statistical Area 2' Census Units ('SA2s'). SA2s usually contain 2000-4000 residents (1000-3000 in rural areas). In DSCM, the dataset includes one row for each choice alternative (SA2) per location choice (each offender's latest offence). To avoid the computational burden of processing the large datasets produced when including all 2153 SA2s (e.g., almost 37 million rows for the residential burglary calibration dataset with n = 17,054 offences), we used importance sampling to create a smaller set of

SA2 alternatives for each offence (Ben-Akiva & Lerman, 1985; Curtis-Ham et al., 2021; McFadden, 1977). The smaller set included the chosen SA2, all SA2s with any activity nodes within 5km of the SA2 boundary and 10 SA2s randomly selected from the remaining SA2s.[1] The SA2s included in resulting dataset were a median of 1.2km^2 land area (interquartile range 0.84-2.2 km^2).

Outcome variable

The outcome variable reflected the location choice for each offender's latest offence, being which SA2 was chosen for the offence. In the dataset for analysis, it is indicated by a dichotomous variable indicating, for each SA2 in each offender's set of SA2s, whether it was chosen or not.

Predictor Variables

The predictor variables were a series of dichotomous variables indicating the presence (1) or absence (0) of an activity node of a given category (e.g., weekly) of a given attribute (e.g., frequency), within each of 6 distance bands of the SA2, resulting in 28 categories x 6 distance bands = 168 variables. The attributes and categories are listed in Table S1.1 and were relative to the outcome variable offence (e.g., within 1-2 days of the offence, or prior crime of the same type as the offence). See Figure 1 in the main text for further definitions of the attributes but note that some adjustments were made to how the categories were coded in GP-SMART as described in Supplementary Material S2 below. The distance bands were: within the same SA2, between 0 and 200m outside the SA2 boundary, then 200-500m, 500m-1km, 1km-2km and 2km-5km. SA2s were coded using only the activity nodes in the distance band containing the nearest activity node(s). For example, if there were no activity nodes in the SA2, but two nodes within 0-200m, the SA2 was coded: 0 for all attribute categories in the 'same SA2' distance band; 1 for applicable attribute categories in the 0-200m distance

[1] The 2018 SA2 shapefile and metadata were downloaded from https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/. Eighty-three SA2s made up of bodies of water were excluded from the analysis.

band (and 0 for attribute categories not applying to the two nodes in that band); and 0 for all attribute categories for the remaining distance bands.

Table S1.1 Categories for each activity node attribute used in the DSCM models

Factor	Category
Frequency	Weekly or more
	Monthly
	Yearly or less
Recency and duration	1-2 days
	3 to 30 days
	1 to 12 months
	1 to 5 years
	Over 5 years
Behaviour similarity	Prior crime of same crime type
	Prior crime of similar crime type
	Prior crime of not similar crime type
	Prior crime experienced as a victim/witness of same crime type
	Prior crime or incident experienced as a victim/witness of similar type
	Prior crime or incident experienced as a victim/witness of not similar type
	Other activity node
	Behaviour similarity unknown
Location similarity	Same location type
	Unknown similarity
	Different location type
Hour similarity	Within 0 to 2 hours
	Within 3 to 6 hours
	Within 7 to 12 hours
Day similarity	Same day
	One day difference
	Two to three days difference
Season similarity	Same season
	One season difference
	Two seasons difference

Covariates

To isolate the relationship between activity node presence and crime location by controlling for the presence of crime opportunities, we included a covariate indicating the number of potential targets in the SA2. The covariates were specific to the crime types, given their different targets and are described in Table S1.2. All opportunity variables were downloaded from the New Zealand Statistics Census and Business Demography data repository (http://nzdotstat.stats.govt.nz/).

Table S1.2 Opportunity variable definitions and data specification

Offence	Variable	Census/Business demography specification
Residential Burglary	Number of residential dwellings	NZ Census > Number of private dwellings. Census 2013 data was used for reference offences occurring from 2009-2015 and census 2018 data were used for 2016-2018 reference offences because a small number of SA2s' residential populations increased or decreased drastically between Census years. Marlborough SA2s were adjusted as per Statistics New Zealand guidance and corrections available at: http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/b09ae9c0-de19-4418-952b-c9e4cbf7e1a2. For reference, Residential Burglaries were those where the Location Type of the offence was Residential*.
Non-residential Burglary	Number of business units	Business demography statistics > Industry Category = All. Business demography statistics remained consistent over the data period so 2018 was used for simplicity. For reference, Non-residential Burglaries were those where the Location Type was not Residential or Unknown*.
Commercial Robbery	Number of commercial business units	Business demography statistics > Industry Category = G Retail Trade, H Accommodation and Food Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services. M Professional, Scientific and Technical Services, K Financial and Insurance Services, L Rental, Hiring and Real Estate Services, M Professional, Scientific and Technical Services, N Administrative and Support Services, R Arts and Recreation Services, S Other Services. For reference, Commercial Robberies were those where the Location Type of the offence was Commercial*.
Personal Robbery & Extrafamilial Sex Offences	Number of commercial and public business units	Business demography statistics > Industry Category = As for Commercial Robbery plus I Transport, Postal and Warehousing, J Information Media and Telecommunications, O Public Administration and Safety, P Education and Training, Q Health Care and Social Assistance. For reference, Personal Robberies were those where the Location Type was not Commercial or Unknown*.

* Location Type categories were: Residential (e.g., private dwellings and associated outbuildings/grounds, residential construction sites, farms, rest homes), Commercial (e.g., shops, restaurants, bars, entertainment facilities, businesses, factories, malls, offices, gyms, motels, hotels, hostels, campgrounds, commercial construction sites), Public (e.g., school/education, sports facilities/grounds, hospitals and medical clinics, transit stations, police stations, courts, prisons, church/religious facility, community buildings, marae), Street (e.g., streets, roads, footpaths, parks and open spaces, car parking facilities, wharf, harbour, beach, lake, river, sea), Transit (e.g., bus, car, train, boat, ship, plane in transit) and Unknown (i.e., not able to be coded as any of above, e.g., "other", "miscellaneous", "unspecified", "unknown", "online"). Location type was coded using a custom rule applied to a range of indicators because there is no single field in the NIA database that is a reliable indicator of location type (covering the entire data period). The rule was tested against a series of alternatives in order to identify the version that most accurately identified location type, by checking the narrative descriptions of a random sample of 1% of the reference offences of each crime type. The optimal rule achieved 95%-98% accuracy across the reference offence categories (burglary, robbery and sex offences). Reference tables mapping the NIA Scene Type and Address Type codes (e.g., farm, restaurant, hospital) to the present location type categories (residential, commercial, etcetera), and the rule incorporating these and other location type indicators, as an R Script, are available from the corresponding author.

Analysis Approach

We used conditional logit models (McFadden, 1984) to estimate the increase in odds of an SA2 being chosen for crime given the proximity and attributes of nearby activity nodes, operationalised as described above, compared with the absence of any activity node within any of the distance bands (i.e., within 5km of the SA2). The resulting odds ratios used in GP-SMART are provided in the next section.

S2 Odds Ratios for Activity Node Associations

Table S2.1 shows the odds ratios (ORs) for each category of each activity node attribute for each crime type used as adjustment values in GP-SMART. We used the ORs that indicated the increase in crime likelihood given an activity node with that attribute category (e.g., frequency: weekly) in the same neighbourhood (SA2). To simplify the GP-SMART process, we combined the attribute categories indicating dissimilarity of behaviour, location type and timing to create dichotomous similarity variables (same/not same) because the dissimilar categories were largely non-significant or had odds ratios close to 1. For these dichotomised variables we set the ‘not same’ ORs used in GP-SMART to 1.

Table S2.1 Odds ratios used in GP-SMART by attribute, category and crime

Attribute	Category	Res. Burg.	Non-res. Burg.	Com. Rob.	Pers. Rob.	Sex offences
Frequency	Weekly	5.36	7.53	3.92	5.12	4.30
	Monthly	1.22	1.19	1.08	1.24	1.13
	Yearly	1.18	0.97	1.02	1.07	1.36
Recency	1 to 2 days	2.89	2.11	1.49	1.92	4.72
	3 to 30 days	2.97	3.03	2.26	2.51	3.29
	1 to 12 months	1.70	1.58	1.54	1.83	1.55
	1 to 5 years	0.98	0.97	0.96	1.04	1.07
	Over 5 years	0.80	0.73	0.93	0.90	0.79
Duration	1 to 2 days	1.29	1.35	1.35	1.28	1.18
	3 to 30 days	1.20	1.18	1.15	1.24	1.23
	1 to 12 months	1.16	1.13	1.08	1.09	1.13
	1 to 5 years	0.95	0.84	1.07	0.82	0.88
	Over 5 years	0.77	0.86	0.97	0.82	0.60
Behaviour similarity	Same prior offence	1.92	2.13	2.04	2.36	1.98
	Other activity node[a]	1.00	1.00	1.00	1.00	1.00
Location type similarity	Same location type	4.75	3.76	2.72	2.58	4.57
	Not same or unknown location type[b]	1.00	1.00	1.00	1.00	1.00
Day part similarity	Same day period[c]	10.58	4.84	7.68	9.96	17.38
	Not same day period[d]	1.00	1.00	1.00	1.00	1.00
Week part similarity	Same week part[e]	0.66	0.77	0.88	0.95	0.85
	Not same week part[f]	1.00	1.00	1.00	1.00	1.00
Season similarity	Same season	2.19	3.08	1.64	2.24	1.99
	Not same season[g]	1.00	1.00	1.00	1.00	1.00

[a] Combining all other ‘behaviour similarity’ categories.
[b] Combining ‘different’ and ‘unknown’ categories.
[c] Using the ‘within 0 to 2 hours’ category as indicating the same day period.
[d] Combining ‘within 3 to 6 hours’ and ‘within 7 to 12 hours’ categories as indicating a different day period.

[e] Using the ‘same day’ category as indicate same week part.
[f] Combining ‘one day difference’ and ‘two to three days difference’ categories as indicating a different week part.
[g] Combining ‘one season difference’ and ‘two seasons difference’ categories as indicating a different season.

The astute reader may notice that some of the ORs are below 1, indicating negative associations (i.e., decreases in likelihood of being chosen for crime compared with an SA2 with no activity nodes within 5km). Further, some ORs are not consistent with the expectation that categories higher on an attribute’s scale (e.g., more recent, longer duration, greater similarity) would be associated with higher odds of crime than categories lower on the scale. We believe these counter-intuitive results are a product of the data and our coding process. For example, all home addresses were coded as weekly, not similar behaviour, same time and same day, producing multicollinearity. Including multicollinear variables in a model can produce unreliable parameter estimates, even reversed signs (Chen, 2012). However, multicollinearity does not affect the accuracy of a model’s predictions when the parameter estimates are used to predict the outcome variable (Morris & Lieberman, 2018). Similarly, many prior crime, victim/witness and incident nodes were coded as having short durations of one day because there was only ever one event at that location, likely under-representing the true duration of people’s associations with these locations, and leading to the spurious result that nodes only visited briefly (according to the data) were associated with a higher likelihood of crime than nodes visited over longer periods. Although these spurious results make it difficult to draw conclusions about the relative associations of activity nodes’ attributes with people’s crime locations using these data, they do not preclude the use of the ORs to make predictions using the test datasets derived from the same original data, as implemented in GP-SMART.

S3 Wilcoxon Signed-Rank Test Results

Table S3.1 shows the results of the paired sample Wilcoxon signed rank tests comparing the percentile rank of the offender among the suspects between GP-SMART and each of the two baseline methods: distance to the nearest activity node and distance to the nearest home address. As a measure of effect size, we report the rank biserial correlation, which indicates the proportion of test cases where GP-SMART placed the offender in a relatively higher position on the suspect list than the baseline method. A value of 0 would mean GP-SMART ranked the offender higher in 50% of cases; the closer the value to 1, the more cases GP-SMART ranked the offender higher.

Table S3.1 Effect sizes (95% Confidence Intervals) for Wilcoxon signed rank tests comparing offender percentile ranks for GP-SMART with the baseline methods. All p-values <0.001.

Crime type	**Any node baseline**	**Home node baseline**
Residential Burglary	0.39 (0.33-0.45)	0.52 (0.47-0.57)
Non-residential Burglary	0.42 (0.36-0.48)	0.45 (0.39-0.50)
Commercial Robbery	0.48 (0.40-0.56)	0.22 (0.12-0.31)
Personal Robbery	0.43 (0.37-0.49)	0.41 (0.35-0.46)
Sex offences	0.58 (0.53-0.62)	0.27 (0.20-0.34)

References for Supplementary Materials

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.

Bernasco, W. (2021). Discrete spatial choice models. In E. R. Groff & C. P. Haberman (Eds.), *The Study of Crime and Place: A Methods Handbook.* Temple University Press. https://osf.io/639cz/

Chen, G. J. (2012). A simple way to deal with multicollinearity. *Journal of Applied Statistics*, *39*(9), 1893–1909. https://doi.org/10.1080/02664763.2012.690857

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The importance of importance sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-021-09526-5

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Familiar locations and similar activities: examining the interaction of reliable and relevant knowledge in offenders' crime location choices.

McFadden, D. (1977). *Modelling the choice of residential location* (No. 477; Cowles Foundation Discussion Papers). Yale University. https://EconPapers.repec.org/RePEc:cwl:cwldpp:477

McFadden, D. (1984). Econometric analysis of qualitative response models. In P. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 105–142). Elsevier. https://doi.org/10.1016/S1573-4412(84)02016-X

Morris, J., & Lieberman, M. (2018). Multicollinearity's effect on regression prediction accuracy with real data structures. *General Linear Model Journal*, *44*(1), 29–34. https://doi.org/10.31523/glmj.044001.004

APPENDIX F

Co-Authorship Forms

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 2:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science, 9*(1), 1–14. https://doi.org/10.1186/s40163-020-00132-7

Nature of contribution by PhD candidate	Sophie conceptualised the theoretical framework, conducted literature searches, wrote and revised the manuscript and responses to reviewers following feedback from co-authors.
Extent of contribution by PhD candidate (%)	66%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim provided feedback on the theoretical framework, manuscript drafts and responses to reviewers.
Oleg Medvedev	Oleg provided feedback on the theoretical framework, manuscript drafts and responses to reviewers.
Devon Polaschek	Devon provided feedback on the theoretical framework, manuscript drafts and responses to reviewers.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco	W Bernasco	12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	January 13, 2022

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). A national examination of the spatial extent and similarity of offenders' activity spaces using police data. *ISPRS International Journal of Geo-Information, 10*(2), 47. https://doi.org/10.3390/ijgi10020047

Nature of contribution by PhD candidate	Sophie conceptualised the research, conducted literature searches, collected data, designed analyses, conducted analyses, and wrote and revised the manuscript and responses to reviewers, following feedback from Wim, Oleg and Devon.
Extent of contribution by PhD candidate (%)	74%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim provided feedback on analyses, manuscript drafts and responses to reviewers.
Oleg Medvedev	Oleg provided feedback on analyses, manuscript drafts and responses to reviewers.
Devon Polaschek	Devon provided feedback on analyses, manuscript drafts and responses to reviewers.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco	W Bernasco	12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	January 13, 2022

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2021). The Importance of Importance Sampling: Exploring methods of sampling from alternatives in discrete choice models of crime location choice. *Journal of Quantitative Criminology.* https://doi.org/10.1007/s10940-021-09526-5

Nature of contribution by PhD candidate	Sophie conceptualised the research, conducted literature searches, collected data, designed analyses, conducted analyses, and wrote and revised the manuscript and responses to reviewers, following feedback from Wim, Oleg and Devon.
Extent of contribution by PhD candidate (%)	69%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim advised on study design, wrote the equations and accompanying text, and provided feedback on analyses, manuscript drafts and responses to reviewers.
Oleg Medvedev	Oleg provided feedback on analyses, manuscript drafts and responses to reviewers.
Devon Polaschek	Devon provided feedback on analyses, manuscript drafts and responses to reviewers.

Certification by Co-Authors

The undersigned hereby certify that:

❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco	W Bernasco	12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	**January 13, 2022**

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Relationships between offenders' crime locations and different prior activity locations as recorded in police data.

Nature of contribution by PhD candidate	Sophie conceptualised the research, conducted literature searches, collected data, designed analyses, conducted analyses, and wrote the manuscript. Sophie also revised the manuscript following feedback from Wim, Oleg and Devon, and following journal peer review.
Extent of contribution by PhD candidate (%)	73%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim provided feedback on analyses, manuscript drafts and manuscript revisions.
Oleg Medvedev	Oleg provided feedback on analyses, manuscript drafts and manuscript revisions.
Devon Polaschek	Devon provided feedback on analyses, manuscript drafts and manuscript revisions.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco	W Bernasco	12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	January 13, 2022

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 6:

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (under review). Familiar locations and similar activities: Examining the interaction of reliable and relevant knowledge in offenders' crime location choices.

Nature of contribution by PhD candidate	Sophie conceptualised the research, conducted literature searches, collected data, designed analyses, conducted analyses, and wrote and revised the manuscript and responses to reviewers, following feedback from Wim, Oleg and Devon.
Extent of contribution by PhD candidate (%)	72%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim provided feedback on analyses, manuscript drafts and responses to reviewers.
Oleg Medvedev	Oleg provided feedback on analyses, manuscript drafts and responses to reviewers.
Devon Polaschek	Devon provided feedback on analyses, manuscript drafts and responses to reviewers.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco		12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	January 13, 2022

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 7: Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). A new Geographic Profiling Suspect Mapping And Ranking Technique for crime investigations: GP-SMART. J*ournal of Investigative Psychology and Offender Profilin*g. https://doi.org/10.1002/jip.1585

Nature of contribution by PhD candidate	Sophie conceptualised the research, conducted literature searches, collected data, designed analyses, conducted analyses, and wrote and revised the manuscript, following feedback from Wim, Oleg and Devon.
Extent of contribution by PhD candidate (%)	79%

CO-AUTHORS

Name	Nature of Contribution
Wim Bernasco	Wim provided feedback on analyses and manuscript drafts.
Oleg Medvedev	Oleg provided feedback on analyses and manuscript drafts.
Devon Polaschek	Devon provided feedback on analyses and manuscript drafts.

Certification by Co-Authors

The undersigned hereby certify that:

❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Wim Bernasco	W Bernasco	12 January 2022
Oleg Medvedev		13/01/2022
Devon Polaschek	Devon Polaschek	January 13, 2022