Cataloguing Protest: Newspapers, Nexis or Twitter?

Lesley Wood and Dyllan Goldstein

Abstract:

Goal – What is the best source for cataloguing protest? Newspaper sources remain dominant, but social media data and multimedia compilations are tempting. This paper compares three differently sourced catalogues of protest events in Toronto from 15 July – 15 September 2020. The Black Lives Matter and anti-eviction cycles of protest are clearly visible in all three sources, but otherwise reveal very different movements, and cover them differently. While Toronto Star coverage shows Toronto protest as state centred, and tied to a small number of ongoing news stories, other catalogues show a field of protest that is more diverse, irrational, fragmented and international. These observations help to unpack the characteristics of differently sourced event catalogues, and the implications of using them to understand patterns in protest.

Introduction

How do we track protest? Since the 1980s, social movement researchers have used catalogues of contentious gatherings or 'event catalogues.' These systematic compilations of activity offer a way to understand variation in the number of protests, the consistency of characteristics of protests in particular locations, and use of particular tactics, frames, claims or actors. (Tilly 2002) Such catalogues are constructed using relatively uniform procedures to compile a set of descriptions from a delimited set of sources (ibid).

Different sources will reveal very different pictures of protest, but how different? In order to answer this question, this paper compares three protest event catalogues from the same time and location - July 15 - September 15, 2020 in Toronto. It compares an event catalogue that uses the Toronto Star, the newspaper with the largest circulation in Canada, a catalogue sourced from Nexis Uni, an aggregation of over 6000 media sources, with a catalogue sourced from Twitter. All three catalogues capture the most visible movements, but beyond those, revealed different fields of protest, with different characteristics.

Event Catalogues

Event catalogues have three main advantages for understanding the patterns of protest. First, they can offer relatively reliable portraits of protest. Second, their form facilitates historical and comparative research (Tarrow 1996 in Earl et al., 2004), quantitative analysis (Olzak in Earl et al., 2004) and analysis of recurrent processes and mechanisms (Olzak 1989; McAdam, Tarrow and Tilly 2001). Third, they emphasize action, rather than what formal organizations and leaders say that they do (Earl et al., 2004). However, this emphasis on 'events' may also blur important differences amongst events (Biggs 2015). It may also diminish the importance of meaning making, interpretation and less visible aspects of social movement life. Like any method, the transformation of an event to a written account, and then to a catalogue entry loses detail. Overall, the quality of an event catalogue depends on the sources used - and all sources only cover certain things (selection biases), in certain ways (description biases). These biases limit the ability of protest event catalogues to fully represent protest activity.

No event catalogue could ever fully represent the 'reality' or complexity of protest interactions. As Jenkins and Maher (2016) note, there is no comprehensive inventory of all 'real world' protest events. Instead, the best contains the closest approximation to the total set of cases, using a systematic and intentional strategy. Any systematic set of accounts can be transformed into an event catalogue, but as Taehyun Nam (2006) says, "data matter." In their groundbreaking catalogue of contentious events in Great Britain, Charles Tilly and his collaborators attempted to represent contention by combining newspaper sources with records of parliamentary debate[1] . Their subsequent study of violent contention in France expanded the list of sources to include archives, newspapers, general histories, yearbooks, monographs, and other published sources (Tilly 1986 ICPR). This work was labour intensive, but revealed important changes in the form of contention. Others sought to imitate this approach and to study periods after daily newspapers became established, researchers increasingly began to rely on newspaper data as their main source of information on contention.

Newspaper data

Newspaper data is relatively accessible, reliable and consistent as a source of contention. It is also favoured, in part of its relatively broad readership and its representation of the attention of the public and discourse in 'the public sphere' (Oliver and Myers 1999). The advantages of newspaper data are best manifested when manual coding is used rather than machine reading, when full issues are sourced, rather than indexes, and when multiple newspapers are sourced, including both local and national press (Nam 2006; Oliver and Maney 2000). However, like any source, it has its limits. There are two main concerns about the recording biases within newspaper data - selection biases (which events are covered) and description biases (how they are covered (Earl et al., 2004; Smith 2001; McCarthy, McPhail and Smith 1996) .

Selection Biases - which protest events

But even the best newspapers do not cover all protest events. It is difficult to get a 'true' measure of the actual number or type of protest in a particular moment, so we do not know what proportion of events are included. Police reports appear to cover a larger proportion. One landmark study of protests in Washington DC in 1982 and 1991 that compared police reports with newspaper articles in the New York Times and Washington Post and TV coverage on three stations found that the New York Times reported on just 4.1% in 1982 and 1.8% in 1991, with the Washington Post covering slightly more at 7.9% in 1982, and 5.8% in 1991 (McCarthy et al., 1996). Movement generated compilations may also capture more events (Almeida and Lichbach (2003). However police data is difficult to obtain, and activist data is inconsistent. As Earl et al. (2004) argue, the proportion of events covered by newspapers wouldn't be a problem if it were representative.

The selection biases of newspapers have the advantage of being relatively well understood. Coverage is shaped by the availability of resources, reporting norms and editorial concerns (Earl et al., 2004). Relatively limited resources are allocated by editors, and journalists use norms to direct attention to 'newsworthy' events. This evaluation is shaped both by the characteristics of the issue, and of the event.

---

[1] (1) Times of London, (2) Morning Chronicle, (3) Mirror of Parliament, (4) Hansard's Parliamentary Debates, (5) Gentlemen's Magazine, (6) Annual Register, (7) London Chronicle, and (8) Votes and Proceedings of Parliament

As Earl et al., (2004) note that "Factors that influence judgment of an event's newsworthiness include the proximity of the event to the news agency (regionally, see McCarthy et al. 1996, 1999; internationally, see Mueller 1997), the size of the event (Barranco & Wisler 1999, Hug & Wisler 1998, McCarthy et al. 1996, 1999; Oliver & Myers 1999, Oliver & Maney 2000), the intensity or drama of the event (Mueller 1997), violence at the event (Barranco & Wisler 1999), the presence of counter demonstrators or police, sponsorship by social movement organizations (SMOs), or the use of sound equipment (Oliver & Maney 2000)." Gitlin (1980) notes that in the late 1960s, the media were more likely to cover events organized by clear and charismatic leaders (Smith et al 2001). The timing of an event, and its relationship to a current debate or issue also affects the likelihood of its coverage. If an issue is part of a current debate, it is more likely to be reported as part of an "issue attention cycle' or "media attention cycle" (McCarthy, McPhail and Smith 1996). Gitlin (1980) noted that reformist movements are more likely to be covered accurately than revolutionary ones, and issues are more likely to be covered than ideologies. Kligo et al. (2019), noted that certain movements are covered more regularly than others. They found that in 16 newspapers, anti-racist struggles were covered less thoroughly than other movements.

Particular types of newspapers are biased in particular ways. Writing in 2000, Oliver and Maney found that local and regional newspapers had more selection biases than national newspapers. They also noted that because protest has become more institutionalized since the 1970s, has become less newsworthy and thus, the rate of coverage has declined. However, looking at 1982 until 1991, McCarthy et al., (1996; 1999) found newspaper biases in the US were relatively stable, and similar findings were found internationally (McCarthy et al. (2008).

It is unclear whether changes to newspapers since 2000 have further changed the level of protest coverage. There has been a great deal of ink and bytes spilled over the decline of the newspaper. Subscriber numbers and print circulation are down, so too are advertising dollars and the number of permanent staff (Pew 2012, Schlesinger and Doyle 2015). Newspapers increasingly rely on wire services for stories (Bouman 2018). However, readership appears to be up, as people access news in new ways, mostly online, as newspapers have become one type of source in a complex, multimedia ecosystem (Black Press 2019).

Description Biases - how protests are covered

The biases in how protesters are described are less well understood (Earl et al., 2004). Past research has shown that newspapers cover protest activity in particular ways. Newspapers tend to cover the hard facts of events (who, what, when, where, why) accurately, or omit them altogether. (ibid,). It appears that given the newsworthiness of repression, arrests are covered accurately, while claims are often omitted (Smith et al., 2001) However, the 'soft news' - the interpretive elements or framing of movements is often more dependent on the journalists framing or the perspective of those interviewed. The need to justify the newsworthiness of coverage may correspond with why protesters are often described as deviant or violent (Gitlin 1980, Kligo and Harlow 2019), and spectacular elements of protest are emphasized. Elements that are not easily observable, such as intention, meaning and motivation are often missing. Overall, the most significant description bias in newspaper coverage is omission, missing data. When compared with observers - found that newspapers covered only most frequent activities within protests (Oliver & Myers 1999).

The coverage itself may be changing. Tanikawa (2017) argues that there is a move away from traditional event-centred news articles, compared with 25 years ago. In competition with online news outlets, newspapers are now serving as analytical and/or in-depth news sources. The effect of these changes on protest coverage have yet to be explored. Nonetheless, the consistency routines of news production means that the biases keep newspaper data reliable and relatively consistent, if limited.

The Dream of Better Sources

Back in 1987, Fransozi noted that, despite the well known biases of newspapers, there was no real alternative. Almost twenty years later, Earl et al., (2004) found that while recognizing the important limitations of description biases for particular tasks, the reliability of the 'hard facts' within newspapers meant that they remained the best data available. However, digital technology has its attractions. Google searches and web scans offer quick insight, but the ever-changing, and unknown algorithms make such searches unsystematic and unreliable. A better option has been to use platforms whose logic is traceable, like Factiva or Nexis Uni. These platforms allow researchers to search hundreds or even thousands of periodicals, television and radio broadcasts, newswires, and blogs alongside newspapers for particular keywords, capturing a broader swath of protest activity (Nam 2006). Machine learning offers the potential to get even more. For thirty years - Kansas Event Data System (KEDS) has used machine assisted approaches to political event data. More recently, Alex Hanna developed Machine Learning Protest Event Data Systems (MPEDS). Such datasets are ambitious, and can capture the ebbs and flows of articles about protest, but to date, they have a more difficult time dealing with repetition, and multiple articles on the same event, making them unwieldy as sources for event catalogues.

Social media sources like Facebook and Twitter offer potential as sources for event catalogues. Twitter is a researcher's favourite. The microblogging site has 330 million monthly active users and 145 million daily active users (Nov 2020), with 7.6 million in Canada. It is most often used to study the patterns of online activity of movements, looking for the ebb and flow of particular issues (frequency) - rise of particular hashtags, #metoo, #blacklivesmatter or mapping the structure of discursive networks. (interrelations). However, there has been less work done on its coverage of offline activity. However, as we will do here, it could be manually examined or digitally 'scraped' or searched and compiled, looking for particular words or hashtags.

While news media coverage is produced by third party observers (journalists) to report on social movement activity for their readership, Tweets are produced by journalists, police and bystanders, but also by social movement participants. Activists tweet to promote their claims, to promote events in advance (Morozov 2009c), to coordinate activity on the ground (Hughes et al., 2010, Earl et al., 2013), to display their alliances, to share event information after events and to discuss political events and campaigns. (Earl et al., 2013; Huang & Sun, 2013; Lotan et al., 2011; Theochardis 2014').'' Twitter allows activists to communicate directly in their own words, sharing their own frames, and details neglected by others. Activists use Twitter to share additional material that they favour or oppose, amplifying sympathetic media coverage, and critique opposing views. These diverse uses means that the content of tweets about protest activity will be less consistent in form. Like newsmedia, tweets may include details about protest activity (actors, tactics, goals, targets, times, dates, and locations). When compiled, the multiple perspectives that compose Twitter could offer an alternative to news media coverage, and provide a richer, multidimensional source for event catalogues. However we do not yet,

understand what this might look like. To this end, this paper seeks to map protest in Toronto during the summer of 2020 using three lenses, that of the Toronto Star, Nexis Uni and Twitter.

The Case

Toronto is the provincial capital of Ontario and the largest city in Canada, and the fourth largest city in North America with over 2,731,571 (2016) in the city proper, and over 6 million in the Greater Toronto Area. It is a diverse city, with over 200 ethnic groups, and over 50% of its population identifying as a visible minority. Its culture reflects its history of British colonialism, its proximity to the US and its many waves of immigration. Unlike many US cities, where 'white flight' transformed the city in the 1960s and 1970s, the core of the city continues to include both single family and apartment units. Since the 1970s, the downtown area has gentrified, pushing newer, poorer and racialized residents to the inner suburbs. Like other global cities, it hosts the financial, real estate, investment, arts and technology sectors, and is relatively politically progressive.

Before the CoVid19 pandemic had become clear, there had been a strong multi-issue, but trade union led movement against the Conservative Provincial government of Doug Ford. The CoVid19 pandemic squelched most public protest in early 2020, but by June, a strong local movement had emerged against anti-Black policing and racism, part of a global response triggered by the police murder of George Floyd in Minneapolis. By July 2020 this wave of protest was slowing, infection numbers were lower, and a range of movements returned to the streets. [2]

Data Collection

A basic shared process was used to construct all three catalogues. For each, the keywords: protest, rallied, demonstrate and Toronto were used to search electronic databases (catalogue 1 & catalogue 2) or scan Twitter (catalogue 3) for material produced on events between July 15th and September 15th. The generated results were then read by the authors, and the results that did not meet the criteria of a protest event are identified and removed from the data set. The cleaned dataset was then manually coded and its relevant information entered into an event catalogue[3]. These are then analyzed in order to better understand the characteristics of the three catalogues as they portrayed protest activity in Toronto between July 15 and Sept 15 2020.

Following Tilly's definition of a contentious event, press conferences and indoor meetings are not included. Each event may include multiple actors, doing different things, including counter-protesting[4]. If coverage involves separate groups in separate locations, even when held simultaneously they are

---

[2] Toronto police reported that 452 demonstrations had occurred in 2020 from January until November 10th, an average of 1.43 events per day (personal communication), which, if averaged, would mean 86 events during the 8 weeks.

[3] Event catalogue columns - event number, date, type, issue 1, issue 2, location 1, location 2, multiple cities? , overall description of type of protest, overall description of claimant, overall description of tactic, size, formation 1, formation 1 tactic, formation 2, formation 2 tactic, formation 3, formation 3 tactic, formation 4, formation 4 tactic, duration, protester violence, target 1, target 2, target 3, counter protest? size of counterprotest, counter-form 1, counter-form 1 tactic, counterform 2, counter-form 2 tactic, describe conflict between sides, police mentioned? name of police, police size, police tactics, charges laid, # arrests, # held, # injured, pepper spray, tear gas, projectiles, fences, other barricades, source 1, source 2, source 3

[4] The definitions of a contentious event are taken from Charles Tilly (2002, 2008, 2013), with the exception of his criterion of a minimum size. These catalogues include protest events of any size.

considered separate events. Multiple articles may contribute to a single event. Reference to events that have not yet happened are not included in the catalogues.

Catalogue 1 - Toronto Star

The Toronto Star was searched using the Nexis Uni and 663 articles were identified, which were coded to become a catalogue of 18 events.

Catalogue 2 - Nexis Uni

From the potential 6000 + sources in Nexis Uni, Toronto protests are covered by 37 sources, in $6,686$ articles. The sources included twelve newspapers, three TV stations, one radio station, and six wire services. The most frequently linked sources were *680 News* (a local radio station); *Global TV*, *Toronto Sun* (tabloid newspaper), and Postmedia (online news). Many of these sources included similar pieces and references. This converted into a catalogue of 40 events. If the events that were covered only by the Toronto Star are removed, this became a catalogue of 34 events.

Catalogue 3 - Twitter

A Python based application scraped posts (or "tweets") from Twitter. This was done bi-weekly to work around data access limitations associated with the free Twitter developer account. The 2433 resultant tweets were automatically uploaded into an excel spreadsheet that included four columns: Twitter handle, location, date and tweet content. Often the tweets retweeted links to articles from twenty other media sources. The most frequent were the *Toronto Star, CP24 (TV) CBC News* tv/radio (10 events), *Citytv* (6 events), *Post-Millenial* (alt-right; 6 events), as well as other television and newspapers, including ethnic media from Guyana Hong Kong and Bulgaria, and youtube, reddit, and organizational websites. The linked pieces are considered part of the 'Twitter data' results and thus included in the catalogue. This produced a catalogue of 80 events, which, when the events covered in the Toronto Star and Nexis Uni were removed, became a catalogue of 50 events.

Method and Findings

Each catalogue was examined to answer the following questions about protest in Toronto between 15 July and 15 September 2020.

- How many protests were there?
- What were the issues being protested?
- What formations were protesting? (organizations, groups, communities)
- What were the targets of the protest?
- Where were the protests located?
- What size were the protests?
- What were the reactions of the police?

In combination, these three catalogues reveal 89 distinct protest events that occurred in Toronto between July 15 and September 15, 2020. Unsurprisingly these different sources reveal different amounts and kinds of protest activity in Toronto in the late summer of 2020.

Number of Events

As Table 1 shows, Twitter data reveals 80 events, far more than the Toronto Star (18) would suggest.

Table 1 - Number of Events

	July 15 - Sept 15	*Proportion of 89 events*
Toronto Star	18	20%
Nexis Uni minus T Star	34	38%
T. Star + Nexis	40	45%
Twitter	80	90%
Twitter minus T. Star and Nexis	50	56%
Total Number of distinct events	89 events	100%

Toronto's Summer of 2020 - The Issues

Reassuringly, all three data sources agreed that the movement against anti-Black racism and racist policing, and the struggles around housing and homelessness were the most visible protests that summer. Beyond these two movements, the different sources revealed distinct pictures of protest in the city.

Catalog 1 -The Toronto Star shows the wave of protest against anti-Black racism and racist policing was absolutely dominant, making up 55% of the protests, There are two other multi-event movements, around homelessness and evictions (3) and labour conditions and contracts in a CoVid context (3). There is coverage of one protest in solidarity with indigenous land defenders, and another in support of migrant workers. No right wing, or diasporic protest appears in this catalogue.

Catalog 2 - Nexis' data These 37 sources (with the Star data removed) also show anti-Black racism and defunding the police as the top issue being protested (10 events). Like the Star coverage, we see struggle around housing and homelessness including three anti-eviction protests, and two protest-counterprotest confrontations around the housing of homeless people in midtown hotels. But unlike the Star coverage, we see protests around immigration (3), human rights in China (3), two anti-mask protests and far more coverage of protests on other issues. These are: animal rights, schools, climate justice, India, drunk driving, democracy in Belorusse, Israel, Pakistan, Nigeria and Mauritius, the Iranian involvement in the shooting down of a jetliner, and the treatment of indigenous land defenders, and indigenous people more generally. Interestingly, we don't see any coverage of labour issues.

Catalog 3 - The Twitter data (and its retweeted media links) confirm the centrality of the cluster of protests around anti-Black racism and policing (14 events), and housing and homelessness (8 events plus 2 anti-homeless/pro safety protests and 1 pro-eviction landlord protest). Like Catalogue 2, the Twitter dataset shows numerous protests against China (8), and immigration (3) - but increases the visibility of anti-mask protests, as well as animal rights protests (3), schools opening (2), drunk driving (1) and gender based violence (1). While catalogue 2 showed diasporic protests from eight communities, Twitter reveals thirteen; including events connected to politics in Belorusse (4), Pakistan/Balochistan (3), Mauritius (1), Iran 2), Israel (1) and Nigeria (1), and covered in Twitter exclusively - protest events around India (2), Bulgaria (1), Japan (1), Azerbaijan (1), Russia (1), Turkey (1), and Guyana (1). Twitter also reveals right wing protests including an anti-sin protest, an anti-Islam protest, and 2 protests against child sex trafficking (Qanon).

What formations were protesting? (formal organizations)

Newsmedia and social media coverage often describe protesters using general categories, like 'protesters', "supporters," "students," or 'demonstrators.' However, if an organization is well known, it is more likely to be named. Given the above accounts, it should not be surprising that Black Lives Matter is the most frequently identified organization in Toronto in all three catalogs during this period. The other named organizations featured in all three catalogues are the Afro-Indigenous Rising Collective, Not Another Black Life and the Canadian Union of Public Employees (Hospital Workers). Sources vary in their consistency and frequency in naming organizations. The Toronto Star, (catalog 1) names an average of one organization per event while Catalogs 2 and 3 name only .76 organization per event. This confirms both journalistic standards, and the research that finds that mainstream newspapers tend to get the basic facts right, even if they select a smaller proportion of the events.[5]

---

[5] Catalog 1- few named organizations, mostly general category of protesters, supporters of BLM, or parents. But the 17 who were named in the 18 events are: Afro-Indigenous Rising Collective (2), Beachers for Black Lives, Black Lives Matter (3), Doctors For Defunding the Police. , Forward Through Progress, Migrant Workers Alliance for Change (Migrant Students United), No Pride in Policing Coalition, Not Another Black Life, Ont Association of Optometrists, Ontario Coalition Against Poverty, Ontario Council of Hospital Unions (Canadian Union of Public Employees - 2), Parents of Black Children. Peoples Defence, Parkdale Organize, South Asian Women's Rights Organization, Toronto East Anti-Hate Mobilization, Vaughan African Canadian Association. (1 org per event)

Catalog 2 - In addition to many categories including ethnic categories, there were 25 named groups. Afro Indigenous Rising (AIR Collective), Association of Families of Flight PS752, Baloch National Movement, Bangladeshi Migrant Rights Alliance, Belarussian Canadian Alliance, Black Flag, Black Lives Matter, Canada Pashtun Council, Canada-Hong Kong Link, Federation of Metro Tenants, Goodwood Tenants Union, Keep Your Rent Toronto, Migrant Rights Network, Migrant Worker Alliance of Canada, Not Another Black Life (3), Ontario Council of Hospital Unions, a division of the Canadian Union of Public Employees (2), Pashtun Tahaffuz Movement, Philadelphia Flyers, Sikh's for Justice, South Asian Women's Rights Organization, Support Center for Azerbaijani Women, Toronto Pig Save, World Sindhi Congress and Peoples Defence. 25 in 39 events; ie. (.64 group per event).
Catalog 3

The distinctive function of news media sources vs. social media data reveal themselves in the descriptive biases. While most Twitter coverage retweets news media sources, the tweets themselves rarely name organizations, instead relying on photos, or video coverage of the events and sometimes using hashtags associated with a particular campaign. Almost 70% of the tweets that did not link to a secondary source, did not name any organizations. This difference in the content, may also be tied to the combination of particular media and this search strategy. Because all three sources used the same terms, and relied on the use of 'protest', 'rallied' or 'demonstrate', Twitter data may be less likely to combine these terms with the names of organizations than other sources.

Targets

In the summer of 2020, to whom are protesters directing their claims? Each catalogue revealed a different set of targets[6]. Although many targets were shared between catalogues, each data set contains a different distribution of targets, as well as targets not identified elsewhere. Comparison was based on the proportion of events that contained a given target. Events with multiple targets in the same category (ex. two police targets) are only counted once.

Catalogue 1 - The Toronto Star identified 11 different targets, 2 (18%~) of which were not identified in the other catalogues. The vast majority (9/11, 81%) of identified targets within this catalogue were government targets, all of which were domestic government targets. Coverage predominantly focused on events with police targets, which were identified in 10 (56%) events. The next most named targets were the municipal government (7 times, 39%), followed by the provincial government (5 times, 28%). at least one form of domestic government target was found in every event in the catalogue, with the federal government being named only once.This catalogue continues the previously observed trend of the Toronto Star covering, which found that reported protests focused on public and private targets predominantly at the municipal and provincial level.

---

51 named organizations in the 86 events - ACORN (2), AIR Collective (2), Anakbayan Toronto, Animal Save, Baloch National Movement, Bangladesh Minority Rights Alliance, Beachers for Black Lives, BikePOC, Black Lives Matter (3), Black Players Coalition of MLS, Canada Pashtun Council, Canada-Hong Kong Link (3), Christs Forgiveness Ministries, CIL Canada, Direct Action Everywhere, Falun Gong, FMTA, Forward Through PRogress (2), Fur Free Toronto, Guyana National Democratic Council, Hindu Forum Canada, Hugs over Masks (2), Jane Finch Action Against Poverty, Japanese Canadians for Social Justice, Keep Your Rent Toronto, Kurdish Popular Council in Toronto, Migrant Students United, Mt. Pleasant/Eglinton Community Group, MWAC (2), Not Another Black Life (3), Older Women's Network, Ontario Council of Hospital Unions, Ontario Parent Action Network, Operation Underground RAilroad, Parkdale Organize, Pashtun Tahaffuz Movement and World Sindhi congress, PEGIDA Canada, Peoples Defence TO, PMLN Canada, PMLN Youth, POOF, Rising Tide, Sikh's for Justice, Stand With Hong Kong, SURJ Toronto, TFC BIPOC Fan Coalition, Toronto Football Club, Toronto Harm Reduction Network, U of T HK Extradition Law Awareness Group, UFCW. .59 groups per event.

[6] Targets are institutions, organizations or individuals. Of course, ideas and practices are also targeted, but those are excluded from this data, as per standard event catalogue practice.

Catalogue 2 - This catalogue identified 21 different targets, 7 (33%) of which were not identified in the other catalogues. There was a fair amount of diversity in the kinds of targets identified. It favored events targeting the provincial government which was featured in 11/21 total events (52%). Police were the next most prominent target (7, 33%), followed by foreign governments, Municipal government, and the Federal government all with 6 events each (28%). This catalogue featured two events (10%) targeting non-federal government international groups (in this case Armenian citizens and the UN). Contrasting from Catalogue 1, there was a large number (8 events, 38%) of events targeting specific political figures. The coverage only slightly favored domestic political figures with 4 targets being foreign and 5 being domestic[7]. This catalogue is similar to the previous one in its overall focus on protests with domestic targets, however it includes significantly more foreign targets, as well as more individually named targets.

Catalogue 3 - This, largest, catalogue identifies 38 targets, 25 (66%) of which were not identified in other catalogues. Overall, this catalogue had the greatest diversity in targets. What is striking is, that in these 80 events, foreign governments were identified most often. 26 events (32.5%) targeted at, least one foreign government, with the Chinese government being named most often[8]. The provincial government was the next most targeted, being named in 20 events (25%). The federal government was identified 16 times (20%), followed by the police with 10 events (12.5%). Politicians were the last group to be widely targeted, with a total of 8 events (10%). 7 of the events targeting politicians targeted domestic politicians, and only 1 event targeted a foreign politician. The municipal government was the last widely named target, with 7 events (8.75%). The remaining targets can be described as non-government international targets (2 times, 2.5%), supremacist ideologies (2 times, 2.5%), and 7 others that were each only identified once each.

All three identify unique targets, but Twitter and Nexis data draw from multiple perspectives and include the greatest diversity of protest targets. Provincial government and police targets were the most consistently named in all three catalogues. Given the significance of the movement against anti-Black policing, and an unpopular Conservative provincial government, this is not a surprise. The Toronto Star had overwhelmingly local targets, the majority of which are municipal government, with the rest being private organizations or groups. Twitter and Nexis Uni differ most from the Toronto Star in their inclusion of international targets.[9] These sources include content from international newsmedia, helping to explain this difference. The way targets were identified also varied between catalogues. The Toronto Star and to a lesser extent Twitter, when naming a government target tended to name organizations as a whole rather than specific ministries or individuals. This contrasts with Nexis Uni's inclusion of partisan data which tended to include the names of specific political figures.

Other differences and similarities in protest coverage
All three catalogues included the same information on protester arrests. However, there were differences in the inclusion of size and location data. Catalogue 1 from the Toronto Star does have more complete data on the size of protest events, 83% of the Toronto Star events (Catalog 1) included information on the

---

[7] One of the events targeting political figures named two different individuals. For the domestic foerign breakdown they are counted twice, but for the total number of events focusing on political figures they are counted once. This is why there is a discrepancy between the number of political figures in the breakdown versus the total.

[8] China being identified 10 times, with the next closest being Belorouse and India both being identified 3 times.

[9] for example, Muslims and sinners were both named as the targets of protests found on twitter.

size of protest, whereas less than 35% of the other catalogues did so. This catalogue also included protests across the city, challenging expectations that Twitter's decentralized structure might cover suburban events more effectively.[10]

Data Analysis

Selection bias

The three catalogues paint very different pictures of protest. Each show the struggle against anti-Black racism and to defund the police manifested in Toronto during the summer of 2020, as it did elsewhere in North America. They also show the struggles against evictions and homelessness, exacerbated by CoVid 19. However, beyond these two movements, the catalogues of protest reveal stark differences. The Toronto Star shows us a small, relatively contained field of protest, with named organizations, a strong labour movement, with protest directed primarily to provincial government or police targets.

In contrast, Nexis shows more movements, and movements that are not as tied to the provincial or local government. Diasporic communities play a larger role, and right wing protest is visible. These features are even stronger in the version of Toronto protest revealed by Twitter. Here, protest is even more multi-ethinc and more conflictual than the other versions. Twitter shows that right wing movements, diasporic communities and relatedly, those targeting foreign governments are part of the mix. In the Twitter data, diasporic protest makes up 44% of protest events. These events mostly involve a single community, but they also coordinate. Most notably during this period are the protests against China, which bring together people from Hong Kong, with Indians, Iranians, Tibetans, Taiwanese, Pakistanis and Bangladeshis.

Description bias

Although the overlap and interplay between different sources means that the differences between the catalogues are less than might be expected, we see that the more consistent style of journalism within the Toronto Star coverage offers the most detail, naming organizations, targets, size, targets and location more consistently, and doing so with a more consistent tone. This catalogue shows protest as more state-centred, more routine and more organized than it does through the catalogues based on Nexis or Twitter.

This becomes apparent if we look at the coverage of a single protest event on 23 July 2020 that was included in all three catalogues. The Toronto Star described it in an article entitled, "Groundbreaking ceremony disrupted by protesters targeting Mayor John Tory." The article names the organizations, the target and the demands of the campaign. The police are not mentioned. In contrast, Catalog 2's event entry incorporates data from five articles, two from the Toronto Sun, another from the Ottawa Sun, one in Gates of Vienna newstext blog and a fourth from Global News (TV). Through these multiple lenses, protest appears violent and chaotic. The Sun and the Gates of Vienna emphasize the attempts of the police to stop the protesters, and give coverage to the Mayor's perspective on the protest. Like the Toronto Star, the Sun articles include tweets from the activists, but don't interview them directly, nor include their

---

[10] Cat 1 - T. Star - 18 events, 7 downtown (39%), 3 west end, 2 midtown, 2 east end. 2 ontario wide, 1 north york; Cat 2 - 21 downtown (75%), 1 west end, 3 midtown, 3 east end; Cat 3 - 71 locations listed, 48 downtown (68%), 5 west end, 5 east, 4 midtown.

demands. The Twitter data in catalogue 3 incorporates information from two sympathetic tweets, and their associated links. The first is from @NarcityCanada, a Canadian online media source established in 2015 and aimed at millennials. It shares an article entitled, "Tory Claps Back At Protestors After They Stormed His Toronto Event Appearance (VIDEO)" that itself includes retweets of an activist press release from tenant rights protestors from Parkdale Organize and People's Defence Toronto responding to passage of Ontario Bill 184 (Kundacina 2020). The second tweet from @bay_wop says, "Tory needs to go. Ford needs to go. All the homeless people in Toronto and those facing eviction should go back to Tory's multi million dollar condo and protest there and DON'T ALLOW HIM IN until he grows a set of nuts and takes a stand against homelessness," and then retweets a CP24 television piece entitled "Police escort mayor into waiting vehicle after protesters concerned about evictions interrupt groundbreaking ceremony (Fox 2020)." This coverage doesn't name the organizations and has a lengthy interview with the mayor who argues that these protesters are not trustworthy. Each of these tweets tells a different story. The first, leverages the perspective of the activists, the second one is contradictory, the tweet supports the perspective of the activists, but recirculates footage that is less supportive.

The portrayals of this event using three different sources illustrate how difficult it is to speak about distinct sources and their biases, given how much these media are interwoven (see also Mare 2013). Although this integration muddies the difference between particular sources, it also ensures that the basic facts of the events can be accessed using different routes. However, it also makes it more difficult to untangle the ways that biases are influencing the coverage. This protest event is described quite differently. One catalogue can show it as an organized intervention within a larger campaign, while another shows it as a violent disruption.

This example cannot fully account for the different descriptive biases of each catalog, it reveals the challenges of trying to synthesize multiple voices in order to represent a protest event or protest activity more generally.

Observations

Protests in Toronto look very different through each of these catalogues and these differences matter. The newspaper data from the Toronto Star in Catalogue One is the most consistent and allows us to track and compare changes in actors, actions and relationships. However, its dependence on a small number of news stories, and its emphasis on formal organizations reinforce a national, state-centred perspective that marginalizes other types of protest activity. Its neglect of diasporic and right wing protest activity mean that research that uses it as a source may miss important elements of political life. Catalogue Two's Nexis Uni compilation includes more varied and partisan voices, some of which follow journalistic norms of balance, and detail while others who do not. It also includes international, national and local media - from television to blog postings. The combination of sources, each with their own biases make such compilations inconsistent with unknown representativeness (Jenkins and Maher 2016). The catalogue shows a more polarized and fragmented polity, with numerous disconnected issues, and movements. Twitter's accessibility for activists shapes its content, as users can retweet other content to amplify their claim. This catalogue shows a globalized, multiethnic polity of disconnected communities. However, although the Twitter coverage includes more and different events, it is unclear

how to use it most effectively as a datasource. It is possible that the search terms of 'protest' and 'demonstrate' may only capture a fraction of the protest coverage on Twitter, given that these terms are rooted in the third-person perspective of the mainstream media. In contrast, Tweets by organizers or participants tend not to use the language of 'protesters' but either the second person 'we' perspective in situ or a more specific (and strategic) identity like 'tenants' or 'students' or 'family members'.

Conclusions

The summer of 2020 will be remembered for the powerful movement against anti-Black racism and policing. But sources matter. Toronto's streets also saw a strong housing justice movement. Beyond that the different sources painted different pictures. The best source for a protest event catalogues depends on the goal of the research. If one seeks to understand the changes or comparisons of a particular social movement, particularly a movement that makes demands on government authorities, or to understand changing tactics, formal organizations and campaigns, newspaper data retains its gold-star status. However, if a project is trying to represent the breadth of protest activity in a place and time, or seeks to include informal or temporary formations, interactions amongst left and right, and the transnational dimensions of protest, Twitter, and to a lesser extent, Nexis Uni have potential. They allow us to utilize the benefits of online data, to trace offline movements. However, their biases need to be better understood. Each of these sources has its limits, and understanding them will allow us to better understand the past and future of patterns of protest.

References

Almeida, Paul, and Mark Lichbach. "To the Internet, from the Internet: Comparative media coverage of transnational protests." *Mobilization: An International Quarterly* 8, no. 3 (2003): 249-272.

Becker, Hila, Mor Naaman, and Luis Gravano. "Beyond trending topics: Real-world event identification on twitter." In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1. 2011.

Barranco, José, and Dominique Wisler. "Validity and systematicity of newspaper data in event analysis." *European sociological review* 15, no. 3 (1999): 301-322.

Biggs, Michael. "Size matters: Quantifying protest by counting participants." *Sociological Methods & Research* 47, no. 3 (2018): 351-383.

Black Press Media Staff. "Don't stop the presses: Report finds Canadian newspaper readership at all-time high." https://www.vicnews.com/business/dont-stop-the-presses-report-finds-canadian-newspaper-readership-at-all-time-high/ (2019)

BOUMANS, Jelle et al. The Agency Makes the (Online) News World go Round: The Impact of News Agency Content on Print and Online News. **International Journal of Communication**, [S.l.], v. 12, p. 22, apr. 2018. ISSN 1932-8036. Available at: <https://ijoc.org/index.php/ijoc/article/view/7109>. Date accessed: 19 Feb. 2021.

Earl, Jennifer, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan, and Ashley Arlotti. "This protest will be tweeted: Twitter and protest policing during the Pittsburgh G20." *Information, communication & society* 16, no. 4 (2013): 459-478.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. "The use of newspaper data in the study of collective action." *Annu. Rev. Sociol.* 30 (2004): 65-80.

Fisher, Dana R., Kenneth T. Andrews, Neal Caren, Erica Chenoweth, Michael T. Heaney, Tommy Leung, L. Nathan Perkins, and Jeremy Pressman. "The science of contemporary street protest: New efforts in the United States." *Science advances* 5, no. 10 (2019): eaaw5461.

Fox, Chris. "Police escort mayor into waiting vehicle after protesters concerned about evictions interrupt groundbreaking ceremony." CP24.com 23 July 2020.

"https://www.cp24.com/news/police-escort-mayor-into-waiting-vehicle-after-protesters-concerned-about-evictions-interrupt-groundbreaking-ceremony-1.5035950"

Grieco, Elizabeth. "U.S. Newspapers Have Shed Half of Their Newsroom Employees since 2008." Pew Research Center. Pew Research Center,. https://www.pewresearch.org/fact-tank/2020/04/20/u-s-newsroom-employment-has-dropped-by-a-quarter-since-2008/. (2020)

Hanna, Alex. "Mpeds: Automating the generation of protest event data." (2017).

Herkenrath, Mark, and Alex Knoll. "Protest events in international press coverage: An empirical critique of cross-national conflict databases." *International Journal of Comparative Sociology* 52, no. 3 (2011): 163-180.

Hug, Simon, and Dominique Wisler. "Correcting for selection bias in social movement research." *Mobilization: An International Quarterly* 3, no. 2 (1998): 141-161.

Jenkins, J. Craig, and Thomas V. Maher. "What should we do about source selection in event data? Challenges, progress, and possible solutions." *International Journal of Sociology* 46, no. 1 (2016): 42-57.

Kilgo, Danielle K., and Summer Harlow. "Protests, media coverage, and a hierarchy of social struggle." *The International Journal of Press/Politics* 24, no. 4 (2019): 508-530.

Kundacina, Anja "Tory Claps Back At Protestors After They Stormed His Toronto Event Appearance (VIDEO)" *Narcity* 23 July 2020. https://www.narcity.com/en-ca/news/toronto/covid19-evictions-in-toronto-protest-had-john-tory-clapping-back-at-demonstrators

Leetaru, Kalev. "The possibility of global data sets." *Journal of International Affairs* 68, no. 1 (2014): 215-219.

Mare, Admire. "A complicated but symbiotic affair: The relationship between mainstream media and social media in the coverage of social protests in southern Africa." *Ecquid Novi: African Journalism Studies* 34, no. 1 (2013): 83-98.

McAdam, Doug, Sidney Tarrow and Charles Tilly. *Dynamics of Contention.* Cambridge University Press: New York, 2001.

McCarthy, John D., Clark McPhail, and Jackie Smith. "Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991." *American sociological review* (1996): 478-499.

Nam, Taehyun. "What you use matters: Coding protest data." *PS: Political Science and Politics* 39, no. 2 (2006): 281-287.

O'Brien, Thomas. *Using Event Analysis to Examine Environmental Protest*. SAGE Publications Ltd, 2017.

Oliver, Pamela E., and Gregory M. Maney. "Political processes and local newspaper coverage of protest events: From selection bias to triadic interactions." *American Journal of Sociology* 106, no. 2 (2000): 463-505.

Oliver, Pamela E., and Daniel J. Myers. "How events enter the public sphere: Conflict, location, and sponsorship in local newspaper coverage of public events." *American journal of sociology* 105, no. 1 (1999): 38-87.

Olzak, Susan. "Analysis of Events in the Study of Collective Action." *Annual Review of Sociology* 15 (1989): 119-41. Accessed February 9, 2021. http://www.jstor.org/stable/2083221.

Rubado, Meghan E., and Jay T. Jennings. "Political consequences of the endangered local watchdog: newspaper decline and Mayoral elections in the United States." *Urban Affairs Review* 56, no. 5 (2020): 1327-1356.

Schlesinger, Philip, and Gillian Doyle. "From organizational crisis to multi-platform salvation? Creative destruction and the recomposition of news media." *Journalism* 16, no. 3 (2015): 305-323.

Smith, Jackie, John D. McCarthy, Clark McPhail, and Boguslaw Augustyn. "From protest to agenda building: Description bias in media coverage of protest events in Washington, DC." *Social Forces* 79, no. 4 (2001): 1397-1423.

Steinert-Threlkeld, Zachary C. "The Future of Event Data Is Images." *Sociological Methodology* 49, no. 1 (2019): 68-75.

Tanikawa, Miki. "What is news? What is the newspaper? The physical, functional, and stylistic transformation of print newspapers, 1988–2013." *International Journal of Communication* 11 (2017): 22.

Theocharis, Yannis, Will Lowe, Jan W. Van Deth, and Gema García-Albacete. "Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements." *Information, Communication & Society* 18, no. 2 (2015): 202-220.

Tilly, Charles. *The contentious french*. Harvard University Press, 2013.

Tilly, Charles. "Describing, measuring, and explaining struggle." *Qualitative Sociology* 31, no. 1 (2008): 1-13.

Tilly, Charles. "Event catalogs as theories." *Sociological Theory* 20, no. 2 (2002): 248-254.