

## CROWDSOURCING SUBJECTIVE PERCEPTIONS OF NEIGHBOURHOOD DISORDER: INTERPRETING BIAS IN OPEN DATA

REKA SOLYMOŠI\*, KATE J. BOWERS and TAKU FUJIYAMA

*New forms of data are now widely used in social sciences, and much debate surrounds their ideal application to the study of crime problems. Limitations associated with this data, including the subjective bias in reporting are often a point of this debate. In this article, we argue that by re-conceptualizing such data and focusing on their mode of production of crowdsourcing, this bias can be understood as a reflection of people's subjective experiences with their environments. To illustrate, we apply the theoretical framework of signal crimes to empirical analysis of crowdsourced data from an online problem reporting website. We show how this approach facilitates new insight into people's experiences and discuss implications for advancing research on perception of crime and place.*

Key Words: crowdsourcing, disorder, new forms of data, perception of crime and place, signal crimes, signal disorder

### *Introduction*

In the era when over 6.2 exabytes of global mobile data traffic is generated each month (Cisco 2016) it is inevitable that open source 'big data' plays an increasingly important role in the advancement of research in the social sciences (Preis *et al.* 2013). This 'big data' movement has generated much interest in potential applications to study crime and disorder (Williams *et al.* 2016). In truth, it appears unlikely that many data used in criminological research would meet the 'volume' requirement necessary to qualify as 'big', according to standard definitions (Kaisler *et al.* 2013). However, the strength of these emerging sources of data does not necessarily come from their size. There are plenty of data sets now being produced as a result of people's online activities that show promise in offering new lines of enquiry in social science, in particular concepts related to crime and disorder which we review below. However, these studies conceptualize such data as an econometric measures of crime and disorder issues, and, we argue, miss an important quality of such information. A central aim of this article is to re-conceptualize the use of open source data produced by online collaborative effort (crowdsourced data) in representing theoretical concepts in criminology, by focusing on the *subjective* bias inherent in their mode of production.

The particular contribution of this article is to outline a new approach for measuring people's perception of crime and safety using crowdsourced data. We propose that a strength of crowdsourced information is that it can represent a measure of *what matters to a community*. We suggest that the underlying bias in what gets reported through such

\*Reka Solymosi, Room 4.53, Williamson Building, School of Law, University of Manchester, Oxford Road, Manchester M13 9PL, UK; [reka.solymosi@manchester.ac.uk](mailto:reka.solymosi@manchester.ac.uk); Kate Bowers and Taku Fujiyama, University College London, London, UK.

crowdsourced data collection techniques actually provides a filter of what communities deem *subjectively* important. We illustrate this by applying signal crimes framework to conceptualize data collected from online problem-reporting website. By considering the subjectivity and bias present in the data generation process when conceptualizing the meaning behind such data, we gain novel insight into people's experiences with disorder. This approach is further transferable to other areas of research on people's subjective perceptions about their environments.

### *Background*

The application of 'big data', in particular social media communications as source of research material for criminology is becoming more and more recognized. For example, Williams *et al.* (2016) found an association between aggregated twitter data and police-recorded crime data in London. Tweets have also been used to estimate the ambient population which can act as a more accurate denominator for crime rates (Malleon and Andresen 2015). Further, O'Brien *et al.* (2015) used data from a 311 hotline system in Boston to develop measures of 'broken windows'. These approaches take great steps towards making use of the wealth of available data for crime research. However, these approaches consider these data to represent a true measure of the phenomena they are framed to depict. While biases with the data in terms of representativeness are discussed, they are referred to as limitations, which detract from their value. By moving to consider the *mode of production* of much of this data, it becomes possible to examine the experiences and perceptions of the people generating it.

'Crowdsourcing', a portmanteau of 'crowd' and 'outsourcing', represents a means for tapping into group intelligence on large scales. Crowdsourced data are produced by large numbers of individuals contributing content to a central repository. One well-known example is Wikipedia ([www.wikipedia.org](http://www.wikipedia.org)), an encyclopaedia pulling together the knowledge of many contributors to provide a reference source freely available to all (Surowiecki 2005). What is novel about these projects is that they do not rely on one person to work or collect data until they meet certain requirements. Instead, anyone can participate as much or as little as they are willing to. Then, the crowd's participation adds up to a complete output (Surowiecki 2005).

Researchers have employed the methodology of crowdsourcing for data collection with great success. In a project from 2007 to 2014, over one million people participated in classifying images of galaxies (Haklay 2015). In Germany, scientists collaborated with 5,000 people to capture over 17,000 samples of mosquito, resulting in the discovery of an invasive species with implications to public health (Haklay 2015). There are similar examples across many domains of academic research.

While such data are available for use by researchers (if open data) this data-collection approach is not one-sided; it can also serve to collect data for use by the participants themselves. Crowdsourced data has been used to lobby for changes in participants' neighbourhoods, contributing to a reversal of the traditional top-down approach to the creation and dissemination of geographic information (Goodchild 2007).

However, the utility of crowdsourcing as a data collection framework does not only lie in its ability to gather large volumes of data (perhaps even big data). Value also lies in the participation of these motivated individuals. These sorts of motivations mean that

people participate in crowdsourcing initiatives when they want to highlight issues they consider problematic, and hope to instigate change. This introduces a bias in the data, which, we argue, can be used to make inferences about people's subjective perceptions and experiences of their environments. To date, the implications of this subjective filter for the study of these data in criminology have not been explored. They are often discussed as a limiting factor, for example in regards to slanting services towards those who have more active voices, and are over-represented in such data (O'Brien *et al.* 2015). However, in this paper we aim to illustrate the utility of such biases in the data by considering the crowdsourced mode of production. We hope to achieve this by exploring participation in a problem-reporting website, similar to 311, fixmystreet.com. The next section will describe this crowdsourced data set, and apply the theoretical framework of signal crimes, to contextualize the subjective bias in a meaningful way for research into people's experiences with place and disorder.

### FMS

Fixmystreet ([www.fixmystreet.com](http://www.fixmystreet.com) (hereon referred to as FMS)), is a web and mobile application for reporting environmental issues, run by the not-for-profit organization mySociety. FMS was created to enable citizens to report potholes, broken streetlights and other problems in their area easily, in order to get them fixed (MySociety 2016). Using the website, citizens are able to locate their problem on a map to provide exact coordinates, choose a category for their report, give it a title, and provide a brief description. The report is logged with the time and date of reporting, and the name of the person submitting the report (unless they remain anonymous). Once the report is submitted, it is displayed on the application website, and a copy is emailed to the responsible local authority. As of 2010, local councils such as Bromley Borough Council in North-West London have integrated the platform into their own website. By providing this platform, FMS facilitates crowdsourced data collection of issues which people encounter in their day-to-day activities. Similar sites also exist in other countries, making this approach transferable and reproducible internationally (Worth 2011). Therefore, while this article focuses on a case study from London, UK, it is possible to undertake such study in many settings worldwide.

FMS constitutes a specific subset of crowdsourced data called volunteered geographical information (VGI), where people collect geo-data (information with a geographical component) (Goodchild 2007; Haklay 2013). Such data include only what people perceive, *subjectively* to be problematic enough to report, thereby introducing a bias into *what* gets reported. This is most often discussed as a limitation when considering such data to represent econometric characteristics of a neighbourhood. However, in this article, it is exactly this subjective bias that we hope to exploit for drawing inferences from such data. To do so, we first specify a theoretical framework. It is important that exploration of such data should be theory driven (Williams *et al.* 2016).

### Signal Crimes

People's perception of places influences all spheres of social and economic activity. Fear of crime in particular affects quality of housing (Ceccato 2012), operational and

financial decisions made by businesses (Casten and Payne 2008), and people's willingness to cycle or walk through an area (Kelly *et al.* 2011; McDonald *et al.* 2010; Mitra *et al.* 2010). In order to draw conclusions about crime risk, people rely on cues in the environment. Research emphasizes the causal effect of physical incivilities on fear of crime, implying that people consult indicators of incivility and neighbourhood disorder when assessing safety in the environment (Kohm 2013; Lewis and Maxfield 1980; Wilcox *et al.* 2003).

One approach to frame the link between disorder and perception of crime and place comes from the signal crimes perspective. This suggests that people draw inferences about their environment based on certain signs of disorder which act as 'signals'. People subjectively interpret the disorder as something problematic, which evokes a negative interpretation of the area (Innes 2004; 2014). A major theoretical advancement of the signal crimes perspective is the emphasis on the *subjectivity* of these signals. 'Not everyone will tune into the same set of signals, nor will they necessarily interpret a signal in the same way' (Innes 2004: 352). Drawing on the wider social scientific literature on risk perception, signal crimes makes sense of how and why different instances of disorder and crime are rendered meaningful by people.

Definitionally, disorder covers any breach of prevalent norms and conventions that are disturbing or troubling. Physical disorder refers to the material detritus of anti-social behaviour and incivilities (Innes 2014). Examples include litter, criminal damage, vandalism, and graffiti (Donoghue and Colover 2011). *Signal* disorders are those that individuals judge to pose a potential threat. These are qualitatively different from unimportant and meaningless information that can be effectively ignored and treated as mere background noise to the conduct of everyday life. For example, interviews conducted by Innes (2014) reveal how people use both the visual nature of the disorder and repeated encounters with it to interpret something as a signal. This repeated encounter is something that is unique to the subjective 'perceiver'. Evidently, signal disorders are the features of an environment that people ascribe meaning to.

Beyond qualitative interviews however, there have not been many options to map the presence of these signals, in order to investigate their fluctuation in place and time. Where quantitative measures of disorder do exist, they tend to either miss this subjective interpretation element entirely, or focus on the subjective attitude only without anchoring it to a specific experience or event. To illustrate this, we now consider the two main approaches to measuring presence of incivilities on a larger scale.

One approach measures perceived disorder (the level of disorder people think is present in their area) (Davenport 2010), by asking people about disorder in their neighbourhood using cross-sectional surveys. An example of this approach can be found in the Crime Survey for England and Wales. The corresponding question asks respondents to rate the extent to which they believe that disorder is a problem in their local area (see Brunton-Smith (2011) for an example of utilising this measure).

This measure falls victim to an issue which has been discussed in relation to the measurement of fear of crime; such questions are better suited to capture overall attitudes and anxieties, as opposed to everyday experiences with an issue (Gray *et al.* 2008a; 2008b). Using these measures to represent people's actual encounters with disorder may overestimate (or underestimate) the extent of the issue. To compensate for this, measurements of fear of crime have shifted emphasis to anchoring these questions to actual experiences (Gray *et al.* 2008a; Solymosi *et al.* 2015). This is not yet true

of perceived disorder questions which fail to consider the frequency and intensity of experience.

Another issue is the lack of spatial specificity of these questions. Measuring perception at a neighbourhood level might mask low-level variation in signal disorders. Disorder and incivilities are likely to vary *within* neighbourhoods. For example, Weisburd *et al.* (2012) found hot spots of physical disorder showed significant within-area variability. This implies that signal disorders are not evenly dispersed in neighbourhoods. Therefore, it is important to consider signal disorders at the smallest possible scale in order to un-erroneously associate them spatially with other elements of the environmental backcloth.

The second approach is to measure instances of observed disorder, usually through systematic social observation (SSO). In SSO, surveyors cover a specified area and record observed instances of incivilities (Sampson and Raudenbush 1999). While this approach reflects instances of disorder rather than generalized attitudes, it relies on the interpretation of the researcher only. This is a major limiting factor in terms of collecting data about *signal* disorders as it misses the identification of an issue as problematic by the passing perceiver. The signal crimes framework emphasizes that it is ‘the situated context in which any signifier is located, together with the characteristics of the audience members, that shapes the construction of meaning’ (Innes 2004: 352). As discussed earlier, in order for something to be a signal disorder, it needs to be interpreted as such (Innes *et al.* 2009). By simply logging all instances of observable disorder, SSO does not capture this subjective element. In areas where such fieldwork is conducted in congruence with interviews about signal disorders, a gap between observed and perceived levels of disorder is often detected (Innes 2014).

Finally, both common approaches to measurement suffer a limitation on the resolution of the temporal information collected. Surveys ask about the general neighbourhood area and rarely make a distinction between *when* a person considers the specific disorder to be an issue. SSO on the other hand is limited to hours when the researchers are working. For example, in Sampson and Raudenbush (1999) the surveyors worked between 7 am and 7 pm, missing out on recording disorder in the environment that might occur, or be more observable, during the darkness hours, for example (Davenport, 2010). The other time-insensitive feature of SSO is its inability to account for the effect of repeated exposure to something. As mentioned earlier, Innes (2014) found evidence that repeated exposure is a major contributing factor to an instance of disorder being interpreted as a signal. Therefore, such distinctions are important to make.

#### *Crowdsourcing as an Alternative Measure*

One possible avenue for collecting signal disorder data lies in the methodology of crowdsourcing. As discussed, the reports submitted to FMS contain a bias within them. They would not, like SSO, represent a collection of all signs of disorder. Instead, they are filtered to include only those people saw problematic enough to report. However, they do not represent people’s generalized evaluations of their environments either; each report is anchored to a specific instance of physical disorder encountered by the person reporting.

The underlying concept of participatory mapping methodology provides a platform for people to report concerns in their neighbourhoods. It is possible to capture observed

disorder (since people report about what they actually encounter), and also perceived disorder (since people prioritize reporting things they consider problematic issues). By looking into the application of participatory mapping at community level, [Innes \*et al.\* \(2009\)](#) found it has potential to aid police interventions focused on reducing the fear of crime. Specifically, it can enable the police to focus their resources on problems in particular locations that are functioning as the key drivers of neighbourhood insecurity ([Innes \*et al.\* 2009](#)). FMS data is both a participatory mapping exercise, and an online open source data set so we can use it to empirically explore whether it can represent the experiences of motivated perceivers coming across instances of disorder.

The remainder of this article will describe the data, considering the biases in reporting, and then compare with the two traditional measures of disorder described above (SSO and questionnaires). Finally, we will use the data to illustrate small-scale spatial and temporal variation in experiences with disorder, highlighting novel insight gained by crowdsourced data.

### *Data*

While the reports made on FMS are available to view individually on the web page hosted by MySociety, this does provide a form that is readily accessible for analysis. One way to acquire such data is by *scraping* the data. Web scraping refers to developing and running an application that processes the HTML of a web page to extract data for manipulation. In the case of FMS, the first author wrote a script using Java programming language to open each report, save the relevant information, close the report and move on to the next one. This script iterated through the hundreds of thousands of reports made on FMS, compiling the data in format where it was easily usable for research purposes. It is important to note that this was done with the full permission of MySociety. This method collected 5 years worth of data that included the following for each report:

- Latitude and Longitude
- Topic of report (e.g. 'Graffiti' or 'Litter')
- Time and date when the report was made
- Name of person reporting (if given, 'anonymous' if not)
- Detailed description of the report

The resulting database contained 276,656 usable entries for the United Kingdom after data cleaning. In the analysis that follows, two assumptions are made: that the location information provided with the report is the true location where the perceiver encountered the issue, and that the time when the report was submitted reflects approximately the time of encounter. To address these, we next explore the geographical and temporal reliability of the data.

### *Assumptions*

In terms of spatial accuracy, FMS allows people to submit, to point level, the location of the issue they are reporting. This location needs to be accurate in order for the problem to be addressed. A characteristic of VGI is that content creators are also consumers, and

therefore have vested interest in providing accurate information. Because of this, there exists a self-regulatory behaviour regarding the validity of the data, where people will strive to ensure the accuracy of the information they provide (Marjanovic *et al.* 2012). People self-regulate, because, in order for the local authority to be able to address the issue being reported, it is in the content creator's interest to provide accurate locations. This assumption is reinforced in the case of FMS data by looking at the proportion of raised issues marked as 'closed' by the council. At the time of writing, Bromley Council have marked 99.69 per cent of their cases as fixed.<sup>1</sup> Although this percentage is lower in other local authorities, it is good indication that the spatial information is reliable enough for the council to be able to locate and fix the reported issues.

To establish the accuracy of the assumption that time of reporting represents the time of experience with the sign of disorder, we consider the case of reporting broken streetlights. Broken streetlights are more noticeable during hours of darkness. Therefore, if people report issues when they experience them, reports of broken streetlights should be more prevalent during the night than the day. Comparing the proportion of reports during daylight and night-time hours reveals that a higher percentage of reports are about streetlights during hours of darkness than during daylight (Figure 1). A chi-square test demonstrates a significant relationship between whether the report was made in hours of darkness or daylight and whether it concerned broken streetlights ( $\chi^2 = 389.22$ ,  $df = 1$ ,  $p$ -value  $< 2.2e-16$ ). Whilst this is an indirect analysis of this assumption, we argue that it allows maintaining that the time of reporting roughly reflects the time of encounter with what is being reported.

### *Reporting Behaviour*

We now have some evidence that FMS data likely reflects *when* and *where* people encounter problems worth reporting, so we can look to spatial and temporal patterns in reports. While the temporal pattern in reporting shows steady increase year-on-year, the spatial patterns show a non-equal distribution of reporting. Some London Boroughs replaced their own online forms with FMS, affording it a form of legitimacy, indicating to residents that complaints on the site are taken seriously. The boroughs of Barnet and Bromley incorporated FMS into their sites, and as a result, the majority of reports come from these two boroughs (Figure 2).

Another interesting feature of the data is the fine-grained temporal cycle in reporting. Examining within-day fluctuation in reporting is reminiscent of looking at people's daily activity patterns—with far less reporting during night-time, when most people are asleep (Figure 3).

We can also examine the nature of the issues being reported. Figure 4 groups reports into 27 topics, giving a general idea of what people submit complaints about. The main two categories under which reports were submitted were 'Pavement or road issues' and 'Litter', which together contain 46 per cent of all reports.

It is also important to consider who the contributors are. Although FMS does not collect demographic information, some people leave their names, providing us with

<sup>1</sup>Excluding those labelled as 'new problems' which they may not have had time to address yet.

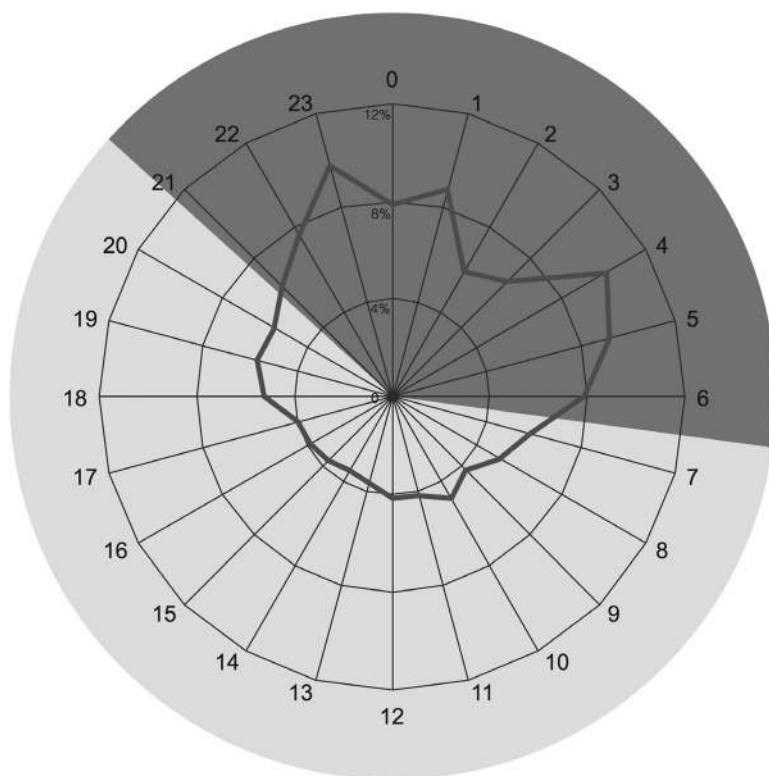


FIG. 1. Percentage of all FMS reports about streetlights

some useable intelligence. A feature of crowdsourced data is that a few users contribute the majority of the content (Howe 2006; Surowiecki 2005). This has disadvantages, which are discussed later, but the great benefit comes from all the contributions being recorded and compiled into one output. In this way, the researcher does not lose input from anyone, just because they choose to only participate once.

The data contain a total of 276,656 reports, of which 166,870 (60 per cent) were submitted anonymously, leaving 109,786 (40 per cent) named reports, which were sent by only 48,065 unique individuals. If distributed equally, this would mean just over two reports made per (name-leaving) person. But of course, it is not at all distributed evenly amongst these contributors. Of these, the top 1 per cent sent in one fourth of all reports. On the other hand, 73 per cent of people (who left a name) contributed only one response. In fact the median number of responses is 1, even though there were some very active users reporting over 800 issues. It is possible to represent this inequality in distribution using a Lorenz curve, used typically to graph wealth inequality. A perfectly equal distribution would be depicted by the straight line  $y = x$  (Gastwirth 1972; Lorenz 1905). Figure 5 shows the proportion of reports assumed by people using FMS to report.

The corresponding Gini coefficient of 0.51 represents the ratio of the area between the line of perfect equality and the observed Lorenz curve, to the area between the line of perfect equality and the line of perfect inequality (Gastwirth 1972). The closer



## Distribution of FMS reports across London boroughs

□ < 0 Std Dev  
 ■  $\geq 4.00$  Std Dev



FIG. 2. Map to highlight the extent to which councils that adopt FMS are over-represented in reports compared to all others

the coefficient is to 1, the more unequal the distribution is (Zeileis *et al.* 2012). This result points immediately to the existence of ‘super contributors’ in the data. In crowdsourcing literature, this phenomenon of ‘participation inequality’ has been noted, and observed to follow a more or less 90-9-1 rule (Stewart *et al.* 2010). Grouping people into three categories (super contributors providing 90 per cent, contributors 9 per cent, and outliers 1 per cent), Stewart *et al.* (2010) note that super contributors are highly motivated in their participation. In the context of FMS reports, we can understand this as a group of people who actively monitor their environments, and report any issues they come across. These people are not category-specific either. Of the 1,024 people who left three reports or more, only 205 reported in only one category (20 per cent), while the majority reported in two or more. We will return to this inequality in the discussion.

While the reports do not collect demographic information, it is possible to draw conclusions based on names where they were left with the report. Using data from credit card and birth certificate information, Longley *et al.* (2015) developed a way to infer gender and age from first names. This inferential procedure is by no means perfect (Lansley and Longley 2016). However, such an approach is a viable means of assigning characteristics to individuals about whom only their name is known. Names are particularly successful as a means of estimating gender (Lansley and Longley 2016). Therefore, it will be used to explore gender differences in FMS reporting here.

Men submitted 24.5 per cent of all reports, with 67 824 reports made with typically male first names, while women submitted only 8.6 per cent of all reports ( $n = 23,825$ ).

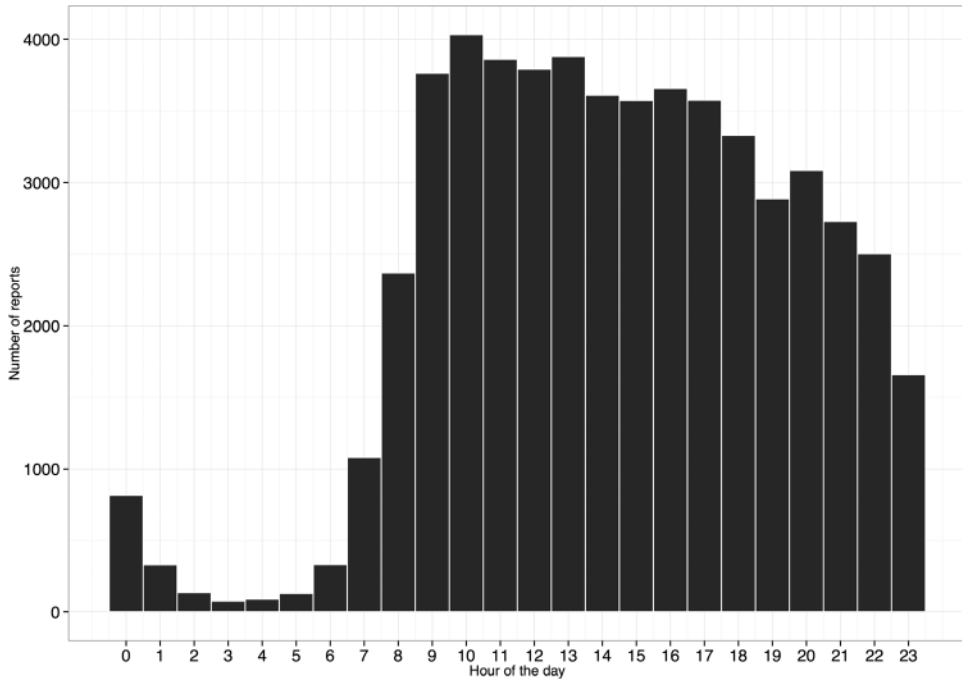


FIG. 3. Temporal trends in reporting reflects daily cycle in all activities

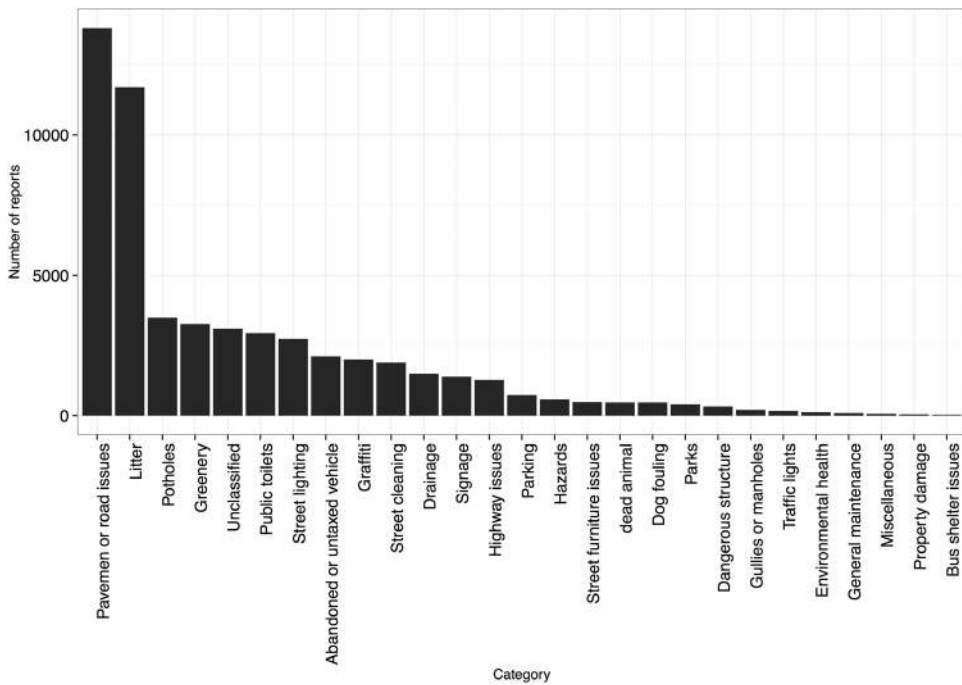


FIG. 4. Overarching categories FMS reports fit into

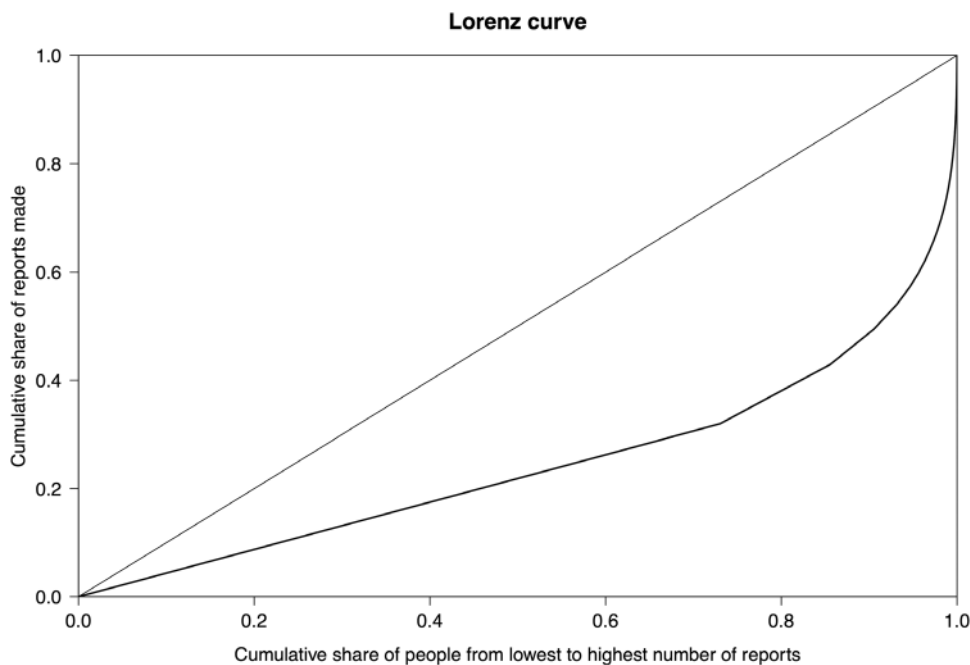


FIG. 5. Lorenz curve of FMS participation

For the rest of the reports, gender could not be inferred from their name. Of these, most (over 90 per cent) were submitted anonymously, while the other 10 per cent provided an initial or an obvious pseudonym, such as 'concerned citizen'. These reports were classed as 'unknown' in terms of gender.

Overall, we see more reports from men than women. There are two potential reasons for this; the first that men submit more FMS reports than women do, and the second that when men make reports, they are more likely to leave their full name. However, drawing the conclusion that men leave more reports would be supported by previous research into biases in other crowdsourced data (Budhathoki 2010; Haklay 2010). Such biases are important to keep in mind when interpreting results from this data, and will be discussed later in the article.

Looking at differences in reporting across categories shows that there is a significant gender difference ( $\chi^2 = 4822.301$ ,  $df = 58$ ,  $p$ -value  $< 0.001$ ). The standardized residuals (shown in Table 1) reveal that reports about parking, abandoned vehicles, graffiti, highway issues, hazards, and carriageway defects were most likely to be reported anonymously. Reports about dog fouling, greenery and litter were more likely to come from women or anonymous reporters, while reports about dead animals, parks and public toilets were more likely to come from women as a single group. People who left their name were more likely to report street cleaning issues than anonymous reporters. And finally, reports in unclassified, potholes, pavement or road issues were more likely reported by men.

On first glance it appears that men are more likely to report in categories related to driving (potholes and road problems), whereas women report more in categories related to walking (parks, dead animals, dog fouling, litter). There is great potential

TABLE 1 *Chi-square standardized residual table of number of reports in FMS categories by gender (anything greater than 2 or below -2 is a meaningful difference)*

	Female	Male	Unknown
Abandoned or untaxed vehicle	-6.14	-22.14	23.9
Bridge issues	-0.78	-0.42	0.85
Bus shelter issues	2.06	-0.24	-1.01
Carriageway defect	-2.67	-4.07	5.31
Dangerous structure	1.92	-1.86	0.56
Dead animal	4.75	-6.52	3.12
Debris	-0.31	-2.9	2.84
Dog fouling	9.31	-19.37	12.15
Drainage	-2.19	-1.31	2.5
Environmental health	1.4	-2.14	1.12
General maintenance	3.22	-3.29	1.08
Graffiti	-6.27	-5.38	8.65
Greenery	6.29	-14.29	9.31
Gullies or manholes	-0.8	-3.13	3.33
Hazards	-1.2	-6.58	6.73
Highway issues	-2.04	-7.13	7.73
Litter	4.89	-26.84	21.62
Miscellaneous	2.58	2.6	-3.91
Parking	-6.88	-22.75	24.9
Parks	3.66	-3.6	1.1
Pavement or road issues	-0.22	5.77	-5.15
Potholes	-4.73	23.63	-18.78
Property damage	-0.12	-1.61	1.54
Public right of way	0.5	3.16	-3.19
Public toilets	3.64	-5.67	3.01
Signage	2.27	3.11	-4.2
Street cleaning	3.44	8.37	-9.7
Street furniture	-1.13	0.55	0.17
Streetlighting	-2.37	1.42	0.11
Traffic lights	-2.16	2.09	-0.62
Unclassified	-3.5	38.19	-32.82

for exploring this further in research in the role of gender and the use of public space (Table 1). The findings are in line with results from anti-social behaviour (ASB) victims survey data from [Innes \(2014\)](#) who found that women, on average, attend more carefully to physical disorder signals. Furthermore, interviews by [Innes \(2015\)](#), also imply that women attend more to physical disorder, whilst men scan more for the potential for violence ([Innes 2015](#)) (or potential for some harmful consequence to them or their property such as a vehicle).

We emphasize these inequalities in reporting by area and gender and the presence of the super-contributors, as these are biases in the data that must be considered in its interpretation. We maintain that these biases can be used to learn more about different people's experiences with disorder in the environment and can be interpreted and studied, rather than dismissed as limitations. The bias of *what* gets reported is another strength of this crowdsourced data, applied to the study of disorder through the signal crimes framework. The next section addresses this bias by comparing with other disorder data, before we move on to consider the dynamic, spatio-temporal information gained by mapping this data as representations of perceived signal disorders.

*Subjective Experiences with Disorder*

Earlier, we suggested that the bias of *what* is reported on FMS means that these data represent not an econometric measure of disorder in a neighbourhood, but issues that people *subjectively* evaluate as problematic, in line with the signal crimes framework. Therefore, we hypothesize that this measure will not directly reflect SSO or questionnaire measures of disorder, but will instead present something new. To explore this, we use SSO and questionnaire data available for the London borough of Camden, and compare their features against the crowdsourced data.

The Camden case study area covers approximately 22 square kilometres in inner London, home to almost 210,000 residents. In socio-economic terms, it is one of the most polarized boroughs with some of the wealthiest areas in England as well as some of the most deprived. Recorded crime levels are above the average for London (Camden Council Sites [Team 2012](#)). With these characteristics, this borough presents a good representation of various land uses and populations.

The data for the traditional ‘questionnaire’ measure of disorder come from the Metropolitan Police Service Public Attitudes Survey (PAS). The PAS is an annual survey running since 1983 with the objective of eliciting Londoners’ perceptions of policing needs, priorities and experiences ([BMG Research 2014](#)). It is made up of face-to-face interviews at the homes of respondents, selected from a random probability sample of residents in each of the 32 boroughs across London ([Mayors Office for Policing and Crime 2016](#)). Approximately 1,067 interviews per month are carried out, equating to 100 interviews per Borough per quarter ([BMG Research 2014](#)).

The data for traditional SSO measure comes from Camden Council local authority. It contains systematically collected data about instances of disorder in the environment, collected by monitoring officers, who patrol the borough. While they record different types of disorder, they were willing to share reports of litter. They typically log between 500 and 1,000 reports of litter per month. Accordingly, to enable direct comparison, the rest of this analysis will focus on FMS reports and PAS responses concerning litter in particular.

Serendipitously, litter features prominently in the signal crimes narrative. [Innes \(2014\)](#) found that ‘the dumping of litter signalled to residents that an area is “deteriorating”’ (p. 29). Litter is also the most commonly reported environmental issue that can be considered an instance of disorder. Furthermore, it impacts upon a lot of people, but fairly diffusely (by contrast, for example, being ‘intimidated and pestered’ is not something many encounter, although those who do are more intensely affected by it) ([Innes 2014](#)). Evidently, litter is something many people experience, and has the potential to be interpreted as a signal disorder, yet will often not necessarily be interpreted as so. Therefore, an SSO measure of litter can be hypothesized to over-estimate the extent of signal disorder encounters with the issue. For clarity, [Table 2](#) summarizes the sources of data used in this section, with some details.

*Relationship between complaints, SSO data and survey responses*

We begin by examining the spatially weighted strength of relationships between FMS complaints and the other measures. We used a spatial error model in order to account for (spatially correlated) covariates that, if left unaccounted for, would affect inference. The unit of analysis was the 133 neighbourhoods within Camden. These are defined

TABLE 2 *Data sources*

Data source	Description	Sample size and time period	Strengths	Weaknesses
Metropolitan Police Public Attitudes Survey data	Survey question regarding the perceived problem of litter in respondent's neighbourhood	1,174 responses to this question over 3 years (Jan 2012–Dec 2014)	Represents residents' perceptions of the problem	Potentially subject to sampling bias and recall bias
Systematic social observation	Council monitoring officer reports of litter from patrol	46,262 logged incidents over 4 years (Jan 2011–Dec 2014)	Give an objective measure of incidents of littering in the environment	SSOs are still a sample reflecting what is present at the time of survey
FMS data on litter	Reports about litter made to <a href="http://fixmystreet.com">fixmystreet.com</a>	11,710 reports, in just over 4 years (Jan 2011–Dec 2014)	Openly available crowdsourced data that covers all of London. Represents a potential indicator of signal disorder—those the public care enough about to report	Possible bias due to some individuals more likely to make complaints to the council about issues that cause them concern. Further bias introduced through 'digital divide'

as lower super output areas (LSOAs) which are geographical regions 'designed to be more stable over time and consistent in size than existing administrative and political boundaries. LSOAs comprise, on average, 600 households that are combined on the basis of spatial proximity and homogeneity of dwelling type and tenure' (SturGIS *et al.* 2014). We consider each LSOA to be one neighbourhood.

Perceived levels of litter are derived from answers of LSOA residents to the PAS question 'How much of a problem is rubbish or litter lying around?', with responses coded 1–4, from 'Not a problem at all' (1) to 'Very big problem' (4). Aggregate scores for each neighbourhood were calculated by taking the median of the responses from residents, in line with Likert scale questionnaire analysis best practice (Clason and Dormody 1994; Johns 2010; Boone and Boone 2012).<sup>2</sup> Higher scores mean that the neighbourhood residents perceive rubbish to be more of a problem.

Table 3 shows four models. For each traditional measure, we examine both its relationship with the other traditional measure and with the FMS data. The *R*-squared, log likelihood, and AIC values can be used to compare the models. We find that FMS data is not significantly associated with either SSO or Questionnaire measures of disorder. Neither does it perform better in predicting perceived disorder measured with questionnaires than SSO, nor in predicting observed disorder measured with SSO than questionnaires.

There may be external issues that affect these results, for example the level of participation in online reporting platforms, or other biases which affect representativeness of the FMS data. However, with these results we cannot say that FMS data purely reflects

<sup>2</sup>The median rather than the mean is taken, because this is not a composite score, rather the answer to only one question, so it is an ordinal variable of ranked qualitative answers.

TABLE 3 *Spatial regression results*

Dependent variable	Predictor variable	Coefficient	<i>P</i> -value	<i>R</i> -squared	Log likelihood	AIC
SSO	FMS	11.18	0.11	0.13	-922.79	1849.58
	Questionnaire	11.30	0.77	0.12	-923.99	1852
Questionnaire	SSO	0.01	0.73	0.04	-112.62	229.24
	FMS	-0.01	0.89	0.39	-112.67	229.33

either observed or perceived levels of disorder. Instead, as theorized above, it potentially measures something new, which when explored in greater detail, could yield insight into people's everyday experiences and subjective perceptions of their environments.

#### *Exposure to Signal Disorders as a Function of Routine Activities*

In the previous section, we established that FMS data, conceptualized using signal crimes framework as people's encounters with signs of disorder they consider problematic, show something not captured either by traditional perception surveys, nor objective SSO observers. Therefore, at least using the case of Camden, it would be incorrect to use it as an econometric indicator of disorder at neighbourhood level. Instead, it is telling us something different about people's experiences with disorders which affect them, as they go about their routine activities. To showcase the full utility of FMS interpreted this way and the benefit of its spatial and temporal granularity, we now explore spatio-temporal features of these data.

To do this, we subset the data again, from the all 26 categories in which reports were made (Figure 4), to all instances of environmental antisocial behaviour (enviro-ASB). Enviro-ASB includes any antisocial behaviour act where the incident is not aimed at an individual or group but targets the wider environment, e.g. public spaces/buildings (Police 2015). Also called incivilities, examples include litter, criminal damage, vandalism and graffiti (Donoghue and Colover, 2011). These are the issues that can be interpreted as signal disorders with potentially harmful consequences on health and well-being (Sampson and Raudenbush 2004). We take this subset as using the entire FMS data set would conflate a number of distinct types of problem and incivilities are, through their very nature, likely to be more sensitive to particular settings, and are most likely to represent signal disorders. In the following analysis, we therefore classify this subset of reports as 'signal disorders' and use the remaining non-disorder reports as a comparison group.

In total, about 30 per cent of reports are about incivilities. Temporally, incivility reports peak at 7 am on weekdays only (Figure 6), when they make up 40 per cent of all FMS reports. This means that at 7 am on weekdays, people encounter higher rates of signal disorders. This is contrary to what might be expected, based on images of increased fear 'after dark'. To investigate, we turn to the wealth of information included in FMS reports. From the detailed descriptions, it appears that the majority of these are litter complaints, about overflowing bins and fly-tipped items. The narrative descriptions included with FMS reports reveal that these reports are made by people who are waking up to go to work, and encountering signs of activity that took place in

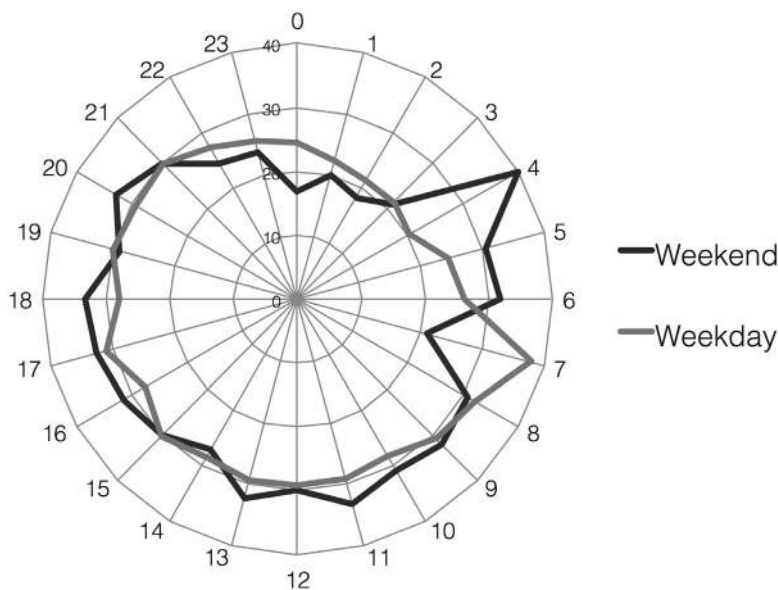


FIG. 6. Percentage of all FMS reports about incivilities

the same location, but at a different time (Figure 7). They see signs of another activity in the space their routine activity pattern takes them through but is incongruent with their current use of this space, and interpret these as a signal disorder, attributing meaning which can result in heightened fear or anxiety. The finding that people make inferences about use of space based on artefacts of its previous use is very interesting, and will be taken up in the discussion.

Besides the fine-grained temporal resolution, the data also allow us to map at micro-geographical level where people are more or less likely to encounter problematic instances of disorder. Using Gi\* to identify areas where local patterns differ from the overall study area (using street segments as the unit of analysis), we can see significant clusters of street segments (those in red) where there is a high proportion of reports concerned signal disorders (Figure 8).

With the fine-grained temporal data, it is further possible to see how these clusters of segments shift over time with changes in routine activities. For example we can consider the changes in travel patterns during the day. Based on travel patterns in London the day can be split into six main groups: early morning (4 am to 7 am), am peak (7 am to 10 am), inter-peak (10 am to 4 pm), pm peak (4 pm to 7 pm), evening (7 pm to 10 pm) and night (10 pm to 4 am) (Transport for London 2014). Figure 9 shows how the clusters of segments where people are encountering a higher proportion of signal disorders varies between these time periods.

Interestingly, Figure 9 appears to demonstrate that there is more dispersion and differentiation in high clusters during the day, perhaps representing differences in land use and non-home based routine activities at these times. Such information can be used to identify places and times where people are more likely to experience perceptions of insecurity as they go about their everyday experiences.





FIG. 7. FMS report about ‘... overflowing litter, street drinkers, litter in the street and rats’

### *Discussion*

The previous sections illustrate the many layers of information available in crowd-sourced data. We have argued that it is important to consider the biases in such data not only as limitations, but as additional layers to be explored. By conceptualizing the bias of *what* gets reported from the signal crimes perspective, we gain insight into people’s everyday experiences with disorders and short-term spatial and temporal fluctuations in these experiences. The benefit of place-based approach to study crime (Groff *et al.* 2010; Weisburd *et al.* 2012) could be applied to place-based study of *perception* of crime, with such new data representing subjective experiences.

Our findings have implications both for contributing to signal crimes theory, and for the broader topic of interpreting the bias in crowd-sourced data as an additional layer of information that has a conceptual interpretation. Regarding signal crimes theory in particular, our exploration of variation in FMS data highlights the importance of the



FIG. 8. Figure IV: Gi\* map of the proportion of reports that are incivilities

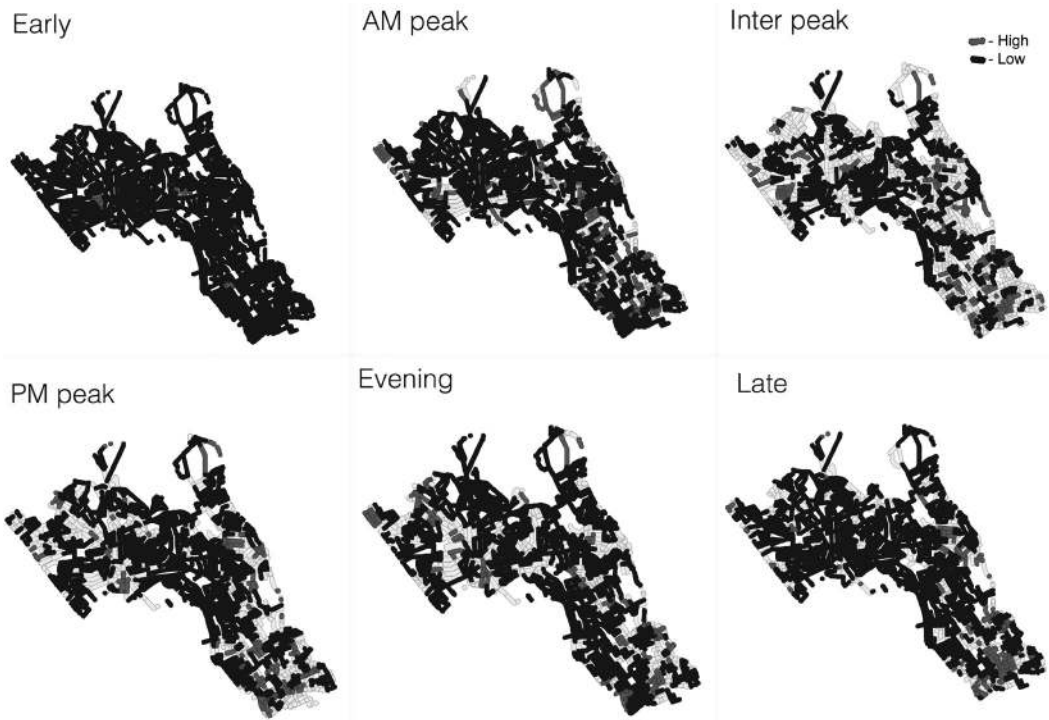


FIG. 9. Variation in hotspots of signal crimes within the day

individual perceiver. This is in accordance with the unique perspectives of those interviewed by Innes (2014) when attributing meaning to disorder, but has the advantage of scaling up the evidence on this using secondary data, and of specifically situating signal crime incidents in space and time. Individual differences play a role in all types of human behaviour (Wortley 2011), and even within-person changes, such as the psychological state of the perceiver (Jackson 2015) can impact the interpretation of a signal. FMS data demonstrate that the experience of a signal disorder is the result of the convergence of the appropriate situational factors with appropriate individual factors in the person interpreting the signal. It is a generated output, much like fear of crime, in that it does not exist on its own, it is only realized when someone comes across to experience it as such.

Interpreting the signal crime as a product of the ‘person–situation interaction’ (Wortley 2011) in a particular place and time has a number of implications in terms of the next steps necessary in furthering our understanding, and in terms of prevention. For example, it is not yet clear whether certain types of setting are likely to elicit a certain kind of reaction in most perceivers or in more limited number of cases. In other words, to what degree does the individual interpretation have more or less weight in the labelling process than objective features such as the setting, the time and the type of disorder? This balance will likely vary over situation, but has implications for dealing with disorder. For example, is it important to work with vulnerable populations, problem places or both?

There is also a broader commentary about interpreting the bias inherent in the mode of production of crowdsourced data through applying theoretical frameworks, such as signal crimes. This article has demonstrated the advantages of operationalizing this bias, and treating it as strength of such new sources of data, rather than a limitation. Using the signal crimes framework enabled the interpretation of FMS data as a representation of people’s subjective experiences. Likewise other theoretical frameworks which examine the role of human interpretation could be applied to crowdsourced data. Such exploration is not limited to FMS data—there are many other emerging data sources that could be usefully analysed. For example, it could be possible to consider the motivations behind tweeters tweeting, to draw inferences about the meaning of the data on a more nuanced level.

In practical terms, the crowdsourced FMS data can help identify times and places where ‘hand over’ of a place occurs between different ‘users’, which we suggest can raise anxieties. This could plausibly help tailor intervention. For example, if we know that those who are out at night tend to retire from an area at 4 am, and the new users don’t appear until 7 am, the signals left behind could be cleaned up in the time before the new users of the space arrive. In fact when it comes to addressing perceptions of insecurity, it may be as effective to prevent a ‘perceiver’ from seeing something they can interpret as a sign as it would be to prevent the act that created the signal in the first place. Mapping these experiences in the micro-scale geographical resolution could highlight hotspots of concern.

However, if such applications are to be considered, the other biases in the data must be kept in mind. Care must be taken to avoid issues such as slanting services towards those who are over-represented in this data. Earlier, we demonstrated the potential ‘over-influence’ of super-contributors. While we explored the bias of *what* is reported, it is also important to consider *who* these contributors are. The sample of participants

is entirely self-selected, leaving out those who do not participate so extensively, but might still be affected (Budhathoki 2010; Haklay 2010). Longley (2012) comments that 'self-selection is an enemy of robust and scientific generalisation, and crowdsourced consultation exercises are likely to contain inherent bias' (Longley 2012: 2233). Beyond self-selection issues in general, an entire body of work has explored the impacts of the 'digital divide' (for an example see Yu (2006)), which refers to certain socio-economic groups being overrepresented in data generated on-line (Malleson and Andresen 2015). Gender bias has been found, showing that men tend to participate more so in such activities than women. Further work on VGI participation has also shown a divide in participation along many socio-demographic variables. Employed people, those people between the ages of 20 and 50, and those with a college or university degree are most likely contributors (Budhathoki 2010; Haklay 2010). Looking into what contextual factors influence participation in Open Street Map, (Mashhadi *et al.* 2013). Mashhadi *et al.* (2013) find that, socio-economic factors such as population density, dynamic population, distance from the centre and levels of poverty all play a role. These factors are important to keep in mind when reporting findings based on analysis of such data.

A limitation of the current study comes from having subset the data to litter reports only, which was necessary to triangulate the data while ensuring the same concept was measured. Different relationships might have been found with disorder data relating to graffiti for example. However, the data presented here and the relationships described are for one context only and serve as a first exploration of the data in this way. The intention is to pave the way for the application of this approach to other crowdsourced data, as well as to encourage replication to help determine the generalisability of some of the findings. As FMS and similar on-line open data become more established in other countries, international comparison should become a possibility.

A further avenue for research is to use openly available online data sources to aid with gaining insight into other elements of people's perception of their environments. For example in their study utilizing community intelligence, Innes *et al.* (2009) asked participants to plot on the maps the boundaries of what they consider to be their neighbourhoods. Perhaps mapping participation in various online forums could generate inferences about geographical location of communities with shared interests, which could help further guide community engagement strategies.

It is inevitable that the next decade will see the emergence of new sources of open access data, which should be ripe for exploration in social sciences and criminology. In assessing the viability of new forms of data as a source of insight into criminological concepts it is important to consider all layers of data and their possibilities. A solid theoretical approach and appropriate domain knowledge are key to be able to frame and make sense of these data in a meaningful way (Williams *et al.* 2016). A necessary step in doing this involves the application of a relevant theoretical framework, as well as the triangulation and exploration of the new data with existing more traditional sources. Here, we have demonstrated this process for the measurement of issues related to disorder. It appears that complaints data, collected through crowdsourcing websites such as FMS, represents a measure that cannot be situated as either entirely objective and entirely subjective measures of disorder, and can be argued to represent signal disorders in this sense. That is, issues that are both observable and cause enough of a reaction in the beholder to act to report them.

This article provides a template for using crowdsourced data to study people's perception of crime, disorder and place at a resolution at which data were previously unavailable. By framing the subjective bias of what gets reported as something that encompasses both the tangible experiential aspect as well as the emotional bias of interpreting something as a signal, we can explore when and where people come across disorders that matter to them during their routine activities. Just like the shift towards considering crime as something that varies with the situational context in place and time, framing perception of disorder this way can lead to new, situational approaches to reduction that have been missed by traditional approaches.

Overall we merely scrape the surface of the potential depth of insight to be gained from crowdsourced data. Using such information to supplement current knowledge on people's perceptions from traditional survey methods could bring new insight such as the ones highlighted in this article, and bring a situational perspective to the study of the subjective perceptions of crime and place. Whilst not yet qualifying as the use of 'big data' approaches, we are optimistic that the use of new open source crowdsourced data opens up opportunities for new theoretical and practical insights in social science research.

#### *Funding*

This work was supported by the Engineering and Physical Sciences [EP/G037264/1] as part of UCL Security Science DTC.

#### ACKNOWLEDGEMENTS

Many thanks to Kieran Widdowson from Camden Council, Liam Fenn and Colleen Reddin from MOPAC, and Dr Rebecca Rumbul from mySociety for access and permissions to the data used in this article.

#### REFERENCES

- BMG Research. (2014), *Public Attitude Survey 2013–2014*, available online at [http://doc.ukdataservice.ac.uk/doc/7048/mrdoc/pdf/7048\\_pas\\_tech\\_report\\_2013\\_qtr\\_35\\_with\\_map.pdf](http://doc.ukdataservice.ac.uk/doc/7048/mrdoc/pdf/7048_pas_tech_report_2013_qtr_35_with_map.pdf) (Accessed 10 August 2017).
- BOONE, H. N. J. and BOONE, D. A. (2012), 'Analyzing Likert Data', *Journal of Extension*, 50: 30.
- BRUNTON-SMITH, I. (2011), 'Untangling the Relationship Between Fear of Crime and Perceptions of Disorder Evidence from a Longitudinal Study of Young People in England and Wales', *British Journal of Criminology*, 51: 885–99.
- BUDHATHOKI, N. R. (2010), *Participants' Motivations to Contribute Geographic Information in an Online Community*. University of Illinois at Urbana-Champaign.
- CASTEN, J. A. and PAYNE, B. K. (2008), 'The Influence of Perceptions of Social Disorder and Victimization on Business Owners' Decisions to Use Guardianship Strategies', *Journal of Criminal Justice*, 36: 396–402.
- CECCATO, V. (2012), 'The Urban Fabric of Crime and Fear', in V. Ceccato, ed., *The Urban Fabric of Crime and Fear*, 3–33. Netherlands: Springer.
- Cisco. (2016), 'Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper', available online at <http://www.cisco.com/c/en/us/>

- solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html (accessed 3 January 2017).
- CLASON, D. L. and DORMODY, T. J. (1994), 'Analyzing Data Measured By Individual Likert-Type Items' *Journal of Agricultural Education*, 35: 31–35.
- DAVENPORT, R. A. (2010), *Incivilities, Crime and Social Order: The Role of Repeat Experience*. University of Sheffield.
- DONOGHUE, J., & COLOVER, S. (2011), 'Anti-Social Behaviour and the Courts in England and Wales Report of Findings', 44 (March): 1–34.
- GASTWIRTH, J. L. (1972), 'The Estimation of the Lorenz Curve and Gini Index', *The Review of Economics and Statistics*, 54: 306–316.
- GOODCHILD, M. F. (2007), 'Citizens as Sensors: The World of Volunteered Geography', *GeoJournal*, 69: 211–21.
- GRAY, E., JACKSON, J. and FARRALL, S. (2008a), 'Reassessing the Fear of Crime', *European Journal of Criminology*, 5: 363–80.
- . (2008b). *Researching Everyday Emotions: Towards a Multi-disciplinary Investigation of the Fear of Crime*, Vol. 3. Universitätsverlag Brockmeyer, available online at [http://eprints.lse.ac.uk/21538/1/Researching\\_everyday\\_emotions\\_\(LSERO\).pdf](http://eprints.lse.ac.uk/21538/1/Researching_everyday_emotions_(LSERO).pdf)
- GROFF, E. R., WEISBURD, D. and YANG, S.-M. (2010), 'Is It Important to Examine Crime Trends at a Local "Micro" Level?: A Longitudinal Analysis of Street to Street Variability in Crime Trajectories', *Journal of Quantitative Criminology*, 26: 7–32.
- HAKLAY, M. (2010), 'How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets', *Environment and Planning B: Planning and Design*, 37: 682–703.
- . (2013), 'Citizen Science and Volunteered Geographic Information', in D. Sui, S. Elwood and M. Goodchild, eds., *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (Vgi) in Theory and Practice*, 105–22. Springer.
- . (2015), *Citizen Science and Policy: A European Perspective*. The Wodrow Wilson Center, Commons Lab.
- HOWE, J. (2006), 'The Rise of Crowdsourcing', *Wired Magazine*, 14: 1–4.
- INNES, M. (2004), 'Signal Crimes and Signal Disorders: Notes on Deviance as Communicative Action', *The British Journal of Sociology*, 55: 335–55.
- . (2014), *Signal Crimes: Social Reactions to Crime, Disorder, and Control*. Oxford University Press.
- . (2015), "'Place-ing" Fear of Crime', *Legal and Criminological Psychology*, 20: 215–7.
- INNES, M., ABBOTT, L., LOWE, T. and ROBERTS, C. (2009), 'Seeing like a Citizen: Field Experiments in "Community Intelligence-Led Policing', *Police Practice and Research*, 10: 99–114.
- JACKSON, J. (2015), 'Cognitive Closure and Risk Sensitivity in the Fear of Crime', *Legal and Criminological Psychology*, 20: 222–40.
- JOHNS, R. (2010), *Likert Items and Scales. Survey Question Bank: Methods Fact Sheet*, 1.
- KAISLER, S., ARMOUR, F., ESPINOSA, J. A. and MONEY, W. (2013), 'Big Data: Issues and Challenges Moving Forward'. in 46th Annual Hawaii International Conference on System Sciences, HICSS 2013, 995–1004.
- KELLY, C. E., TIGHT, M. R., HODGSON, F. C., and PAGE, M. W. (2011), 'A Comparison of Three Methods for Assessing the Walkability of the Pedestrian Environment', *Journal of Transport Geography*, 19: 1500–8.

- KOHN, S. A. (2013), 'Spatial Dimensions of Fear in a High-Crime Community : Fear of Crime or Fear of Disorder ?', *Canadian Journal of Criminology and Criminal Justice*, 51: 1–30.
- LANSLEY, G. and LONGLEY, P. (2016), 'Deriving Age and Gender from Forenames for Consumer Analytics', *Journal of Retailing and Consumer Services*, 30: 271–8.
- LEWIS, D. A. and MAXFIELD, M. G. (1980), 'Fear in the Neighborhoods: An Investigation of the Impact of Crime', *Journal of Research in Crime and Delinquency*, 17: 160–89.
- LONGLEY, P. A. (2012), 'Geodemographics and the Practices of Geographic Information Science', *International Journal of Geographical Information Science*, 26: 2227–37.
- LONGLEY, P. A., ADNAN, M. and LANSLEY, G. (2015), 'The Geotemporal Demographics of Twitter Usage', *Environment and Planning A*, 47: 465–84.
- LORENZ, M. O. (1905), 'Methods of Measuring the Concentration of Wealth', *Publications of the American Statistical Association*, 9: 209–19.
- MALLESON, N. and ANDRESEN, M. A. (2015), 'The Impact of Using Social Media Data in Crime Rate Calculations: Shifting Hot Spots and Changing Spatial Patterns', *Cartography and Geographic Information Science*, 42: 112–21.
- MARJANOVIĆ, S., FRY, C. and CHATAWAY, J. (2012), 'Crowdsourcing Based Business Models: In Search of Evidence for Innovation 2.0', *Science and Public Policy*, 39: 318–332.
- MASHHADI, A., QUATTRONE, G. and CAPRA, L. (2013), 'Putting Ubiquitous Crowd-Sourcing into Context', in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 611–22. ACM.
- Mayors Office for Policing and Crime. (2016), *Metropolitan Police Public Attitudes Surveys, 2000–2015. [data collection]*. 4th edn. available online at [ckan.data.alpha.jisc.ac.uk/dataset/7048](https://ckan.data.alpha.jisc.ac.uk/dataset/7048) (accessed 27 January 2016).
- MCDONALD, N. C., DEAKIN, E. and AALBORG, A. E. (2010), 'Influence of the Social Environment on Children's School Travel', *Preventive Medicine*, 50(Suppl 1): S65–8.
- MITRA, R., BULIUNG, R. N. and ROORDA, M. J. (2010), 'Built Environment and School Travel Mode Choice in Toronto, Canada', *Transportation Research Record: Journal of the Transportation Research Board*, 2156: 150–9.
- MySociety. (2016), 'MySociety About Page', available online at <https://www.mysociety.org/about/> (Accessed 10 August 2017).
- O'BRIEN, D. T., SAMPSON, R. J. and WINSHIP, C. (2015), 'Ecometrics in the Age of Big Data', *Sociological Methodology*, 45: 101–47.
- POLICE, M. (2015), 'What is Anti-Social Behaviour?', available online at <https://www.met.police.uk/advice-and-information/antisocial-behaviour/> (accessed 10 August 2017).
- PREIS, T., MOAT, H. S. and STANLEY, H. E. (2013), 'Quantifying Trading Behavior in Financial Markets Using Google Trends', *Scientific Reports*, 3. doi: 10.1038/srep01684.
- SAMPSON, R. J. and RAUDENBUSH, S. W. (1999), 'Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods', *American Journal of Sociology*, 105: 603–51.
- . (2004), 'Seeing Disorder: Neighborhood Stigma and the Social Construction of "Broken Windows"', *Social Psychology Quarterly*, 67: 319–42.
- SOLYMOSI, R., BOWERS, K. and FUJIYAMA, T. (2015), 'Mapping Fear of Crime as a Context-Dependent Everyday Experience That Varies in Space and Time', *Legal and Criminological Psychology*, 20: 193–211.
- STEWART, O., LUBENSKY, D. and HUERTA, J. M. (2010), 'Crowdsourcing Participation Inequality: A SCOUT Model for the Enterprise Domain', In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 30–3.

- STURGIS, P., BRUNTON-SMITH, I., KUHA, J. and JACKSON, J. (2014), 'Ethnic Diversity, Segregation and the Social Cohesion of Neighbourhoods in London', *Ethnic and Racial Studies*, 37: 1286–309.
- SUROWIECKI, J. (2005), *The Wisdom of Crowds*. Anchor.
- Camden Council Sites Team. (2012), *Camden Site Allocations: Local Development Framework*, available online at [https://www.camden.gov.uk/ccm/cms-service/download/asset?asset\\_id=2817197](https://www.camden.gov.uk/ccm/cms-service/download/asset?asset_id=2817197) (accessed 10 August 17).
- Transport for London. (2014), *London Travel Demand Survey*, available online at <http://content.tfl.gov.uk/travel-in-london-report-8.pdf> (accessed 10 August 2017).
- WEISBURD, D. L., GROFF, E. R. and YANG, S. M. (2012), *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*. Oxford University Press.
- WILCOX, P., QUISENBERRY, N. and JONES, S. (2003), 'The Built Environment and Community Crime Risk Interpretation', *Journal of Research in Crime and Delinquency*, 40: 322–45.
- WILLIAMS, M. L., BURNAP, P. and SLOAN, L. (2016), 'Crime Sensing with Big Data: The Affordances and Limitations of using Open Source Communications to Estimate Crime Patterns', *British Journal of Criminology*, 57: 320–340.
- WORTH, J. (2011), *Fix My Street—anywhere!* Jon Worth Euroblog.
- WORTLEY, R. (2011), *Psychological Criminology: An Integrative Approach*. Taylor & Francis.
- YU, L. (2006), 'Understanding Information Inequality: Making Sense of the Literature of the Information and Digital Divides', *Journal of Librarianship and Information Science*, 38: 229–52.
- ZEILEIS, A., KLEIBER, C. and ZEILEIS, M. A. (2012), *Package "ineq."* Comprehensive R Archive Network.