Behavioural Case Linkage with Solved and Unsolved Crimes

Tonkin, M., Woodhams, J., Bull, R., & Bond, J. W.
Abstract

Offender behaviour is used to distinguish between crimes committed by the same person (linked crimes) and crimes committed by different people (unlinked crimes) through behavioural case linkage. There is growing evidence to support the use of behavioural case linkage by investigative organisations such as the police, but this research is typically limited to samples of solved crime that do not reflect how this procedure is used in real life. The current paper extends previous research by testing the potential for behavioural case linkage in a sample containing both solved and unsolved crimes. Discrimination accuracy is examined across crime categories (e.g. a crime pair containing a car theft and a residential burglary), across crime types (e.g. a crime pair containing a residential burglary and a commercial burglary), and within crime types (e.g. a crime pair containing two residential burglaries) using the number of kilometres (intercrime distance) and the number of days (temporal proximity) between offences to distinguish between linked and unlinked crimes. The intercrime distance and/or the temporal proximity were able to achieve statistically significant levels of discrimination accuracy across crime categories, across crime types, and within crime types as measured by Receiver Operating Characteristic (ROC) analysis. This suggests that behavioural case linkage can be used to assist the investigation, detection and prosecution of prolific and versatile serial offenders.

Keywords: Linkage analysis; comparative case analysis; behavioral case linkage; unsolved crime; serial crime
Investigators, such as the police, frequently attempt to identify groups of so-called linked crimes that have been committed by the same offender [1,2]. The advantages of doing so include increasing the likelihood of successful detection and prosecution of the person/s responsible by combining the evidence collected across several investigations [2]. Investigators have a variety of techniques from forensic science available to them to assist in this task, such as DNA matching, fingerprint analysis, bitemark analysis, and the analysis of footwear impressions [3-6]. In the context of linking crime, these techniques are used to identify a statistically significant degree of match between the physical material obtained at two or more crime scenes, which allows the investigator to infer that the same person/s were responsible. However, such physical material is not always available to link crimes, particularly since some offenders may be quickly learning to hide their forensic identity when committing crime (e.g. by wearing gloves, masks and, in the case of sexual offences, condoms). For example, only 12% of crime scenes examined in England and Wales during 2004/2005 yielded DNA evidence and only 6% yielded samples that were suitable to be loaded onto the National DNA Database [7]. Also, even when this material is available to link crimes, processing this information can be time-consuming and expensive, which leaves many investigative organisations without the resources to process large volumes of physical material on a regular basis [8].

For these reasons investigative organisations use behavioural evidence in addition to or in the absence of physical evidence to link crime [1,9-11]. This requires an investigator to, first, determine whether a stable pattern of behaviour exists across the various crimes (behavioural consistency) and, second, whether that pattern is sufficiently different from the patterns displayed in other crimes (behavioural distinctiveness) to infer that the same person might be responsible [11,12].
A growing body of research has developed to support the principles of consistency and distinctiveness that underpin behavioural case linkage. This research has tested whether offender crime scene behaviour can be used to distinguish between crimes committed by the same person (linked crimes) and crimes committed by different people (unlinked crimes). It has been found that certain types of offender behaviour can be used to achieve statistically significant levels of discrimination accuracy in this task. In particular, the number of kilometres and the number of days between offences have been shown to be particularly effective with samples of burglary [13-16], robbery [17] and car theft [18]. In terms of violent and sexual crime, behaviours that reflect the level of pre-offence planning, the offender’s methods of controlling the victim/witnesses, and target characteristics have all been effective with samples of homicide [19,20], rape/sexual assault [1,2,21-23] and robbery [17].

This research, however, has restricted itself to testing behavioural consistency, distinctiveness and discrimination accuracy with samples that are homogenous in terms of crime type (i.e. only one type of crime, e.g. residential burglary is studied in isolation) [24]. This is problematic because the majority of offenders (particularly the most prolific) tend to commit a variety of different types of crime rather than specialising in individual types [25-28]. Consequently, investigators are regularly faced with apprehending versatile offenders who have committed crime series that contain several different types of offence. Tonkin et al. [24], therefore, investigated whether simple measures of offender behaviour could be used to link crimes from different crime types and categories, which would help investigators to detect and prosecute versatile serial offenders more successfully. In their study, Tonkin and colleagues [24] used the Home Office counting rules for recorded crime, which provide a national standard for recording crime amongst police forces in England and Wales. At the time of the study, these rules recognised 156 individual types of crime that were grouped into
nine crime categories. Tonkin et al. [24] tested the ability of offender behaviour to
discriminate between linked and unlinked crimes at three different levels. The first level,
“within crime types”, examined discrimination accuracy between crimes of the same type
(e.g. could offender behaviour be used to determine whether two residential burglaries were
committed by the same person or not?). This replicated the way in which previous research
had tested the principles of behavioural case linkage [e.g. 13,17,18]. The second level,
“across crime types”, tested discrimination accuracy with crimes that were from the same
broad category of crime but of different specific types (e.g. a residential burglary and a
commercial burglary). The third level, “across crime categories”, examined discrimination
accuracy with crimes that were from completely different crime categories (e.g. a residential
burglary and a rape). No previous research had tested discrimination accuracy across crime
types (level 2) or across crime categories (level 3). Using the number of kilometres and the
number of days between offences (referred to as the intercrime distance and temporal
proximity, respectively) it was possible to discriminate between linked and unlinked crimes
at all three levels of analysis, with linked crimes tending to occur significantly closer together
in space and time than unlinked crimes. Importantly, the level of accuracy that was achieved
when linking across crime types and across crime categories was comparable to that achieved
when linking within crime types. These findings, therefore, provide guidance on conducting
behavioural case linkage with crime series that contain several different types of offence,
which has the potential to contribute significantly to the detection and prosecution of versatile
serial offenders.

However, it is important to note that a degree of error should be expected when
linking crime using offender behaviour (both within and across crime types/categories). This

---

1 Violent offences (containing 38 individual crime types), sexual offences (containing 31 types), burglary
offences (containing 7 types), drug offences (containing 4 types), robbery (containing 2 types), theft or handling
offences (containing 16 types), fraud or forgery (containing 16 types), criminal damage offences (containing 11
types), and other offences (containing 31 types).
is indicated by the fact that the Area Under the Curve (AUC)\(^2\) values achieved in this
research fall below the maximum value of 1.00; for example, the intercrime distance has
achieved AUCs ranging from 0.72 to 0.97 with samples of burglary and car theft [13-16,18]
and a combination of offender behaviour has achieved AUCs ranging from 0.70 to 0.96 with
samples of homicide, rape/sexual assault, and commercial robbery [1,17,19,21]. This
variation in the AUC is unsurprising since an offender’s behaviour may vary across a series
of crimes (thereby causing errors in the linkage process) due to any number of situational
factors, such as the presence or absence of victim resistance, or as a consequence of the
maturation and learning that can come with criminal experience (see [11] for a review).
Consequently, practitioners must expect both false positive and false negative errors when
linking crime behaviourally. This does not mean, however, that behavioural case linkage is
worthless; it simply means that researchers and practitioners should be realistic in their
expectations of case linkage effectiveness.

In summary, there is a growing body of work to suggest that behavioural case linkage
has the potential to work in a reliable manner, provided the correct behaviours are used and
the findings are interpreted with an appropriate degree of caution\(^3\). But, this work can be
criticised for its reliance on samples of solved crime. In a real-life scenario, an investigator
would be highly unlikely to link a solved crime with another solved crime because there
would be no investigative benefit. Consequently, the existing research has not tested
behavioural case linkage in a way that reflects its use in practice (where behavioural case

\(^2\)The AUC has been used by some researchers of behavioural case linkage to provide an estimate of
discrimination accuracy in the linkage task. The AUC can range from 1.00 (indicating perfect positive
prediction) to 0.00 (indicating perfect negative prediction), with an AUC of 0.50 indicating a chance level of
discrimination accuracy.

\(^3\) However, it should be noted that some researchers have questioned the existence of offender behavioural
consistency altogether, thereby casting doubt on the potential for behavioural case linkage [e.g. 29,30]. But, the
weight of evidence suggests that behavioural consistency and distinctiveness do exist at a level that allows
statistically significant discrimination accuracy in this task.
linkage is conducted with unsolved crimes). Indeed, the estimates of discrimination accuracy derived from samples of solved crime may be unrealistically high compared with what we would expect to see in real life, because crimes that are solved may be more consistent and distinctive than those that remain unsolved [12,13]. For example, an offender who commits crime in a very restricted geographical area may be easier to apprehend than a more transient offender. If this were true, solved crime series would tend to have shorter intercrime distances than unsolved crime series, thereby leading to an artificially high AUC value when discrimination accuracy was tested with a sample composed solely of solved crime. It is, therefore, important to test discrimination accuracy under conditions that more closely reflect how behavioural case linkage is practised in reality.

In reality, there are two scenarios in which behavioural case linkage would be conducted [12]. In the first, an investigator would be presented with a specific crime (often a solved crime) and they would be asked to search amongst their databases for unsolved crimes that were linked to this particular index offence (reactive case linkage). In this situation the investigator would be required to link a solved crime with an unsolved crime (referred to as solved-unsolved links). In the second scenario the investigator would not have a specific index offence, but would instead be asked to search proactively through the organisation’s databases to identify potential linkages (proactive case linkage). Typically, this would involve linking an unsolved crime with another unsolved crime (unsolved-unsolved links). In these two scenarios there is no situation where an investigator would be required to conduct case linkage with two crimes that have already been solved (solved-solved links). Empirical research into behavioural case linkage would, therefore, be more realistic if it sampled both solved and unsolved crimes and only tested discrimination accuracy with solved-unsolved links and unsolved-unsolved links (thereby excluding the solved-solved links that have traditionally been used in previous research). This represents a fundamental change in
methodology, which may provide a more realistic evidence base upon which to develop the practice of behavioural case linkage.

To facilitate such an approach, it has been suggested that researchers could use samples of solved and unsolved crime that have been linked via DNA evidence [12,31]. This would allow the necessary solved-unsolved and unsolved-unsolved links to be established. Only one existing study of behavioural case linkage has been able to collate a sample of DNA-linked series, which was a study of 22 serial rapists and their offences in South Africa [32]. A combination of 114 offender behaviours⁴ was able to facilitate a statistically significant level of discrimination accuracy (AUC = 0.88). These findings suggest that behavioural case linkage has the potential to work when applied to samples that contain both solved and unsolved crimes, thereby providing further support for its use in practice. However, Woodhams and Labuschagne’s [32] study shares the same limitation of previous research, which is that it only considers the possibility of linking within crime types.

The current paper, therefore, reports an examination of behavioural consistency, distinctiveness, and discrimination accuracy with a sample that contained both solved and unsolved crimes that were linked via DNA evidence and had been committed by a group of versatile serial offenders. In particular, this research tested whether Tonkin et al.’s [24] recent findings could be replicated in a sample containing both solved and unsolved offences. This provided a more realistic insight into the potential contribution of behavioural case linkage in practice.

Method

⁴ The behaviours tested by Woodhams and Labuschagne [32] included sexual behaviours (such as whether the victim was penetrated anally), approach behaviours (such as whether the victim was approached by the offender in private or in public), and control behaviours (such as whether the victim was physically restrained in any way). The full list of behaviours is reported by Woodhams and Labuschagne [32].
The data

An electronic search of Northamptonshire Police databases was made to identify all crimes that had been linked to a common offender via DNA evidence between 2005 and 2009\(^5\). Northamptonshire is a county in the East Midlands of England, with an approximate area of 2,364 km\(^2\) and a population of approximately 687,300 in 2010. From this sample, all offenders were identified who had committed at least two crimes during the study period (with at least one of their crimes being classed as unsolved). It was necessary to include at least one unsolved crime for each offender so that when the linked crime pairs were formed (as described in the Design and Procedure section) there were no solved-solved links included in the tests of discrimination accuracy. This ensured that the current research methodology was closer than previous research to the real life situation in which case linkage is conducted. One hundred and thirty two offenders were identified who had committed two or more offences between 2005 and 2009 within Northamptonshire. It has been common in previous research to select a constant number of offences per offender to ensure that prolific offenders (who may display unusually high or low consistency and distinctiveness in their offending behaviour) do not exert an undue influence on the analyses [e.g. 1,17,22]. Consistent with Tonkin et al.’s [24] original methodology, two offences per offender were randomly selected (ensuring that at least one of these offences was classed as unsolved), thus creating a sample of 264 solved and unsolved offences committed by 132 offenders. One hundred and ninety five (74\%) of these offences were unsolved and 69 were solved (26\%)\(^6\).

\(^5\) As with Tonkin et al. [24], a number of crime types were excluded because they do not typically have definite offence locations and times, which makes it difficult to calculate meaningful intercrime distance and temporal proximity values (e.g. crimes included under the Home Office categories of ‘drugs offences’, ‘fraud or forgery offences’, and ‘other offences’ were excluded).

\(^6\) For the purposes of this study, crimes that were classed as “detected” on Northamptonshire Police databases were considered to be solved, whereas those classed as “undetected” were considered unsolved. A crime is classed as detected when there is sufficient evidence for the case to be submitted to the Crown Prosecution Service (the CPS are the government department responsible for prosecuting criminal offences in England and Wales). It should be noted that CPS policy does not allow a DNA match between a suspect and a crime scene to
The geographical location of each offence (stored as an x,y coordinate) and the time and date that the offence was reported to the police were extracted for analysis.

**Design and procedure**

The methodology described by Tonkin et al. [24] was followed. Three different types of crime pair were created: 1) Across Crime Categories; 2) Across Crime Types; 3) Within Crime Types (see Table 1). Cross-category pairs contained two crimes from different Home Office categories of crime (e.g. a residential burglary was paired with a rape). Cross-type pairs contained two crimes from the same broad category of crime but of different specific types (e.g. a residential burglary was paired with a commercial burglary). Within-type pairs contained two crimes from the same broad category of crime and of the same specific type (e.g. a residential burglary was paired with a residential burglary). Both linked and unlinked crime pairs were created for each of the three different types of pair, thus creating a total of six crime pair subsets (see Table 1). The linked crime pairs contained two crimes that had been committed by the same offender, whereas the unlinked pairs contained two crimes that had been committed by different offenders. In this methodology the unlinked crime pairs represented the degree of behavioural consistency/distinctiveness one would expect to see through chance, so any consistency/distinctiveness observed in the linked pairs above and beyond that in the unlinked pairs was evidence to support the potential viability of behavioural case linkage. For each crime pair the number of kilometres (the intercrime distance) and the number of days (temporal proximity) separating the two offences was calculated. These procedures prepared the data for analysis.

---

be submitted as evidence of guilt without corroborating evidence. Thus, it was possible in this study for two crimes to be linked via DNA evidence (and, therefore, considered to be committed by the same person), but for one of these crimes to be classed as solved (i.e. detected) and the other to be classed as unsolved (i.e. undetected).
Data analysis

The six crime pair subsets were compared statistically to determine whether they differed in the intercrime distance and temporal proximity. Statistically significant differences between the linked and unlinked subsets (i.e. linked cross-category versus unlinked cross-category; linked cross-type versus unlinked cross-type; linked within-type versus unlinked within-type) demonstrated the potential for behavioural case linkage to work in a statistical sense when compared with chance. If these statistical differences were similar in magnitude (as measured by the effect size) and there was no difference between the three linked pair subsets in terms of the intercrime distance and temporal proximity, this would suggest that the potential for linking crime was comparable across crime categories, across crime types, and within crime types [24].

To examine discrimination accuracy, binary logistic regression analysis was conducted using a Leave-One-Out (LOO) classification. Three regression analyses were run at each level of analysis (i.e. across crime categories; across crime types; within crime types), two of which were direct regressions that examined the independent ability of intercrime distance or temporal proximity to predict linkage status and one of which was a stepwise regression that examined the combined predictive value of intercrime distance and temporal proximity. The stepwise regression, therefore, identified the most simple and efficient statistical model that maximised discrimination accuracy with these data. These nine regression analyses (3 regressions X 3 levels of analysis) produced predicted probability values that were used to conduct Receiver Operating Characteristic (ROC) analyses. ROC analysis provided an index of discrimination accuracy called the AUC, which can range from 0 (indicating complete inaccuracy in the discrimination task) to 1 (indicating complete
accuracy). Typically, AUC values of 0.50 to 0.70 are considered to indicate a low level of
accuracy, values of 0.70 to 0.90 are moderate, and values of 0.90 to 1.00 are high [33].

When testing discrimination accuracy, it is important that the estimates of accuracy
have a reasonable likelihood of generalising beyond the sample studied [34]. This is
particularly important in the current area of research, where the aim is to develop findings
that can be used to guide future investigations. Tonkin et al. [24] used split-half cross-
validation to test the generalisability of their findings. However, this method only splits the
data once, which increases the likelihood of a good (or bad) fit occurring through chance,
particularly when sample size is small [35,36]. To overcome this potential limitation, cross-
validation was provided by the LOO classification method in this study, which performs
multiple splits of the data. As explained by Woodhams and Labuschagne [32], this method
involves removing each crime pair from the sample one at a time and the remaining data are
then used to develop a logistic regression model, which is subsequently applied to the
extracted pair to produce a predicted probability value. This pair is then returned to the
dataset and the procedure repeated with the next pair until a probability value has been
calculated for all pairs [32]. These probability values were subsequently used to produce
ROC curves that indicated discrimination accuracy. Consequently, the logistic regression
models are developed and tested on different samples, which provides reassurance that the
estimates of discrimination accuracy have a reasonable likelihood of generalising to future
crimes in this jurisdiction. However, split-half validation was also performed on the data and
a further nine ROC curves constructed to examine whether the two methods of cross-
validation produced comparable results, thereby increasing confidence that the current
findings could be reliably compared with those of Tonkin et al. [24].
It should be noted that non-parametric statistics were appropriate throughout the analyses due to the distributions of intercrime distance and temporal proximity values departing significantly from normal ($p < .05$), as indicated by Kolmogorov-Smirnov tests.

**Results and Discussion**

**Statistical comparisons**

Two Friedman’s ANOVAs indicated significant differences across the six crime pair subsets in terms of intercrime distance, $\chi^2 (5) = 61.77$, $n = 32$, $p < .001$, and temporal proximity, $\chi^2 (5) = 31.96$, $n = 32$, $p < .001$. The results of post-hoc comparisons (Bonferroni corrected $\alpha = 0.008$) and effect size calculations are presented in Table 2. From this table we can see that all three linked subsets contained shorter intercrime distance and temporal proximity values than their unlinked counterparts ($p < .001$ with large effect sizes; [37]), except for the difference in temporal proximity between linked and unlinked cross-type pairs ($p > .05$ with a small to medium effect size; [37]). When the three linked subsets were compared with each other, none of these comparisons reached the adjusted alpha value ($p > .01$ with small to medium effect sizes; [37]).

These findings suggest that behavioural case linkage has the potential to work reliably when compared with chance using the intercrime distance and the temporal proximity to link across crime categories, across crime types, and within crime types. Furthermore, it seems that the level of discrimination accuracy is comparable across the three levels of analysis. However, it should be noted that the temporal proximity may not be able to facilitate significant discrimination accuracy across crime types. These findings are generally consistent with those reported by Tonkin et al. [24], which suggests that the basic findings
observed with a sample of solved crimes are replicated in a sample containing both solved and unsolved crimes.

[INSERT TABLE 2 HERE]

ROC analyses

To examine discrimination accuracy, eight ROC curves were produced (see Table 3). A ROC curve was not constructed for the stepwise regression model across crime types because the stepwise analysis only included the intercrime distance in the final model (i.e. the addition of temporal proximity did not improve discrimination accuracy). Consequently, the AUC value was identical for the stepwise and intercrime distance models at this level of analysis.

[INSERT TABLE 3 HERE]

With the exception of temporal proximity across crime types, all models were statistically significant. Also, all models achieved moderate levels of discrimination accuracy, except temporal proximity across crime types and crime categories [33]. But, the intercrime distance achieved larger AUC values than temporal proximity when linking across crime categories and across crime types. These findings indicate that certain types of offender behaviour appear to demonstrate the requisite levels of consistency and distinctiveness to facilitate the linking of crimes across categories, across types, and within types. Overall, the basic conclusions from this study are consistent with those reported by Tonkin et al. [24], but the size of the AUC values differed in some instances (see Table 4). Most notably, the intercrime distance achieved a lower level of discrimination accuracy within crime types in
the current study compared with the findings of Tonkin et al. [24]. Also, the level of accuracy achieved using the temporal proximity was lower across crime types in the current study. Both of these differences were confirmed statistically ($p < .05$) using ROCKIT 0.9B © (University of Chicago, Chicago, IL, United States), which has been used to compare AUC values in previous studies of case linkage [13,14,16,17]. However, the other estimates of discrimination accuracy were similar across the two studies ($p > .05$), although it should be noted that the lower level of discrimination accuracy in this study for the intercrime distance across crime types approached significance ($p = .08$).

When the analyses were repeated using the split-half method of cross-validation there were only minor differences in the size of the AUC values obtained (Mean difference in the AUC = 0.04), with none of these differences reaching statistical significance (as indicated by overlapping confidence intervals; [19]). This provides reassurance that the current findings can be reliably compared with those of Tonkin et al. [24], despite the use of different cross-validation methods.

In some instances, previous research with solved crime may, therefore, have over-estimated the potential for offender behaviour to link crimes. This partially validates the concerns expressed by researchers of case linkage [12,13] and suggests that the degree of linkage error may be larger than previously thought. However, it is important to note that — despite the reduction in discrimination accuracy that was observed in the current study — the majority of models were statistically significant, which indicates that behavioural case linkage has the potential to be a viable procedure in practice.

7 The specific ROC figures can be obtained from the first author upon request.
Practical applications

Given the statistically significant levels of behavioural consistency, distinctiveness, and discrimination accuracy that were observed in the current study, it is important to discuss how these findings might be applied in practice. These findings could be used to develop a statistical tool that would support the linking of crime. For a given dataset of crimes, this tool would calculate the intercrime distance (or the temporal proximity if geographical information were unavailable) for each pairwise combination of crimes in the dataset (e.g. the geographical distance between crime 1 and crime 2 in the dataset, the distance between crime 1 and crime 3, and so on). These distances would then be arranged in order from highest to lowest, thereby providing the investigator with a prioritised list of potentially linked crimes. This would help to overcome one of the main challenges faced by investigators at the early stages of case linkage, which is that they often face a large volume of crime and it is simply not feasible to conduct individual analysis for every crime. The linkage tool proposed here would, therefore, provide the investigator with an evidence-based and time-efficient method for prioritising the investigation of certain crimes. This filter-based approach is similar to that suggested by other researchers [13,38], but importantly these findings indicate that this approach could be extended beyond series that contain just one type of crime to series that contain a variety of different offences. This would allow those prolific and versatile offenders who are often responsible for a disproportionate amount of crime to be investigated in a more efficient and successful manner.

However, it is clear from the findings reported in this study and previous studies that behavioural case linkage which is based on such tools will be associated with a degree of error. Indeed, the AUC values reported in the current study were below the maximum value of 1.00, which indicates that both false positive and false negative errors were observed when
using the intercrime distance and temporal proximity to identify linked and unlinked crime pairs. Investigators should, therefore, be cautious when applying behavioural case linkage in practice and should always present their conclusions with appropriate caveats.

**Theoretical implications**

In addition to the practical applications, there are theoretical insights that can be gleaned from the present study. In particular, the moderate levels of discrimination accuracy that were achieved using the intercrime distance indicate that the offenders in this sample tended to offend in relatively restricted geographical areas that did not overlap considerably with those of other offenders [13]. This suggests that the offenders in this sample had somewhat well-defined offending territories within which they preferred to offend. These findings extend those observed with specific crime types, such as residential burglary, commercial robbery, and car theft [14,17,18] to series comprising a mixture of crime types, thereby suggesting that the offenders in the present sample tended to offend in broadly similar geographical regions from one crime to the next, regardless of what type of crime they were committing. This supports several seminal models of spatial behaviour that assume generic psychological processes are involved in the production of criminal spatial behaviour, irrespective of crime type [39-41]. In terms of explaining why the offenders preferred to offend in certain geographical areas, the literature on geographical profiling is relevant. Specifically, this research has demonstrated that offenders tend to show a preference for committing crime in areas that are personally significant (e.g. they live, work, or socialise in these areas) [42-44]. One might, therefore, predict a degree of individual difference from one offender to the next in the areas that are personally significant, which would explain why somewhat distinct offending territories emerged for the offenders in this sample.
However, it is a well-known finding within the criminology literature that certain locations and certain targets appeal to the majority of offenders when they are choosing places to commit crime [45,46]. It is, therefore, unsurprising that there was a degree of overlap in the offending territories of each offender in the current study (as indicated by AUCs below 1.00). Nevertheless, it is clear that there was sufficient criminal opportunity for the offenders to develop statistically distinctive criminal domains.

Another interesting finding from this study was that the AUC for temporal proximity was larger within crime types than it was across crime categories/types, whereas the AUC for intercrime distance remained relatively stable across the three levels of analysis. Thus, the temporal distance between offences tended to be smaller when crimes of the same type were committed compared with crimes from different types and categories. These findings suggest that the offenders in this sample may have engaged in periods of short-term specialisation, whereby they showed a preference for committing one type of offence over others for a restricted period of time, but subsequently changed this preference for a different crime type (due perhaps to “local life circumstances”, such as social and romantic relationships, and employment and living circumstances [47,48]- the reason for these potential changes cannot be determined from these data, however). If this interpretation is correct, it would fit with recent criminological theory that indicates long-term versatility in offending behaviour but also periods of short-term specialisation [e.g. 47,48].

Limitations and future directions for research

Having considered the implications of the findings, it is important to consider the limitations of this study. First, the analysis was restricted to two crimes per offender. While this decision was necessary to replicate the methodology of Tonkin et al. [24] and to prevent prolific offenders having an undue influence on the analyses [1,17,22], researchers have
recently argued that this approach does not replicate real life case linkage because investigative databases will contain crime series that vary in length [32]. Consequently, future research should explore the impact of including all offences in an offender’s series, as this may provide a more realistic test of behavioural case linkage.

Second, it is important to recognise the potential biases that arise from using a sample of DNA-linked crimes. This methodology naturally meant that the sample consisted solely of offenders who left DNA evidence at the scenes of their crime. Arguably, these offenders may differ from those that do not leave physical evidence, which raises doubts about the generalisability of these findings to the wider offender population. One might argue, for example, that offenders who leave DNA are more impulsive and opportunistic and engage in less pre-offence planning than offenders who do not leave DNA. These are characteristics that might be associated with spatial and temporal offender behaviour.

Third, the current methodology used scene-to-scene DNA matches as the basis for determining whether crime pairs were linked or unlinked. This is in contrast to previous research, which has used detection status to make these decisions [e.g. 13,15,24]. This is potentially a limitation if the offender’s DNA was found at a crime scene for some reason other than that s/he committed the crime. For example, maybe a suspect’s DNA was left at the victim’s house because s/he is a friend of the victim, rather than the perpetrator. In a situation such as this, a crime pair classed as linked in the current study would in fact be unlinked. This would introduce noise into the analyses. However, the pattern of findings observed in this study does not suggest that noise has been systematically introduced as a result of the methodology. If this were the case, one would expect to see a significant reduction in discrimination accuracy across all statistical models tested in this study when compared with Tonkin et al.’s [24] findings. In reality, the majority of statistical models
achieved comparable levels of discrimination accuracy to those obtained by Tonkin et al. [24].

Fourth, the current study only considered two very simple measures of offender behaviour. While the simplicity of the intercrime distance and temporal proximity is actually a strength (because they would be easy to implement in practice and would require very few changes to the way in which data are stored by investigative organisations — unlike other behavioural approaches), future research must explore whether different types of offender behaviour can be used to link across crime categories and crime types. This may, however, require a slightly different methodology to that utilised in the current study. For example, researchers might focus on certain types of crime that share particular behavioural features, such as robbery, rape and murder, which all contain elements of victim-offender interaction, control and escape behaviours (e.g. the use of a weapon, methods of victim restraint, and attempts to conceal one’s identity from the victim). Indeed, a research study investigating the potential to link across rapes and robberies using such behaviour is already underway at the University of Birmingham, UK. This may help to overcome the obvious difficulty posed by studying a diverse range of crimes that contain often very different types of offender behaviour. Alternatively, future work might consider developing behavioural themes from offender crime scene behaviour that would subsequently be used to examine discrimination accuracy across crime categories/types. These themes could be developed statistically using techniques such as multidimensional scaling and cluster analysis [e.g. 20,29,31] or researchers might consider using theoretical models that are designed to apply to a wide variety of crime types, such as the Narrative Action System (NAS) model [49,50]. Regardless of the approach used, research such as this is important because investigators will otherwise have no guidance in situations where geographical and temporal behaviour is either
unreliable or absent (e.g. where a victim has been drugged or knocked unconscious and is unable to recall where and when the offence took place).

Fifth, while the sample size in this study compares favourably to some previous case linkage research [e.g. 15,16,17], the sample should still be considered relatively small. The limitations on sample size arose primarily due to the low recovery rate of DNA at the scenes of crime (as discussed in the introduction), which precluded the inclusion of many unsolved crimes in this study. This demonstrates a limitation of the methodology used in the current study because, despite sampling data across the whole of Northamptonshire’s electronic crime records, only 132 offenders were identified for analysis. Thus, future research with samples of solved and unsolved crime that have been linked via DNA may have to consider sampling from larger, more urbanised forces and combining datasets from several forces to yield datasets that are large enough to support the necessary analyses.

Sixth, the analyses reported in this study (and, in fact, those reported in many previous studies of case linkage; [e.g. 1,13-24]) have used chance as the benchmark for judging whether a particular statistical model has the potential to support case linkage in practice. This allows us to conclude that the statistical models reported in the current study can achieve discrimination accuracy that is greater than chance, but it does not allow us to conclude that these models are better than the methods currently available to the police (i.e. the discrimination accuracy achieved by experienced crime analysts). Indeed, crime analysts might be expected to perform at a level that exceeds chance due to their experience of crime, criminal behaviour and case linkage. Future research should, therefore, compare the statistical models reported in this study with the discrimination accuracy of experienced crime analysts using mock case linkage tasks [e.g. 51,52]. This will provide further insight into the potential practical value of these findings.
Conclusions

Previous research has not tested behavioural case linkage in a way that reflects how it is used in practice, but the use of solved and unsolved offences in this study provides a more realistic test of discrimination accuracy. This study, therefore, provides a stronger evidence base upon which to develop the practice of behavioural case linkage. Furthermore, these findings provide investigative organisations with guidance on how they can identify linked crime series that contain several different types of offence, which has the potential to contribute significantly to the investigation, detection, and prosecution of prolific and versatile offenders.
References


http://www.urban.org/UploadedPDF/411697_dna_field_experiment.pdf


Table 1

*A Summary of the Six Crime Pair Subsets Included in the Analyses*

<table>
<thead>
<tr>
<th>Crime Pair Type</th>
<th>Linkage Status</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across Crime Categories</td>
<td>Linked</td>
<td>This subset includes crime pairs that contain two crimes from different Home Office* crime categories that have been committed by the same offender.</td>
<td>A personal robbery committed by offender 1 was paired with a burglary in a dwelling also committed by offender 1.</td>
</tr>
<tr>
<td></td>
<td>Unlinked</td>
<td>This subset includes crime pairs that contain two crimes from different Home Office categories that have been committed by different offenders.</td>
<td>A burglary in a dwelling committed by offender 1 was paired with the theft of a motor vehicle committed by offender 2.</td>
</tr>
<tr>
<td></td>
<td>Linked</td>
<td>This subset includes crime pairs that contain two crimes from the same Home Office crime category that are of different specific crime types. These two crimes have been committed by the same offender.</td>
<td>Personal robbery committed by offender 1 was paired with a commercial robbery also committed by offender 1.</td>
</tr>
<tr>
<td>Type</td>
<td>Description</td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Unlinked</td>
<td>This subset includes crime pairs that contain two crimes from the same Home Office crime category that are of different specific crime types. These two crimes have been committed by different offenders.</td>
<td>A shoplifting offense committed by offender 1 was paired with a theft from a vehicle committed by offender 2.</td>
<td></td>
</tr>
<tr>
<td>$n = 32$ pairs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linked</td>
<td>This subset includes crime pairs that contain two crimes committed by the same offender that are of the same specific crime type.</td>
<td>Two personal robbery crimes committed by offender 1 were paired.</td>
<td></td>
</tr>
<tr>
<td>$n = 53$ pairs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Crime Types</td>
<td></td>
<td>Two burglaries in a dwelling committed by different offenders were paired.</td>
<td></td>
</tr>
</tbody>
</table>

*The Home Office is the government body responsible for setting definitions of crime that police in England and Wales follow.*
### Table 2

**Post-Hoc Comparisons of the Six Crime Pair Subsets in terms of the Intercrime Distance and Temporal Proximity**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Intercrime Distance</th>
<th>Temporal Proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LCrCat-UCrCat</strong></td>
<td>( Z = 5.08, n = 47, p &lt; .001^*, r = 0.74 )</td>
<td>( Z = 3.49, n = 47, p &lt; .001^*, r = 0.51 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LCrCat} = 1.73; \text{UCrCat} = 18.31 )</td>
<td>( \text{Median (Days): } \text{LCrCat} = 199.00; \text{UCrCat} = 427.00 )</td>
</tr>
<tr>
<td><strong>LCrTyp-UCrTyp</strong></td>
<td>( Z = 3.80, n = 32, p &lt; .001^*, r = 0.67 )</td>
<td>( Z = 1.66, n = 32, p &gt; .05, r = 0.29 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LCrTyp} = 1.77; \text{UCrTyp} = 13.90 )</td>
<td>( \text{Median (Days): } \text{LCrTyp} = 287.50; \text{UCrTyp} = 377.50 )</td>
</tr>
<tr>
<td><strong>LWi-UWi</strong></td>
<td>( Z = 4.72, n = 53, p &lt; .001^*, r = 0.65 )</td>
<td>( Z = 4.76, n = 53, p &lt; .001^*, r = 0.65 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LWi} = 3.09; \text{UWi} = 19.70 )</td>
<td>( \text{Median (Days): } \text{LWi} = 119.00; \text{UWi} = 522.00 )</td>
</tr>
<tr>
<td><strong>LCrCat-LCrTyp</strong></td>
<td>( Z = 0.88, n = 32, p &gt; .05, r = 0.16 )</td>
<td>( Z = 1.65, n = 32, p &gt; .05, r = 0.29 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LCrCat} = 1.73; \text{LCrTyp} = 1.77 )</td>
<td>( \text{Median (Days): } \text{LCrCat} = 199.00; \text{LCrTyp} = 287.50 )</td>
</tr>
<tr>
<td><strong>LCrCat-LWi</strong></td>
<td>( Z = 2.10, n = 47, p = 0.04, r = 0.31 )</td>
<td>( Z = -0.42, n = 47, p &gt; .05, r = -0.06 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LCrCat} = 1.73; \text{LWi} = 3.09 )</td>
<td>( \text{Median (Days): } \text{LCrCat} = 199.00; \text{LWi} = 119.00 )</td>
</tr>
<tr>
<td><strong>LCrTyp-LWi</strong></td>
<td>( Z = -0.06, n = 32, p &gt; .05, r = -0.01 )</td>
<td>( Z = -1.66, n = 32, p &gt; .05, r = -0.29 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Median (km): } \text{LCrTyp} = 1.77; \text{LWi} = 3.09 )</td>
<td>( \text{Median (Days): } \text{LCrTyp} = 287.50; \text{LWi} = 119.00 )</td>
</tr>
</tbody>
</table>

* Denotes significance at the adjusted alpha level (\( \alpha = 0.008 \)).
a LCrCat = Linked Cross-Category pairs; UCrCat = Unlinked Cross-Category pairs; LCrTyp = Linked Cross-Type pairs; UCrTyp = Unlinked Cross-Type pairs; LWi = Linked Within-Type pairs; UWi = Unlinked Within-Type pairs.
Table 3

Receiver Operating Characteristic (ROC) Results for Behavioural Case Linkage Across Crime Categories, Across Crime Types and Within Crime Types using the Intercrime Distance and Temporal Proximity

<table>
<thead>
<tr>
<th>Case Linkage</th>
<th>Feature</th>
<th>AUC(^a) (SE)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across Crime Categories</td>
<td>Intercrime Distance</td>
<td>0.86 (0.04)**</td>
<td>0.78-0.93</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.67 (0.06)**</td>
<td>0.56-0.78</td>
</tr>
<tr>
<td></td>
<td>Stepwise</td>
<td>0.88 (0.04)**</td>
<td>0.80-0.95</td>
</tr>
<tr>
<td>Across Crime Types</td>
<td>Intercrime Distance</td>
<td>0.79 (0.06)**</td>
<td>0.67-0.90</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.53 (0.08)</td>
<td>0.39-0.68</td>
</tr>
<tr>
<td></td>
<td>Stepwise(^b)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Within Crime Types</td>
<td>Intercrime Distance</td>
<td>0.77 (0.05)**</td>
<td>0.68-0.86</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.77 (0.05)**</td>
<td>0.68-0.86</td>
</tr>
<tr>
<td></td>
<td>Stepwise</td>
<td>0.83 (0.04)**</td>
<td>0.75-0.91</td>
</tr>
</tbody>
</table>

\(^*\) \(p < .05\); \(^**\) \(p < .01\); \(^***\) \(p < .001\)

\(^a\) Area Under the Curve (AUC) values of 0.50 - 0.70 are considered low, values of 0.70 - 0.90 are considered moderate, and values of 0.90 - 1.00 are high [33].

\(^b\) A stepwise model was not reported across crime types because the combination of intercrime distance and temporal proximity did not lead to statistically improved discrimination accuracy compared with the single-feature regression models.
Table 4

*Receiver Operating Characteristic (ROC) Results Reported by Tonkin et al. [24]*

<table>
<thead>
<tr>
<th>Case Linkage</th>
<th>Feature</th>
<th>AUC(^a) (SE)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Across Crime</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categories</td>
<td>Intercrime Distance</td>
<td>0.88 (0.03)***</td>
<td>0.82 - 0.95</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.67 (0.05)**</td>
<td>0.57 - 0.78</td>
</tr>
<tr>
<td></td>
<td>Stepwise</td>
<td>0.88 (0.04)***</td>
<td>0.82 - 0.95</td>
</tr>
<tr>
<td><strong>Across Crime</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Types</td>
<td>Intercrime Distance</td>
<td>0.90 (0.03)***</td>
<td>0.84 - 0.97</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.74 (0.05)***</td>
<td>0.64 - 0.83</td>
</tr>
<tr>
<td></td>
<td>Stepwise(^b)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Within Crime</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Types</td>
<td>Intercrime Distance</td>
<td>0.91 (0.03)***</td>
<td>0.84 - 0.97</td>
</tr>
<tr>
<td></td>
<td>Temporal Proximity</td>
<td>0.74 (0.05)***</td>
<td>0.64 - 0.84</td>
</tr>
<tr>
<td></td>
<td>Stepwise(^b)</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

\(^{**p < .01; ***p < .001}\)

\(^a\) Area Under the Curve (AUC) values of 0.50 - 0.70 are considered low, values of 0.70 - 0.90 are considered moderate, and values of 0.90 - 1.00 are high [33].

\(^b\) A stepwise model was not reported across crime types or within crime types because the combination of intercrime distance and temporal proximity did not lead to statistically improved discrimination accuracy compared with the single-feature regression models.