



Understanding the social construction of juvenile delinquency: insights from semantic analysis of big-data historical newspaper collections

Yu Zhang¹ · Adam Davies²  · ChengXiang Zhai²

Received: 25 October 2023 / Accepted: 5 February 2024
© The Author(s) 2024

Abstract

Massive historical newspaper collections contain rich information about the historical development of social issues and constitute a unique resource for studying the social construction of issues such as juvenile delinquency. However, manual analysis of millions of pages of newspaper articles is infeasible. In this paper, we propose a suite of computational methods, including cross-context lexical analysis, dynamic semantic analysis, and valence analysis, to facilitate the study of historical social construction. We apply these methods to ProQuest Historical NewspapersTM collection in the period of 1790–2006 to study the social construction of juvenile delinquency over this period. Our results show that the proposed methods are effective in revealing insights regarding the social construction of juvenile delinquency, leading to a better understanding of this complex issue and specific hypotheses for further study. Overall, our study shows the great promise of leveraging natural language processing techniques for analyzing historical news data to study social construction of societal issues.

Keywords Juvenile delinquency · Social construction · Historical newspapers · Natural language processing · Word embeddings · Computational semantics

Yu Zhang and Adam Davies have contributed equally to this work.

✉ Yu Zhang
yuzh@csufresno.edu

✉ Adam Davies
adavies4@illinois.edu

¹ Department of Criminology, California State University, Fresno, Fresno, CA, USA

² Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, IL, USA

Introduction

Viewed from the perspective of social construction, criminality is not inherent in an act or a person but a social phenomenon defined by a social reaction [1]. It constitutes a social process wherein behavior is determined and classified as a social issue, such as crime or juvenile delinquency. This process of social categorization is an integral aspect of social construction. However, the manner in which social construction is enacted in society remains somewhat ambiguous. Critical constructivism posits that knowledge and phenomena are socially constructed through an ongoing dialogue involving culture, institutions, and historical contexts [2]. How can we effectively describe and explore this dialogue in our research? For instance, what forms of data can we employ to elucidate the social process involved in constructing the concept of juvenile delinquency? A significant challenge emerges in sourcing systematic data that adequately measures this social process as it manifests in everyday social life [3]. It becomes imperative to discern the dialogue that shapes the definitions of “juvenile” and “delinquency” within specific social contexts. Additionally, national statistics in England did not begin disclosing the ages of offenders until 1834 [3], even though the issue of juvenile delinquency dates back much earlier. Digital newspaper archives offer potential resources for data.

News outlets function as platforms through which social contexts can be comprehended. Individuals rely on news to gain insights into the social realm and make informed choices. Throughout this process, the societal contexts presented in news stories are assimilated by the public, regardless of whether they align with underlying realities. Over time, news contributes to the construction of social reality [4]. For instance, some scholars posit that journalists or news organizations may gravitate towards reporting certain “unusual” stories, thereby presenting a skewed reality aimed at shocking, entertaining, or exciting readers [5].

In this paper, we understand the social construction of juvenile delinquency as formulated by Hydén [6] to unveil the process by which social problems are shaped and evolve: juvenile delinquency represents a social construction that exploits the societal implications of age to differentiate between children and adults, under the assumption that children should not be subjected to trials in adult criminal courts or correctional facilities. The origins of how the term “juvenile delinquency” became entrenched in human society remain somewhat enigmatic, given that comprehensive statistics on juvenile offenders were not published until the mid-1830s. Historical newspapers provide a valuable avenue for examining the emergence of juvenile delinquency as a significant social concern, including when, where, and why it was recognized as such. Drawing from these news articles, we can observe how the descriptions of juvenile delinquency have transformed over different time periods. By performing semantic analysis of historical newspapers, we study linguistic shifts in related terminology and investigate the inception and evolution of the term “juvenile delinquency” throughout history. This approach allows us to assess the gradual development of behaviors associated with juvenile delinquency and how they are recognized as societal issues within the context of social construction. We aim to understand the history and process of how juvenile delinquency evolved as a

social problem, particularly before the establishment of juvenile courts in the early 20th century. The central research questions driving this investigation are:

1. How has the meaning of juvenile delinquency changed over time?
2. How have public attitudes towards juvenile delinquency changed over time?
3. How can the answers to (1) and (2) inform our understanding of how juvenile delinquency has been constructed as a social problem?

Although our work is driven by specific questions about juvenile delinquency, the approach to studying social construction that we develop in this work is not limited to juvenile delinquency—our exploratory framework and the suite of computational methods we employ here are, in principle, applicable in studying the historical social construction of any other concept of interest. Our primary contributions are as follows:

- We propose an exploratory framework for studying historical social construction by leveraging current computational techniques in text mining and natural language processing.
- We extend several of these techniques to improve their suitability in studying historical social construction. For instance, we extend the Cross-Context Lexical Analysis (CCLA) methodology to operate at the level of aggregated concepts rather than individual terms, and further develop an approach to generate explanations of historical semantic dynamics revealed by CCLA.
- We carry out the first computational analysis of the social construction of juvenile delinquency, yielding important insights that can serve as a foundation for future study. For example, we find that there is a substantial shift in the semantics of juvenile delinquency and associated concepts in the period of 1870–1890 (immediately preceding the institution of juvenile courts in the US), and that this period marks the start of an increasingly positive construction of juvenile delinquency over time (relative to crime in general).

A brief history of juvenile delinquency

Juvenile delinquency, a term coined after 1800, traces its roots back to a long-standing history of delinquent behavior among young individuals [7]. The transition from agrarian to commercial economies during the late Middle Ages in Europe triggered a shift in societal dynamics. Urban migration surged as rural populations sought survival in cities, resulting in declining living standards and heightened poverty. Economic hardships led families to struggle in retaining control over their children, often resulting in child abandonment and increased misconduct [8].

To address this predicament, communities introduced the binding-out system, a practice where children were placed as apprentices within other families. As the challenges mounted, dedicated institutions like the London Bridewell emerged in 1553 to house homeless children, later evolving into the House of Correction in

1555 [9]. Similar establishments proliferated across England, providing a platform for young offenders to work and reintegrate into society upon release.

The 1800s marked a significant turning point as juvenile delinquency surged in Britain due to worsening poverty. Historical British Parliament records illuminate the extent of this issue [10]. This trend persisted, with a notable increase in juvenile offenders evident in British Home Office statistics from 1846 [11]. This era saw the emergence of the reformatory movement, advocating specialized treatment for young offenders. These efforts culminated in the establishment of the Juvenile Court in England, later renamed the Youth Court, with the capacity to address both criminal and noncriminal juvenile cases [12].

In early North American colonization, the family unit played a crucial role in juvenile social control [13]. However, by the 18th century, poverty eroded family bonds, leading communities to place children with other families under the doctrine of *loco parentis*, effectively turning these children into indentured servants within their new households. Alternatives like almshouses were provided for those unable to find suitable families. Notably, this migration pattern extended beyond Europe, as children from overcrowded Bridewells and poorhouses in England were transported to the Americas as indentured servants [12].

The industrial revolution in North America further reshaped the landscape, dismantling family ties and spurring juvenile misbehavior [12]. Legislation was introduced to address this growing issue, leading to the establishment of houses of refuge. These institutions utilized apprenticeships to occupy young individuals until they reached the age of legal adulthood. However, as the 19th century dawned, economic difficulties persisted, prompting an influx of youths to urban areas brimming with both leisure activities and criminal opportunities.

In response, innovative strategies were devised to combat the challenge of juvenile delinquency. During the mid-19th century, governmental authorities assumed control of juvenile delinquency institutions from philanthropic groups, giving rise to refuge and reform schools. These institutions often sentenced children to remain until reaching the age of majority or achieving reform. The cottage or family system emerged between 1857 and 1860, dividing youths into units with distinct schedules and activities. Yet, despite these endeavors, the success of reform schools remained uncertain [13].

The trajectory of juvenile delinquency and its social implications continued to evolve. As the efficacy of reform schools came under scrutiny, juvenile misbehavior persisted, prompting a Federal committee's establishment in 1961 to address the issue. Sociologists were encouraged to experiment with prevention projects rooted in diverse sociological theories [12]. Both community-based and agency-run programs were implemented, but their outcomes were modest [14]. A distinctive link between juvenile delinquency and poverty emerged, underlining the necessity of wealth and power redistribution within society. However, the history of juvenile justice unveiled a persistent cycle of responses that often failed to address systemic problems, as the dominant social class wielded social institutions to maintain the status quo, complicating efforts to effectively rehabilitate delinquents [12].

This historical analysis provides a broad overview of juvenile delinquency, but it falls short of revealing how the social significance of juvenile delinquency has

evolved over time and how public perceptions of it have shifted. Within our analysis, our goal is to identify corresponding patterns in our newspaper data. This societal process, encompassing the definition, categorization, and response to juvenile delinquency and juvenile justice, culminates in the social construction of juvenile delinquency as a societal concern. Through semantic analysis, our aim is to clarify how the linguistic patterns within newspaper articles illustrate the formation of juvenile delinquency.

Social construction theory of juvenile delinquency

Juvenile delinquency stands as a prominent social problem [15]. Within the realm of sociology, a social problem is characterized as a condition or behavior that is widely perceived as harmful [16]. Such issues encompass undesirable and detrimental circumstances such as crime, delinquency, and poverty, prevalent within human society. But how does a social condition evolve into a social problem? Is a social problem merely a tangible existence, or is it a concept forged through societal construction? In the domain of sociology, a social problem consists of both objective and subjective dimensions [17]. The objective aspect postulates that a social problem represents a reality—an objective condition stemming from inherent dysfunctions within a society. Yet, the objective facet alone is insufficient, for social conditions must gain acknowledgment as harmful by the public to attain the status of a social problem. The subjective aspect of a social problem concerns how a problem is defined. Different societies may interpret harm differently. Blumer [18] contended that a society only recognizes a social problem if it acknowledges its existence. This process commences with divergent interests, intentions, and objectives that interact to ascertain whether a condition assumes the role of a social problem. From these conflicts emerges a specific condition that society identifies, addresses, and ultimately defines as a social problem. These collective definitions constitute a repetitive and selective process, where various detrimental social conditions vie for societal recognition, with only a few emerging as genuine concerns (see Fig. 1).

This process derives its theoretical basis from the perspective of social constructionism. According to this view, a human-created world exists in contrast to the external world [19]. The world takes shape, evolves, and becomes a social creation through everyday thoughts and actions—an ongoing process of social construction [20]. The essence of social constructionism lies in the collective societal perception of a target population or situation, reinforced through social processing [21, 22]. This social processing, encompassing government policies, can acknowledge, legitimize, or reshape social constructions [23]. Additional factors involved in this process include politics, culture, socialization, history, the media, literature, religion, and more [24].

Throughout this process, interest groups contend with one another and more powerful factions to gain societal recognition for certain conditions, elevating them to the status of social problems. Strategies may encompass leveraging emotions, political figures, and mass media. For instance, influential groups can draw attention to

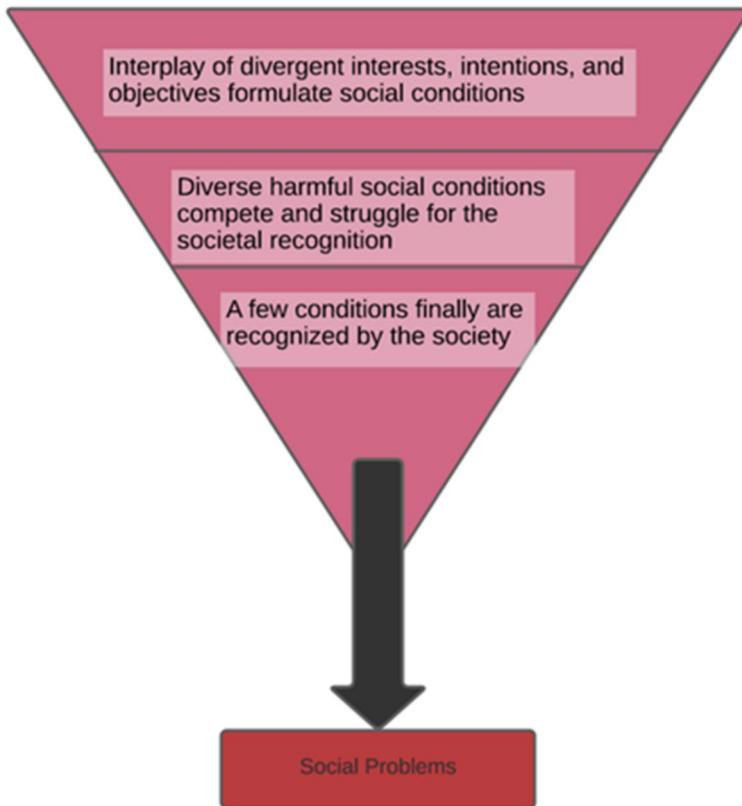


Fig. 1 The funnel of social problem processing

particular conditions by disseminating newspaper stories infused with agitation, concerns, and sensationalism. These themed narratives, reiterated across different reports, both reflect and shape these conditions, eventually cementing them as social problems. The mass media, including newspapers, thus offers a means to conduct a historical analysis of social construction.

Research has delved into how mass media constructs crime and justice historically [25]. Serving as a crucial platform for social and political struggle, mass media projects the official reality of crime and delinquency, encompassing shifts in historical and social contexts related to race, class, and gender relations [26, 27]. However, pressures stemming from political and cultural hegemony could distort reality within mass media [26]. Marginalized groups, including the impoverished, the working class, women, and people of color, might find their voices silenced in mass communication. News serves as a tool wielded by the powerful [28–30], functioning as an ideological mechanism [30–34]. It unveils how commonplace thought processes contribute to structures of dominance [25] and the process of social construction.

In this paper, we analyze a massive historical newspaper collection in order to study the history and process of the social construction of juvenile delinquency. Before the establishment of the juvenile justice system in the nineteenth century, no distinct legal category existed for deviant youth behavior. The journey of how and when juvenile delinquency emerged as a social problem captivates our inquiry. Do social problems result from the intrinsic malfunctions of society, or do they emerge from a competitive funnel where various harmful social conditions vie for recognition? For example, the linguistic stability of the term “reformatory schools” during a specific era reflects the classification and processing of children’s misbehavior as a societal reality during that period. The way law enforcement and communities address such misbehavior, whether through incarceration or community programs, unveils multifaceted struggles among societal factions. Consequently, the frequency of “reformatory schools” usage may fluctuate across different decades. How can we delve into the intricacies of these dynamics? We address these questions by harnessing natural language processing techniques to analyze historical newspaper data, as elaborated below.

Methodology

Manual analysis of hundreds of years of newspaper data is infeasible because of the high-volume information, difficulty in searching the historical newspapers, and challenges in defining and quantifying relevant measures for studying social construction. Fortunately, the availability of digital forms of historical newspapers (e.g., the Library of Congress and ProQuest Historical Newspapers™)¹ and the recent progress in machine learning-driven techniques for analyzing such data for social science [35, 36], particularly in the domains of natural language processing [37, 38] and information retrieval [39], have opened up an exciting opportunity for using computational algorithms to assist researchers in analyzing historical newspapers to study the social construction of juvenile delinquency.

Dataset

The primary dataset we used for this research is a subset of the ProQuest Historical Newspapers™ (PHN) collection, which we received through the Cline Center for Advanced Social Research and the University of Illinois Library. This subset contained ~58.2 million segmented newspaper article files from 25 newspaper titles worldwide.²

Analysis methods

To examine and understand the social construction process of *juvenile delinquency*, we focus on answering the following research questions:

¹ Accessible via <https://www.loc.gov/> and <https://about.proquest.com/en/products-services/pq-hist-news/>, respectively.

² Note that, at the time of writing this paper, the full PHN dataset contains 52 M digitized pages from 68 historical newspaper titles, many of which contain multiple segmented newspaper articles. Of those newspaper titles, only 25 titles are present in the subset we received.

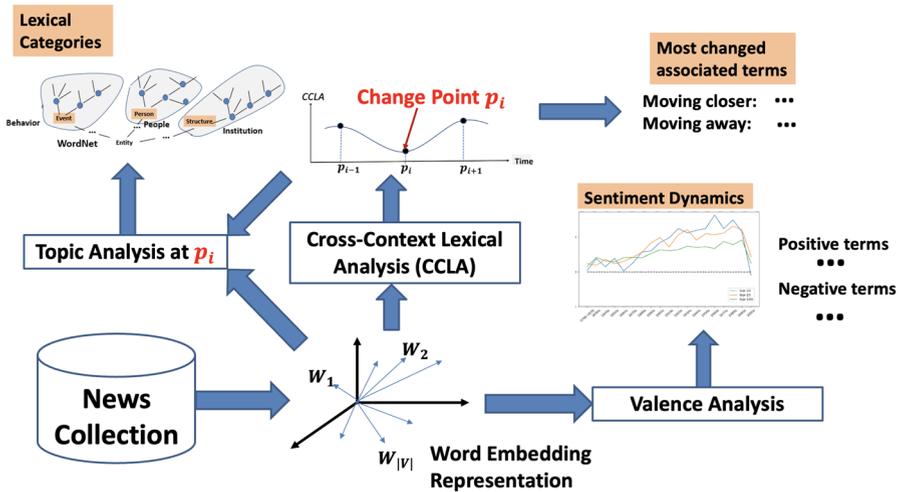


Fig. 2 A schematic of our analysis modules. Arrows $X \rightarrow Y$ denotes that the output of X is used as an input for Y , and orange headings denote endpoints of the analysis (i.e., results that are not used as inputs to any module)

1. How has the meaning of juvenile delinquency changed over time?
2. How have public attitudes towards juvenile delinquency changed over time?
3. How can the answers to (1) and (2) inform our understanding of how juvenile delinquency has been constructed as a social problem?

Addressing these questions requires methods for semantic analysis of meanings of terms in text data, semantic associations of related terms, and how the meanings of a concept vary across different contexts (e.g., different regions or time periods). We also need to address the challenge that a concept such as juvenile delinquency can be expressed using many different terms and thus we must aggregate information about all those variations of terms to fully represent a concept. Finally, in order to analyze public attitude, we must also be able to analyze the valence (positivity or negativity) of sentiment in each context.

Below, we discuss how we address these challenges by leveraging semantic analysis techniques based on word embedding representation and developing new methods suitable for answering our research questions. First, we describe the basic analysis methods, including the word embedding representation and how it can be used to quantify semantic associations of words, term aggregation methods to capture variations of the expression of the concept juvenile delinquency, and lexical categorization to facilitate interpretation of associated terms of a concept. Building on these basic analysis methods, we proceed to discuss additional analysis techniques that can be used to answer our research questions by discovering historical semantic dynamics and analyzing public sentiment.³

³ It is important to note that newspapers are not necessarily representative of the sentiment, attitudes, or perspectives held by the public. However, as PHN is a wide-ranging collection of tens of millions of newspaper articles across hundreds of years, several countries, and dozens of publishers, we consider it a sufficiently diverse collection to serve as an (imperfect) proxy for public sentiment.

The overall process of analysis is illustrated in Fig. 2, where we see that the first step is to obtain a semantic representation of all words using latent vectors (i.e., word embeddings), which then provides a basis for all the subsequent analysis steps, including specifically (1) Cross-Context Lexical Analysis for discovering the change points, where we saw significant changes of the semantics of the concept juvenile delinquency, (2) Topic Analysis around a change point to facilitate semantic interpretation with meaningful lexical categories, and (3) Valence Analysis to reveal the sentiment dynamics around a change point and some specific positive/negative terms that can facilitate interpretation. Below we will explain all the component techniques in detail.

Semantic analysis using word embeddings

We may model the semantic change of *juvenile delinquency*-related terms over time by training word2vec [40] models on each decade of text data in PHN, denoted $C_i, i \in [1, k]$,⁴ giving us a term vector v_t for every term t occurring in C_i .⁵ Such embedding representations map each term to a vector in a latent semantic space such that semantically associated terms have similar vectors, allowing us to compute terms' semantic association by comparing their corresponding term vectors. That is, given terms t, s , we compute the semantic association of t and s in the word embedding model for a given decade C_i using cosine similarity over the corresponding term vectors v_t, v_s , i.e., $\text{sim}(t, s|C_i) = \cos(v_t, v_s)$, enabling us to measure how semantic associations between various terms change over time by comparing $\text{sim}(t, s|C_i)$ and $\text{sim}(t, s|C_j)$ for different decades C_i, C_j .

Term aggregation

The concept of *juvenile delinquency* (both as a behavior and as individuals engaging in it) has historically been expressed using a variety of terms such as “young criminality” or “juvenile offenders”.⁶ As our goal is to study juvenile delinquency at the concept-level, not term-level, it is necessary to develop analysis methods for aggregating observations over multiple terms. We begin by collecting a set of “JUDEL terms” which have historically been used to express this concept: first, we consulted a domain expert to provide a set of known *seed terms* $T_0 = \{t_i\}_{i=1}^n$ (including, e.g., “juvenile delinquency”, “young offenders”, and “cosh boys”), then computed a list of the k -nearest neighbor terms by aggregating the term-level neighbors

⁴ E.g., contexts C_2, C_3 consists of all documents from the years 1830–1839 and 1840–1849, respectively. In the special case of C_1 , we use all documents from the years 1790–1829 because PHN has far fewer documents per year until ~1830; but all remaining $C_i, 1 < i \leq k$ consist of documents from only a single decade of the PHN dataset.

⁵ By default, terms are individual words; but we also used the Python Gensim Phrases module (see <https://radimrehurek.com/gensim/models/phrases.html>) to automatically detect common multi-word phrases (e.g., *juvenile delinquency*) by collocation, and treated these phrases as individual terms when training the word embedding model.

⁶ That is, we are interested in 2 categories of terms: the first denotes the behavior of juvenile delinquency (e.g., “juvenile delinquency” itself, but also including other terms like “young criminality”, “youthful crime”, etc.), and the second denotes individuals engaging in such behavior (e.g., “juvenile delinquents”, “young criminals”, etc.).

with respect to each term and normalizing by term frequency. That is, for relative term-frequency measure $\text{freq}(t|C) = \frac{\text{freq}(t|C)}{\sum_{t' \in T} \text{freq}(t'|C)}$ and query term q , we compute aggregated similarity scores between T and q in context C as:

$$\text{sim}(q, T|C) = \sum_{t \in T} \text{freq}(t|C) \cos(t, q|C)$$

This similarity measure may be used to rank each $q \in V_C$ (where V_C is the vocabulary of the word-embedding model trained on C) by its similarity to T . We generate a list of aggregated nearest-neighbor terms T'_i as the top- k terms with the highest aggregated similarity to T_0 in context C_i . We manually inspect the top 25 terms (i.e., $k = 25$) for each context C_i partitioned by decade i with $i \in [1, m]$ for m number of decades in the dataset to find additional terms $T_{i,\text{JUDEL}} \subseteq T'$ referring to juvenile delinquency rather than related, non-coreferential terms (e.g., “reformatories” appears at or near the top of the nearest-neighbors lists for each decade from 1850–1910, but does not refer specifically to juveniles engaging in delinquent behaviors or such behaviors themselves), expanding the list of terms from the original seed terms T_0 to the final list of “JUDEL terms” $J = \cup_{i=1}^m T_{i,\text{JUDEL}} \cup T_0$.

Lexical Categories

In order to better understand the kinds of associations revealed by generating k -nearest neighbor lists with respect to JUDEL in a given period, we can decompose the set of neighbor terms hierarchically using WordNet [41], positioning each term in a hierarchy of lexical categories. For example, we may categorize the terms according to whether they refer to *juvenile delinquency*-related events or behaviors (e.g., “juvenile delinquency”, “youth crime”), to individual juvenile persons engaging in delinquent behaviors (e.g., “juvenile delinquents”, “young criminals”), or to the penal institutions in which these individuals serve sentences for their criminal behavior (e.g., “penitentiaries”, “reformatories”).

Discovering historical semantic dynamics

To address our first research question about how the semantic meaning of juvenile delinquency change has changed across the years, we need a tool to perform a comparative analysis of semantics across different contexts such as different time periods. For this purpose, we leverage the existing Cross-Context Lexical Analysis method [42], and develop techniques to extend it to concept-level analyses and better interpret the results it yields.

Cross-context lexical analysis

The cross-context lexical analysis (CCLA) approach was initially proposed by Massung [42] as a general approach for quantifying the degree of similarity of a term’s meaning/usage in different contexts. Given a term that we are interested in studying (e.g., juvenile delinquency) and two contexts (e.g., two adjacent time periods) that we want to compare in terms of their coverage of the term, CCLA works by first computing the similarity of a query term with respect to other terms in the

first context to form a context-specific term similarity profile for the query term, repeating this process for the second context, and comparing the similarity of both profiles across the two contexts.

Formally, let $V_C = \{t_1, \dots, t_{|V_C|}\}$ be the set of all terms in our vocabulary for context (decade) C , and let $T \subseteq V_C$ be any subset of these terms that we wish to consider in our analysis (detailed further below). Given a particular term $t_i \in V_C$, the term similarity profile of term t_i in context C is defined as the vector $v_T(t_i, C)$ of similarity values between term t_i and all other terms in context C , i.e.,

$$v_T(t_i, C) = (\text{sim}(t_i, t_1|C), \dots, \text{sim}(t_i, t_{i-1}|C), \text{sim}(t_i, t_{i+1}|C), \dots, \text{sim}(t_i, t_{|V_C|}|C))$$

Intuitively, the term similarity profile of a term shows what kind of terms are most strongly associated with a target term t_i in the context C . If term t_i is “juvenile delinquency”, the profile would indicate what kind of (other) terms are often used together with “juvenile delinquency” in the context of C . Those semantically associated terms thus reflect how the term t_i is discussed in the context of C . Since low similarity values are not as meaningful (e.g., we can learn more about the meaning of a word by looking at the semantics of its most closely associated terms than less associated terms [43]), we are generally only interested in the top- k terms in the similarity profile that have the highest similarity values; so for our study, we find the k -nearest neighbor term embeddings for a query term in two contexts C_1, C_2 , and use the union of these terms to construct the term similarity profile in each period, effectively truncating the vocabulary to terms which are nontrivially associated with the query term in either period. Specifically, we find two sets of terms S_1, S_2 that occur in the vocabularies for both contexts (i.e., $S_1, S_2 \subset V_{C_1} \cap V_{C_2}$) and whose corresponding vectors are in the top- k most similar terms to query term t in contexts C_1, C_2 (respectively), and compute term profiles $v_T(t, C_1), v_T(t, C_2)$ over the union of these terms $T = S_1 \cup S_2$. CCLA is the degree of semantic consistency of query term t between contexts C_1, C_2 , measured by the cosine similarity of the two neighbor term profiles of term t between the two contexts:

$$\text{CCLA}(t|C_1, C_2) = \cos(v_T(t, C_1), v_T(t, C_2))$$

We employ this approach by computing the CCLA score for each pair of subsequent decades C_i, C_{i+1} and plotting a time series of these scores across all decades to analyze the semantic (in)stability of the term over time (see Fig. 5). Of particular interest for our purposes are periods of rapid change in the meaning of the query, which correspond to “dips” in the CCLA time series.

Aggregating CCLA

Since the original CCLA is designed for analyzing semantic changes of just one term, we need to extend it to cover all the variants of the terms describing the concept of juvenile delinquency.

In prior work [36], we approached term aggregation in Cross-Context Lexical Analysis (CCLA) by replacing all terms related to a concept (e.g., for the general concept of *juvenile delinquency*, these terms include “juvenile delinquency”, “young offenders”, “juvenile delinquent”, etc.) with an aggregator token (e.g., “JUDEL”)

in text pre-processing. This allowed for a more comprehensive historical semantic analysis of a general concept, particularly one that is referred to using many different words over time, than would be possible by examining any individual term. However, this also means that the influence of each distinct term in the aggregator term is conflated, making it impossible to add, remove, or perform CCLA with respect to any individual constituent term without retraining all word embedding models across all contexts (decades). To address this concern, we propose an alternative approach to term aggregation by taking the frequency-weighted average of the CCLA scores of all terms of interest in each time period. Weighting scores by frequency means that terms which are most often used to describe the target concept are more influential in the aggregated CCLA score than those that are used only a few times in a given time period. That is, for JUDEL terms J , contexts C_1, C_2 , and frequency $f(t) = f(t, C_1) + f(t, C_2)$, the aggregated CCLA score is given by:

$$\text{CCLA}(J|C_1, C_2) = \sum_{t \in J} \text{freq}(t) \cdot \text{CCLA}(t|C_1, C_2)$$

We applied aggregated CCLA to the entire period of the data by first segmenting the data into a segment of a decade each and then use CCLA to compute a score of any two adjacent decades and then plot the curve to examine where there is the greatest change. The CCLA method can reveal the time points when the semantics of *juvenile delinquency* have changed significantly. Those time points can then guide us to further understand the public attitude towards *juvenile delinquency* in the adjacent periods and provide evidence for or against the competing social construction hypotheses discussed above.

Explaining changes

To better understand the potential changes of public attitude, we identified the neighbor terms whose association with JUDEL has changed most significantly during a given period of interest from C_i to C_{i+1} , i.e., terms $t : \text{sim}(J, t|C_{i+1}) \gg \text{sim}(J, t|C_i)$. However, simply knowing that a given neighbor term t changes in association with JUDEL does not tell us *how* or *why* it changed; so in order to better understand these terms' relationship with JUDEL, we extract mutual neighbor terms t' that have a high similarity with both JUDEL and the neighbor term t by creating a "mutual neighbors list" ranking each t' by its joint similarity with both J and t . One option for computing this joint similarity measure is to simply take the weighted average of the similarities $\text{sim}(J, t'|C)$ and $\text{sim}(t, t'|C_i)$ and adjusting the balance between similarity with J and t using parameter $\lambda \in (0, 1)$, i.e.,

$$\text{joint} - \text{sim}(t'; t, J|C) = \lambda \cdot \text{sim}(t, t'|C_i) + (1 - \lambda) \cdot \text{sim}(J, t'|C)$$

We found that doing so often yields terms which are highly related to either J or t , but not both; and that, for practical purposes, modulating λ simply selects between these two options (e.g., a smaller or larger λ "selects" J or t as the priority, respectively). As we are interested in finding "mutual neighbors" t' which are associated with *both* t and J , we experiment with an alternative formula where the linear weighted average above is substituted with a weighted harmonic mean, which

heavily preferences t' that are associated with both t and J over those associated with only one of them, yielding the following joint – sim measure:

$$\text{joint – sim}(t'; t, J|C) = \frac{1}{\frac{\lambda}{\text{sim}(t, t'|C)} + \frac{1-\lambda}{\text{sim}(J, t'|C)}}$$

Analyzing public sentiment

Our second research question is about how the public attitude towards juvenile delinquency might have changed over time. To answer this question, we need to go beyond what CCLA can generate to further analyze the sentiment of the coverage of juvenile delinquency. Specifically, CCLA can only indicate which periods saw the greatest change in the usage and context of JUDEL terms over time, but it does not provide a precise characterization of this change. The word clustering approach described earlier can facilitate understanding of the public attitude towards JUDEL from specific perspectives such as People, Institution, and Events/Actions. However, to test the social construction hypothesis, we need to further understand their positive and negative social construction throughout history by analyzing the sentiment of the content covering juvenile delinquency in the periods where its semantics has changed.

To this end, we apply the *valence analysis* approach developed by Charlesworth et al. [44] to characterize the positive and negative valence of juvenile delinquency in each decade, reflecting changes in public attitude regarding this topic. This technique works by comparing a set of query terms Q representing a social group of interest (in our case, *juvenile delinquency*) to an inventory of terms T with human-annotated valence scores $\text{valence}(T)$ that capture the average “pleasantness of a stimulus” experienced by human subjects when they encounter the term [45],⁷ finding the terms $T_Q \subset T$ which are most associated with Q in C_i using a word embedding model, and computing the average valence of these most-associated terms. However, it is not enough to directly define T_Q as the most-associated terms via sim, as this may yield a list of terms that is more representative of a general social or demographic category of which Q is a member (e.g., gender, race, ethnicity, class, etc.), biasing the results toward the category and not the specific group indicated by Q . To correct for this, a *comparison group* Q' in the same category is used as a foil to normalize for such category-level associations. Finally, we may compute the valence of Q in context C_i as T_Q , the average valence of the top – k most-associated valence terms to Q in C_i normalized for associations with Q , using the mean average-cosine score (MAC) over each term $q \in Q$:

⁷ Following Charlesworth et al. [44], we use the Warriner et al. [45] inventory, which is the largest and most comprehensive of its kind: it includes ~304K total valence ratings for ~14K high-frequency English words (yielding an average of ~22 ratings per term, collected from ~2K participants in total), where each term is rated on a scale of 1 to 9, and the valence of each term $\text{valence}(t)$, $t \in T$ is averaged over the responses of each participant. For example, the lowest-scoring term is t_{\min} = “pedophile”, with $\text{valence}(t_{\min}) = 1.26$; and the highest-scoring term is t_{\max} = “vacation”, with $\text{valence}(t_{\max}) = 8.53$. Terms not present in the Warriner et al. [45] inventory are excluded from our analysis.

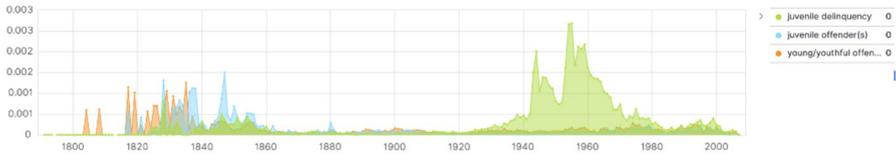


Fig. 3 Normalized document frequencies, where the y-axis is the proportion of articles each year that contain each term

$$T_Q = \arg \operatorname{top} - k_{t \in T} (\operatorname{MAC}(t, Q|C_i) - \operatorname{MAC}(t, Q'|C_i))$$

and computing $\operatorname{valence}(Q) = \operatorname{avg} \operatorname{valence}(T_Q)$, the average valence of T_Q . We perform this valence analysis with $Q = J$ and Q' as the set of JUDEL terms J with age-denoting words removed (e.g., “juvenile delinquency” and “young offender” are substituted for “delinquency” and “offender”, respectively) to examine the general positivity or negativity in news media coverage of *juvenile delinquency* relative to criminality in general, and observe how this sentiment changed over time.

We have so far discussed how we can use computational methods to address the first two research questions. To address the third research question and understand how the changes in 1 and 2 relate to the construction of juvenile delinquency as a social problem, we will synthesize the results generated from answering the first two questions and analyze them with consideration of any known historical activities that happened during those periods (e.g., establishment of the juvenile court). This will be further discussed in detail in the next section, where we report all the results and attempt to extract some overall insights about the social construction of juvenile delinquency.

Results

Overview of historical dynamics

To see the overall dynamics of *juvenile delinquency*, we first examine the raw frequency of the terms expressing this concept (across a small subset of high-frequency “JUDEL terms”, as discussed above).

Figure 3 shows the normalized document-frequency of the most prominent (highest-frequency) JUDEL terms over time in the ProQuest data. Overall, we see that there are two “peak-frequency” periods: 1800–1850 and 1940–1970, showing interesting changes of terms describing JUDEL. Through ~ 1835 , young offenders was the most common term, followed by juvenile offenders from ~ 1835 through ~ 1860 , when all terms plateau at a low relative frequency. After ~ 1930 , juvenile delinquency rapidly moves to become the dominant term used to refer to the

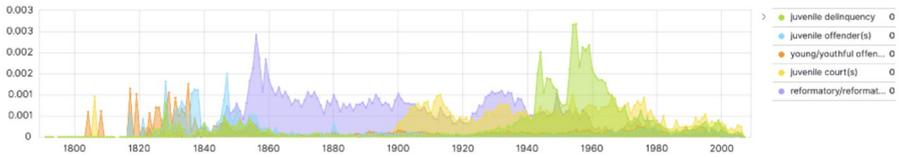


Fig. 4 Normalized document frequencies including “juvenile court” and “reformatory/reformatories”

associated concept, and this trend continues up to the present day. We also see that in this period, the term juvenile delinquents appears to be replacing the term young offenders. Until ~1860, relative frequencies of the terms visualized in Fig. 4 vary substantially until ~1860 to ~1930, at which point there is a clear plateau (with only brief, minor deviations like that of juvenile offender(s) in ~1880) until ~1930, when juvenile delinquency rapidly moves to become the dominant term used to refer to the associated concept, and remains so through at least ~1980.⁸ Additionally, the relative frequency of juvenile delinquency from 1945–1965 is greater than that of all other terms at any other period, and the other JUDEL terms remain at a low frequency throughout this period. Thus, 1945–1965 is the period in which the concept of juvenile delinquency was discussed the most, and this discussion relied far more on the specific term juvenile delinquency than other terms (unlike earlier periods).

There are a few simple observations we can make on the basis of this frequency information. First, in the period of 1800–1835, “young offenders” was the most common term, followed by “juvenile offender”. Young offenders simply adds the term “young” to “offender”, vaguely naming a group of people who committed a broad category of offensive behaviors, but does not clearly distinguish specific categories of offensive behavior (e.g., it could be criminally offensive, or merely socially offensive without any implication of unlawfulness). Additionally, it is important to note that “young” did not clearly distinguish between children and adults, as the specific categorization of juvenile crime versus adult crime had not yet emerged (as discussed above). During this time, the term of “juvenile delinquency” started emerging. On September 16 in 1825, the Manchester Guardian newspaper started using the term “juvenile delinquent” to refer those young people who committed bad

⁸ I.e., there was no clear equivalent term to “juvenile delinquency” prior to ~1930, as there was not a common term used to describe the category of behavior independently of the individuals engaging in it (e.g., “juvenile offender”). While we observe the first use of the term in our dataset in 1828, it did not enter widespread use until ~1930, when it supplanted the use of the term “juvenile offender”.

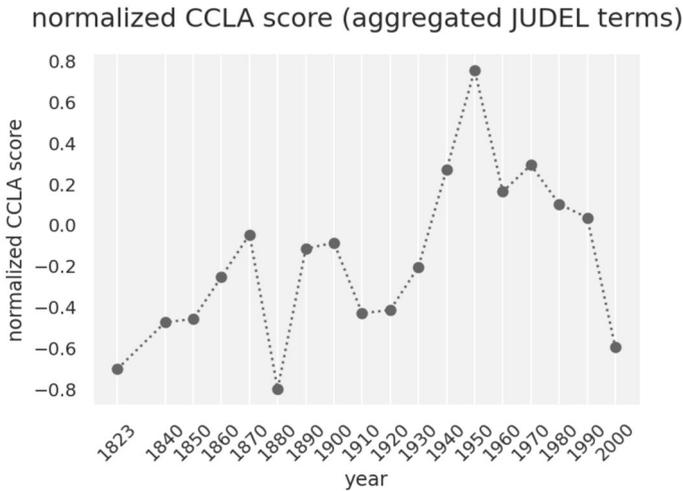


Fig. 5 Time series of normalized, aggregated CCLA scores across all JUDEL terms

behaviors. In 1828, the Manchester Guardians (March 7) was the first in our dataset to use the term “juvenile delinquency”. In 1825 through 1860, the appearance of the term of juvenile indicated the gradual coining of new term to describe the specific group. Compared to “young”, “juvenile” not only refers to young age, but also means physiologically immature or undeveloped. This term change was laying a foundation to build a separate juvenile court, by forming views that youth are less culpable than adults, more capable of change and rehabilitation, and more deserving of protection from the harsh and punitive conditions of the adult criminal justice system [46]. After the juvenile court was established around 1900, the term “juvenile delinquency” quickly became the dominant term used to refer to such behavior. This evolution in terminology and its alignment with the concept of a distinct court system reflect a shifting societal perspective on youthful offenders and the need for a specialized approach to address their actions.

Change point discovery

To discover the specific time periods where the semantic changes happened, we applied CCLA to the list of JUDEL terms (described above) during the last few centuries. The aggregated CCLA score across all decades is shown in Fig. 5.

Recall that the CCLA score indicates the semantic similarity of terms associated with JUDEL between two adjacent decades. Thus a higher CCLA score (y-axis value) means that the two adjacent decades have similar discussions of JUDEL, whereas a low score means that there is more difference between the terms associated with JUDEL in those two decades. These CCLA scores are intended to capture the collective semantic shift of the JUDEL terms throughout history. More precisely, they represent the similarity of the meaning of these terms between

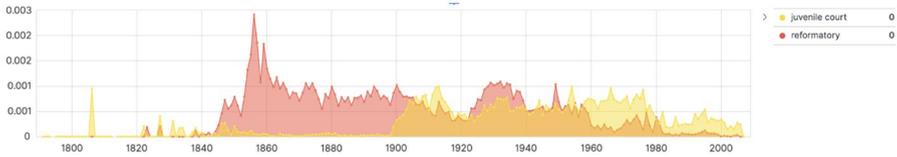


Fig. 6 Normalized document frequencies of “juvenile court” and “reformatory”

different contexts – in this case, the contexts are time periods, and the “similarity in meaning/representation” can be understood as semantic stability over time. Thus, the “peaks” correspond to periods of relative semantic stability (e.g. in 1890–1910 or 1940–1960), and the “valleys” (e.g. in 1870–1890 or 1910–1930) correspond to periods of change.

From the results in Fig. 5, we see that the period of most rapid change (i.e., lowest point on y-axis) and period of sharpest change relative to the periods immediately before and after (i.e., biggest “dip”) is from 1870–1890, corresponding to the most significant⁹ semantic change in the concept of juvenile delinquency, and represents the greatest break from the periods before and after.

CCLA tells us that the subjective perception and discursive context of juvenile delinquency (on the part of US news media reporters), and likely its construction as a social problem changed more during 1870–1890 than any other period. Criminal institutions like reformatories were not infrequently discussed during this period (see Fig. 6), but the pattern revealed by CCLA is not simply a matter of how often juvenile delinquency (or those engaging in such behavior) was discussed (as shown in Fig. 4). Rather, it appears that discussions of juvenile delinquency shifted substantially prior to the changes in penal institutions around ~1900 (i.e., the institution of juvenile courts), which itself preceded the transition to juvenile delinquency as the term-of-choice for describing the concept that it does today and the corresponding increase in discussion of this concept overall from ~1945–1965. Thus, US news media shifted its representation of juvenile delinquency well before corresponding changes in criminal law or frequency of media coverage occurred. It is possible that this shift corresponds to a changing social construction of juvenile delinquency, precipitating later changes in law and popular nomenclature. However, this does not tell us precisely how the social construction of juvenile delinquency changed in 1870–1890. We further analyze the changes in this period in the following section.

⁹ By “significant”, we mean that, given CCLA normalization, the y-axis directly corresponds directly to the statistical significance (z-score of a normal distribution) of JUDEL CCLA score relative to CCLA scores across all terms that occur in all periods – or in our case, a randomly sampled subset of several thousand such terms from each period, which serves as a more computationally efficient approximation in order to compute this quantity.

	1860s	1870s	1880s	1890s
0	reformatories	reformatories	reformatories	hardened_criminals
1	reformatory	vagrancy	criminals	reformatory
2	young_offenders	criminals	reformatory	criminals
3	criminals	vagrants	offenders	reformatories
4	juvenile_delinquent	reformatory	habitual_drunkards	young_offenders
5	vagrancy	offenders	young_criminals	juvenile_criminals
6	convicts	paupers	young_offenders	young_criminals
7	vagrants	habitual_drunkards	juvenile_criminals	juvenile_delinquents
8	young_criminals	lunatics	lunatics	incorrigible
9	penitentiaries	prisons	penitentiaries	youthful_offenders
10	juvenile_criminals	vagrant	commitments	truancy
11	youthful_offenders	drunkenness	workhouses	habitual_drunkards
12	refuges	offences	prisons	penitentiaries
13	committals	truants	vagrants	offenders
14	poorhouses	penitentiaries	jails	jails
15	juvenile_delinquents	jails	paupers	youthful_offender
16	juvenile_offenders	juvenile_criminals	truancy	petty_offenses
17	habitual_criminals	drunkards	incorrigible	lunatics
18	orphanages	truancy	infanticide	vagrants
19	lunatics	asylums	criminal	felons
20	gaols	offenses	asylums	feebleminded
21	corporeal_punishment	dipsomaniacs	incarcerate	penal
22	offenders	inmates	offenses	unconvicted
23	idiots	pauper	convicts	corporal_punishment
24	commitments	punishments	juvenile_offender	prisons

Fig. 7 The top-25 aggregated nearest neighbors to JUDEL surrounding the “change point” in 1870–1890 (bookended by a decade on each side for comparison). Columns are decades and rows are the most-associated terms in each period listed in descending order

Semantic analysis of articles in the changing period

How can we better interpret the rapid semantic change in 1870–1890? One way to do so is by examining the neighbor terms that changed the most in their association with JUDEL (i.e., the collection of JUDEL terms, as described above) in each period.

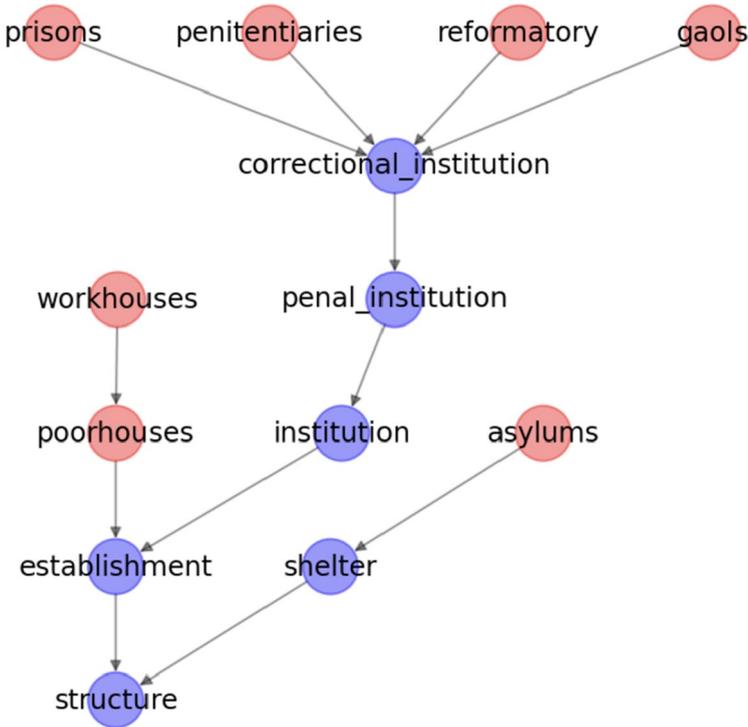


Fig. 9 The *structure* WordNet sub-taxonomy of JUDEL neighbors, following the same format as Fig. 8. (In this case, we understand *structure* as being a proxy for *institution*, as each of the neighbors are either descended from *institution* or are of its parent, *structure*.)

semantic network using WordNet [41], allowing us to break down neighbor terms by word category:

These visualizations suggest some potentially interesting interpretations of people’s attitudes. We observe that this network allows us to partition nearly all terms into a small number of (non-overlapping) categories, like “person”, “behavior”, and “institution”. (See Section A for details on how and why we suggest this partition.) Below, we present a manual analysis of several articles retrieved from PHN, each of which includes at least one JUDEL term and one of the semantic neighbors, to provide additional context on how neighbor terms relate to *juvenile delinquency*. We do not correct OCR (Optical Character Recognition) errors present in newspaper article texts when providing excerpts, to give an appropriate representation of the quality of texts used in analyses, as our intention is to maintain the authenticity of the original content (including any imperfections introduced during the digitization process).

Terms associated with institutional view

In 1860, three articles of the New York Times and three articles of the New York Daily Tribune discussed a bill to set up a reformatory school for juveniles. These articles collectively underscored society’s acknowledgment of the necessity

to regulate and reform juvenile delinquency through dedicated institutions. The legislative action signified society's growing recognition of JUDEL as a pressing concern, leading to a need for legal institutions to regulate it (Fig. 9).

On 2/17/1860, a New York Times article stated that the legislature "introduced a bill to incorporate the Society for the Reformation of Juvenile Delinquents in New-York." We can find more details regarding how reformatories were used for children of different ages in New York Daily Tribune (2/9/1860): "Children between the ages of 7 and 14, who by in writing of their parents or guardians may be surrendered or intrusted to it for preservation or reformation. Children between seven and fourteen of age who may be committed to the care of sues Corporation as idle or children by order of any magistrate empowered to make committal of such children. Children of like age who may be transferred at the option of the Superintendent or Overseers of the Poor from any county or other public Poor-House or Alms-Route to Corporation. The Trustees or Managers shall have power to the children committed to their care at such employment and cause them to be instructed in such branches of knowledge as shall be suitable to their years and they shall have power to bind out toe said children with their consent as apprentices or servant daring their minority."

In addition, The Irish Times and Daily Advertiser (1/7/1861) stated reformatory attendance should be reinforced: "the spread of juvenile delinquency partly owing to the want of proper in dust rial correctional and Reformatory Schools and partly to the want of authority in magistrates to command the attendance of children in such schools." From these articles, we can see multiple institutions emerged to regulate juvenile delinquency-related issues. On 6/7/1862, The Observer described how juveniles were sentenced to reformatory school¹⁰: the judge "sentenced [juvenile A] and [juvenile B] to be detained in the Feltham Industrial School for Middlesex for 3 years and [juvenile C] to be imprisoned for 14 days in the House of Correction and afterwards detained in the Hume in the East Reformatory Bow for four years the parents of the boys to contribute towards their maintenance." These articles emphasized the need to reinforce reformatory attendance due to the spread of juvenile delinquency. They attributed this spread to the lack of appropriate correctional and reformatory schools and insufficient authority for magistrates to enforce attendance. This demonstrates a growing awareness of the role of reformatory schools in addressing the rise of juvenile delinquency.

The news also praised the impacts of these reformatory schools. Compared with the detrimental effects of conventional jails, these reformative efforts could provide youth with education and trade skills, empowering them to lead honest lives upon release. On 1/7/1861, The Irish Times and Daily Advertiser noted that "In England Reformatories have already lessened juvenile crime 42 per cent with certainty of still further decrease. As general rule the juvenile criminals are the victims [...] They are taught and trained in crime with much more care and perseverance than others are trained to virtue The young criminal was sent to jail to be inoculated with crime of all descriptions He came out of jail infinitely worse because of his punishment Some boys scarcely thirteen years of age have been thirty times imprisoned and have spent more time within prison at their home or even in the streets."

¹⁰ We replace the names of the respective juveniles with A, B, and C to preserve anonymity.

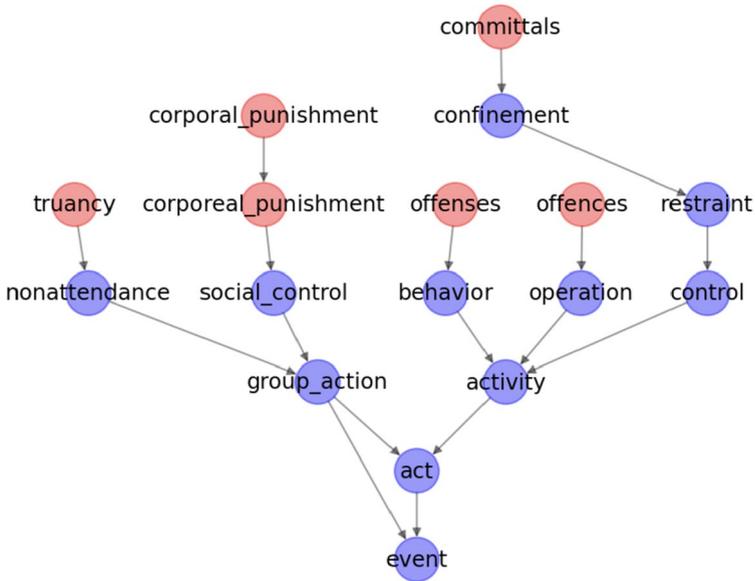


Fig. 10 The *event* WordNet sub-taxonomy of JUDEL neighbors, following the same format as Fig. 8. (In this case, we understand *event* as being a proxy for *behavior*, as the event types in question pertain to either delinquent activities or subsequent responses.)

On 7/8/1861, The Irish Times and Daily Advertiser noted the effect of reformatory facilities for youth: “Good and wise men thought that to place these in jail and to mix them with old offenders was to place them in school of crime. The young came out thoroughly impregnated with evil and ten times more dangerous than before. Reformatories were instituted. In these the young criminal is trained to trade he is taught the most useful rudiments of education at the end of his period of detention he leaves the Reformatory with trade in his hands and knowledge in his head. He need not be forced by poverty to have recourse to crime.”

On 9/27/1862, The Observer noted that “the number of young offenders in the 62 certified reformatories increased in the course of the year 1861 from 3,803 to 4,337 including 186 placed out on license and not yet finally discharged. Now the commitments of persons under 10 will be found to have decreased since 1856 about 43% in England allowing for increase of population. The number steadily diminished from 1850 to 1860 but in 1861 increased above 9% over the previous year and the number of adult commitments increased still more. Of the results of reformatories in the diminution and prevention of crime.”

However, the reformatories could have high costs. On 7/30/1861 The Irish Times and Daily Advertiser stated, “The Grand Jury of the county of Antrim have refused to enter upon any agreement to pay the managers of Reformatory Schools two shillings weekly for each of the criminal juveniles sent from their county to Reformatory [...] juvenile criminals are under restraint and are taught trades or agriculture considerable staff of skilled teachers as well as wardens is required and when the numbers in Reformatory are small the expense per head is greatly increased. The

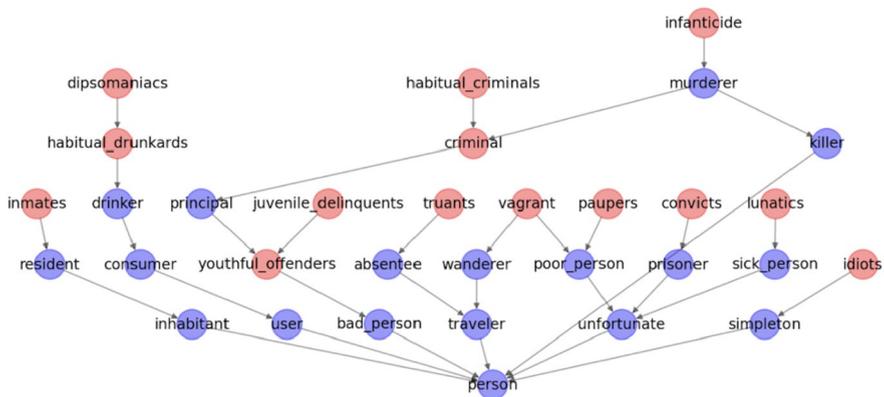


Fig. 11 The *person* WordNet sub-taxonomy of JUDEL neighbors, following the same format as Fig. 8

true way to the comparative cost of Reformatories is to take into account the loss occasioned by habitual thieves and the gain to the community by training thorn to take to honest labour.” This data-driven analysis presented a nuanced perspective on the impact of reformatories and their contribution to the reduction of crime rates.

These discussions surrounding the establishment and functioning of institutions to regulate juvenile delinquency highlight its transformation into a recognized social problem during this era. The legislative actions, operational details of reformatories, and debates over costs collectively underscore the societal concern surrounding JUDEL and the strategies being employed to address it.

Terms associated with behavioral view

Compared to institutional views, we found fewer articles related to behavioral views. On 7/27/1861 The Observer described a “boy was charged with stealing 41 of lead belonging to the rifle corps using the butt at Wormwood Scrubs. Mr Bird solicitor attended to prosecute and said not only did these juvenile delinquents run danger of being shot but in raking for the lead they destroyed the embankments. Notwithstanding previous warning the prisoner was again discovered and on being searched 41 weight of spent bullet were found in his possession. Mr Dayman severely lectured the boy and told him he would inflict the nominal fine of one shilling for the present offence but he or any others were brought there on similar charge [...] will go to prison for seven. Bird however paid the shilling and the prisoner was discharged.” This article provides insight into the handling of specific behavioral instances of delinquency, illustrating the judiciary’s response to such cases (Fig. 10).

On 8/23/1862, The Observer stated “young man named [removed] aged 18 who was said to be well connected was charged with stealing horse and phaeton from Mr Watson livery stable keeper at Highbury.” This incident points to a broader spectrum of juvenile delinquency, extending beyond thefts to more complex criminal activities. The fact that the young man in question is described as well-connected adds a layer of complexity, suggesting that delinquent behavior could be observed

across various societal strata. However, the article lacks further elaboration on the circumstances or motivations behind the crime, leaving the specifics of the behavioral aspect relatively ambiguous. It is important to note that, while these articles provide valuable insights into behavioral aspects for juvenile delinquency, they represent a limited subset of cases and are not comprehensive in addressing the spectrum of behaviors or underlying causes. They do, however, indicate an emerging recognition of the need to discuss and address delinquent behaviors in addition to institutional approaches during the given time period (Fig. 11).

Terms associated with personal view

Who were those juvenile delinquents? Newspaper articles also provide insights into how society perceived and discussed juvenile delinquency from the perspective of the individuals involved, including discussions on the circumstances surrounding young offenders, their motivations, and the proposed solutions to address the issue.

Some articles underscore the adverse impact of their surroundings on these young individuals, revealing a clear correlation between upbringing and delinquency. It is suggested that the absence of proper guidance and support played a pivotal role in their engagement in criminal activities. On 10/22/1861, The Manchester Guardian described that those kids “are without home- and parent- that many roam the streets unchecked and at untimely hours of the night that several are tempted to crime by the ease with which they can dispose of stolen property and by the homes which are provided for them in low lodging-houses that very many have homes which do not deserve the vice and crime.” On 6/7/1862, The Observer stated one reason of juvenile delinquency: “juvenile offenders were moral plague in the metropolis and he believed that in nine cases out of ten such children were driven to crime by the neglect of their parents.” On 2/26/1860, New York Daily Tribune described the connection between intoxication and young men: “here was not one which was not clearly traceable to indulgence in intoxicating liquors. The criminals too were all young men.” Furthermore, the articles offer glimpses into the influence of external factors on these individuals, shedding light on the intricate interplay between personal choices and external stimuli in shaping the conduct of juvenile offenders. On 3/27/1963, The Irish Times and Daily Advertiser stated “the waste and loss caused to juvenile thieves have been estimated at seven million year. Ignorance is the fertile mother of crime and the sums expended on education are an investment to present treble cost in prosecuting juvenile depravity.” On September 16 in 1825, the Manchester Guardian newspaper first time used the term “juvenile delinquent” to refer those young people who committed bad behaviors. It pointed out the “importance of some institution where young who are not thoroughly hardened in having the means of reformation supplied to them.” It then called to support these institutions as they can help put “juvenile delinquents [...] under proper regulations, be furnished with employment and with instruction and thus enabled ultimately to regain that good character which their former misconduct had destroyed.” Thus, these institutions can make sure them “were not further contaminated in the prison [...] with older and more hardened criminals” and “obtaining an honest livelihood” after being released.

	1870s	1880s	1890s
0	drunkards (+229)	infanticide (+228)	corrigible (+291)
1	beneficiaries (+228)	juvenile_offender (+279)	vagabondage (+267)
2	misdemeanors (+351)	flogging (+317)	
3		malefactors (+187)	
4		incarceration (+205)	
5		imprisonment (+240)	

Fig. 12 “Most-changed neighbor terms” table from the relevant period

In 1828, the Manchester Guardians (March 7) reported “in London the number of juvenile delinquents was at present so enormous [...] no less than six shops that have been robbed [in one week].” In 1837, the Manchester Guardians (October 31) indicated that the cause of delinquency was more of education. They found that “young criminals display more astuteness intelligence [...] It is usual to attribute juvenile delinquency to want of education when in truth it is the result of [...] immoral training and vicious instruction devised by the enemies of social order to supply what ought to be provided by the government.” In 1838, the Manchester Guardians (August 10) realized the parental social responsibility of juvenile delinquency, as it noted “the lapse of children into criminal habits is in most cases attributable to the neglect of parents and in others to their avocations compelling them to leave their offspring for great portion of the day without protection look to the general extension of infant schools for the humbler classes as most preventative of juvenile delinquency.” In 1844, the Manchester Guardians (May 3) claimed “the consequence of sending juvenile offenders to prison” was that they “become associates and companions of hardened thieves and [...] they too frequently come out with their vicious propensities strengthened and confirmed to enter on new course of crime [...] the prison thus unfortunately become the school of crime instead of reform.”

Regarding potential remedies, the articles accentuate the necessity of reformation, education, and support to deter subsequent criminal behavior. Reformatories surface as a viable approach to address the challenges faced by these young individuals. The articles commended the positive outcomes associated with reformatories, affirming their capacity to equip young offenders with the means to lead productive lives and steer clear of criminal trajectories. The traditional criminal justice system was deemed as not fit for those young delinquents. The Manchester Guardians (July 21, 1846) were saddened that “those who beginning as juvenile delinquents become habitual criminals are by our present miserable system of prison discipline ever reformed and how hard it is for those who do repent to regain respectable position in society and to earn their bread by honest labour.” As a result, the government committee sought “the establishment of an asylum for juvenile delinquents [...] to confer most upon community and feel assured that it will be the means of reclaiming many from their early errors and enabling them to enter society with the opportunity as well as the desire of regaining their lost position” (The Manchester Guardians, May 3, 1844).

	1860s	1870s	1880s	1890s
0	lunatics	paupers	charity	charitable_institutions
1	poorhouses	criminals	charities	indigent
2	idiots	pauper	benevolent	orphans
3	paupers	charitable_institutions	charitable	institutions
4	asylums	charities	indigent	institution
5	pauper	charitable	orphans	charities
6	inmates	hospitals_dispensaries	institutions	friendly_societies
7	inmates_workhouses	orphans	widows	juvenile_offenders
8	pauper_lunatics	orphan_asylums	juvenile_offenders	hardened_criminals
9	schoolmasters	indigent	young_criminals	reformatory
10	orphans	charity	reformatories	young_criminals
11	almshouses	benevolent	offenders	reformatories
12	almshouse	eleemosynary	solitary_confinement	youthful_offenders
13	orphan	institutions	criminals	juvenile_criminals
14	indigent	institution	habitual_drunkards	criminals
15	deaf_dumb	persons	reformatory	juvenile_delinquents
16	benevolent	individuals	punishments	offenders
17	institution	unincorporated	truancy	jails
18	charitable	reformatories	juvenile_criminals	truancy
19	inmate	juvenile_offenders	penitentiaries	habitual_drunkards
20	unfortunates	youthful_offenders	workhouses	vagrants
21	pensioners	females	punishment	corporal_punishment
22	institutions	drunkards	prisons	truants
23	charitable_institutions	young_criminals	jails	penitentiaries
24	eleemosynary	reformatory	offences	punishments

Fig. 13 Mutual neighbors list for JUDEL and “beneficiaries”

Bills were proposed to establish special institution. The Manchester Guardians (July 2, 1844) noted “to remedy this evil the bill proposes to empower justices of the peace to establish houses of refuge where young offenders on being released from prison may go and learn some useful trade In such institutions as these the juvenile delinquent will be as much as possible removed from temptation and the influence of bad example and will thus have the best opportunity that can be afforded him of retrieving his character.” The U.S. also paid more attention to similar institutions for juvenile delinquency: in 1844, the New York Daily Tribune (July 1) called “If our Legislatures in their wisdom would make provisions for similar Institutions for reclaiming and educating the juvenile delinquents in our Urge cities and towns it would be most important means of saving our country from the vices which degrade so large portion of the population of Europe.” Around 1850, new system reform for

juvenile offenders was proposed as The Observer (June 25, 1848) discussed “the reformation of juvenile offenders by means of religious and industrial training.”

In sum, these articles provide a multifaceted perspective on juvenile delinquency as seen through the lens of the individuals implicated. They underscore the vulnerable circumstances, dearth of guidance, and external influences that contributed to their delinquent conduct. The suggested solutions, including reformatories mirror society’s aspiration to confront these underlying factors and extend assistance to young individuals, facilitating their reintegration into society as responsible members.

Changes of associated terms

In Fig. 12, we display the terms which changed most in their association with JUDEL in each decade. That is, they are the set of all terms that meet the following 3 criteria (for the period of 1870–1900):

1. It was in the top 500 most related terms to JUDEL in the prior decade (e.g., for a term in the 1880s column, it was in the top-500 most related terms in both the 1870s and the 1880s).
2. Its position in the top-500 most-related terms list jumped by more than 95% of terms in that period.
3. Its position in the current decade is in the top-64 most related terms (so its “jump” from the previous period to the current period meant that it “landed” in the top 64).

The number in the parentheses indicates the number of positions by which a term moved up in the ranked list of all the associated terms as compared with the previous decade. For example, the term “drunkards” moves from being the [n -th] most closely associated term in the 1860s to being the [$(n - 229)$ -th] most associated term in the 1870s, suggesting that the incidence of juvenile drunkenness (or some other association between juvenile delinquents and alcohol consumption, e.g., “drunkard” parents with delinquent children) increased substantially during this period.

To interpret the relationship between JUDEL and each of these terms, we generate a “mutual neighbors list” following the methodology described in Section 4.2 to interpret the relationship between JUDEL and each term. For example, the top mutual neighbors between JUDEL and “beneficiaries” in each decade between 1860–1900 are provided in Fig. 13 (with similar figures for a selection of other terms displayed in Appendix B.1). Recall from Fig. 12 that the decade where “beneficiaries” suddenly becomes highly associated with JUDEL is in the 1870s, so we may compare the mutual neighbors list in the 1860s to the succeeding 3 decades to help us determine why this is the case. We observe that, in the 1870s, terms like “charitable institutions” and “charities” start to trend towards the top of the list, and this pattern solidifies in the 1880s and 1890s. Thus, we may infer that the relationship between JUDEL and “beneficiaries” in this period is likely characterized by juvenile delinquents becoming the beneficiaries of charities and associated institutions,

providing one possible explanation for the increased association between JUDEL and “beneficiaries” in this period.

However, this analysis does not always lend itself to such clear or straightforward conclusions (see Appendix B.1); and in any case, it is important to confirm any tentative interpretations of the changing social construction of JUDEL and its relation to associated concepts of interest by consulting their use in actual documents from the relevant periods. We looked at the top term in each period and attempted to understand the relation of the term and JUDEL through individual newspaper articles. Using these terms we identify relevant articles and had the following observations.

Drunkard

When we observe the articles from the 1870s that include both JUDEL terms and “drunkard”, we find that they are discussed in similar contexts regarding the institution of novel penal institutions associated with both kinds of behavior (e.g., juvenile reformatories and inebriate asylum). However, the courts were relatively harsh. Chicago Daily Tribune (1/9/1873) wrote “The Governor suggests that the law under which Juvenile offenders are sent to the Reform School until they reach the age of 18 results in the discharge of many who attain that age [...] and he proposes that the Courts be to sentence juveniles until they are 21 years old. This is open to the objection, so well stated by one of the Judges of the Supreme Court, that it punishes the boy for a petty offence by a sentence covering from six to ten years, from which he attains manhood with a certificate of graduation from a house of correction. The Governor recommends, as an additional State charity, the establishment of an Inebriate Asylum, to which the victims of that disease may voluntarily retire, and to which drunkards, if dangerous, maybe committed. The suggestion is a wise and humane one.”

The New York Times (3/26/1873) emphasized that those kids should be sent to reformatory, and juvenile delinquents should not be put in the same jail as old drunkards: “There have often been herded together in one small and fetid the old and young, black and white, the unfortunate virtuous young girl and the hardened woman of the streets, the old drunkard and the little vagrant. Such dens have become schools of crime and pauperism and the children who have been sent forth from them have come back again in a few years. The reforms throughout the State in this administration, for the past fifty years, have been in the direction of withdrawing these various subjects from the county control. First, the juvenile criminals were removed from county jails and placed in State reformatories.” Similarly, the San Francisco Chronicle (4/20/1874) noted that they should “prevent many youths from being trained up for the maturity of crime by finding appropriate quarters in the jail.” Otherwise, “a corral in which infants, boys, girls, drunkards, maniacs, unfortunate offenders against city ordinances, blasphemers, prostitutes; and hardened criminals are herded together like hogs is the sty.”

Boston Daily Globe (7/15/1873) connected those kids with a family background of drunkards: “The young offenders spring almost entirely from the poor, from the untrained outcasts, such as neglected orphans and the children of drunkards. It is to the lack of sympathy for them [...] that we may look for the cause of their errors.”

The increasing number of drunkards is one reason. The Manchester Guardian (12/18/1872) stated “the large number of cases of drunkenness had increased from 28 percent of the total number of committals in 1868 to 42% in 1872. The total number of committals to this prison in 1872 for drunkenness was 2784 Year by year there had been an increasing number of female drunkards in proportion to the number of males.”

These articles shed light on the concerns and proposed solutions of the time regarding the intertwining issues of juvenile delinquency, alcoholism, and penal institutions. The discussion reflects societal efforts to address these challenges and reform the approach to handling young offenders. We performed a similar analysis for all the other terms in the table by searching for articles that match both a target term (e.g., “beneficiaries”) and at least one JUDEL term and reading the top-ranked articles to understand why those terms were mentioned together with JUDEL. While we were not able to find clear patterns in all the cases, we present our findings for a selected subset of these terms below.

Infanticide

The term “infanticide” was used mainly because alleged infanticide cases were often associated with young/juvenile people (e.g., young mothers who were alleged to have murdered their children that were born as the result of “seduction”). Infanticide has long been viewed as one type of crime/delinquent behavior. For example, in early 1800s, Hartford Daily Courant (6/11/1862) described a story: “A woman calling herself Mrs. Laahcoil an Indian, ins arrested on Friday for causing the death of a named Mary Hamilton, by an abortion. She had sent the body away [...] and the police followed it and then searched her [...] Two other girls were found there in a dying condition from similar criminality. All these young women were victims of the seducers and respectably conceded. The woman Laalicoil has once been tried for an and acquitted [...] Two of her accomplices, a negro, an physician, are also in jail.”

This continues in the late 1800s. The Washington Post (5/15/1882) described that “a young colored woman, was yesterday permitted in the Criminal Court to withdraw her plea of not guilty to the indictment against her for Infanticide and substitute one of guilty of manslaughter. Justice Wylle sentenced her to the Erie county New York penitentiary for two years.” Statistics also show infanticide was one major type of crime committed by youth in the 1800s. San Francisco Chronicle (2/8/1891) stated “Recent statistics show that twice as grave crimes are committed by persons from 30 to 40 years old. In the last year, minors are charged with manslaughters 3 parricides 3 and 44 cases of infanticide.”

The historical usage of the term “infanticide” reveals an interesting relationship between infanticide cases and young or juvenile individuals. This correlation can be attributed to the specific contexts surrounding infanticide incidents, particularly those involving young mothers who resorted to such extreme actions due to circumstances like seduction and societal pressures. Throughout history, infanticide has been viewed as a specific type of crime and delinquent behavior that often intersected with the vulnerability of young individuals.

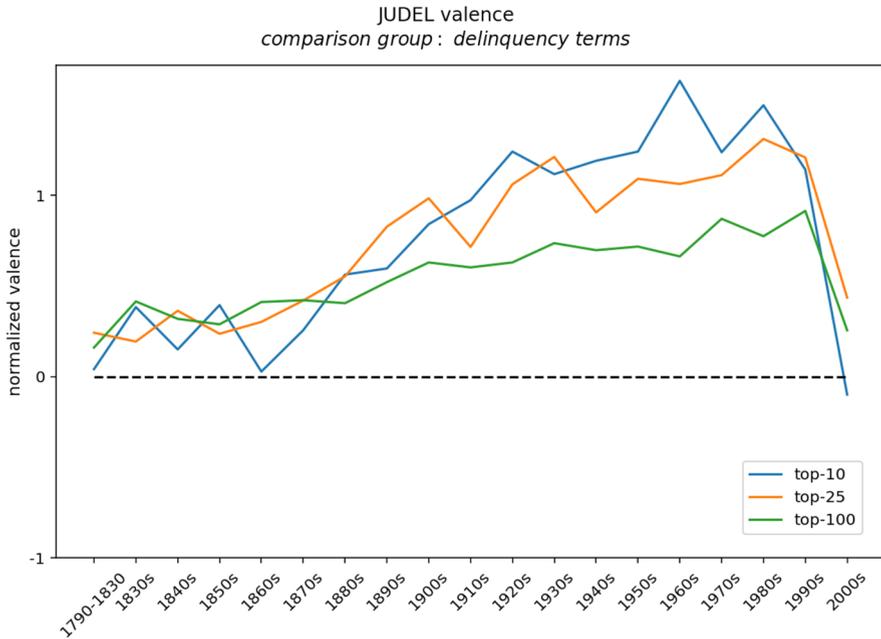


Fig. 14 Valence analysis time series using only “trait” terms. See Section 4.3 for methodological details

Corrigible

The term “corrigible” was used mainly to describe things that could be corrected. From 1870 to 1899, there were only four articles that contain both corrigible and at least one JUDEL term. The term specified in these four articles was “incorrigible” instead of “corrigible”, as the OCR system incorrectly separated “in” and “corrigible”. One article used incorrigible to describe a young kid. Los Angeles Times (6/17/1892) “Clty Marshal Lorbeer took a young In-corrighible before Judge Morton yesterday. He had run away from the Loj Angeles Orphan Asylum and rode up to Pomona on a Southern Pacilc freight train. The lad, only 10 years of age, hunted up Mr. Armour, with whom he lived two years ago, and told him that he paid a brakeman 84 to take him to Arizon but the brakeman put him off here. He then began to tell more lles than he could keep up with and he was taken before Judge Morton; where it was learned that the young fellow [...] had on several occasions ran away, and he had in some way been neglected by his parents. A relative came up front Garvanza, and at her request, the boy was sent to the reform school at Whittier. He was taken to and will be taken to the school today.”

	1860s	1870s	1880s	1890s
0	cooperative (1.22)	cooperative (1.22)	cooperative (1.22)	cooperative (1.22)
1	crabby (-1.49)	steady (1.26)	antisocial (-1.45)	unrealistic (-1.19)
2	tidy (0.92)	crabby (-1.49)	detached (-1.11)	moderate (0.36)
3	steady (1.26)	moderate (0.36)	benevolent (0.70)	neat (1.48)
4	bland (-0.93)	detached (-1.11)	stable (0.11)	tidy (0.92)
5	analytical (0.73)	stable (0.11)	moderate (0.36)	refined (0.82)
6	detached (-1.11)	artificial (-0.94)	sociable (1.07)	reliable (1.75)
7	sociable (1.07)	respectable (1.17)	pleasant (1.71)	charming (1.56)
8	moderate (0.36)	tidy (0.92)	charitable (1.78)	crabby (-1.49)
9	glum (-1.76)	sociable (1.07)	steady (1.26)	direct (0.54)
Avg Valence	0.03	0.26	0.56	0.60

Fig. 15 List of top-10 trait terms associated with JUDEL (by decreasing association), with each term's normalized valence in parentheses

Analysis of public attitude: positive vs. negative construction

Finally, to measure the valence of public sentiment regarding *juvenile delinquency*, we follow the method described in Section 4.3 to compute the valence of the aggregated JUDEL embedding over time, as visualized in Fig. 14.

This approach requires defining a comparison group, a social group in a similar category (e.g., gender, race, ethnicity, social class, etc.) which is used as a foil to normalize terms that relate to the general category, and not the specific group—otherwise, the valence of such category-general terms may obscure the valence of the group-specific terms, making it difficult to distinguish between the positive or negative social construction of the social category versus that of the individual group. The comparison group that we define for JUDEL is the set of JUDEL terms with age-denoting words removed (e.g., “juvenile delinquency” and “young offender” are substituted for “delinquency” and “offender”, respectively), which we refer to as “delinquency terms”.

The quantification of valence enables us to examine to what extent the reporting of JUDEL was positively or negatively characterized in each time period and how the attitude might have changed over time. From the results in Fig. 14, it appears that JUDEL had the most negative valence in the 1860s, which rapidly spiked in the period from the 1870s through the 1880s, and remained consistently high through all remaining periods.¹¹ This tells us that news media sentiment associated with JUDEL went from being relatively similar to that of crime overall to being

¹¹ That is, with the exception of the final period from 2000–2010. However, PHN contains substantially less data from this period, as it extends only through the end of 2006; so it is not clear how seriously the decline in the early 2000s should be taken. We leave the question of whether this trend persists in the context of a larger, updated text corpus with sufficient news data from this period for future work.

considerably more positive following ~1870. Notably, the period of rapidly increasing positive valence is the same period as we observe the most substantial semantic change in JUDEL identified by CCLA in ~1870–1890, indicating that the semantic shift revealed by CCLA in this period is likely characterized by the rapidly increasing positive valence in the same period.

This shift in sentiment could have laid the groundwork for the establishment of juvenile courts in the early 1900s. The increasing positive valence and evolving semantic usage of JUDEL reflect a growing societal inclination towards more compassionate and rehabilitative approaches for addressing young offenders. Taking this interpretation into account, the establishment of the first juvenile courts in 1900 could potentially be attributed to a preceding 10–20 year period marked by increasingly sympathetic or lenient societal sentiments. This may have influenced the formation of institutions aimed at addressing juvenile delinquency with a more compassionate approach. Thus, the emergence of juvenile courts could be seen as a response to the changing narrative around juvenile delinquency, shaped by a desire to provide specialized and considerate treatment for young individuals involved in criminal activities.

We may decompose the valence scores into the list of the top- k most-related trait terms and their respective valences used to compute the valence scores (displayed in Fig. 15 for $k = 10$). These terms provide limited support for the “sympathetic coverage” hypothesis outlined above, but leave substantial room for interpretation. For example, the highest-valence terms in the 1880s (which sees the greatest jump in average valence scores – note that these are the same values as visualized in Fig. 14, “top-10”) are “charitable” (1.78) and “pleasant” (1.71), which do not appear in the top-10 list of the preceding 2 decades. Favoring the hypothesis, one could argue that JUDEL’s association with “pleasant” may indicate that juvenile delinquents are less often seen as being malevolent and deserving of punishment, and the association with “charitable” could be linked to an increasing number of charitable acts or institutions serving them (i.e., a direct manifestation of individual or societal sympathy, respectively, towards these delinquents). However, the opposite argument could be made as well: “antisocial” (−1.45) also makes its first appearance on this list in the 1880s, which indicates a conception of juvenile delinquents as being less cooperative or ethical members of society, and by extension, potentially less prone to rehabilitation or deserving of sympathy. Thus, additional work is required to further explore the “sympathetic coverage” hypothesis.

Discussion

We employ a variety of visualizations and methodologies to uncover how public perceptions of juvenile delinquency evolved, displaying the social construction of juvenile delinquency and leading to the formulation of the juvenile justice system. Below, we describe our findings with respect to each of the research questions outlined in Section 4.

How has the meaning of juvenile delinquency changed over time?

By examining the normalized document frequencies of key terms related to juvenile delinquency over centuries, prominent changing time periods were revealed. These periods show shifts in the terms used to describe juvenile delinquency, indicating changing public attitudes and discourse over time. In the early 1800s, terms like “young offenders” and “juvenile offenders” were used without clear definitions or specific behaviors associated with them. From 1800 to 1825, news sources did not distinctly differentiate between juvenile and adult offenders, indicating that society had not yet perceived juvenile delinquency as a distinct issue or social problem. The emergence of the term “juvenile delinquency” around the mid-1800s marked a turning point. The Manchester Guardian newspaper, in 1825, started using the term “juvenile delinquent” to describe young individuals engaged in bad behaviors. 1825 through 1860, juvenile delinquency was coined to describe a specific group: they were youth who committed delinquent behaviors. The gradual transition from “young” to “juvenile” reflected a shift in the understanding of youth as not only age-related but also as psychologically immature or undeveloped. Finally, using CCLA, we found that the period in which the meaning of juvenile delinquency most rapidly changed is in the period of ~1870–1890.

How can this semantic shift inform our understanding of how juvenile delinquency has been constructed as a social problem?

The evolution of terminology played a pivotal role in paving the way for the establishment of a distinct juvenile court system. The transition from ambiguous terms like “young offenders” to the more precise term “juvenile delinquency” was instrumental in reshaping the perception that young individuals were less accountable than adults, more capable of transformation, and merited a different approach within the criminal justice framework. Following the establishment of the juvenile court around 1900, the term “juvenile delinquency” rapidly gained prominence as the primary label for describing related behaviors. This shift was mirrored by both legal institutions and public attention, which began to emphasize the behavioral aspect.

Additionally, the fact that CCLA shows the greatest semantic shift in 1870–1890, preceding the development of the first juvenile courts in the US by less than 10 years, may not be a coincidence. It is possible that the changing meaning of juvenile delinquency in the news media may have reflected deeper changes in the underlying social conceptualization of juvenile delinquency, which themselves may have paved the way for the emergence of the juvenile courts. We used information retrieval techniques to find relevant articles from this period in order to “zoom in” and better understand what changes might have occurred at this time. Our analysis of these results indicates that these social changes contribute to the construction of juvenile delinquency both semantically and socially in this period, but they are currently inconclusive.

How have public attitudes towards juvenile delinquency changed over time?

Using topical analysis, we identified three types of views that display the social construction of juvenile delinquency among newspaper articles. First, our analysis of institutional views underscores the growing societal recognition of juvenile delinquency

as a social concern (e.g., the establishment of reformatory schools emerges as a key response to address juvenile delinquency). Legislative actions, operational details, and discussions about the costs and benefits of these institutions reveal society's acknowledgment of the need to intervene in the lives of young offenders through education and vocational training. The contrast between the rehabilitative approach of reformatories and the detrimental effects of conventional jails become evident, presenting a shift from punitive measures to more holistic interventions. This shift represents an important transition in how society approached and understood juvenile delinquency, focusing on reform rather than mere punishment. Second, our examination of behavioral views provides insights into the specific behaviors associated with juvenile delinquency: while these discussions are limited in scope, they exemplify the broader criminal activities that some young individuals were involved in during the given period. News articles highlight the judiciary's approach to handling these specific instances of delinquent behavior, showcasing the legal system's evolving responses to such cases. Finally, our exploration of personal views delves into the individual lives of those involved in juvenile delinquency. The analysis reveals how societal perspectives evolved regarding the underlying factors influencing delinquent behavior: the news articles offer glimpses into the circumstances that led young individuals to engage in criminal activities, including family backgrounds, lack of education, and environmental influences. The recognition of external factors influencing these behaviors add complexity and nuance to the public's understanding of juvenile delinquency.

Our valence analysis findings illustrate the changing emotions associated with juvenile delinquency over time. In the 1860s, it was viewed quite negatively, but from the 1870s through the 1880s, sentiments towards it sharply improved and remained consistently positive thereafter. Notably, this period of notable positive shift aligns closely with the era of significant semantic change in JUDEL identified by CCLA between 1870 and 1890, suggesting a potential link between shifting societal perceptions around juvenile delinquency and the establishment of specialized institutions like juvenile courts.

How can these changes in public attitudes inform our understanding of how juvenile delinquency has been constructed as a social problem?

The increasingly positive sentiment preceding the establishment of these courts may reflect a changing societal attitude towards embracing more compassionate and rehabilitative approaches for addressing young offenders. It is plausible that this evolving sentiment influenced the creation of institutions tailored to the needs of juvenile delinquents. Therefore, the emergence of juvenile courts can be interpreted as a response to the shifting narrative surrounding juvenile delinquency, driven by a desire to provide specialized and empathetic intervention for young individuals entangled in criminal activities.

Using concept search, we identified relevant news articles to further explore how news media described juvenile delinquency. These articles highlight the historical context and perspectives on the interrelated issues of juvenile delinquency, alcoholism, and penal institutions during the late 1800s. Analyzing articles from that period reveals the common discussions and proposed solutions surrounding these topics. Articles from the 1870s that include both JUDEL terms and "drunkard" show a parallel discussion about

the establishment of new penal institutions for both juvenile offenders and individuals struggling with alcoholism. This suggests a perceived link between certain environmental factors and the emergence of delinquent behavior among young individuals. The term “infanticide” reveals a consistent pattern of association with young or juvenile individuals. Infanticide cases often involved young mothers who resorted to extreme actions due to circumstances like seduction, reflecting a specific type of crime intertwined with the vulnerability of youth. In contrast, the term “incorrigible” is used to draw the line between correctable and uncorrectable individuals or behaviors, highlighting the potential for change or reform in certain individuals (particularly those who are still in their developmental stages). Overall, these historical articles provide insight into the societal concerns of the time and the approaches taken to address juvenile delinquency, substance abuse, and other related issues. The discussions reveal evolving perspectives and solutions for handling these challenges and reforming the treatment of young offenders.

Related work

News data have been used to study many foundational topics in social science such as gender [47], politics [48], and culture more broadly [49]. Our work adds to this line of research with the exploration of the social construction of juvenile delinquency, which has not been studied before. Our results are generated using the underlying analysis system we created in [36]; but in this work, we develop several novel analysis techniques and integrate them with the previous system, including our approach to word embedding term aggregation and joint similarity metric; and we also extend the previous system in several ways, such as nearest-neighbor list functionalities and associated WordNet visualizations and all valence analysis work. Furthermore, while our previous work reported some preliminary results in studying the depiction of juvenile delinquency in historical newspaper texts, in this work we conduct a far more thorough and detailed investigation of the historical social construction of juvenile delinquency using both the existing analysis functionalities and novel techniques developed for this work.

The computational techniques we apply in this work build on several categories of existing work. First, WordNet [41] has been broadly applied as a resource for studying relations between words based on the paths connecting them on WordNet, perform word sense disambiguation [50, 51], text categorization [52] and text clustering [53]. We used WordNet to categorize the terms associated with juvenile delinquency automatically into meaningful high-level categories such as people and institution to facilitate interpretation of semantic dynamics of juvenile delinquency. Word embeddings [54, 55] have also been widely used to perform semantic analysis of text data, including, e.g., construction of a concept lexicon [56], text classification [57], and the discovery and quantification of group stereotypes [58]. Our use of word embeddings is mainly for analysis of associations of terms with our target concept, *juvenile delinquency*, to study how it has been socially constructed over time in news media. The CCLA analysis method was first proposed in [42] using word embeddings, but the original method could only be used to analyze individual keywords. We extended it to support concept analysis where a concept can be expressed with multiple terms

with consideration of the uncertainty of the terms expressing a concept. Finally, our valence analysis method builds on the approach developed by Charlesworth et al. [44], where the authors used historical word embeddings to analyze representations of social groups in literary content and measured the positive versus negative valence of these representations over time. There are several important differences between their experiments and ours: while they performed comparative analysis of different demographic groups using historical literary data (Google Books), our analysis compared juvenile delinquency to criminality in general using historical newspapers (PHN), and also integrated key findings from CCLA in order to build a more comprehensive picture of how the valence results relate to broader patterns in social construction.

Conclusion and future work

We conducted the first in-depth study of social construction of juvenile delinquency by using tools from text mining and computational semantics to analyze a massive collection of historical news data. From these articles, we explored changes in the meaning and use of “juvenile delinquency” and other related terms, analyzing their semantic dynamics using the CCLA method and discovering the principal “change point” that occurred between ~1870–1890. Following this discovery, we analyzed the associated terms with the concept of juvenile delinquency around this period to reveal more specific details. Specifically, we extracted the terms that were introduced in that period when discussing juvenile delinquency and used WordNet to categorize those words into three categories (people, institutions, and behaviors), which helped us to interpret the discussion of juvenile delinquency. We further developed an embedding-based semantic distance analysis to extract additional “mutual neighbor” terms that can help explain why certain terms increased in their association with juvenile delinquency. Those results enabled us to further manually examine related articles to gain insights about the reporting of juvenile delinquency. Finally, we analyzed the sentiment of the terms associated with juvenile delinquency over time, finding that the change point identified by CCLA corresponds to a rapid change towards a less negative construction of JUDEL (relative to crime in general). With these computational methods, we were able to obtain some interesting insights about the social construction of juvenile delinquency that can facilitate further study of this problem.

The computational techniques we employed in this study have offered a rapid and efficient approach to navigate through extensive collections of newspaper articles, enabling us to uncover the evolving narratives and perceptions surrounding *juvenile delinquency* over the span of centuries. Through the application of CCLA and valence techniques, we delved deeper into the semantic nuances, revealing the chronological evolution of these narratives and their alignment with prevailing public sentiments. Our utilization of computational methods exemplifies a promising avenue for investigating complex social phenomena by harnessing the vast troves of historical newspaper data. The methods we used include word-embedding based semantic representation, concept modeling using term aggregation, cross-context lexical analysis (CCLA; including an extension of CCLA for concept-level analysis that we proposed), WordNet-based concept interpretation, and valence analysis. All

these techniques can be potentially applied to support research on the social construction of other concepts using historical news data.

Our work also reveals some limitations of the current natural language processing techniques, which can be improved in the future. First, while the lexical approaches that we have used are general and powerful, they were not able to help researchers further interpret the results directly. As such, we relied on the manual search for relevant articles. In the future, it would be useful to further improve the system that we have used to better support users in searching for relevant content and interpreting the results. We anticipate that the recent development of large language models (LLMs) can be leveraged to automate the summarization of relevant content, thus enabling more effective/efficient interpretation of the insights obtained using our computational methods. Another challenge is the OCR errors unique to historical newspaper data: the OCR tools that we have used were not able to correct many errors, causing serious challenges for humans to digest the content of the old news articles in the collection. In the future, it would be useful to further explore more advanced computer vision techniques (as well as, e.g., leveraging LLMs) to improve the quality of OCR.

Although the focus of the paper is on juvenile delinquency, the computational methods we used and the way we combine those methods are quite general, and can thus be directly applied to study other social science issues. In the future, we are planning to implement a publicly accessible system with all these methods available for social science researchers to use via a graphic interface, building on our previous work [36], which we hope will allow more social science researchers to use the methods we developed and deployed in this work in order to study the historical social construction of other concepts of interest to the broader social science community.

Appendix A: WordNet categorization

We categorize the JUDEL neighbors in the semantic network visualized in Fig. 8 by partitioning them into the set of $n = 3$ categories (parent nodes) that best satisfy the following criteria:

1. Categories must have non-overlapping JUDEL neighbors as children.
2. Categories should be defined at the appropriate level of abstraction to provide both internal semantic consistency and meaningful distinctions between categories.

(1) is necessarily true of any valid partition, but (2) is subjective. However, it is nonetheless important for meaningful categorization—without it, the partition would simply consist of a single category, “entity”, followed by $n - 1$ singletons. As such, the partitioning into the three categories discussed in Sect. 6.3 is a combination of objective and subjective criteria: the original hypernym graph (automatically generated by querying WordNet for hypernyms) and criteria (1) are objective, but the number of categories to analyze (i.e., setting $n = 3$) and the appropriate level of abstraction under criterion (2) were subjectively determined to yield a meaningful analysis.

While it is possible to, e.g., set $n = 4$ (yielding a fourth category of “attribute”), or to determine that the term “abstraction” (a parent hypernym of both the analyzed

“event” category and hypothetical “attribute” category) is at a suitable level of abstraction under criterion (2), we elected to set $n = 3$ and define a more concrete categorization (under criterion (2)). This follows our subjective determination that the potential fourth category (“attribute”) was insufficiently conceptually uniform (e.g., we did not find it clear why “commitments” and “drunkenness” should fall into the same category), and that “abstraction” was simply too abstract as a category.

Appendix B: supplemental results

“Explanation Lists”

See Figs. 16, 17 and 18.

	1860s	1870s	1880s	1890s
0	criminals	criminals	criminals	criminals
1	vagrancy	punishments	offenders	young_offenders
2	corporeal_punishment	habitual_drunkards	young_criminals	corporal_punishment
3	convicts	offenders	habitual_drunkards	incorrigible
4	penitentiaries	drunkenness	young_offenders	offenders
5	gaols	prisons	punishment	punishments
6	pauper	offenses	penitentiaries	petty_offenses
7	convict	garotting	solitary_confinement	malefactors
8	punishments	crime	infanticide	prisons
9	punishment	unconvicted	incorrigible	offenses
10	felons	offence	punishments	crimes
11	imprisonment	crimes	offenses	prison
12	drunkenness	offense	offender	trivial_offenses
13	offence	punishment	malefactors	heinous_offense
14	solitary_confinement	magistrates	crime	incarceration
15	incorrigible	prisoners	imprisonment	culprits
16	sentence	deterrent	offence	crime
17	offender	cruelty	crimes	punishment
18	crime	solitary_confinement	juvenile_offenders	offender
19	offense	offender	birching	solitary_confinement
20	refractory	malefactor	misdemeanors	drunkenness
21	hogging	infanticide	floggings	sentencing
22	habitual_drunkenness	incarceration	whippings	punishment_meted
23	flogged	murderers	offense	malefactor
24	sentencing	gibbets	culprits	punished

Fig. 16 Mutual neighbors list for JUDEL and “flogging”

	1860s	1870s	1880s	1890s
0	convicts	criminals	criminals	hardened_criminals
1	convict	offenders	young_criminals	prison
2	prisons	prison	solitary_confinement	incarcerated
3	prison	crime	punishment	penitentiary
4	penitentiary	culprits	incarcerate	crimes
5	sentences_commuted	unconvicted	prison	commitment
6	imprisonment	offence	prisoners	prisoners
7	punishment	offense	penitentiary	lunatic
8	bridewell	penitentiary	offender	crime
9	gaol	prisoners	commitment	solitary_confinement
10	convicted	commitment	imprisonment	punishment
11	offence	punishment	crime	young_criminal
12	committed	transgressors	offence	inmate
13	penal_servitude	solitary_confinement	confinement	convict
14	crimes	offender	misdemeanant	incarcerations
15	imprisonments	incarcerated	offense	sentencing
16	culprits	sentenced	committal	committal
17	imprisoned	imprisoned	incarcerated	youngers
18	commitment	jail	durance_vile	culprit
19	solitary_confinement	murderer	murderers	committed
20	insane	culprit	convict	felon
21	sentence	felon	accomplices	murderer
22	expiating	malefactor	committed	abduction
23	sentence_commuted	sentencing	brutalities	durance_vile
24	sentenced	heavily_ironed	inmate	jail

Fig. 17 Mutual neighbors list for JUDEL and “incarceration”

	1860s	1870s	1880s	1890s
0	vagrancy	vagrancy	offenders	penal
1	offences	offences	penitentiary	offenders
2	penitentiary	offences	solitary_confinement	offences
3	rogue_vagabond	criminal	punishment	petty_offenses
4	convict	punishments	offenses	bridewell
5	prison	penal	criminal	offences
6	penal	misdemeanour	offences	punishments
7	felony	prison	prison	penitentiary
8	sentences_commuted	offence	offender	prison
9	convicted	convicted	prisoners	flogging
10	offence_punishable	offense	incarceration	trivial_offenses
11	penal_servitude	penitentiary	punishments	incarcerated
12	misdemeanour	commuted	offence	incarceration
13	punishment	felony	correction	commitment
14	offence	prisoners	penal	vagrancy
15	committed	misdemeanor	commitment	prisoners
16	imprisonments	treason_felony	magistrates	solitary_confinement
17	prosecutions	punishment	vagrancy	punishment
18	punishable	commitment	misdemeanour	offender
19	imprisoned	correction	misdemeanors	convict
20	solitary_confinement	committed	confinement	sentencing
21	sentence	commuting	punishable	incarcerations
22	sentence_commuted	solitary_confinement	offense	correction
23	sentenced	offender	felony	committal
24	commuted	incarcerated	bastardy	punishment_meted

Fig. 18 Mutual neighbors list for JUDEL and “imprisonment”

Acknowledgements This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC) through allocation TG-SES170016. We thank the XSEDE Extended Collaborative Support Service (ECSS) program. This work also made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign. This work also utilized resources supported by the National Science Foundation's Major Research Instrumentation Program, grant number 1725729, as well as the University of Illinois at Urbana-Champaign. We thank the Cline Center for Advanced Social Research and the University of Illinois at Urbana-Champaign for providing access to ProQuest Historical NewspapersTM. Finally, we express our appreciation to Alan Craig, Sandeep Puthanveetil Satheesan, and Bhavya for both their feedback on this draft and their contributions to the text analysis system used in this work (as described in our earlier publication, Satheesan et al. [36]).

Data availability Statement The text dataset used to generate all results is the ProQuest Historical NewspapersTM (PHN) collection, a proprietary dataset which we received through the Cline Center for Advanced Social Research and the University of Illinois library, as detailed in Section 5. For more information about PHN, please refer to the following webpage: <https://about.proquest.com/en/products-services/pq-hist-news/>.

Declarations

Conflict of interest statement On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Becker, H. (1963). *Outsiders: Studies in the sociology of deviance* (New York). Art Worlds (Berkeley)
2. Kincheloe, J. L. (2005). *Critical Constructivism Primer*. P. Lang. New York, NY.
3. King, P. (2006). *Crime and Law in England, 1750–1840: Remaking Justice from the Margins*. Cambridge University Press.
4. Adoni, H., & Mane, S. (1984). Media and the social construction of reality: Toward an integration of theory and research. *Communication Research*, 3, 323–340. <https://doi.org/10.1177/009365084011003001>
5. Gans, H. (1979). *Deciding What's News*. New York: Random House.
6. Hydén, L.-C. (1993). The social construction of juvenile delinquency: Sailing in cold or hot water. *Young*, 1(3), 2–10.
7. Griffiths, P. (2017). Juvenile delinquency in time. In: *Becoming Delinquent: British and European Youth, 1650–1950*, pp. 35–52. Routledge.
8. Sanders, W.B. (2018). *Juvenile Offenders for a Thousand Years: Selected Readings from Anglo-Saxon Times to 1900*. UNC Press Books.

9. Allderidge, P. (1979). Management and mismanagement at bedlam, 1547–1633. In C. Webster (Ed.), *Health, Medicine and Mortality in the Sixteenth Century* (pp. 141–164). Cambridge: Cambridge University Press.
10. Papers, B.P. (1817). British Parliament 1817. Police of the Mktropoliis. Hansard 1803–2005. <https://api.parliament.uk/historic-hansard/commons/1817/jul/07/police-of-the-mktropolis>
11. Office, H. (1846). Criminal Tables for the Year 1845.-England and Wales.
12. Krisberg, B. (2005). *Juvenile Justice: Redeeming Our Children*. Thousand Oaks, CA: Sage Publications.
13. Hawes, J. M. (1971). *Children in Urban Society: Juvenile Delinquency in Nineteenth-Century America*. New York: Oxford University Press.
14. Norman, S. (1972). The youth service bureau: A key to delinquency prevention. National Council on Crime and Delinquency
15. Crosby, O. (2017). Factors That Contribute to Juvenile Delinquency.
16. Jamrozik, A., & Nocella, L. (1998). *The Sociology of Social Problems: Theoretical Perspectives and Methods of Intervention*. Cambridge University Press.
17. Thio, A., & Taylor, J. (2011). *Social Problems*. Jones & Bartlett Publishers.
18. Blumer, H. (1971). Social problems as collective behavior. *Social Problems*, 18, 298–306.
19. Searle, J. R. (1995). *The Construction of Social Reality*. Simon and Schuster.
20. Berger, P., & Luckmann, T. (1966). *The Social Construction of Reality*. Garden City, NY: Anchor.
21. Link, M. W., & Oldendick, R. W. (1996). Social construction and white attitudes toward equal opportunity and multiculturalism. *The Journal of Politics*, 58(1), 149–168.
22. Schneider, A., & Ingram, H. (1993). Social construction of target populations: Implications for politics and policy. *American political science review*, 87(2), 334–347.
23. Schneider, A. L., & Ingram, H. M. (2005). *Deserving and Entitled: Social Constructions and Public Policy*. Suny Press.
24. Schneider, A., & Ingram, H. (1993). Social construction of target populations: implications for politics and policy. *The American Political Science Review*, 87, 334–347.
25. Barak, G. (2013). Media, Process, and the Social Construction of Crime: Studies in Newsmaking Criminology. Routledge.
26. Barak, G. (2007). *Violence, Conflict, and World Order: Critical Conversations on State-Sanctioned Justice*. Lanham, MD: Rowman & Littlefield.
27. Barak, G. (2007). Doing newsmaking criminology from within the academy. *Theoretical Criminology*, 11(2), 191–207.
28. Herman, E., & Chomsky, N. (1988). *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.
29. Bagdikian, B. G. (2004). The new media monopoly. *Theoretical Criminology*, 11(2), 191–207.
30. McChesney, R. (2004). The Problem of the Media: U.S Communication Politics in the 21Century. Monthly Review Press, New York.
31. Fishman, M. (1978). Crime waves as ideology. *Social Problems*, 25(5), 531–543.
32. Fishman, M. (1980). *Manufacturing the News*. Austin, TX: University of Texas Press.
33. Hall, S. (1979). Culture, the media and the 'ideological effect'. In J. Curran, M. Gurevitch, & J. Woollacott (Eds.), *Mass Communication and Society*. London: Sage Publications.
34. Entman, R., & Rojecki, A. (2000). *The Black Image in the White Mind: Media and Race in America*. Chicago: University of Chicago Press.
35. Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24, 395–419.
36. Sathesan, S.P., Bhavya, Davies, A., Craig, A.B., Zhang, Y., & Zhai, C. (2022). Toward a big data analysis system for historical newspaper collections research. In: Proceedings of the Platform for Advanced Scientific Computing Conference. PASC '22. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3539781.3539795> .
37. Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis lectures on human language technologies*, 10(1), 1–309.
38. Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2021). Recent advances in natural language processing via large pre-trained language models: A survey. arXiv preprint [arXiv:2111.01243](https://arxiv.org/abs/2111.01243).

39. Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: a Practical Introduction to Information Retrieval and Text Mining*. Morgan & Claypool.
40. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* **26**.
41. Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.
42. Massung, S.A.: Beyond topic-based representations for text mining (2017). UIUC Dissertation
43. Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, 10–32.
44. Charlesworth, T. E., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from google books. *Proceedings of the National Academy of Sciences*, *119*(28), 2121798119.
45. Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, *45*, 1191–1207.
46. Abrams, L. S. (2013). Juvenile justice at a crossroads: Science, evidence, and twenty-first century reform. *Social Service Review*, *87*(4), 725–752.
47. Geertsema-Sligh, M. (2019). Gender issues in news coverage. *The International Encyclopedia of Journalism Studies*, 1–8.
48. Ashfaq, M., Shahid, N. --A., & Zubair, J. (2022). New media and political protest: Framing analysis of news content on twitter. *Asian Journal of Comparative Politics*, *7*(4), 1177–1189.
49. Raftee, A., Spooen, W., & Sanders, J. (2023). Framing similar issues differently: a cross-cultural discourse analysis of news images. *Social Semiotics*, *33*(3), 515–538.
50. Voorhees, E.M. (1993). Using wordnet to disambiguate word senses for text retrieval. In: *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 171–180.
51. Li, X., Szpakowicz, S., & Matwin, S. (1995). A wordnet-based algorithm for word sense disambiguation. In: *IJCAI*, *95*, 1368–1374
52. Elberrichi, Z., Rahmoun, A., & Bentaalah, M.A. (2008). Using wordnet for text categorization. *International Arab Journal of Information Technology (IAJIT)* **5**(1).
53. Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using wordnet and lexical chains. *Expert Systems with Applications*, *42*(4), 2264–2275.
54. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* **26**.
55. Pennington, J., Socher, R., & Manning, C.D. (2014). Glove: Global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.
56. Park, D., Kim, S., Lee, J., Choo, J., Diakopoulos, N., & Elmqvist, N. (2018). Conceptvector: Text visual analytics via interactive lexicon building using word embedding. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 361–370. <https://doi.org/10.1109/TVCG.2017.2744478>
57. Selva Birunda, S., & Kanniga Devi, R. (2021). A review on word embedding techniques for text classification. *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA, 2020*, 267–281.
58. Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), 3635–3644. <https://doi.org/10.1073/pnas.1720347115><https://www.pnas.org/doi/pdf/10.1073/pnas.1720347115>