# Demanding justice in the Cloud: An analysis of punitive attitudes in social media for traditional and cyber-enabled crime.[[1]](#footnote-1)

Citizens' attitudes to crime and punishment significantly impact the functioning of the criminal justice system and crime prevention policies. Traditionally, such research has relied on data from public opinion surveys, which is limited by their high cost and cross-sectional nature. Funded by the seed corn funding of the Centre for Digital Trust and Society at the University of Manchester this article delves into the potential of utilising New and Emerging Forms of Data (NEFD) to understand punitive attitudes in social media. Collecting crime-related news stories from the mainstream media in the UK and analysing users' responses, we show that the discussion about punishment and punitiveness occupies a significant proportion of the conversations surronding crime.

## Introduction

For some time now, Social Sciences have been studying punitive attitudes, which refer to perceptions of what kind of punishment we believe is legitimate and valid to apply when a crime occurs. Research has shown that citizens' opinions and preferences for punishment carry serious consequences. For example, they can influence decisions made by politicians and governments regarding the design and implementation of crime prevention policies that are ineffective or even counterproductive (Enns 2014; Jennings et al. 2017) and even highly costly in economic terms (Jaitman et al. 2017; Welsh 2018).

In the UK, even more than 50 years after the abolition of the death penalty, it continues to be supported by 40% of the population (Kirk 2022a). When considering specific crimes (murder of a child, acts of terrorism, or multiple murders), the death penalty is supported by most of the population (ibid). Furthermore, approximately two-thirds of Britons believe that the sentences received by offenders convicted of crimes are not sufficiently severe (Kirk 2022b). This prevailing public opinion regarding punishment takes place in a context of growing economic uncertainty (KPMG 2022) and an increase in violence (pre-pandemic) that had been steadily declining since the mid-1990s (Office for National Statistics, 2021; Office for National Statistics, 2022). Furthermore, the recent covid-19 pandemic and subsequent lockdown measures affected social interactions and criminal strategies in particular, multiplying cyber-dependent and cyber-enabled crime (Buil-Gil et al. 2021), raising the question about the effects of this type of crime on citizens' attitudes.

But what are these punitive attitudes and how can we gain insights about them? Over the past three or four decades, criminologists have tried to address these two questions by relying on conventional methods generally based on scales and surveys. What are the limits associated with using these data sources? What other alternatives are available, and what possibilities and limitations do they present? This article aims to provide a brief overview of these non-traditional data alternatives and the so-called computational criminology field. Along the way, it aims to showcase some initial empirical findings of the project we developed in partnership between the University of Manchester and Cardiff University.

## "Hang him, then deport the whole family"

After recognising the importance of citizens' punitive attitudes, a key question is how to observe them effectively. Initially, studies only included very general measures of attitudes towards sentencing and punishment, usually based on a single survey question with minimal contextual information (for example referring to support for the death penalty or the general increase in sentences). This type of design tends to overestimate the punitive attitudes of individuals, as it fails to provide information regarding other types of measures (such as rehabilitative or restorative ones) and does not consider the characteristics of the crime, the victim or the offender (Stalans, 2013).

Over time, measurement strategies became more sophisticated. For example, the use of vignettes has been increasingly adopted in the field, allowing not only to control the characteristics of the crime, but also to manipulate them within experimental contexts (Socia et al. 2019; Horstman et al. 2021). Also, batteries of questions have been applied to enhance the validity of the measures of punitive attitudes through more sophisticated psychometric analysis (Maguire and Johnson 2015; Armborst 2017). Despite these advances, measuring punitive attitudes in this way faces some challenges.

The first problem with these alternatives is the limited flexibility of scales applied in surveys which often fail to capture individuals' opinions, moods and emotional reactions towards punishment validly and reliably (Smith et al. 2017). When respondents are required to express their opinion on sensitive or controversial topics like punishment, there is a social desirability bias and in many cases the response is tailored to what the interviewer or the community expects (Stephens-Davidowitz 2018). The second problem is that these surveys involve, at best, a static and rather costly to implement snapshot that cannot capture the real-time dynamics and volatility of citizen opinion over weeks and months (Williams et al. 2020). Finally, the third problem is that survey-based measures of punitiveness are focused on atomised individuals and thus hardly capture existing networks of communication, exchange and discussion (Edelmann et al. 2020).

So, what is the solution? This is where computational criminology steps in to complement and tackle the limitations of these conventional measures through the use of New and Emerging Forms of Data (NEFD). What exactly are these NEFDs? Basically, it is a grandiose name for multimedia data digitally captured through various types of digital connections (Radanliev and De Roure 2023). The expansion of personal Internet usage and its application to multiple kinds of processes changed how individuals inform and communicate, generating an exponential growth of available data on online behaviours that greatly interest social sciences (Edelmann et al. 2020; Williams et al. 2020). NEFDs provide detailed information on the development of social relationships in large populations as they occur, unlike traditional datasets typically collected by social scientists (Hofman et al. 2021; Radanliev and De Roure 2023). However, it is only recently that criminologists have begun to exploit these data and start studying phenomena such as social disorder, policing, online crime, or racial hate crimes (O’brien et al. 2015; Smith et al. 2017; Lynch 2018; Williams et al. 2020). As far as we know, no one has ventured into investigating punitive attitudes in this way. In this blog note, we want to outline the potential of this type of data and the associated techniques drawing on the initial outcomes of the study we are carrying out in the United Kingdom.

## The black box

In order to analyse punitive attitudes on Twitter, the following data was collected: First, 20 news portals with a presence on Twitter were selected based on their relevance according to public opinion polls and number of followers. From these 20 news portals, all tweets related to crime were downloaded through a set of keywords (e.g., "homicide," "assault," "robbery," "robbery," etc.) through the year 2022. These keywords were derived from exploratory analysis of news media. All quotes or responses made by users to these news items were downloaded together with the originals tweets. This latter dataset holds relevance for our analysis, as it groups user reactions to crime news. This dataset is anticipated to yield punitive reactions, which, in one way or another, call for an increase in the cost of crime (i.e., that committing a crime should have higher costs for offenders through harsher penalties).

After acquiring this data and following the removal of spam and bots, the first step involves attempting to summarise this information. Qualitative analysis of every tweet would be a daunting task, and here is where Topic Modeling strategies prove invaluable. But what exactly is Topic Modeling? Topic Modeling is an analysis technique in the Machine Learning (or Artificial Intelligence - AI) toolkit. This analysis technique is based on a series of sequential steps or instructions where the machine will try to optimise a specific task by adjusting values within a statistical model based on the variation in an indicator which is iteratively employed as a quality criterion (akin to what can be called an "algorithm" in the context of data science). More specifically, topic modelling constitutes a non-supervised learning technique. Unlike in a supervised classification task, we are not giving examples to the algorithm regarding the expected outcome, - but instead, we are giving it data and a goal. In this case, the purpose is categorising (or clustering) the textual information. How is this accomplished? By identifying words that commonly co-occur within documents (in this context, tweets) and that do not tend to occur in other contexts (Blei et al. 2003; Blei 2012).

Why are we doing this? By employing this approach we can accomplish the coding of our documents of interest in an automatic and inductive maner (Mohr and Bogdanov 2013). By detecting these words that distinguish discussion topics, we can identify the main themes of naturally occurring communications within social media, specifically in the "red chronicle" context. Thus, we can answer how crime is discussed in social media, especially to what extent punishment and punitiveness are emerging concerns. All this analysis, in addition, is executed through a beautiful, replicable, automated, and cost-free software package.

If all this seems confusing, don't abandon us (yet): many of these points will probably become clearer with analysis we present next.

## Topic 11

Following the preceding stage, we embarked on the analysis phase. Based on a series of quantitative indicators, a model with 15 topics was chosen. The distribution of these topics and their keywords are presented below (Figure I). For the sake of consistency and to avoid counting different words that refer to the same concept, words were trimmed to their root forms (e.g.: "criminal" and "criminals" are included in the model as "crimin").

Figure 1 Keywords by topic



Figure 2 Topic proportion based on the classification of tweets



How should we interpret these figures? The first graph illustrates the most important words for each topic - those whose presence most strongly predicts the likelihood of the tweet being part of that specific topic. Meanwhile, the second graph (Figure II) shows the distribution of tweets associated with each of the topics. Delving into the intricacies of each topic is beyond the purview of this blog note (and perhaps exceeds what the audience is inclined to read). We will focus on one topic in particular, namely topic 11, which appears to support the initial intuition: punishment is relevant to users' reaction to news about crime on Twitter. And there is more evidence that bolsters this assertion:

First, we looked at the 10 keywords and expanded this inquiry to encompass up to 30. In addition, we adjusted the criteria for selecting the words to analyse those pertinent to each topic (as illustrated in Figures 3 and 4). We again find that a substantial number of these central words seem to be associated with the idea of punishment (e.g.: "sentenc", "prison", "justic") and anger in response to criminal activities (e.g.: "evil", "fuck", "sad").

Figure 3 Top 30 most important words for Topic 11



Figure 4 Top 30 most important words for Topic 11 (relevance adjusted)



Secondly, we engaged in the qualitative analysis (less grandiloquently put: we read and interpreted) of those tweets classified in this topic. As an illustration, let's consider a random sample of 10 tweets (mentions of users - @ are omitted):

* *People don't get it... neither does Amanda*
* *He will get everything that coming to him in jail , believe that ... he will suffer the rest of his life ...*
* *what sort of justice is this?*
* *Just a small shove to remove someone that was intent on causing a breach of the peace.*
* *Jeasus wept! The cretins the parasites!*
* *Finally faced justice, now Harry's family can have some sort of closure.*
* *I hope he has a shank as he will get shagged in jail*
* *There are some horrible people around,l hope they get caught.*
* *Some ugandan citizens suffer at the hands of these human rights abusers, and have cried to all responsible bodies to raise their voices but nothing is done. Imagine paying taxes to get such a treatment. #UgandaIsBleeding #VisitUganda*
* *Bring back Death penalty, enough is enough. Cases like this should be automatic.*

A few elements to consider from this preliminary analysis: firstly, it becomes evident that Topic Modeling's classification of punitive tweets is not flawless, as the punitive tweets appear to be semantically connected to other types of expressions associated with the implementation of justice. However, this topic does seem to accurately encapsulate the discussion and discourses surrounding the punishment of offenders, although it does not consistently imply punitive attitudes. Secondly, retributive components are observed in the punitive expressions - indicating the desire for the offender to suffer as a consequence of having committed a crime. This attribute seems to be a distinctive feature of the British case when compared with the Uruguayan scenario, where punitive expressions seem to be aligned with more emotional reactions without a clear indication or articulation of the objectives of punishment. Third, it is noteworthy that this discussion accounts for slightly less than 13% of the debate, being one of the most relevant topics around criminal news.

## Users and time

The analysis extends beyond the topic description and can use the information surrounding the tweets (e.g., user details, the date and the characteristics of the original news, etc.) to delve into their temporal dynamics.

A distinctive feature of the expression of punitive attitudes within social media is their variability over time. Analysing data regarading punitive attitudes coming from YouGov[[2]](#footnote-2), we observe relative stability across surveys, a trend that diverges from the more fluctuating nature of punitive expressions on Twitter. Of course, it is crucial to acknowledge that these are two radically different populations (as Twitter users are not a representative sample of the United Kingdom, let alone those who follow press portals, and even more so, those who actively participate in online discussions). Nonetheless, beyond the relatively small size of this group, it is also inherently influential, and we do not know precisely how this noise affects society and institutions. The relative stability of attitudes reported by surveys should not lead us to conclude that the system is impervious to the influence stemming from specific mobilisation occurring in social media triggered by significant events such as violent or sexual crimes.

Figure 5 Top 5 topics by date



Conversely, at the user level, slightly over 20% of the analysed users (those who comment on these news items) participate in the discussion regarding punishment, and within this group, the most active 5% contribute to 22% of the responses. Punitive reactions are not evenly distributed across the different user profiles. Who are the most punitive? What attributes define them? What insights can we learn from their behaviours? These questions also seek to be addressed in subsequent stages of this project.

Additionally, as anticipated, this kind of reaction seems to intensify in response to news linked to violent crimes such as homicides, murders, and rapes. This alignment mirrors findings from the conventional literature on punitive attitudes (McCorkle 1993), and while seemingly evident, it offers a promising indicator of concurrent validity. In the future, trying to understand in greater detail what types of crimes generate particularly punitive and violent reactions in the public could yield exciting insights.

Finally, it is worth noting that this information aligns to a considerable extent with the analysis carried out in a vastly different context, as it is the case of Uruguay (Ezquerra forthcoming[[3]](#footnote-3); Ezquerra and Trajtenberg 2023).

## Virtual tridents and torches

This blog note is a first step, somewhat timid in terms of analysis. But we also believe that it is promising. This first intuition, which seems to be shared by many of us watching what happens on social media regarding public opinion and crime, appears to be supported by the data. Twitter - in this case - is a means of emergence, communication and dissemination of punitive attitudes. This result opens a new field for the analysis of the problem. This data, in interaction with traditional data, can help us understand elements that until now have been very difficult or costly to measure: for example, the effect of the networks with which individuals interact on punitive attitudes, the impact of specific events on attitudes, the emotional characteristics linked to the desire for punishment, and we could go on.

In this project, carried out at the University of Manchester and Cardiff University, we have also developed an automatic classifier for detecting punitive attitudes with sufficient accuracy to track such communications on a large scale. In future publications, the focus will be on understanding who produces such communications, in reaction to what, and whether these fluctuations affect the overall functioning of the system. This initial article aims to show the tools' potential and provide a general overview of the underlying problem.

The punitive attitudes of individuals can have perverse effects on the system's functioning. Computational criminology has the potential to be a tool that allows the monitoring of public opinion and a better understanding of the phenomenon at a low cost. These techniques can be used by society, the political system, and academia to avoid punitive overreaches that may be not only cruel or unjust (something that may not convince many) but also ineffective and undermine the legitimacy of the institutions.

## References

Armborst, A. 2017. How fear of crime affects punitive attitudes. *European Journal on Criminal Policy and Research* 23(3), pp. 461–481. doi: 10.1007/s10610-017-9342-5.

Blei, D.M. 2012. Probabilistic topic models. *Communications of the ACM* 55(4), pp. 77–84.

Blei, D.M., Ng, A.Y. and Jordan, M.I. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), pp. 993–1022.

Buil-Gil, D., Miró-Llinares, F., Moneva, A., Kemp, S. and Díaz-Castaño, N. 2021. Cybercrime and shifts in opportunities during COVID-19: a preliminary analysis in the UK. *European Societies* 23(sup1), pp. S47–S59. doi: 10.1080/14616696.2020.1804973.

Edelmann, A., Wolff, T., Montagne, D. and Bail, C.A. 2020. Computational social science and sociology. *Annual Review of Sociology* 46, pp. 61–81.

Enns, P.K. 2014. The Public’s Increasing Punitiveness and Its Influence on Mass Incarceration in the United States. *American Journal of Political Science* 58(4), pp. 857–872. doi: https://doi.org/10.1111/ajps.12098.

Ezquerra, P. and Trajtenberg, N. 2023. *Usando datos no tradicionales para explorar un concepto tradicional: como podemos usar Twitter para entender la discusion sobre castigo en la opinión pública en Uruguay*. Available at: http://www.razonesypersonas.com/2023/05/usando-datos-no-tradicionales-para.html.

Hofman, J.M. et al. 2021. Integrating explanation and prediction in computational social science. *Nature* 595(7866), pp. 181–188. doi: 10.1038/s41586-021-03659-0.

Horstman, N.J., Bond, C.E. and Eriksson, L. 2021. Sentencing domestic violence offenders: A vignette study of public perceptions. *Journal of interpersonal violence* 36(21–22), pp. NP11916–NP11939.

Jaitman, L. et al. 2017. The Costs of Crime and Violence: New Evidence and Insights in Latin America and the Caribbean. Available at: https://publications.iadb.org/en/costs-crime-and-violence-new-evidence-and-insights-latin-america-and-caribbean [Accessed: 11 May 2023].

Jennings, W., Farrall, S., Gray, E. and Hay, C. 2017. Penal Populism and the Public Thermostat: Crime, Public Punitiveness, and Public Policy. *Governance* 30(3), pp. 463–481. doi: 10.1111/gove.12214.

Kirk, I. 2022a. *Britons don’t tend to support the death penalty… until you name the worst crimes | YouGov*. Available at: https://yougov.co.uk/topics/politics/articles-reports/2022/03/30/britons-dont-tend-support-death-penalty-until-you- [Accessed: 12 October 2022].

Kirk, I. 2022b. *Criminal sentencing is too soft, say two-thirds of Britons | YouGov*. Available at: https://yougov.co.uk/topics/politics/articles-reports/2022/03/30/criminal-sentencing-too-soft-say-two-thirds-briton [Accessed: 12 October 2022].

KPMG. 2022. *UK economy marred by uncertainty while high inflation takes its toll on global growth prospects - KPMG United Kingdom*. Available at: https://home.kpmg/uk/en/home/media/press-releases/2022/09/uk-economy-marred-by-uncertainty-while-high-inflation.html [Accessed: 12 October 2022].

Lynch, J. 2018. Not even our own facts: Criminology in the era of big data. *Criminology* 56(3), pp. 437–454.

Maguire, E. and Johnson, D. 2015. The structure of public opinion on crime policy: Evidence from seven Caribbean nations. *Punishment & Society* 17(4), pp. 502–530. doi: 10.1177/1462474515604385.

McCorkle, R.C. 1993. Research note: Punish and rehabilitate? Public attitudes toward six common crimes. *Crime & Delinquency* 39(2), pp. 240–252.

Mohr, J.W. and Bogdanov, P. 2013. Introduction—Topic models: What they are and why they matter. *Poetics* 41(6), pp. 545–569. doi: 10.1016/j.poetic.2013.10.001.

O’brien, D.T., Sampson, R.J. and Winship, C. 2015. Ecometrics in the age of big data: Measuring and assessing “broken windows” using large-scale administrative records. *Sociological Methodology* 45(1), pp. 101–147.

Office for National Statistics (UK). 2021. *Homicide rate per million population in England and Wales from 2002/03 to 2020/21*. Available at: https://www.statista.com/statistics/318385/homicide-rate-england-and-wales/ [Accessed: 12 October 2022].

Office for National Statistics (UK). 2022. *Number of police recorded violence against the person offences in England and Wales from 2002/03 to 2021/22*. Available at: https://www.statista.com/statistics/288256/violent-crimes-in-england-and-wales/ [Accessed: 12 October 2022].

Radanliev, P. and De Roure, D. 2023. New and emerging forms of data and technologies: literature and bibliometric review. *Multimedia Tools and Applications* 82(2), pp. 2887–2911. doi: 10.1007/s11042-022-13451-5.

Smith, G.J.D., Bennett Moses, L. and Chan, J. 2017. The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach. *The British Journal of Criminology* 57(2), pp. 259–274. doi: 10.1093/bjc/azw096.

Socia, K.M., Rydberg, J. and Dum, C.P. 2019. Punitive Attitudes Toward Individuals Convicted of Sex Offenses: A Vignette Study. *Justice Quarterly* 0(0), pp. 1–28. doi: 10.1080/07418825.2019.1683218.

Stephens-Davidowitz, S. 2018. *Everybody lies: What the internet can tell us about who we really are*. Bloomsbury Publishing.

Welsh, B. 2018. *Costs and benefits of preventing crime*. Routledge.

Williams, M.L., Burnap, P., Javed, A., Liu, H. and Ozalp, S. 2020. Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology* 60(1), pp. 93–117. doi: 10.1093/bjc/azz049.

1. This article reproduces the article made with homologous data for Uruguay in “Razones y Personas” (Ezquerra and Trajtenberg 2023) http://www.razonesypersonas.com/2023/05/usando-datos-no-tradicionales-para.html [↑](#footnote-ref-1)
2. https://yougov.co.uk/topics/politics/explore/topic/Death\_Penalty?content=all [↑](#footnote-ref-2)
3. This project is part of the doctoral project of Pablo Ezquerra, PhD candidate at Cardiff University. His project analyzes in depth the punitive attitudes expressed in social media comparatively between the UK and Uruguay. The main objective of this thesis is not only to detect punitive attitudes on Twitter, but also to analyze their effects and determinants, as well as the specificities of each country. [↑](#footnote-ref-3)