This is the author accepted manuscript of a published article. Please refer to this work as follows (APA 7th):

Snaphaan, T., Hardyns, W., Pauwels, L. J. R., & Bowers, K. (2024). Rating places and crime prevention: Exploring user-generated ratings to assess place management. *Computers, Environment and Urban Systems*, *109*, Article 102088. https://doi.org/10.1016/j.compenvurbsys.2024.102088

Rating places and crime prevention: Exploring user-generated ratings to assess place management

Thom Snaphaan[a,b*], Wim Hardyns[a,c], Lieven J. R. Pauwels[a] & Kate Bowers[d]

[a] Department of Criminology, Criminal Law and Social Law, Ghent University, Universiteitstraat 4, 9000 Ghent, Belgium
[b] Centre of Expertise Safe and Resilient Society, Avans University of Applied Sciences, Hugo de Grootlaan 37, 5223 LB s-Hertogenbosch, the Netherlands
[c] Institute of Environment and Sustainable Development, University of Antwerp, Sint-Jacobstraat 2, 2000 Antwerp, Belgium
[d] Department of Security and Crime Science, University College London, 35 Tavistock Square, WC1H 9EZ London, UK
* Corresponding author: Thom.Snaphaan@UGent.be

Abstract

This study assesses how the quality of place management (measured with user-generated ratings from Google Places) is related to crime occurrences at specific settings and whether specific crime types are related to specific types of places. In 50 randomly sampled neighborhoods in Ghent (Belgium) and London (United Kingdom), we analyzed Google Places data as a proxy measure for the quality of place management at the street segment level. We used hurdle models to examine the effects for both the prevalence and frequency of crime at micro places, and to deal with excess zeros in the data. User-generated ratings of places provide a useful place-level indicator for place management that are related to crime. However, contextual differences are found between Ghent and London. For London, the results suggest that higher quality of place management has a protective effect on crime occurrences at the street segment level. This study indicates the importance of exploring new and emerging data sources as unique measurement opportunities to enhance insight in crime prevention mechanisms, and also acknowledges its limitations. For the first time from a large-scale empirical perspective, this study suggest that improving place management at specific places might be an effective intervention to guard against crime.

Keywords

Big data, Crime concentrations, Google Places, Place management, Situational crime prevention

1 Introduction

Why do a small portion of places in a larger area (e.g., a city) experience most of the crime and most of the places only little or no crime? Why is a particular business a crime hot spot but other similar facilities not? These practical questions regarding the spatial patterning of crime are central to environmental criminology (Bruinsma & Johnson, 2018) and particularly the criminology of place (Sherman et al., 1989; Weisburd et al., 2012). Crime is concentrated at micro places to such an extent that Weisburd (2015) postulates a "law of crime concentration at places." Related, Wilcox and Eck (2011) state that only a small minority of facilities within all risky facilities experience the majority of offenses, also known as the "iron law of troublesome places." Beyond these descriptive observations, one of the key questions remains *why* crime is so concentrated.

Prior research assessed the potential mechanisms that affect crime at micro places from various perspectives. Weisburd et al. (2016, p. 48) put forward ten explanations of why crime is concentrated if one ignores random variation: (1) poor physical design, (2) lack of informal social control, (3) inadequate guardianship, (4) excessive reporting of crime, (5) repeat victimization, (6) inadequate handling, (7) large number of offenders (crime attractors), (8) large number of targets (crime generators), (9) hot products, and (10) inadequate place management. The four functions of place management (i.e., organization of space, regulation of conduct, control of access, and acquisition of resources) provide an overarching opportunity theory to include the other nine explanations of why crime is concentrated. Place management is a quite simple and long-time often overlooked (Douglas & Welsh, 2020; except from some recent studies, e.g., Gomory & Desmond, 2023; Lee et al., 2022; Tucker & O'Brien, 2023), but powerful explanation of why some (proprietary) places experience high levels of crime and why others do not. It also explains why within the same facility type, crime concentrates at a small proportion of the facilities (Eck et al., 2007) – i.e., the "iron law of troublesome places." Place management is also a mechanism that explain the processes that keep crime out of most places (Eck, 2019), and, hence, provide more valid explanation of crime concentrations than for example simply land use (Lee et al., 2022). Despite the powerful (hypothesized) explanatory power of this explanation of crime concentrations, empirical evidence is rather scarce.

New and emerging technologies provide new data sources – also known as big data, which are in fact new measurement opportunities to enhance insights in a wide range of social issues. Buyalskaya and colleagues (2021) therefore state that we are currently in the "golden age of social science" (p. 1). One of these opportunities are provided by Location-Based Services

(LBS). As mentioned in the quote from Kuralarasan and Bernasco (2022, p. 674) above, this large-scale spatially referenced information from location service providers such as Google (e.g., Google Maps, Google Street View) is becoming increasingly important in environmental criminology (Snaphaan & Hardyns, 2021; Vandeviver, 2014). A part of Google's service that has not yet often been utilized in crime-related research is Google Places. This service, and in particular the Place Details request, provides detailed information about specific places, including user ratings and reviews.

In this study, we use Google Places' ratings of non-residential facilities as a measure of place management. The aim of this article is to assess the mechanism of place management in relation to crime distributions at micro places, from an empirical perspective. Therefore, we examine the types, presence, ratings, and diversity of place types at the street segment level in Ghent (Belgium) and London (UK) in relation to five specific crime types, consisting of violent, property and hybrid types of crime, and a form of physical incivilities.

2 Background

2.1 Place management theory

As an "outgrowth of routine activity theory" (Lee et al., 2022, p. 1012), place management theory provides explanations for the assertion that operators of places, due to their roles can have significant impact in reducing and preventing crime, but some of them fail (Eck & Madensen, 2018). It is built on the contemporary version of routine activity theory that states that crime occurs when *offenders* and *targets* converge in *places* where all three potential controllers (respectively handlers, guardians, and managers) are absent, negligent, or ineffective (Eck, 2010). Place management mechanisms explain why a few proprietary places are heavily involved in crime, and the processes that keep crime out of most places (Eck, 2019; Weisburd et al., 2016).

Place management is defined as "a set of four processes that owners, their employees, and others use to organize the physical and social environment of a location so that the functions of the place can be carried out" (Madensen, 2007, p. 19). The four processes that are referred to, are the organization of space, regulation of conduct, control of access, and acquisition of resources. These processes, also known under the acronym ORCA, explain why places with poor organization of space, inept regulation of conduct, a lack of control of access or servicing the wrong clientele, and an inability to garner resources are more criminogenic than others.

An owner or operator of a place is considered as a place manager. The role of a place manager will vary to some degree depending on the place, but their common interests is to assure the smooth functioning of the place (Eck & Madensen, 2018; Weisburd et al., 2016). Facilities are homogeneous sets of proprietary places (e.g., all restaurants or all bars; Madensen & Eck, 2012). Several facility types have been implicated as generating or attracting crime within the communities where they sit in, respectively referred to as crime generators and crime attractors (Brantingham & Brantingham, 1995). While these risks among different types of facilities can vary, risks among proprietary place types, representing the same facility, also vary dramatically. Research demonstrates that a small proportion of facilities produce a large proportion of the crimes experienced by the entire group (Eck et al., 2007).[1]

Eck (2019) states that there are three ways in which poor place management can influence wider areas. First, when poorly managed places concentrate, areas with more poorly managed places will have more crime than areas with few poorly managed places. This is a direct result of the simple sum in aggregating high-crime locations. Second, high crime places can be part of networks of crime-involved locations that create crime hotspots (Madensen et al., 2017). Third, empirical research shows that places with high levels of crime inside their property radiate out to the close surroundings of the places they are associated with (Bowers, 2014; Groff, 2014). Reversely, Linning and Eck (2021) argue that "good place management often radiates safety beyond property boundaries, down streets, and across wider areas." (p.43). When considering place management at the street segment level, the continuous presence of people at the streets (also known as *the constant succession of eyes on the street*) is considered to be relevant in keeping street blocks safe (Eck et al., 2023). A mix of types of businesses and shops create an "intricacy of sidewalk use, bringing with it a constant succession of eyes" (Jacobs, 1961, p. 50) that will be beneficial in terms of preventing crime. Although some empirical tests of place management are emerging (Gomory & Desmond, 2023; Lee et al., 2022; Tucker & O'Brien, 2023), the empirical research on place management is still in infant stages; nevertheless, measurement opportunities are increasingly available.

2.2 Points of interest, ratings, and reviews

With the advent of the so-called Web 2.0 (O'Reilly, 2005) and particularly the rise of social media platforms and location-based services, the possibilities that are available to share experiences and opinions on physical places in a digital manner have increased. There are

[1] This empirical observation is known as the "iron law of troublesome places" (Wilcox & Eck, 2011).

currently plenty of points of interest (POI) directories for capturing user-generated ratings and reviews on places (Jiang et al., 2015). These 'volunteered' data can be generated both proactively and retroactively (Thakuriah et al., 2017). Retroactively user-generated content is generated when secondary sources of user-submitted (interaction) data published through the web, social media, and/or other tools that represent daily activities online are processed. Examples of these data are Twitter data and Foursquare data. Proactively user-generated content is created when users voluntarily and consciously generate data on ideas, help to solve problems, and/or report on events, disruptions or activities that are of social and civic interest. Examples of these data are OpenStreetMap data and Google Places data.

A particular type of proactively user-generated content are ratings and reviews of places. These ratings and reviews of places can be considered as electronic word-of-mouth (eWOM) communication. The producers and consumers of this type of information typically have a high degree of overlap, with individuals contributing to both also being known as 'prosumers' (Ritzer et al., 2012). The questions about who contributes to these platforms and what their motivations are, are widely covered in prior research, mostly tailored to specific platforms. It is beyond the scope of this contribution to go into detail on all these individual platforms, however, generally it is found that contributors are not representative of the general population (e.g., Mashhadi et al., 2013; Solymosi et al., 2018) and there are notable differences within the group of contributors, i.e. participation inequality (e.g., the existence of super-contributors; Stewart et al., 2010). The (intrinsic and extrinsic) motivations to contribute to these platforms are quite diverse. Intrinsic motivations found to be related to the desire to contribute are, among others, altruism (Baruch et al., 2016; Budhathoki, 2010), the desire to support the service provider and future customers (Parikh et al., 2014; Yoo & Gretzel, 2008), personal reputation (Cheung & Lee, 2012) or more collectivist motivations, such as belonging to a community (Baruch et al., 2016). Extrinsic motivations are, among others, related to the goal of the platform (Budhathoki, 2010) or personal experiences (Parikh et al., 2015). Besides these coverage issues, the extent to which ratings and reviews are accurate, relates (apart from processing errors and biases) to both the measurement processes (e.g., measurement error) and the representation processes (e.g., selection error) (see Snaphaan et al., 2024). Therefore, it is important to assess the quality of the volunteered geographic information (Senaratne et al., 2017).

2.3 A valid measure of place management?

The Total Error Framework (Amaya et al., 2020) provides guidance in the assessment of specific errors in the process from theoretical concept to a measurement and representation of

it. Prior research showed that this framework is applicable to relevant concepts in criminology and criminal justice (Snaphaan et al., 2024). Although a lot is still unknown about the errors in Google Places data, theorizing about its use (with a view to representation) and the measurement of the concepts of interest helps in enhancing our understanding of the validity of the measures.

On the measurement side, the main question is to what extent Google Places ratings are a valid representation of (the quality of) place management. Therefore, different forms of validity assessments can be used (see for example Drost, 2011 for an overview). The Google Places star rating system is used by users to review and rate their experience at specific non-residential places (businesses). Assessing face validity[2], the ratings of Google Places provide a rather abstract measure of the quality of place management. The qualification of how a business performs appears to be an appreciation of the extent to which the functions of a place can be carried out or, in other words, the smooth functioning of a place (see Section 2.1). Since the four processes of place management (i.e., ORCA) refer to general processes of the management of facilities and given that Google Places ratings capture a general (single) measure of the experience at those facilities, we assume that Google Places ratings reflect a general measure of the quality of place management at facilities. This study aims to expand this assessment from an empirical perspective, as it aims to take first steps in testing the convergent validity by confronting the obtained measure of place management with crime as an outcome variable.

Regarding the representation of facilities and the appreciations of its functioning by means of Google Places data, several observations can be made. The study of Smarzaro et al. (2017) shows that Google Places has far more coverage of places than other platform, which is advantageous from a representation perspective. Facebook Places, Foursquare and Yelp had respectively 5.8, 15.5 and 23.8 percent of the number of (valid, i.e. de-duplicated and categorized) places compared to Google Places. In our study, for example, Yelp (representing mainly restaurants) yielded 240 results for Ghent in total, while our sample of 50 neighborhoods yielded already 620 eating facilities from Google Places. The representativeness of the appreciation of the functioning of places or, in other words, to what extent the ratings and reviews that people leave are representative of actual place management, is an important question that has not yet been addressed in prior research.

[2] Face validity is understood as how a measure or procedure appears and whether it is considered, at the surface, to represent the construct it is supposed to be measuring (Drost, 2011).

2.4 The current study

This study focuses on place management and its effect on wider areas, namely street segments. A street segment is a microgeographic unit that is often used in spatiotemporal studies of crime (Weisburd et al., 2012). Street segments are considered theoretically meaningful behavioral spaces and people are likely to have 'line of sight' along a street segment (Oberwittler & Wikström, 2009; Taylor, 1997). It is also convenient from an analysis viewpoint – it represents both crimes within places and those external, which helps to situate risky facilities in their locational context. Additionally, in this study, it is argued that place management can be measured by means of new and emerging data sources, particularly Google Places' ratings of POIs. Since Google Places only relate to non-residential facilities, the focus of the present study is on place management in non-residential facilities. In this study, we test the following four hypotheses:

H1 The first hypothesis tested in this study is more generally embedded in crime opportunity theories, that state that land use has an influence on crime in the sense that facilities provide opportunities for crime (e.g., Kinney et al., 2008). Therefore, H1 states that **on street segments with more non-residential facilities (i.e., count of Google Places per street segment) there will be more crime.**

H2 Place management theory also states that good place management radiates safety across wider areas. Therefore, the second hypothesis (H2) is that **higher quality of place management (i.e., Google Places' ratings) will have protective effects for crime (i.e., the higher the rating, the lesser the crime problem)**.

H3 As mentioned before, place management theory states that when poorly managed places concentrate, areas with more poorly managed places will have more crime than areas with few poorly managed places. The third hypothesis (H3) is, accordingly, **on street segments with a higher proportion poorly managed places there will also be a higher concentration of crime**.

H4 Last, place management theory states that a constant succession of eyes is a mechanism for crime prevention (Eck et al., 2023; Jacobs, 1961). From this perspective, the fourth hypothesis (H4) states that **the more diverse the place types on a street segment are, the lower the crime at that street segment will be, because of variable opening times of different types of non-residential facilities**.

3 Data and methods

3.1 Description of the study area and units of analysis

This comparative, empirical study is conducted in Ghent (Belgium) and London (United Kingdom). The city of Ghent has an area of 156 km² and has 264,666 inhabitants in 2022. For London, the study area was narrowed to the 12 Inner London Boroughs[3]. This area is 319 km² and had 3,404,300 inhabitants in 2021 (most recent Census). We selected neighborhoods (201 statistical sectors for Ghent, Belgium and 1,731 Lower Super Output Areas for Inner London, United Kingdom) in both cities to increase comparability. The analyses were conducted at the street segment level (see Table 1.1 in Appendix 1 for descriptive statistics).

3.1.1 Selection of neighborhoods

To make a representative selection[4] of neighborhoods within each city, we used several socio-demographic and socio-economic variables at the neighborhood level.[5] We applied a hierarchical clustering method (based on the squared Euclidian distance and using Ward's method; Everitt et al., 2011), ensured cohesiveness between the clusters by visually inspecting the dendrogram, and compared the differences between the resulting clusters with an analysis of variances (ANOVA) to ensure significance of the differences between the clusters. We yielded four clusters for Ghent and five clusters for London. We sampled 50 neighborhoods per city, relative to the number of neighborhoods within the clusters from the population (i.e., the total area). Figure 1 shows the spatial distribution of the selected neighborhoods. A detailed description of the sampling procedure and summary statistics of the clusters can be found in Appendix 2. We only used the neighborhood-level for sampling purposes that were necessary in line with the regulations imposed by the data provider (see Section 3.2.2).

[3] Camden, Greenwitch, Hackney, Hammersmith and Fulham, Islington, Kensington and Chelsea, Lambeth, Lewisham, Southwark, Tower Hamlets, Wandsworth, Westminster.

[4] Stratification is considered to be a useful technique to improve representativity and turned out to have the potential to improve the precision of estimates (Bethlehem, 2009, p. 101).

[5] Population density, relative distribution of age groups, relative distribution of one person households, relative distribution of birth nationality, and an index of social rent per 100 households. Data from London are from the most recent (2011) Census at the time of selecting neighborhoods. Data from Ghent are openly available ('Gent in Cijfers'; https://gent.buurtmonitor.be/) and data from 2021 have been used.

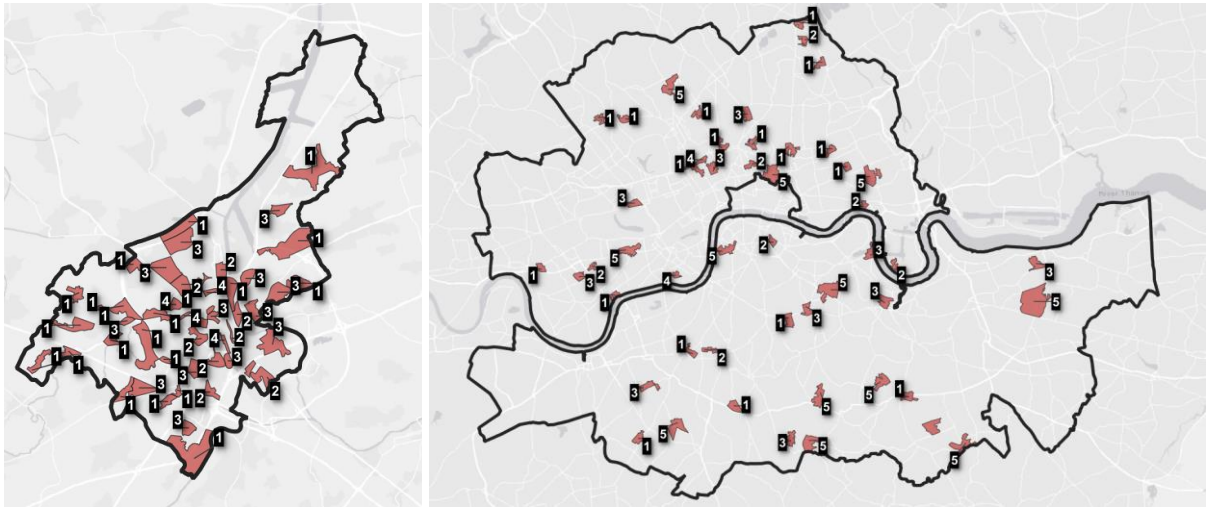

Figure 1. Selected neighborhoods within Ghent (left) and Inner London (right; labels contain cluster numbers).

3.2 Data sources and measurement of key constructs

3.2.1 Crime data

Police-registered crime data from 2022 are used to measure crime at micro places. These data were provided by the Local Police of Ghent (Belgium) and the Metropolitan Police for Greater London (UK). These data were provided in raw form at the record level, with specific time (date) and place (x,y-coordinates) indications. Crime data from the Metropolitan Police for Greater London were 'masked,' i.e. attributed to the x,y-coordinate representing the center of a street segment. The geolocations are registered by the police systems. It is useful in terms of assessing reliability to determine the comprehensive of the spatial referencing of data. For Ghent, we only obtained information on missing geolocations for residential burglary for the city as a whole. In this case, 0.65 percent of the crime data has missing x,y-coordinates and was therefore not included in the analysis. For London, we obtained all crime data from the Metropolitan Police spanning the Greater London area and for this area 3.58 percent of the relevant crime incidents lack x,y-coordinates.

We selected different and specific crime types, with a view to improving the comparability of this study (see Appendix 3). We therefore included specific property crime types (i.e., Business and community burglary, and Residential burglary), a specific violent crime type (i.e., Violence against the person), a crime type that can be seen as both property and violent crime (i.e., Aggressive theft) and a form of physical incivility (i.e., Criminal damage). Figure 2 shows the monthly counts of crime for these categories across the study period for the two cities.

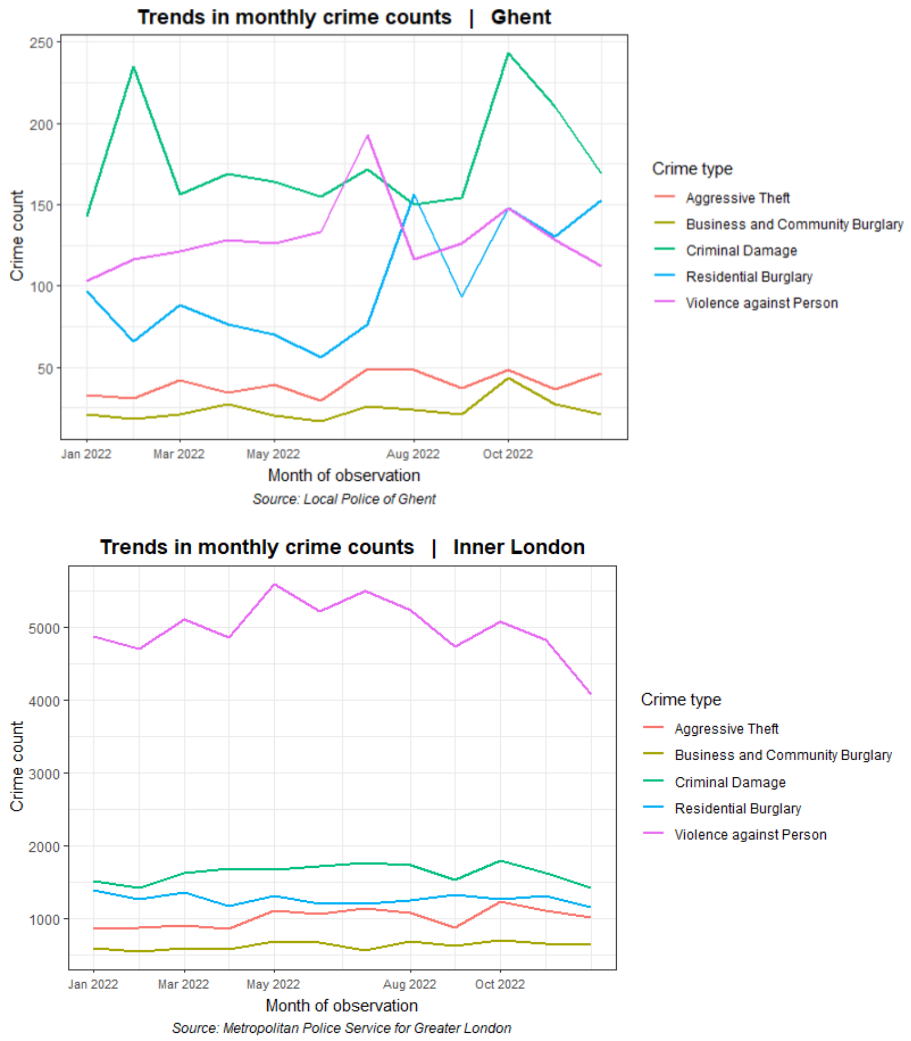

Figure 2. Trends in monthly crime counts for Ghent (above) and Inner London (below).

3.2.2 Google Places data

In this study, Google Places are used as a volunteered data source to extract geo-referenced points of interest and user ratings for those particular places. Google Places is part of the Google Maps Platform, which has several Application Programming Interfaces (henceforth: API(s)), categorized by Maps, Routes and Places (Google, 2022a). Google charges API requests but provides an initial free credit of the equivalent of $200 usage every month (Google, 2022b). Although there are open-source LBSs, none of these have a wider coverage and include both points of interest as well as user ratings and reviews than Google Places. For example, OpenStreetMap does not include user ratings and reviews, and TripAdvisor and Yelp have less coverage in terms of number of place types than Google Places.

Google Places can be seen as a large-scale data source for non-residential proprietary places. Google Places come with additional information of those proprietary places, such as place types, opening hours, price level, ratings, and reviews (Google, 2022c). Google Places data are

compiled from a variety of sources: publicly available information, licensed data from third parties, users who contribute information, and information based on Google's interaction with a local place or business (Google, 2023a). Users can leave both a rating and a review, or just a ratings (1 to 5 stars), however, many of the Google Places ratings lack a textual review. The ratings can be considered a general appreciation of the users experience at a particular place. In this study, ratings of places function as a proxy of place management at facilities. Review ratings can relate to many elements linked to places. For example, in restaurants they may focus on the standard of the food, in retail places the stock, prices or service, in institutions the proficiency of the personnel. We acknowledge that directly linking this to general place management makes an assumption, but we argue that experience at place will be heavily related to the proficiency of the management and return to this assumption in our conclusions.

We extracted the data via the Places API and API calls were made by using the R-package *googleway*.[6] An API request results in 20 Google Places and refreshes a further two times with a 'pagetoken' which returns the next 20 results (i.e., the next page), if present, from the previously executed search. This results in a maximum of 60 Google Places per query, which equals three API requests (the initial search, plus two times the additional search with 'pagetoken'). For this reason, one needs to make a trade-off regarding the scale of the search area (i.e., a radius) which is prompted, and the parsimony needed to make the most out of the API requests. To ensure that we collected all Google Places in the areas of interest and with respect to a predefined number of places (to avoid high costs for using the API and prevent us from violating Google's Terms of Service that does not allow for "bulk downloading", see Google, 2020), we first selected a number of neighborhoods (see above). Second, we used a stepwise method to obtain all data in the areas of interest: the quadtree decomposition method (Samet, 1984; see also Martí et al., 2019). This method recursively divides the radiuses into four quadrants and, if necessary, the partial quadrants are again subdivided into four sub-quadrants until the number of Google Places obtained is less than 60, hence, all Google Places in that area have been retrieved (see Figure 3).

[6] https://cran.r-project.org/web/packages/googleway/index.html

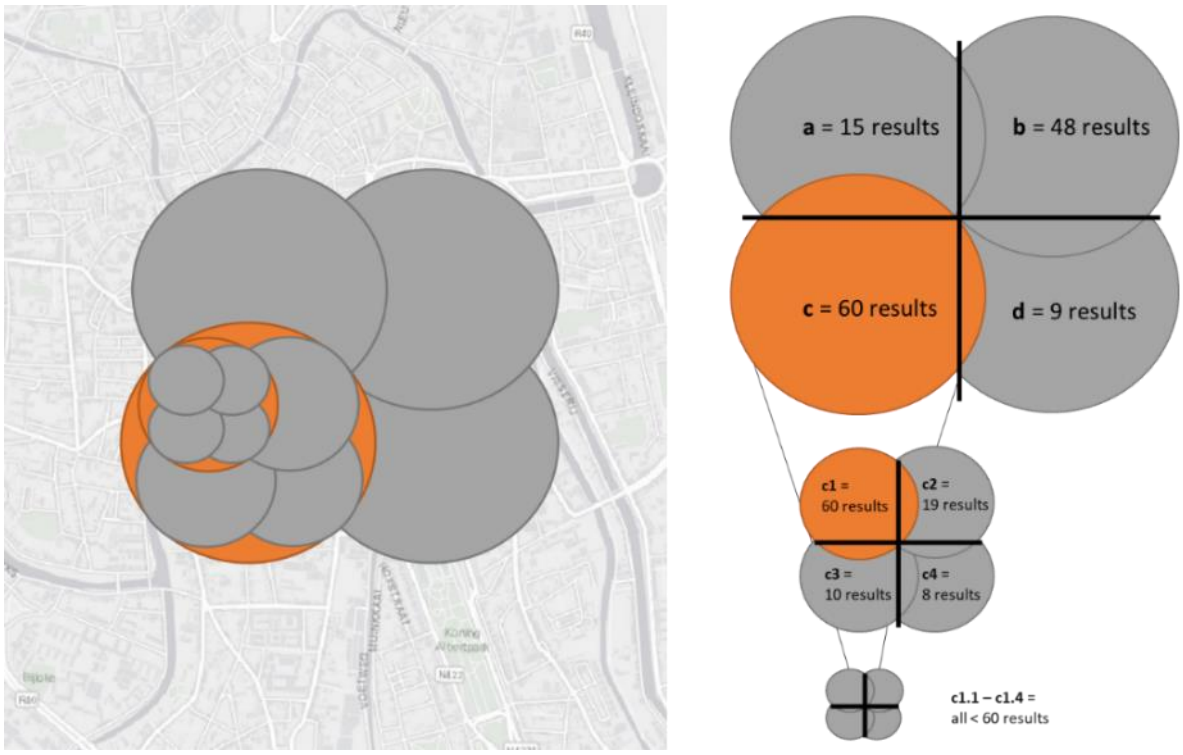

Figure 3. The principle of the quadtree decomposition method in obtaining Google Places data.

Following prior research (Martí et al., 2019, 2022), we have used the Land Based Classification Standards (LBCS) categories from the American Planning Association (2022) as benchmark place-type categories. In this study, we focus on facilities, which generally align with the 'Place Types' in Google Places. We used the Google Places with their existing classifications, which means that we did not add to or rectify place types. We categorized place types according to the LBCS function typology of the American Planning Association with a single adjustment that was relevant in the context of our study. In the category 'Food services', we distinguish between food services (i.e., food, restaurant, meal delivery, meal take-away) and drinking services (i.e., bar, cafe, night club). After cleaning the data, in Ghent we found 7,023 mentions of place types in 5,623 unique Google Places with a total of 311,120 ratings. In London we obtained data from places with 12,513 mentions of place types in 9,599 unique Google Places with in total 990,203 ratings. The Google Places data for London were obtained between 20 and 23 June 2022, and the data for Ghent were obtained on 1 July 2022.

3.2.3 Measures

The dependent variables regard police-registered crime counts at the street segment level. We composed measures of crime counts for the individual crime types assessed in this study (see above).

For the **count of Google Places** at the street segment level, we took into account all places categorized according to the LBCS classification. For the subsequent measures, we use the unique Google Places in the LBCS classification, to avoid double-counting. We composed a measure for the **average rating of places** at the street segment level, that were weighted based on the total number of ratings for the given places in the street segment. We also included a **count of Google Places per place type** and a (weighted) **average rating per place type** (categorized based on the LBCS function typology, but also further aggregated, see Table 3.2 in Appendix 3). To estimate the **proportion of low rated facilities**[7] per street segment, we took the decile of places with the lowest rating (3.0 or lower; i.e., 10.10% of the places for London and 3.5 or lower, i.e., 10.34% for Ghent). To quantify the **diversity of place types** at street segments, we have used the Simpson diversity index (Simpson, 1949) – see Eq. (1). Simpson's diversity index takes into account the number of place types present, classified according to the LBCS classification (represented by N which can vary between 0 and 28 for individual street segments), as well as the relative abundance of each category of place types present (represented by n, for example, a street segment might contain n = 3 drinking services). The value of *D* ranges between 0 and 1, with higher values indicating more diversity. To calculate the Simpson diversity index, it is necessary to categorize the population into meaningful groups that express the heterogeneity of the population. We have used the aforementioned LBCS function typology as meaningful groups.

$$D = 1 - \left(\frac{\sum n(n-1)}{N(N-1)}\right) \quad (1)$$

Last, we have included two control variables: a spatial lag term and the length of the street segments. The **spatial lag** term reflects the average number of crimes (per crime type) on street segments within 402 m (i.e., one-quarter of a mile) from the center of the street segment (comparable to the procedure of Weisburd et al., 2016). A spatial lag term is a variable that contains the weighted average of the values of neighboring street segments, where the value of the neighboring street segment is multiplied by the spatial weight (i.e., distance) and then the products are multiplied. By including this spatial lag term, we take into account larger area influences in our models, because according to Tobler's (1970) first law of geography, it is believed that everything is related to everything else, but near things are more related than distant things. Therefore, we computed a spatially lagged variable from inverse distance

[7] Since this variable is estimated from the (also included) ratings of places, and both these variables are included in the explanatory analyses, we checked for multicollinearity. The maximum variance inflation factor (VIF) was 2.21, so we included both variables in the analyses.

weights. We also include a measure of the exact **length of the street segments** as a control variable to take into account the potential variability in the crimes counts on shorter versus longer streets.

3.3 Analytical strategy

First, we conducted descriptive analyses to describe the different included variables. To assess the extent to which crime concentrates at street segments, we use (generalized) Gini coefficients where 1 is maximum concentration and 0 is perfectly equal distribution (Bernasco & Steenbeek, 2017). Second, we conducted correlational analyses to assess whether specific place types are associated with specific crime types. Last, we conducted regression analyses to search for explanatory variables in the Google Places data. For these regression analyses, we have used models that deal with overdispersion. Since the sample variance exceeds the sample mean, and the counts of the outcome variables (crime in crime types) are very strongly skewed, we used negative binomial models. However, there is a high number of excess zeros in the data, that cannot be accommodated properly by the negative binomial model alone; this model would underpredict them. To deal with these data compositions, two-part models[8] are available. These models allow for two different processes: one that drives whether the value is zero or positive (prevalence part) and the other one drives the value of the strictly positive count (frequency part). This terminology (prevalence vs. frequency) is also established within research on crime concentration (see Lee et al., 2017). The hurdle model is selected over a mixed-model, that might be of interest to account for the possible nested structure in the data given the sampling at the neighborhood-level.[9]

A hurdle model (Heilbron, 1994; Mullahy, 1986) assumes that all zero counts are from one structural source with one part of the model models whether the outcome variable is zero or positive (binary; prevalence) and another part using a truncated model for the positive data

[8] Similar to hurdle models, zero-inflated models (Lambert, 1992) are also two-part models. A zero-inflated model assumes that zero counts result from a mixture of two distributions: one where subjects always produce zero counts (structural zeros) and one where subjects are exposed to the outcome but did or did not report the experience of the outcome during the study period (sampling zeros). The structural zeros are for example due to subpopulations of subjects who are not at risk for certain outcomes during the study period.

[9] In selecting a hurdle model over a mixed-model for analyzing our data, both statistical and theoretical considerations are paramount. Statistical significance arises from the presence of excess zeros in the data, necessitating a model that can appropriately account for this phenomenon to avoid biased parameter estimates and erroneous inferences. A hurdle model offers a tailored solution to control for this. On a theoretical level, the micro-level mechanisms inherent in place management suggest that factors on the micro-place level may influence the occurrence of zeros, without the necessity to analyze contextual effects. By capturing the nuanced dynamics at play in place management, hurdle models enhance interpretability and offer a more comprehensive understanding of the underlying mechanisms. Thus, the choice of a hurdle model is not only statistically justified but also theoretically grounded, making it the optimal approach for analyzing our data.

(count; frequency). If the realization is positive, the "hurdle" is crossed, and the conditional distribution of the positives is governed by a truncated at-zero count data model. For the hurdle model we have used the pscl-package in R (Jackman, 2020). From a theoretical perspective, dividing the data in a prevalence model and a frequency model is relevant because different mechanisms may lead to either prevalence or frequency of crime at micro places. The model parameters are estimated by maximum likelihood, where the specification of the likelihood has the advantage that the zero (hurdle) and the count component can be maximized separately (see Eq. (2); Zeileis et al., 2008).

$$f_{\text{hurdle}}(y; x, z, \beta, \gamma) = \begin{cases} f_{\text{zero}}(0; z, \gamma) & \text{if } y = 0 \\ (1 - f_{\text{zero}}(0; z, \gamma)) \cdot f_{\text{count}}(y; x, \beta) / (1 - f_{\text{count}}(0; x, \beta)) & \text{if } y > 0 \end{cases} \quad (2)$$

We used R (Version 4.2.2) and RStudio for most of data processing and the statistical analyses. We have only used SPSS (Version 26) for the hierarchical cluster analysis in selecting the neighborhoods. For the geographical analyses, we used QGIS (3.28 Firenze).

4 Results

The descriptive analyses reveal that crime in London is generally quite concentrated at the street segment level (see Table 1). Except for Violence against the Person in London, the generalized Gini coefficient is calculated, to compensate for the empirical fact that there are more places than crimes in these data. For Ghent, most crime concentration measures are quite low, due to a strong adjustment by the generalized Gini coefficient. Ghent has a relatively low crime count (e.g., 79 commercial burglary incidents) and a larger number of street segments than London, which affects the concentration measure. Descriptive statistics of the variables included in the following analyses can be found in Table 1.2–1.5 in Appendix 1.

Table 1. Crime concentrations at the street segment level (Gini coefficients).

	Ghent, Belgium	London, UK
Aggressive theft	0.468	0.664
Commercial Burglary	0.289	0.657
Criminal Damage	0.549	0.717
Residential Burglary	0.369	0.692
Violence against the Person	0.761	0.883

To gain insight in what specific facility types are associated with specific crime types, correlational analyses were conducted (see upper parts of Tables 2 and 3). The results for London (see Table 2) show that all facility types are significant and positively associated with all individual types of crime that were included in this study, however, the magnitude of the

relationships differs. For London, the strongest associations are found with aggressive theft and commercial burglary (e.g., number of eating facilities and commercial burglary: R=0.229, p<0.001). Looking at the categories of facilities, it appears that drinking facilities, eating facilities, and recreation and accommodation facilities yield the strongest relation with crime at street segments. For Ghent (see Table 3), almost all facility types are significant and positively associated with all crime types included. Here, a strong correlation (R=0.725, p<0.001) is found between the count of drinking facilities and incidents of violence against person at the street segment level. Drinking facilities are also relatively strong correlated with aggressive theft (R=0.336, p<0.001) and commercial burglary (R=0.428, p<0.001). These findings are in line with hypothesis 1 (more non-residential facilities will equate to more crime), however, stronger associations have been found for specific types of places and specific types of crime.

The lower parts of Tables 2 and 3 show the relationships between the average ratings of facility types and crime. Here, the results suggest negative/protective relationships in most cases (i.e., the lower the rating, the higher the crime and vice versa). However, most of the correlations are found to be not significant. For London (see Table 2), significant associations have been found for drinking facilities with aggressive theft (R=-0.149, p=0.013) and commercial burglary (R=-0.142, p=0.018). For the facility type retail sales and services, significant and negative associations have been found for criminal damage (R=-0.103, p=0.008) and violence against the person (R=-0.148, p<0.001). In Ghent (see Table 3), the average rating of eating facilities at street segments are significant negatively associated with aggressive theft (R=-0.304, p<0.001). Ratings of the facility type finance and professional services are significant negatively correlated with aggressive theft (R=-0.192, p=0.005) and violence against person (R=-0.156, p=0.024). The average ratings of public services and transportation at the street segment level are negatively associated with multiple crime types: aggressive theft (R=-0.195, p<0.001), commercial burglary (R=-0.161, p=0.004), criminal damage (R=-0.117, p=0.035) and violence against person (R=-0.205, p<0.001). Last, ratings of the facility type recreation and accommodation is negatively associated with criminal damage (R=-0.197, p=0.002) and residential burglary (R=-0.146, p=0.02). Therefore, there is evidence to corroborate hypothesis 2 (that better ratings have a negative effect on crime), however, notable differences exist, both in magnitude and direction, for different crime and place types within the different cities.

The explanatory regression analyses (see Tables 4 and 5) allow for distinguishing between the prevalence/zero part[10] (that considers whether one or more crimes occurred at the street segment or not) and the frequency/count part[11] (that considers how many crime incidents have occurred in the street segments with crime) of the crime data. We assessed model fit by running negative binomial regression analyses and compared the AIC-values with those from the hurdle models.[12] The model fit of the hurdle models was only better for violence against person in London and for criminal damage in Ghent, however, to account for excess zeros, these two-part models are preferred. The results show notable differences between the models for the different contexts and the different crime types.

Both in Ghent and London, the number of facilities per street segment has a significant positive effect on the *prevalence of crime* (zero part) at the street segment for almost all crime types (except for residential burglary in London). In London, for the violent crime types against individuals (aggressive theft and violence against person), the number of facilities has also a significant and positive effect on the *frequency of crime* at the street segment level (count part). This also goes for criminal damage in Ghent. These findings are in line with hypothesis 1.

In London, the average rating of facilities also have significant effects on the violent crime counts, suggesting that the lower the rating, the more violent crime incidents at that street segment (which is in line with hypothesis 2). The exponentiated estimate of the regression component for aggressive theft and violence against person is respectively 0.45 and 0.55, suggesting that a one unit increase in the average rating is associated with a -55 and -45 percent change in the crime count, holding other variables constant. In these cases (i.e., both violent crime types in London), the results are in line with hypotheses 1 and 2. However, for the other crime types we did not find significant effects. None of the results related to ratings of places in Ghent were significant, however, the direction of the effects is interesting and differs from those in London. Both in London and Ghent, in the prevalence (zero) part, the direction of the effect is in all cases the same: the higher the rating, the lower the chance of having any crime incident. In Ghent, however, in the frequency/count part, we found other directions of the effects for aggressive theft, criminal damage, and residential burglary, suggesting that higher ratings result in more crime incidents at the street segment level. In assessing the concentration

[10] Interpretation of the estimates in the zero model: a one unit change in the predictor variable is associated with a $(\exp(\hat{\beta}) - 1)*100$ percentage increase of the odds (chance) of having any crime incident.

[11] Interpretation of the estimates in the count model: a one unit change in the predictor variable is associated with a $(\exp(\hat{\beta}) - 1)*100$ percentage change of the expected sample count, holding other variables constant.

[12] We examined mixed-effects models to accommodate for nested data, however, we found no noticeable deviations in terms of standard errors and statistical significance.

of places with the lowest rating specifically (those in the lowest decile), none of the effects were significant. Thus, the third hypothesis is rejected.

Regarding the diversity of facilities (i.e., place types) at the street segment level, we found no significant effects for the different crime types in Ghent and London. Thus, we do not find evidence for hypothesis 4. Nevertheless, the direction and magnitude of the exponentiated regression components are noteworthy. In the prevalence (zero) part, most effects are positive, suggesting that higher diversity results in higher chances of having any crime incident. Exceptions are commercial burglary in both contexts (20 percent chance decrease for London and 42 percent chance decrease for Ghent) and violence against the person in Ghent (11 percent chance decrease). In those cases, higher diversity results in lower chances of having any crime at the street segment level. Regarding the frequency (count) part, in the case of a one unit change in the diversity of facilities, we found a 7 and 8 percent decrease of the crime count for London and Ghent respectively. Nevertheless, none of these effects were significant and hypothesis 4 is not supported.

Table 2. Pearson correlation coefficients between crime types and place categories (both number of places per street segment and average rating per street segment in London, UK).

		Aggressive Theft		Commercial Burglary		Criminal Damage		Residential Burglary		Violence against Person	
	N	*R*	Sig.	*R*	Sig.	*R*	Sig.	*R*	Sig.	*R*	Sig.
Number of places per street segment											
Drinking facilities	2,140	**0.147**	***	**0.196**	***	**0.068**	**	**0.071**	***	**0.115**	***
Eating facilities	2,140	**0.135**	***	**0.229**	***	**0.098**	***	**0.078**	***	**0.131**	***
Finance and professional services	2,140	**0.147**	***	**0.194**	***	**0.081**	***	**0.108**	***	**0.104**	***
Manufacturing and contracting	2,140	**0.134**	***	**0.146**	***	**0.119**	***	**0.107**	***	**0.108**	***
Public services and transportation	2,140	**0.107**	***	**0.116**	***	**0.119**	***	**0.089**	***	**0.108**	***
Recreation and accomodation	2,140	**0.179**	***	**0.187**	***	**0.106**	***	**0.158**	***	**0.106**	***
Retail sales and services	2,140	**0.154**	***	**0.155**	***	**0.118**	***	**0.088**	***	**0.118**	***
Average rating per street segment											
Drinking facilities	279	**-0.149**	*	**-0.142**	*	-0.083	n.s.	-0.072	n.s.	-0.113	n.s.
Eating facilities	437	-0.020	n.s.	-0.065	n.s.	-0.008	n.s.	-0.025	n.s.	-0.046	n.s.
Finance and professional services	290	0.012	n.s.	0.019	n.s.	0.030	n.s.	0.000	n.s.	0.015	n.s.
Manufacturing and contracting	278	-0.003	n.s.	-0.042	n.s.	-0.034	n.s.	0.008	n.s.	-0.005	n.s.
Public services and transportation	363	-0.021	n.s.	0.013	n.s.	-0.003	n.s.	0.069	n.s.	-0.040	n.s.
Recreation and accomodation	366	-0.032	n.s.	0.000	n.s.	-0.066	n.s.	-0.056	n.s.	-0.043	n.s.
Retail sales and services	675	-0.037	n.s.	0.009	n.s.	**-0.103**	**	0.054	n.s.	**-0.148**	***

n.s. not significant; * significant at <0.05; ** significant at <0.01; *** significant at <0.001

Table 3. Pearson correlation coefficients between crime types and place categories (both number of places per street segment and average rating per street segment in Ghent, Belgium).

		Aggressive Theft		Commercial Burglary		Criminal Damage		Residential Burglary		Violence against Person	
	N	*R*	Sig.	*R*	Sig.	*R*	Sig.	*R*	Sig.	*R*	Sig.
Number of places per street segment											
Drinking facilities	4,416	**0.336**	***	**0.428**	***	**0.252**	***	0.011	n.s.	**0.725**	***
Eating facilities	4,416	**0.174**	***	**0.262**	***	**0.215**	***	**0.075**	***	**0.251**	***
Finance and professional services	4,416	**0.043**	**	**0.037**	*	**0.174**	***	**0.090**	***	0.029	n.s.
Manufacturing and contracting	4,416	**0.045**	**	**0.059**	***	**0.128**	***	**0.113**	***	**0.040**	**
Public services and transportation	4,416	**0.070**	***	**0.062**	***	**0.230**	***	**0.164**	***	**0.067**	***
Recreation and accomodation	4,416	**0.041**	**	**0.135**	***	**0.185**	***	**0.052**	***	**0.081**	***
Retail sales and services	4,416	**0.116**	***	**0.177**	***	**0.286**	***	**0.101**	***	**0.157**	***
Average rating per street segment											
Drinking facilities	144	-0.008	n.s.	-0.052	n.s.	0.000	n.s.	0.061	n.s.	-0.055	n.s.
Eating facilities	265	**-0.304**	***	-0.056	n.s.	-0.069	n.s.	0.09	n.s.	-0.105	n.s.
Finance and professional services	210	**-0.192**	**	0.007	n.s.	-0.003	n.s.	0.132	n.s.	**-0.156**	*
Manufacturing and contracting	178	-0.038	n.s.	0.012	n.s.	0.101	n.s.	-0.021	n.s.	0.064	n.s.
Public services and transportation	324	**-0.195**	***	**-0.161**	**	**-0.117**	*	-0.024	n.s.	**-0.205**	***
Recreation and accomodation	254	-0.038	n.s.	-0.099	n.s.	**-0.197**	**	**-0.146**	*	-0.033	n.s.
Retail sales and services	575	-0.046	n.s.	-0.043	n.s.	-0.038	n.s.	0.039	n.s.	-0.023	n.s.

n.s. not significant; * significant at <0.05; ** significant at <0.01; *** significant at <0.001

Table 4. Regression estimates, SEs and *p*-values estimated in the hurdle models for each crime type at the street segment level in London, UK (*n*=2,140).

		Aggressive Theft			**Commercial Burglary**			**Criminal Damage**			**Residential Burglary**			**Violence against Person**		
		exp($\hat{\beta}$)	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**
Zero part[a]	Intercept	0.45	0.90	0.373	0.28	1.03	0.217	0.28	0.83	0.123	0.32	0.91	0.211	0.64	0.78	0.567
	Number of places	1.01	0.00	**0.006**	1.02	0.00	**<0.001**	1.01	0.00	**0.017**	1.01	0.00	0.094	1.01	0.00	**0.002**
	Average rating of places	0.64	0.19	**0.019**	0.76	0.22	0.214	0.89	0.17	0.516	0.77	0.18	0.163	0.82	0.16	0.212
	1st decile lowest rating	0.38	0.55	0.079	0.48	0.63	0.244	0.94	0.48	0.894	0.42	0.53	0.099	0.66	0.45	0.344
	Diversity of facilities	1.40	0.43	0.431	0.80	0.46	0.632	1.00	0.37	0.994	1.70	0.42	0.207	1.06	0.34	0.856
	Spatial lag	1.64	0.15	**<0.001**	2.04	0.22	**0.001**	0.93	0.18	0.683	0.84	0.25	0.481	0.91	0.06	0.120
	Length	1.01	0.00	**<0.001**	1.00	0.00	**0.026**	1.00	0.00	**<0.001**	1.01	0.00	**<0.001**	1.00	0.00	**<0.001**
Count part[b]	Intercept	3.11	1.86	0.542	0.77	1.88	0.888	3.58	1.10	0.247	0.79	1.26	0.849	26.74	0.84	**<0.001**
	Number of places	1.01	0.01	**0.045**	1.01	0.01	0.060	1.00	0.00	0.293	1.00	0.01	0.612	1.01	0.00	**0.005**
	Average rating of places	0.45	0.33	**0.017**	0.78	0.35	0.477	0.71	0.23	0.136	0.97	0.23	0.907	0.55	0.18	**<0.001**
	1st decile lowest rating	0.13	1.04	0.051	0.48	1.14	0.519	1.29	0.69	0.712	1.19	0.75	0.820	0.72	0.60	0.583
	Diversity of facilities	2.49	0.98	0.351	2.59	0.91	0.296	1.12	0.54	0.839	0.98	0.58	0.974	0.93	0.45	0.867
	Spatial lag	1.91	0.24	**0.008**	0.99	0.35	0.975	0.88	0.25	0.623	0.91	0.35	0.793	1.03	0.08	0.731
	Length	1.00	0.00	0.848	1.00	0.00	0.697	1.00	0.00	0.794	1.00	0.00	0.164	1.00	0.00	0.063
	AIC		1336.563			1080.353			1624.752			1461.983			**2564.682**	
	AIC of GLM NB		**1335.6**			**1071.5**			**1620.9**			**1458.6**			2578.4	

[a] The zero part of the hurdle model is a binomial regression model with logit link.

[b] The count part of the hurdle model is a truncated negative binomial regression model with log link.

Note: significant p-values (p<0.05) and lower AIC-values are in bold.

Table 5. Regression estimates, SEs and *p*-values estimated in the hurdle models for each crime type at the street segment level in Ghent, Belgium (*n*=4,416).

		Aggressive Theft			**Commercial Burglary**			**Criminal Damage**			**Residential Burglary**			**Violence against Person**		
		exp($\hat{\beta}$)	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**	**exp($\hat{\beta}$)**	**SE**	**sig.**
Zero part[a]	Intercept	0.22	1.56	0.343	0.70	1.84	0.847	0.09	1.55	0.150	0.05	1.32	**0.027**	0.12	1.10	0.058
	Number of places	1.07	0.02	**<0.001**	1.09	0.02	**<0.001**	1.08	0.01	**<0.001**	1.05	0.01	**<0.001**	1.12	0.01	**<0.001**
	Average rating of places	0.63	0.33	0.154	0.50	0.39	0.077	0.93	0.30	0.399	0.95	0.28	0.851	0.95	0.23	0.833
	1st decile lowest rating	0.89	0.85	0.889	0.29	1.08	0.257	1.12	0.70	0.055	0.76	0.70	0.702	1.77	0.56	0.310
	Diversity of facilities	1.17	0.96	0.803	0.58	0.66	0.405	1.35	0.67	0.858	1.08	0.46	0.861	0.89	0.40	0.779
	Spatial lag	28.99	0.93	**<0.001**	39.41	2.85	0.197	12.21	0.64	0.622	77.41	0.94	**<0.001**	1.36	0.11	**0.005**
	Length	1.00	0.00	0.445	1.00	0.00	0.584	1.00	0.00	0.912	1.00	0.00	**0.007**	1.00	0.00	0.130
Count part[b]	Intercept	0.00	23.53	0.654	0.00	18.11	0.400	0.11	1.01	**0.016**	0.00	38.92	0.732	0.00	55.03	0.842
	Number of places	1.06	0.04	0.160	1.15	0.09	0.102	1.05	0.01	**<0.001**	1.00	0.04	0.891	1.05	0.03	0.064
	Average rating of places	1.93	0.98	0.504	0.66	1.75	0.811	1.28	0.21	0.721	1.72	0.68	0.426	0.84	0.59	0.775
	1st decile lowest rating	4.77	2.16	0.470	5.93	3.31	0.591	3.81	0.51	0.832	1.53	1.47	0.772	2.75	1.14	0.376
	Diversity of facilities	2.11	2.61	0.774	16.57	13.04	0.204	0.89	0.36	0.410	1.80	1.47	0.687	0.92	1.03	0.938
	Spatial lag	1.47	2.45	0.875	0.11	6.21	0.720	1.37	0.42	**<0.001**	5.44	2.60	0.514	4.63	0.90	0.090
	Length	1.00	0.01	0.325	1.00	0.01	0.466	1.00	0.00	**0.021**	1.00	0.00	0.329	1.00	0.00	0.880
	AIC		523.3501			392.3597			**1447.759**			805.3064			1163.765	
	AIC of GLM NB		**517.73**			**390.4**			1449.1			**800.3**			**1163.3**	

[a] The zero part of the hurdle model is a binomial regression model with logit link.

[b] The count part of the hurdle model is a truncated negative binomial regression model with log link.

Note: significant p-values (p<0.05) and lower AIC-values are in bold.

5 Discussion and conclusion

This study is – to our knowledge – the first that assesses place management as a mechanism for crime prevention by means of data from Google Places. The aim of this study was to assess the mechanism of place management in relation to crime at micro places, from an empirical perspective. Therefore, we examined the presence, types, the ratings, and the diversity of facilities at the street segment level in Ghent (Belgium) and London (UK) in relation to five specific crime types.

We found partial evidence for our hypotheses. In line with our first hypothesis, we found that almost all facility types are significant and positively associated with all individual types of crime that were included in this study. In the explanatory analysis, we only found significant effects from the number of Google Places on the violent crime types in London. Further, the results for London suggest that for the two violent crime types, the rating of Google Places has a significant protective effect on the crime count at the street segment level. For the other crime types we did not find significant effects, so we only found partial evidence for our second hypothesis. The correlational analysis shows notable differences per facility type and per crime type. Notably, all significant associations are in the (negative) direction, as hypothesized. Regarding the concentration of these specifically deemed poorly rated places, we did not find any significant effect on crime counts of different crime types in both contexts. Therefore, the third hypothesis is rejected. Last, we assessed to what extent the mechanisms of constant succession of eyes (measured by diversity of facilities at the street segment) has an effect on the crime count. We did not find any significant effects, so we could not confirm our fourth hypothesis.

Looking at the descriptive statistics of the variables under study, the differences between the findings in the two different contexts included can possibly be explained by different urban structure, as well as different (intensity) of the use of Google Places. The included 50 neighborhoods in Ghent have twice as many street segments as in London. At the same time, the 50 neighborhoods in London have two and a half time more crime for the included crime types and the Google Places in these neighborhoods have three times more ratings than the Google Places in Ghent.

This exploratory study has several limitations. A number of these relate to the Google Places data itself. First, due to the restriction that it is not allowed to bulk download these data, we decided to draw stratified samples of 50 neighborhoods, equal for both cities. Although we

tailored the procedure to the local context by applying different cluster analyses, the sampling procedure of neighborhoods could have been better tailored to the local context, e.g., by sampling more neighborhoods in London, because there are more neighborhoods in the city than in Ghent. Apart from the general drawbacks of secondary data (e.g., Hox & Boeije, 2005), and although we argue that the ratings might deliver a (rather abstract) measure of place management, much is still unknown about the biases and errors in the data. Google's product director states that their Maps product is used by one billion people every month (Russel, 2019). However, this does not necessarily mean that the resulting data is representative (Boyd & Crawford, 2012; Harford, 2014). We need close consideration of both measurement and representation processes constituting the data at hand. In contrast to conventional data sources, for 'found' data, which were collected without a primary research purpose, it is – *a priori* – not clear what they (can) measure or how such measurement would relate to existing theories and related concepts (Boyd & Crawford, 2012; O'Brien et al., 2015). In this case, it should be acknowledged that for example diversity of facilities (as included in this study) is not a direct measure of 24 hour/day guardianship, but a mere proxy of this concept. Further, it is plausible that there are other mechanisms at work (e.g., that more diversity equals less cohesion, and consequently less collective efficacy to reduce crime).

Another limitation of the Google Places data is that these only can be obtained for the moment of the API request. Unlike for Google Street View, there is no history available here. With a view to the temporal sequence of the hypothesized relation (place management → crime), scholars need to wait after obtaining the Google Places data for the crime data to conduct the study. Further limitations relate to the sampling procedure used to select neighborhoods. Our sampling procedure, although necessary to be justified in terms of Google's Terms of Service and the cost of the data, also has some limitations. Due to the relative scarcity of data, we were not able to assess the number of facilities and average rating at the street segment level per disaggregated place type. Additionally, at the time we have conducted the cluster analyses, the most recent data from London date from the 2011 census. Although we know that structural factors are rather stable over time, these remain outdated figures. Last, the cross-sectional nature and the fact that we have used only crime data for one year is a limitation, since we cannot study causal mechanisms sufficiently in such a short time frame.

Due to this highly exploratory nature of this study, we found interesting avenues for future research. The first category relates to the extent to which Google makes its data available for research. The data are not easily, nor fully accessible for scientific research. In assessing the

functionalities of Google Maps and the information that can be obtained by means of the Google Places API, we have identified a discrepancy between data being collected by Google and data being made available. For example, Google collects data on 'popular times' and 'visit duration' of places (Google, 2023b), but these data are not obtainable by means of a Place Details request (Google, 2022c). Relatedly, however broader, is another data governance issue. Plenty of alternative POI directories exist, resulting in a fragmentation of data. Besides that matching techniques can be used (e.g., Jiang et al., 2015) this is a fundamental data governance issue that provides challenges for future research and practical applications.

The second category of avenues for future research relates to the necessary methodological assessment of the data quality of Google Places data. Scholars may use error frameworks analogues to the Total Survey Error framework (Groves et al., 2009) to assess and quantify the accuracy of the data sources (see Amaya et al., 2020; Kreuter, 2021; Snaphaan et al., 2024). It is also recommended to conduct more comprehensive validity tests, such as by assessing the (concurrent) validity of Google Places data with triangulation against other data sources, such as business registers and/or OpenStreetMap data. Future research should scrutinize what selective forces drive different kinds of people to be in places and to rate a place or not. Additionally, future studies should assess the (methodological) differences between different electronics Word of Mouth platforms, and assess how and to what extent these platforms are used in the general population. Only then, we will advance the knowledge regarding the mechanisms of place management as well as the methodological biases and errors involved.

The third category relate to the substantive objectives of future studies. Google Places data allow for designing (place management) studies at the proprietary level. This was beyond the scope of our contribution and was not reliably possible given the masking process of crime data in London. This also allows for empirical testing of the "Iron Law of Troublesome Places." Further, it would be interesting to explore what people are actually reviewing on Google Places, through a deeper dive into the textual reviews. However, maybe the quantitative raters (assumed to be more general) differ in what they rate to qualitative reviewers (assumed to be more specific) so conceptual differences may be present. It appears that the strongest relations with the ratings of places are with crimes against individuals, rather than crimes against properties. So, it might be the case that individuals who experienced victimization go on to leave a negative rating and/or write a negative review. Future research should address these questions. Relatedly, another interesting inquiry would be to assess whether responding to reviews on Google Places from the business owner (see Salehi-Esfahani et al., 2023) is a proxy for place management,

since this is actually virtual place management of a physical place. However, digital representation might not be representative for physical presence, especially for small and medium enterprises (OECD, 2017, p. 116).

It is plausible that the aforementioned one billion users per month exceed the number of active contributors in terms of ratings and reviewing places. Therefore, it would be interesting to study to what extent the use of Google Maps (and its ratings and reviews of places) actually shape routine activities of people, given the importance of this concept to crime research. After all, we know that "if men define situations as real, they are real in their consequences" (Thomas & Thomas, 1928, p. 572). Examining the coverage of the use of Google Maps (and Places) by the general population, and the motivations to use and contribute (by rating and reviewing) is another avenue for future research.

A final note might be made on implications for prevention practice. The advent of new and emerging data sources brings with it the opportunity of moving towards (but not necessarily reaching) complete coverage rather than sampling (Kitchin & McArdle, 2016). This brings opportunity to those wanting to assess micro-level phenomena at scale. For example, town planners might assess how to deal with risky land uses using Google Place data and crime prevention practitioners with limited resources might encourage better place management in certain parts of a city or neighborhood where there is an indication this might be lacking. For example, business watch interventions might support such locations or, where problems are obviously having an effect on crime and causing serious radiation issues, enforcement of better place management with appropriate monitoring might be required.

References

Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology, 8,* 89-119. https://doi.org/10.1093/jssam/smz056

American Planning Association. (2022). *LBCS function dimension with description*. https://www.planning.org/lbcs/standards/function/

Baruch, A., May, A., & Yu, D. (2016). The motivations, enablers and barriers for voluntary participation in an online crowdsourcing platform. *Computers in Human Behavior*, *64,* 923–931. https://doi.org/10.1016/j.chb.2016.07.039

Bernasco, W., & Steenbeek, W. (2017). More places than crimes: Implications for evaluating the law of crime concentration at place. *Journal of Quantitative Criminology*, *33,* 451–467. https://doi.org/10.1007/s10940-016-9324-7

Bethlehem, J. (2009). *Applied survey methods: A statistical perspective*. Wiley.

Bowers, K. (2014). Risky facilities: Crime radiators or crime absorbers? A comparison of internal and external levels of theft. *Journal of Quantitative Criminology*, *30,* 389–414. https://doi.org/10.1007/s10940-013-9208-z

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

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.

Bruinsma, G. J. N., & Johnson, S. D. (2018). Environmental criminology: Scope, history, and state of the art. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 1–31). Oxford University Press.

Budhathoki, N. R. (2010). *Participants' motivations to contribute geographic information in an online community* [Doctoral dissertation]. University of Illinois. https://core.ac.uk/download/pdf/4825428.pdf

Buyalskaya, A., Gallo, M., & Camerer, C. F. (2021). The golden age of social science. *Proceedings of the National Academy of Sciences*, *118*(5), Article e2002923118. https://doi.org/10.1073/pnas.2002923118

Cheung, C. M. K., & Lee, M. K. O. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems*, *53*(1), 218–225. https://doi.org/10.1016/j.dss.2012.01.015

Douglas, S., & Welsh, B. C. (2020). Place managers for crime prevention: The theoretical and empirical status of a neglected situational crime prevention technique. *Crime Prevention and Community Safety*, *22*, 99–109. https://doi.org/10.1057/s41300-020-00089-4

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105–123.

Eck, J. E. (2019). Place managers and crime places. In *Oxford research encyclopedia of criminology*. https://doi.org/10.1093/acrefore/9780190264079.013.307

Eck, J. E., & Madensen, T. D. (2018). Place management. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford handbook environmental criminology* (pp. 629–663). Oxford University Press.

Eck, J. E., (2010). Places and the crime triangle. In F. T. Cullen & P. Wilcox (Eds.), *Encyclopedia of Criminological Theory*. https://doi.org/10.4135/9781412959193

Eck, J. E., Clarke, R. V., & Guerette, R. T. (2007). Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. In G. Farrell, K. J. Bowers, S. D. Johnson & M. Townsley (Eds.), *Imagination for crime prevention: Essays in honour of Ken Pease* (Crime Prevention Studies, Vol. 21, pp. 225–264). Criminal Justice Press.

Eck, J. E., Linning, S. J., & Herold, T. D. (2023). *Place management and crime. Ownership and property rights as a source of social control*. Springer.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th Ed.). Wiley.

Gomory, H., & Desmond, M. (2023). Neighborhoods of last resort: How landlord strategies concentrate violent crime. *Criminology*, *61*(2), 270–294. https://doi.org/10.1111/1745-9125.12332

Google. (2020, 6 May). *Google Maps Platform Terms of Service*. https://cloud.google.com/maps-platform/terms

Google. (2022a). *Find the right custom map product*. https://mapsplatform.google.com/maps-products/

Google. (2022b). *Pricing that scales to fit your needs*. https://mapsplatform.google.com/pricing/

Google. (2022c). *Place Details*. https://developers.google.com/maps/documentation/places/web-service/details

Google. (2023a). *How Google sources and uses information in local listings*. https://support.google.com/local-listings/answer/9851099?p=how_google_sources&authuser=0&hl=en&visit_id=637539321855538809-1024761200&rd=1

Google. (2023b). *Popular times, wait times, and visit duration*. https://support.google.com/business/answer/6263531?hl=en

Groff, E.R. (2014). Quantifying the exposure of street segments to drinking places nearby. *Journal of Quantitative Criminology, 30,* 527–548. https://doi.org/10.1007/s10940-013-9213-2

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Wiley.

Harford, T. (2014). Big data: A big mistake? *Significance, 11*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2014.00778.x

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Journal of Mathematical Methods in Biosciences*, *36*(5), 531–547. https://doi.org/10.1002/bimj.4710360505

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 593–599). Elsevier Science.

Jackman, S. (2020). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. United States Studies Centre, University of Sydney, Sydney, New South Wales, Australia. R package version 1.5.5, https://github.com/atahk/pscl/

Jacobs, J. (1961). *The death and life of great American cities*. Vintage.

Jiang, S., Alves, A., Rodrigues, F., Ferreira Jr, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, *53*, 36–46. https://doi.org/10.1016/j.compenvurbsys.2014.12.001

Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. G., & Brantingham, P. J. (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment*, *34*(1), 62–74.

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society, 3*(1), 1–10. https://doi.org/10.1177/2053951716631130

Kreuter F. (2021). What surveys really say. *Nature*, *600*(7890), 614–615. https://doi.org/10.1038/d41586-021-03604-1

Kuralarasan, K., & Bernasco, W. (2022). Location choice of snatching offenders in Chennai City. *Journal of Quantitative Criminology*, *38*, 673–696. https://doi.org/10.1007/s10940-021-09514-9

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14.

Lee, Y., Eck, J. E., O, S., & Martinez, N. H. (2017). How concentrated is crime at places? A systematic review from 1970 to 2015. *Crime Science, 6*, Article 6. https://doi.org/10.1186/s40163-017-0069-x

Lee, Y., O, S., & Eck, J. E. (2022). Why your bar has crime but not mine: Resolving the land use and crime – risky facility conflict. *Justice Quarterly*, *39*(5), 1009–1035. https://doi.org/10.1080/07418825.2021.1903068

Linning, S. J., & Eck, J. E. (2021). *Whose "eyes on the street" control crime? Expanding place management into neighborhoods*. Cambridge University Press.

Madensen, T. D. (2007). *Bar management and crime: Toward a dynamic theory of place management and crime hotspots* [Doctoral dissertation]. University of Cincinnati.

Madensen, T. D., & Eck, J. E. (2012). Crime places and place management. In F. T. Cullen & P. Wilcox (Eds.), *The Oxford handbook of criminological theory* (pp. 554–578). Oxford University Press.

Madensen, T. D., Herold, M., Hammer, M. G., & Christenson, B. R. (2017). Place-based investigations to disrupt crime place networks. *The Police Chief*, (April), 14–15.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems, 74*, 161–174. https://doi.org/10.1016/j.compenvurbsys.2018.11.001

Martí, P., Serrano-Estrada, L., Nolasco-Cirugeda, A., & López Baeza, J. (2022). Revisiting the spatial definition of neighborhood boundaries: Functional clusters versus administrative neighborhoods. *Journal of Urban Technology*, *29*(3), 73–94. https://doi.org/10.1080/10630732.2021.1930837

Mashhadi, A., Quattrone, G., & Capra, L. (2013). Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 611–622). https://doi.org/10.1145/2441776.2441845

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*(3), 341–365. https://doi.org/10.1016/0304-4076(86)90002-3

O'Brien, D. T., Sampson, R. J., & Winship, C. (2015). Ecometrics in the age of big data: Measuring and assessing "Broken Windows" using large-scale administrative records. *Sociological Methodology*, *45*(1), 101–147. https://doi.org/10.1177/0081175015576601

O'Reilly, T. (2005, 30 September). *What is Web 2.0*. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1

Oberwittler, D., & Wikström, P.-O. H. (2009). Why smaller is better: Advancing the study of the role of behavioral contexts in crime causation. In D. Weisburd, W. Bernasco & G. J. N. Bruinsma (Eds.), *Putting crime in its place: Units of analysis in geographic criminology* (pp. 36–60). Springer Science Business Media, LLC.

OECD. (2017). *Key issues for digital transformation in the G20*. https://www.oecd.org/g20/key-issues-for-digital-transformation-in-the-g20.pdf

Parikh, A., Behnke, C., Nelson, D., Vorvoreanu, M., & Almanza, B. (2015). A qualitative assessment of Yelp.com users' motivations to submit and read restaurant reviews. *Journal of Culinary Science & Technology*, *13*(1), 1–18. https://doi.org/10.1080/15428052.2014.952474

Parikh, A., Behnke, C., Vorvoreanu, M., Almanza, B., & Nelson, D. (2014). Motives for reading and articulating user-generated restaurant reviews on Yelp.com. *Journal of Hospitality and Tourism Technology*, *5*(2), 160–176. https://doi.org/10.1108/JHTT-04-2013-0011

Ritzer, G., Dean, P., & Jurgenson, N. (2012). The coming age of the prosumer. *American Behavioral Scientist*, *56*(4), 379–398. https://doi.org/10.1177/0002764211429368

Russel, E. (2019, 1 October). *9 things to know about Google's maps data: Beyond the Map*. https://cloud.google.com/blog/products/maps-platform/9-things-know-about-googles-maps-data-beyond-map

Salehi-Esfahani, S., Torres, E., & Hua, N. (2023). Responding to negative reviews? The interplay of management response strategy and service failure type. *Journal of Hospitality Marketing & Management*, *32*(1), 29–49. https://doi.org/10.1080/19368623.2022.2121349

Samet, H. (1984). The Quadtree and Related Hierarchical Data Structures. *ACM Computing Surveys*, *16*(2), 187–260. https://doi.org/10.1145/356924.356930.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, *27*(1), 27–55.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688. https://doi.org/10.1038/163688a0

Smarzaro, R., Lima, T. F. M., & Davis Jr., C. A. (2017). Could data from location-based social networks be used to support urban planning? In *Proceedings of the 26th International World Wide Web Companion* (pp. 1463–1468). http://dx.doi.org/10.1145/3041021.3051700

Snaphaan, T., & Hardyns, W. (2021). Environmental criminology in the big data era. *European Journal of Criminology*, *18*(5), 713–734. http://doi.org/10.1177/1477370819877753

Snaphaan, T., Hardyns, W., & Pauwels, L. J. R. (2024). Expanding the methodological toolkit of criminology and criminal justice with the Total Error Framework. *Journal of Crime and Justice, 47*(1), 112–129. https://doi.org/10.1080/0735648X.2022.2114099

Solymosi, R., Bowers, K. J., & Fujiyama, T. (2018). Crowdsourcing, subjective perceptions of neighbourhood disorder: Interpreting bias in open data. *British Journal of Criminology, 58*(4), 944–967. https://doi.org/10.1093/bjc/azx048

Stewart, O., Lubensky, D., & Huerta, J. M. (2010). Crowdsourcing participation inequality: A SCOUT model for the enterprise domain. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 30–33). https://doi.org/10.1145/1837885.1837895

Taylor, R. B. (1997). Social order and disorder of street blocks and neighborhoods: Ecology, microecology, and the systemic model of social disorganization. *Journal of Research in Crime and Delinquency*, *34*, 113–155.

Thakuriah, P. V., Tilahun, N. Y., & Zellner, M. (2017). Big data and urban informatics: Innovations and challenges to urban planning and knowledge discovery. In P. V. Thakuriah, N. Y. Tilahun, & M. Zellner (eds.), *Seeing cities through big data: Research, methods and applications in urban informatics* (pp. 11–45). Springer.

Thomas, W. I., & Thomas, D. S. (1928). *The child in America*. Knopf.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(Supplement), 234–240.

Tucker, R., & O'Brien, D. T. (2023). Do commercial place managers explain crime across places? Yes and NO (PE). *CrimRxiv*. https://crimrxiv.com/pub/yfdcyxer

Vandeviver, C. (2014). Applying Google Maps and Google Street View in criminological research. *Crime Science, 3*, Article 13. https://doi.org/10.1186/s40163-014-0013-2

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology 53*(2), 133–157.

Weisburd, D., Eck, J., Braga, A., Telep, C., Cave, B., Bowers, K., Bruinsma, G. J. N., Gill, C., Groff, E., Hibdon, J., Hinkle, J. C., Johnson, S. D., Lawton, B., Lum, C., Ratcliffe, J., Rengert, G., Taniguchi, T., & Yang, S.-M. (2016). *Place matters: Criminology for the 21st century*. Cambridge University Press.

Weisburd, D., Groff, E. R., & Yang, S. M. (2012). *The criminology of place: Street segments and our understanding of the crime problem*. Oxford University Press.

Wilcox, P., & Eck, J. E. (2011). Criminology of the unpopular. Implications for policy aimed at payday lending facilities. *Criminology & Public Policy*, *10*(2), 473–482. https://doi.org/10.1111/j.1745-9133.2011.00721.x

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism*, *10*(4), 283–295. https://doi.org/10.3727/109830508788403114

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. http://www.jstatsoft.org/v27/i08/