

Manuscript Under Review

March 2022

Citizen complaints as an accountability mechanism: Uncovering patterns using topic modeling

Francisco Olivos

Lingnan University, Hong Kong

Patricio Saavedra

Universidad de O'Higgins, Chile

Lucia Dammert

Universidad de Santiago de Chile

Acknowledgements

The authors thank Rafael Bravo from reclamos.cl for openly facilitating all the necessary information about the platform. A preliminary version of this study was presented in the working group on computational social sciences at the Chinese University of Hong Kong. The authors are also grateful for the comments made by the editor, three anonymous reviewers, Alexis Sossa, Catalina Ortuzar, and Nicolas Rodriguez.

*Correspondence concerning this article should be addressed to Francisco Olivos, 222 Dorothy Y L Wong Building, Department of Sociology and Social Policy, Lingnan University, Tuen Mun, Hong Kong. Email: francisoolivosrave@ln.edu.hk.

Citizen complaints as an accountability mechanism: Uncovering patterns using topic modeling

Abstract

OBJECTIVES. Citizen complaints are considered by policing researchers as an indicator of police misconduct, and a proxy of police-community relations. Nevertheless, US and EU-based studies tend to focus on sustained complaints as reported by official agencies and officer-based correlates. Using the case of Carabineros, the Chilean militarized police force, this study examines (a) latent topics contained in a large set of complaints against the police on a digital platform, and (b) the change of those topics across time and (c) by complainants' educational level.

METHODS. We use novel computational natural language processing techniques to identify latent themes across the corpus of complaints ($N=1,623$), hosted on an online forum from 2013 to 2020.

RESULTS. Our findings show eight latent themes across the corpus. Among others, these themes were related to police effectiveness, police misbehavior, and a master frame of institutional crisis that has significantly grown over the last year. Additionally, differences in the prevalence of topics by complainants' educational level were also found.

CONCLUSIONS. Thus, the findings contribute to the enterprise of opening the black box of complaints against the police and highlighting opportunities for social accountability in a developing country.

Key words: Complaints, accountability, police, topic modeling, Chile.

Citizen complaints as an accountability mechanism: Uncovering patterns using topic modeling

1. Introduction

In recent decades, the study of citizen complaints against the police has substantially contributed to understanding police misconduct and the relationship between them and communities. The focus has been on the process of filing a complaint and in complaints as outcomes (Ariel, Farrar, and Sutherland 2015; Cao and Huang 2000; Hedberg, Katz, and Choate 2017; Pryor et al. 2019; Schuck and Rabe-Hemp 2016; Smith and Holmes 2003). In the complaints as outcomes approach, the literature leverages complaints in two ways: first as a proxy of police-community relations, and second, as a quantifiable indicator to measure the performance and impact of different reforms. Although it provides relevant insights on the diagnosis of citizens' satisfaction with the police force, the sole count of complaints does not inform concrete courses of action to improve relationships between police and citizens or the heterogeneities contained in those complaints.

Policing researchers, particularly in developing countries, do not usually have access to complaints' content (Dammert 2019; Frühling 2009a). Therefore, the understanding of complaints heterogeneity has been limited. However, digital media offers a firsthand opportunity to directly study citizen complaints and the relational meanings embedded in them. We build on previous studies suggesting the heterogeneity of complaints against the police (McLean 2019; Terrill and Ingram 2016; Torrible 2018) to examine (a) latent topics contained in a large set of complaints against the Chilean police on a digital platform, and (b) the change of those topics across time and (c) by complainants' educational level. The latter is particularly relevant considering the emphasis on the effect of officer characteristics and a glaring gap on understanding the relationship between complaints and citizen-based features (Terrill and Ingram 2016).

The identification of the nature and content of complaints offers substantial avenues for accountability. Traditional analyses of police governance consider accountability as the overseeing of agencies or agents of the state by other state agencies (O'Donnell 1998). However, in contrast to this horizontal accountability, modern approaches also incorporate forms of social accountability that rely upon civic engagement and are intended to complement and enhance conventional accountability

mechanisms (Bonner 2009; Malena, Forster, and Singh 2004; O'Donnell 1998). Following Peruzzotti and Smulovits (2006), public citizen complaints against the police provide another mechanism of social accountability by (1) shaming public officials, and potentially (2) activating horizontal accountability. Moreover, among social accountability mechanisms, the literature on governance makes a further distinction between vertical and diagonal accountability (Lührmann, Marquardt, and Mechkova 2020). While the former is mediated by elections and includes political parties, the latter incorporates non-state actors, media, and civil society. Such mechanisms generate and amplify the information about the government, holding it accountable. Therefore, opening the “black box” of citizen complaints from a public complaints platform could act as a form of diagonal accountability.

Utilizing the case of the Carabineros, the Chilean militarized police force, we supplement the efforts of policing researchers and police agencies in applying computational methods to identify complaints patterns (United Nations 2011). There are two main reasons why the Chilean case in particular has the potential to improve understanding of citizen complaints about their experience with the police. First, several corruption scandals involving high-ranking officers and serious human rights violations (e.g., United Nations, 2019) after the wave of protests in 2019 have negatively affected Carabineros' public image to such an extent that some experts have called for a major restructuring of the Chilean police force (Comisión Independiente de Reforma Policial 2020). This social outburst, and the underlying malaise, are not unique to the Chilean context and its law enforcement. In the last decade, similar scenarios can be depicted across the globe in diverse places such as Colombia, Ukraine, Hong Kong, or countries involved in the Arab Spring. Hence, the understanding of police complaints in a context of social instability can illuminate the relationship between the police and communities in other countries. Second, Chile combines cultural homogeneity and high socioeconomic inequality (Torche 2007). There are no salient symbolic boundaries in terms of language, religion, race, ethnicity, or region. Nevertheless, Chile is one of the most unequal countries in Latin America, and it has increased further over the last years (Rosa, Flores, and Morgan 2020). Therefore, the effect of complainants' socioeconomic status on the content of complaints is brought into sharper relief.

We use novel computational natural language processing techniques (DiMaggio, Nag, and Blei 2013) to identify latent themes in a large corpus of complaints. Previous analyses of police accountability crises (McLaughlin and Johansen, 2002) suggest that public confidence can only be restored by an independent and verifiable police complaints system. To this date, Chile has neither a formal nor an independent police complaint system. This situation contrasts with other countries (e.g., the United Kingdom and Scandinavian countries) where independent institutions are responsible for overseeing the system and handling complaints made against police forces (Holmberg 2019). Therefore, there is a knowledge gap around the content of citizen complaints. In line with the latter, in this study, we provide an analysis of complaints from an independent and open online forum in Chile (www.reclamos.cl) from 2013 to 2020. On the platform, individuals can publicly complain against any public or private organization from all walks of life, including *Carabineros*.

The analysis of the corpus of information compiled from reclamos.cl leads to a threefold contribution. First, as mentioned, many studies have focused on complaints as a signal of police misconduct. Nevertheless, this perspective does not truly provide a detailed account of what citizens complain about. As Boehme and colleagues (2022) point out, most of the literature on community-police relations uses abstract indicators of negative police-community relations, which subsume constructs such as over- and under-policing, among other distinctions. This study examines the latent themes in a publicly available corpus of complainants, identifying nuances and distinctions in the content of citizen complaints. Understanding the nature of complaints is especially relevant if we consider that the literature on Latin American police misconduct has been limited mainly because the local police forces provide no access to the necessary information (Frühling 2009a), including the content of complaints. Therefore, this study uses a bottom-up approach, enabling us to overcome the lack of accountability that characterizes Latin American police forces.

Second, the increasing use of internet platforms offers new channels for monitoring police complaints and exerting diagonal accountability. Although they are not legally sanctioned, digital platforms reduce the costs of filing a complaint (Ba 2020) and provide public pressure on institutions to act (Deseriis 2020). Since it is online, generating public awareness of the events motivating the

complaint is one of the salient incentives to complain through this channel. Complaints from individuals who decide to fill out a form at the police headquarters, make a case in court, or use public media may differ in content because of their access, scope, and costs. This analysis uses a unique corpus of online complaints against the police that has not previously been examined by policing researchers. Moreover, in contrast to the media, citizen complaints platforms are less limited by political and economic interest (Bonner, 2009). In recent analyses of the Chilean case, policing researchers have shown that the media can be strategically used to justify high levels of police violence and oppose criticism of human rights abuses (Bonner and Dammert 2021). This situation is even more critical in countries with a high level of media ownership concentration, as is the case in Chile (Mellado et al. 2012). Thus, digital complaint platforms offer an alternative, independent, external, and civilian mechanism of diagonal accountability.

Finally, the few studies that utilize qualitative methods to reveal the meanings of complaints or complaining processes made by different actors usually rely on small cross-sectional samples of interviews (Galovic et al. 2016; Porter, Prenzler, and Fleming 2012; Reynolds, Fitzgerald, and Hicks 2018). In our analysis, we directly examine a large set of complaints from 2013 to 2020. Apart from reducing the complexity of these digital data, computational methods allow us to both (a) conduct a reproducible analysis and, at the same time, (b) carry out an interpretative exercise to understand relational meanings in citizen complaints.

This study is structured in four sections. First, we provide background information, discussing the recent literature on the potential linkages between citizen complaints and complainants' educational level, and introducing the case of Carabineros and the current institutional crisis. Second, we describe the data extraction procedures and the implementation of the topic modeling. Third, the latent themes are presented, interpreted and explained according to our key predictors. For each latent topic, we provide the narrative of the complaints with the highest probability of being assigned to each of them. Finally, the findings are discussed.

2. Background

2.1. Citizen complaints research

Citizen complaints have been considered as an indicator of police misconduct (Harris and Worden 2012; Lersch 1998; Terrill and Ingram 2016), and some countries have developed formal civilian oversight systems to address the complaint process (Smith 2013; Walker and Katz 2013). The institutional response to these complaints is a crucial element of police legitimacy (Terrill and Paoline 2015), and as such, has motivated a burgeoning corpus of research about different aspects of citizen complaints.

This body of research tends to be focused on four main aspects of complaints against the police. First, policing researchers have examined the process and experience of filing a complaint by citizens (McLean 2019; Walker 1997; Worden, Bonner, and McLean 2018). Researchers argue that complaints are motivated by diverse goals, such as seeking punishment, explanation for the incident, apology, or justice-restoration (Torrible 2018; Walker 1997). However, most of the complaints are not found to have sufficient evidence to justify sanctions against the police officers involved in the events (Dugan and Breda 1991; Lersch 1998; Terrill and Ingram 2016). This outcome shapes complainants' subjective experience and satisfaction with the complaint review process (Worden et al. 2018). Second, there is a large body of existing empirical research on the relationship between complaints and features of police officers, including gender, age, experience, and education, among others (Hassell and Archbold 2010; Hickman, Piquero, and Greene 2000; McElvain and Kposowa 2004). However, a lack of evidence on citizens-correlates of complaints represents a significant gap in this corpus of work, which is discussed in the next section. Third, researchers and policy-makers have used complaints indicators as measures to evaluate the impact of changes in procedures, practices, or more comprehensive police policies (Chalfin et al. 2021; Hedberg et al. 2017; Porter et al. 2012).

The fourth body of research interrogates the nature of the complaints. The studies answering this question suggest that most of the complaints are related to verbal discourtesy and improper use of force (Dugan and Breda 1991; Terrill and Ingram 2016). Although this evidence represents important progress in terms of understanding what citizens are complaining about, most of the analyses are based on the nature of allegations as reported by the police themselves (Dugan and Breda 1991; Harris 2010; McElvain and Kposowa 2004; Smith 2004). In some of the cases, the complaints data used by researchers do “not contain definitions or criteria by which the types were determined” (Harris

2010:2020). Moreover, most of the studies only focused on sustained complaints (Dugan and Breda 1991; Terrill and Ingram 2016), which are a small portion of the total of complaints filed as mentioned above, and leave aside an important resource of information from the perspective of the complainant. In addition, these studies use complaints channeled through the formal complainant's system in developed countries. The nature of complaints made in alternative avenues or in countries without formalized complaints systems is underresearched.

One of the limitations of extant research focusing on the nature of complaints is the access that researchers have to the complaints' content. The analyses are conducted using administrative data or police reports. To the best of our knowledge, there are no studies analyzing the content of the complaints from text-based data. In addition, as mentioned, studies mostly focus on sustained complaints through complaints systems. In this study, we augment existing knowledge about citizen complaints by examining the latent themes on an online corpus of complaints, which are substantial inputs into the understanding of the complainant process.

Regarding other avenues to channel police complaints, researchers have also examined community meetings as an important facet of community policing (Roussell and Gascón 2014; Skogan and Hartnett 1997). In contrast to sustained complaints reported by the police, ethnographic evidence indicates that complaints about police performance are central themes in residents' narratives during these meetings (Skogan and Hartnett 1997). Nevertheless, as Herbert (2006) suggests, improving citizens' oversight through community actions can be futile because neighborhoods rarely organize effectively. Other studies point out that power relationships in these encounters hinder citizens' chances of success (Roussell and Gascón 2014). As shown by Cheng (2020), most complaints are replied with literal silence by the police. Thus, our expectations of effective democratic policing through community meetings should remain modest. In addition, These meetings in Chile are imbued with an ambivalence between infiltration as part of the police intelligence services and social assistance, which limits the community-police partnership (Luneke, Dammert, and Zuñiga 2022). Thus, the analysis of online complaints has the potential to shed light on alternative mechanisms of citizens' oversight with distinctive features, yet has not been examined by policing researchers.

2.2. Education and citizen complaints

Most citizen complaints studies within the literature focus on officer-based correlates (Terrill and Ingram 2016). However, studies about attitudes toward the police could provide substantial insights through understanding the relationship with complainants' educational level. Stratification research defines socioeconomic status (SES) as one's access to financial, social, cultural, and human capital resources (Cowan et al. 2012), and educational level is one of its primary dimensions (Mueller and Parcel 1981; White 1982). Nonetheless, evidence about the stratification of police complaints is inconclusive.

In Chile, disadvantaged individuals tend to hold more negative attitudes toward the police (Dammert 2016; Frühling 2007). Using evidence from the United States, Frank et al. (2016) also show that respondents with lower levels of education express less favorable attitudes toward the police. The authors explain that low SES groups require more assistance from the police as they lack access to private security and present higher levels of deprivation (Dammert 2016). Indeed, Jackson et al. (2013) argue that people from SES groups have a more negative discourse about the police because they live in under-policed areas, experience lower levels of police effectiveness, higher levels of crime victimization, more corruption, and more abuse of power or other forms of over-policing. In contrast, other studies in Germany (Cao 2001) and the United States (Dowler 2002) report opposite effects. However, they do not provide theoretical interpretations of that result.

In one of the few studies in the literature of citizens complaints examining complainants' educational level, Lersch (1998) found that citizens living in areas marked by lower educational level are more likely to file more complaints. Malpractices are more recurrent themes in these complaints than in complaints of citizens living in more educated areas. However, they do not directly examine the educational level of the complainant. As Terrill and Ingram explain, "*there is little research relating to citizen-based correlates of complaints (in comparison with officer based), with much of this work being descriptive in nature as opposed to multivariate*" (2016:156). Thus, a central contribution of this study is to analyze the effect of complainants' educational level on a wide array of inductively generated complaint topics.

On the other hand, dissatisfaction or a negative public perception cannot be equated with complaints. In this regard, research on institutional trust suggests a negative effect of educational level on complaints filing. The central argument of this literature is that citizens with more education are more likely (1) to be better able to identify practices that undermine the functioning of institutions and (2) to be troubled by those practices. Hakhverdian and Mayne (2012) term these effects of education as the accuracy-inducing and norm-inducing functions of education. Kessler (1999) also pointed out that disadvantaged individuals may not complain because they assume they will not be taken seriously despite the more negative interactions and attitudes. Hence, the likelihood of complaining may be higher for more educated individuals.

2.3. The Chilean police force's institutional crisis

After the end of military dictatorships in Latin America in the 1980s, there have been numerous police reform efforts throughout the region (Frühling 2009a, 2009b). Governments sought to create civil police that were independent of the armed forces that held power during the military regimes, strengthening the internal accountability mechanisms, improving police management, and bolstering police-community relations. Although political interests have inhibited the implementation and sustainability of these reforms (Dammert 2019; Hathazy 2013), the Carabineros were one of the very few cases that successfully shook off their negative image of political involvement in gross human rights violations during Pinochet's dictatorship without undertaking radical reforms (Dammert 2009). Thus, police specialists have agreed on the exceptional nature of the Chilean success in the continent (Dammert 2009; Frühling 2007, 2009a; Prado, Trebilcock, and Hartford 2012).

Despite the increasing crime rates in Chile (Prado et al. 2012), the Carabineros succeeded in maintaining high levels of legitimacy. Three are the main explanations the experts have provided for the exceptional Chilean case. First, high levels of trust in the Carabineros depends on the perception of lower rates of corruption in comparison to other police forces in the region (Bonner 2013; Frühling 2007). Indeed, as explained by Dammert (2009), because of its military-style organization, the Carabineros considers corruption as an act that violates their cultural codes. Second, the Carabineros' role is not limited to crime prevention and police control, with their duties also including a wide variety

of social services (Dammert 2009), such as interventions in the aftermath of natural disasters, childhood protection, and community interventions. Third, complying with the law and the maintenance of public order are two central values in Chilean political culture, closely related to the legitimization of militarized institutions within a democratic regime (Frühling 2007). As a consequence, in the 2000s, there was no pressing need to introduce major reforms in a police force with such a high level of legitimacy as the Carabineros (Bonner 2013; Frühling 2007). Indeed, according to policing experts (Bonner 2013), the respect for the Carabineros was grounded more in their effective communication strategy and people's fear of crime than in a comprehensive reform process.

In the last decade, although the Carabineros have undergone no major reform. However, several partial (but significant) reforms have been implemented (Hathazy 2013). For instance, the Carabineros' chain of command changed drastically, from depending directly on the Minister of Defense to receiving orders from the Minister of the Interior and Public Security. Another important modification to the Carabineros is that cases in which a police officer is accused of a crime by a civilian are no longer part of a military court's jurisdiction, as during the Pinochet dictatorship. Such cases are now processed as crimes in civilian courts (Bonner 2013; Pereira 2001). Regardless of the reforms mentioned above, the Carabineros are still a militarized police force with a high level of autonomy from the political authorities, while it remains reluctant to accept any external accountability system open to public scrutiny (Hathazy 2013; Prado et al. 2012).

Despite the favorable scenario in the first decades after the return to democracy, Figure 1 shows that attitudes toward Carabineros have drastically changed in recent years. Since 2000s, citizens have become increasingly dissatisfied with the Carabineros as part of a broader increase in discontentment with government institutions. This provides compelling reasons to examine longitudinal variations on police complaints. Several reasons could drive this change. First, memories of human rights violations by the police that faded in the first decades of the restoration of democracy started to re-emerge due to the repression of a massive wave of protests initiated by student movements in 2006 (see Donoso, 2013, for a review). More recently, several international organizations have reported human rights violations (United Nations 2019) by the police force in repressing the recent wave of protests that started on

October 18, 2019 (Somma et al. 2020). Second, cases of commissioned officers' unlawful enrichment have been made public over recent years (Malone and Dammert 2020). Hence, the improvements made by partial reforms and the transition toward democratic policing in Chile are in jeopardy (Malone and Dammert 2020). As Frühling (2009b) suggests, one of the conditions for successful comprehensive police reforms is a social context conducive to reform, with the current context seeming to be a catalyst for this process. In the following analyses, we seek to examine how the latent themes vary across years and how they may reflect the shift in attitudes toward the Carabineros shown in public opinion polls.

[Figure 1 about here]

In sum, the aim of our analysis is threefold: identification of (a) topics across the whole corpus of complaints, (b) variations across time, and (c) by complainants' educational level.

3. Data and method

3.1. Data

We extracted complaints against the police posted between 2013 and 2020 on the platform (<http://www.reclamos.cl>) using R.¹ Reclamos.cl is an online forum where users can publicly complain against any public or private institution in Chile. It is worth noting that any citizen can upload a post² and the forum offers consulting in quality services to organizations³.

To extract the data for our analyses, we carried out the following procedures. First, all the hyperlinks associated with every single police complaint were gathered. This was facilitated by the design of the website, where complainants are classified by the institution targeted in the complaint. Second, the content of the complaints was extracted using these hyperlinks. A short complaint is shown

¹ Although the forum has existed since 2009, visits to the website and the number of complaints started to show a consistent increase in 2013. We restrict the analysis to the complaints from 2013 to prevent biases due to any potential website or user drift.

² According to the regulations of the forum, the complainants have the freedom to use or not personal identifiers in the main text of the complaint. For the purpose of anonymity, all the potential identifiers in the excerpts have been removed or replaced by fictitious information.

³ During the review process of this study, the forum has incorporated a channel for controversy resolution in partnership with a private company. Thus, complainants and organizations can reach an agreement to resolve the case and remove the complaint from the website.

in Figure 2 for illustration purposes. The extraction included title (A), date (B), complainant's province of origin (C), and main text (D). We collected a corpus of 1623 complaints (hereafter, the documents). On average, the complaints are 187 words long⁴. The distribution of the number of documents per year is shown in Table 1. Despite the scandals of police violence in 2019, the complaints in 2018 were slightly higher than in 2019. One possible explanation is that the social malaise irrupted in the public sphere in 2019 and other means were used to channel complaints from the public.

[Figure 2 about here]

[Table 1 about here]

To have a clear idea of the users, the forum's administration provided us with a general overview of the platform users' demographics based on Google Analytics statistics between 9 September, 2019 and 9 September, 2020. Although it does not exactly match the period under study, it provides insightful information to understand who the users of the platform are. Most of them were females (58%) and young adults (34% between 25 to 34 years old). As expected, the young population is overrepresented in the sample due to the self-selection into digital media. Regarding geographic location, 74% of the users resided in the Metropolitan Region, where 40% of the Chilean population lives.

3.2. Topic modeling preprocessing

Computational natural language processing methods are tools to reduce complexity and identify the principal latent themes among complaints against the Carabineros. In particular, we used Latent Dirichlet Allocation (LDA), a technique of unsupervised classification of text (Blei, Ng, and Michael I. Jordan 2003; Grimmer and Stewart 2013). In LDA, the words in the texts are classified into different latent topics based on the co-occurrence with other words within a document. As mentioned, we considered a single complaint extracted from the platform as a document. The model assumes that co-occurring words across documents reflect latent topics and meanings embedded in them (Turney and

⁴ Previous studies have used topic modeling to analyze documents of similar length, such as tweets (Lee and Jang 2021), open-ended survey responses (Roberts et al. 2014), or scientific abstracts (Sperandeo et al. 2020).

Pantel 2010). Therefore, a topic is a mixture of words, where each word has a probability of belonging to a topic. As a mixed membership model, terms could belong to a different topic or latent theme with a certain probability in each of them⁵. At the same time, documents can be associated with topics with a different prevalence based on the co-occurring words contained in their text.

We utilized an unsupervised approach because supervised techniques require prior theoretical categories that have not been developed for this type of digital data and context of study. In our approach, the topics are algorithmically identified based on word co-occurrence instead of previously defined categories in an initial training set that enables further automated classification of the remaining complaints. Since this is a pioneering study analyzing digital police complaints in a developing country, we opted for a totally inductive approach to identify data-driven categories. In addition, this approach has a higher level of scalability because the inductive analysis can identify emergent topics when new resources of information are incorporated (You et al. 2012), i.e., complaints after 2020 or from different countries. As Grimmer and Stewart (2013) note, supervised and unsupervised approaches are not competitors and can complement each other since they can serve different purposes. Future studies can use our data-driven categorization to produce training sets to analyze similar documents.

The scalability of this exercise also provides opportunities for learning that are not offered by administrative data that classifies complaints according to their type of circumstances⁶. For instance, the Citizens Police Data Project (Institute 2022) makes misconduct data in Chicago publicly available and categorized. This a major leap forward in terms of social accountability. Nevertheless, it relies on pre-established categories that cannot be modified without losses in comparability. For instance, if a new category of complaints emerges. In the case of the unsupervised approach to text analysis, the categories can be updated with new information without relegating those cases to a residual category of unknown cases. Thus, unsupervised approaches can be a valuable complement to administrative data due to its flexibility.

⁵ The ability to capture this polysemy given by different uses of a term is one of the strengths of topic modeling (DiMaggio et al. 2013). Thus, it is common that certain topics share terms, but other co-occurring words facilitate different relational interpretations.

⁶ Important comment made by one of the reviewers during the publication process.

In topic modeling, each of the latent topics may be considered a “*semantic context that primes particular association or interpretation of a phenomenon in a reader*” (DiMaggio et al. 2013:578). After removing⁷ explicit mentions of the Carabineros, the terms ‘tickets’, ‘day’, ‘precinct’, and ‘house’ are among the most common in the corpus (see Figure 3). These procedures could provide insights about the content of the complaints. Nevertheless, instead of counting words, LDA analyzes how these terms relate to each other around latent themes. It follows the cornerstone principle of relationality in modern cultural sociology (Basov, Breiger, and Hellsten 2020; Edelman and Mohr 2018; Pachucki and Breiger 2010), where meaning cannot be understood by analyzing isolated cultural elements, but rather it emerges from the relationship of these elements. Therefore, the co-occurrence of words enables us to capture their relational meaning and goes beyond the sole count of complaints or words.

[Figure 3 about here]

Data cleaning is the first step prior to the analysis. Several procedures were implemented before running the main analysis: (1) all the words were converted to lower case, and (2) punctuation marks, (3) stopwords,⁸ and (4) numbers were removed. After examining the most common terms in each topic, we (5) removed additional custom words that were not accurately identified by the stopword algorithm due to spelling errors or incompleteness of the predefined dictionary.⁹ Thus, the analyses included 1346 unique words. Only words with a frequency higher than 15 times and lower than 400 words were considered. Since the literature does not provide a rule of thumbs for defining a threshold, we tested different thresholds and they showed similar topic solutions.

3.3. Covariates

One of the advantages of topic modeling is that we can estimate the models, controlling for different covariates and estimating their correlations with the topics. Certain variables could explain

⁷ All the analyses were conducted using the *stm* package and its extensions in R with Spanish specifications (Roberts et al. 2019).

⁸ In natural language processing, stopwords are words filtered out in the preprocessing stage of analysis. They include articles, prepositions, and adverbs that are assumed not to provide additional meaning to the text. This procedure is beneficial because it reduces the size of the data for computational load, improves accuracy, and interpretability (Schofield et al. 2017).

⁹ These terms include ‘still’, ‘but’, ‘only’, ‘report’, ‘so’, ‘ahi’, ‘there’, ‘according to’, ‘times’, ‘here’, ‘behind’, ‘time’, ‘two’, ‘then’, ‘while’, ‘previous’, ‘none’, ‘also’, ‘besides’, ‘after’, and ‘opposite.’

topical content or its prevalence. Thus, the identified models will be inferred considering the distribution of the control variables. Although the website does not provide detailed information about the complainant, which is another common limitation in studies using online data (Salganik 2017), four important covariates are incorporated into our models.

First, we included the year when the complaint was submitted to the platform as a factor variable. Thus, we can account for potential non-linear variations of the complaint content across years and over-representation of those years with a higher number of complaints. At the same time, we evaluate those variations to understand how topics across time could reflect the change in attitudes toward the Carabineros.

Second, attitudes toward institutions and the police are socially stratified. The forum does not provide direct information about the socioeconomic status of the complainants. To solve this limitation, we built a proxy of educational level based on the omission of accent marks as the most common spelling mistake that can be attributed to an individual's socioeconomic background in Chilean samples (Bedwell et al. 2014). In addition, studies on computer mediated communication indicate that people with lower educational levels have lower quality of writing than more educated individuals, as shown by inappropriate spelling (Rosen et al. 2010). To elaborate the proxy, we counted the number of words with accents in each complaint (á, é, í, ó, ú), divided by length of the complaint (i.e., the total number of words of the complaint). Thus, individuals with a lower score used less accents than complainants with a higher score, considering the total number of words used. This variable is further recodified into four quartiles, where Q1 represents the lowest educational level and Q4 the highest educational level¹⁰.

Third, in line with the argument regarding crime frequency variation throughout the year (including summer peaks), which can affect complaints against the police (Linning 2015), the month of the complaint accounts for the potential seasonality of complaints. Government statistics show that high

¹⁰ A similar exercise could apply to other diacritic languages that use accent marks (e.g., French, Portuguese, German, Italian). Although there are words of French origin with accent marks in English, they are not common. However, other sub-lexical properties of words can be used for languages that do not rely on accent marks, such as word length.

social connotation crimes increase during summer months (Centro de Estudios y Análisis del Delito 2021).

Finally, all the estimations include fixed effects by province of origin of the complainant. Thus, we control the model for all the subnational-level heterogeneity.¹¹

4. Results

4.1. Topic selection

Topic modeling is used to identify themes across the corpus and assist in interpretation. Although there are no statistical tests for the optimal number of topics and researchers' qualitative judgment is the best criterion (DiMaggio et al. 2013), different indices have been developed to assist in this task. These metrics provide relative information to identify useful topics. Figure 4 displays four indices. The (1) held-out likelihood is a verification process where an uncoded test set of documents is used to test the accuracy of a training set coded by the algorithm. The index indicates the log probability of the topics in the training set correctly identifying topics in the testing set. Higher values indicate that solutions improve greatly up to eight topics. Solutions with more than eight topics only provide little marginal improvement. The same conclusion can be drawn from the (2) analysis of residuals of the comparison between both sets, where a lower value of the index suggests a better fit. Residuals sharply decrease until eight topics, where they start to fall at a smaller rate. We will seek to avoid choosing solutions with a smaller number of topics because they are over-dispersed, as indicated by their higher value. Regarding the (3) semantic coherence, there is a trade-off between semantic coherence and exclusivity of words to topics. This index is maximized when a few topics are identified by very common words. However, with only a few topics we cannot differentiate their content or exclusive words.

[Figure 4 about here]

¹¹ By incorporating province dummies into the topic models, results are robust against, for instance, crime rates and police resources associated with this subnational area and other unobserved province characteristics.

Moreover, the (4) lower bound refers to the lower bound of the marginal log-likelihood, which might be maximized. Thus, the higher the value, the better is the adjustment between the observed data and the solution. Overall, we chose an eight-topic solution because it performs relatively well across these metrics and provides a parsimonious and useful organization of the data for our interpretative exercise.

An alternative strategy to validate our results is to compare them against different potential solutions. For instance, residuals may indicate that 25 topics could also be an accurate number of topics. To compare them, we estimate a multilevel structure of topics using this alternative solution, which is detailed in Table S1 of the supplementary material. It suggests that a higher number of topics are nested in higher-level topics that resemble the 8-topic solution used in our main analysis. In addition, given our sample size, using 8 topics instead of a higher dimensional solution enables us to identify significant differences across groups of documents. Although it is not the aim of this study, further research can examine the multilevel structure of the data in detail (Coppola et al. 2016).

Table 2 displays the 8-topic solution and lists the 15 most probable terms for each topic. Table S2 in the supplementary material reports the 15 words that are both frequent and exclusive to each topic¹². Since each document has a probability of being assigned to each topic, Figure 5 shows the average probability for each topic across documents.¹³ This is an indicator of how prevalent each topic is. We selected one-word labels based on terms that better represent the interpretation of topics described in the following section. Topic 4 has the highest prevalence across documents (0.153), followed by Topics 7 (0.145) and 2 (0.144). Prevalence captures the modal estimate of the proportion of word tokens in a document assigned to the topic under the model (see Roberts, Stewart, and Tingley 2019).

[Figure 5 about here]

¹² While the first metric is the simple matrix of the words within each topic with the highest probability, the exclusiveness of words is calculated by taking the harmonic mean of rank by frequency within the topic and rank by distribution of topic given word $p(z/w=v)$. Both are valid word weightings to make sense of the latent topics based on words describing each topic.

¹³ Topic 7 is not relevant for our purposes and indicates that citizens use this communication channel to request specific information from the police; copies of security training certifications, for example.

[Table 2 about here]

4.2. Interpretation of topics

4.2.1. Procedural misbehavior

Topics 1, 2, and 3 address different police procedural misbehaviors or over-policing. Thus, the complaints are powered by a perception of being treated unjustly. As shown in Table 2, Topic 1 brings together terms referring to the institution's employees and their duties. Officer itself appears as the most probable word in this topic, followed by other organizational terms such as personnel, sergeant, chief, and commissioned officers. At the same time, it includes terms describing actions and events related to police officers' tasks such as service, procedure, and situation. Altogether, these words suggest intra-organizational conflicts.

We closely examined these complaints and confirmed that they represent conflicts with the police as an employer. Most of them are related to officers who were decommissioned due to allegedly arbitrary or unfair decisions carried out by higher-ranking officers. The latter indicates that the complaints are not only about issues between citizens and police officers, but also reflect intra-institutional conflicts. For instance, one of the complaints with the highest prevalence of Topic 1:

“My complaint is against Captain Juan González¹⁴ of sub-precinct Number X, where I worked until August 1, 2013. In his office, he threatened to decommission me because I answered back when he mistreated me in front of citizens during the guard service of that sub-precinct. Since that time, he has been looking for a way to harm me and punish me unfairly, making bad-faith comments to the commissioned officers of the prefecture of sub-precinct Number X, and achieving his goal on August 01, 2013.”

As in this passage, in the complaints with the highest proportion of words in Topic 1, ex-police officers were decommissioned after decisions that were criticized in their complaints. Overall, we confirm that this latent topic focused on conflicts with the police as an employer by reading the ten most

¹⁴ The officer's name is fictitious, and identification of the sub-precinct has been removed for the sake of anonymity.

representative complaints. Unfair processes and arbitrary decisions dominated the narratives in these complaints.

Topic 2 includes terms related to traffic. ‘Vehicle’ and ‘car’ are the terms with the highest probability of belonging to this category, followed by ‘[traffic] offense’, ‘documents’, and ‘driving license.’ It is worth noting that the term ‘ticket’ was removed from the corpus of analyses because it was one of the most frequent terms across complaints. The reading of the most representative complaints suggests that citizens are reporting their disagreement with traffic sanctions. In these complaints, citizens provide a detailed description of the events, arguing that police procedures are unlawful and arbitrary. The following excerpt illustrates one of these complaints:

“I was traveling from Santiago to Cauquenes. In the crossing to take the Conquistadores route, there is a turning lane. I stopped the car and waited for a truck to pass and two minor vehicles before entering the route. Some 150 meters after that, the Carabineros stopped me. I thought it was a routine control, but no. They gave me a ticket because I didn’t stop at the STOP sign at the turning track. Is this the height of abuse, or was I part of a ticket quota? Unfortunately, they are the law, and nothing can be done.”

Topic 3 reflects those complaints involving unfair treatment during police controls. These complaints often involved alleged abuse against complainants and their children, as suggested by the high frequency of the terms daughter or son. As shown by the most representative complaints, procedural misbehavior includes police violence, arbitrary detentions, and minimizing attitudes. Indeed, as shown in Table 2, power is among the most probable terms within the topic, which is another indicator of police misconduct. It is worth noting that most of the events described occur during identity checks. A crucial factor that might explain the latter is that the Carabineros have the power to carry out identity checks on any person over the age of 18 in public places or places of public access. For this topic, a significant number of the complaints refer to abuses of power and citizen maltreatment during identity checks. As one of the complaints described:

“I’m 21 and a student of medical technology. Yesterday at 18:00, I was in Coronel waiting for a bus in Lautaro Street when I saw three ‘flaites¹⁵’ (a young woman and two guys) staring at me. I thought I was going to be mugged. After five minutes, they approached me, and one of them shouted: ‘Hey dude, come here.’ My natural reaction was to run away, but after five minutes, they caught me and handcuffed me. At that moment, I realized three plain-clothes police officers were stopping me. At no point did they show proof of their identity or explain why I was being arrested. I had no idea what was going on. They put me in a police car. When we got to the precinct, I gave them my identity card, and they realized I had no criminal record and they let me go [...] I’m making this complaint through this channel so they won’t be the ones that carry out the investigation on their own behavior.”

The same complainant explained they decided to complain through the platform because the police officers recorded the actions that motivated the complaint as an attempt to avoid the identity check.

4.2.2. Police effectiveness

Topics 4 and 6 refer to police effectiveness in crime control and public order. The importance of these topics in citizen complaints is shown by the number of complaints in Topic 4 regarding unreasonable noise, the most prevalent topic in the corpus. In Chile, any citizen can report unreasonable noise to the police. As illustrated by the most common words related to these topics in Table 2, noise problems are associated with neighbors and the night. The terms are highly prevalent, and the most representative complaints associated with the topic indicate its relationship to under-policing on police-community relations: when complainants reported the existence of nuisance noise to the police, officers did not attend the scene. The complaint may be fueled by feeling neglected by Carabineros. For instance, the following complaint illustrates this situation: *“Good evening, I’ve called [the Carabineros] lots of*

¹⁵ Chilean Spanish slang referring to urban youth from low socioeconomic status, usually associated with rude behavior and delinquency.

times for them to put a stop to noise from partying every Saturday. It starts at midnight on Sunday and ends at 7 or 8 in the morning and not a single carabiniero shows up.”

Topic 6 includes terms regarding parking, such as street, place, wrong, law, and vehicle. This topic suggests that citizens use this channel to report traffic regulation violations, mainly related to parking in unauthorized areas. In some cases, complainants refer to under-policing, but they only report third parties' actions in others. The multilevel structure reported in Table S1 of the supplementary material suggests that this topic shares important commonalities with Topic 2, and they can be potentially aggregated in a more abstract topic referring to “license,” “street,” “infraction,” “car,” and “place.” Nevertheless, we keep them separated because there are also important differences between them. Topic 2 includes complaints by drivers who claim unfairness of tickets received due to traffic violations. In contrast, Topic 6 refers to citizens who witness violations incurred by other fellow citizens, mainly related to parking in non-designated areas.

4.2.3. Household and family

Complaints in Topic 5 combine both police misconduct and ineffectiveness. The events described in the documents with the highest prevalence of this topic take place at the complainant's house and involve household members. The most probable words suggest an interpretation in the same direction—house, residency, daughter, son, partner, among others. In contrast to Topic 3 about police controls, the victim in the complaint includes the complainants themselves and other household members. At the same time, Topic 3 describes stop-and-frisk events and Topic 5 describes situations where the police go to complainants' house. For instance, one of the representative complainants describes how police officers went to their homes due to a report about the incineration of refuse. Regarding the police procedure, the complaint indicates:

“They insisted on opening [the door]. My daughter asked for a warrant, but their insistence and the threat continued. So, my daughter opened the door under pressure, risking being deceived by false police officers, as has happened before. The least the Carabineros could do is respect legal procedures.”

Other complaints included in the same topic only refer to a rather slow, ineffective or inexistent response of Carabineros to alleged or reported crimes: *“Today, my house was burgled, I called them [Carabineros] three times to come so I could report the crime, but nobody showed up.”* Thus, this topic combines both under- and over-policing. In addition, the multi-level structure in Table S1 of the supplementary material suggest commonalities with Topic 3.

4.2.4. Generalized institutional crisis: a master frame

The theme represented by Topic 8 takes a more generalized and abstract perspective on police misbehavior. This topic can be defined as what cultural sociologists call a “master frame” (Snow and Benford 1992), which refers to a set of discourse cues wider in scope and influence than run-of-the-mill frames describing particular events, actors, institutions, practices, or situations. In line with this, it may be argued that complainants view police behaviors through the master frame of a national-level crisis. Even though complaints are related to specific incidents involving the police, citizens tended to evoke a more generalized institutional crisis. Consequently, as shown in Table 2, this frame brings together terms such as ‘Chile’, ‘justice’, ‘form’, ‘people’, and ‘institution.’ In doing this, people might reinterpret the events that led them to post their complaint. For instance, in a representative complaint, a woman makes a link between the complaint’s events and human rights violations, and attributes wider inequalities to the institution:

“For women like my mom and I, our rights are infringed, our human rights are violated. They [Carabineros] play with people and their dignity as human beings. They have to replace people like them [Carabineros] and bring in different police officers. It’s not fair, because they chose the path of serving Chile and the citizens. These Carabineros are responsible for this inequality, and for that, they need criminals and corrupt individuals.”

The same quote shows that complainants even go one step further and provide possible actions and opportunities for change (i.e., “bring in different police officers”). Hence, our analysis strengthens the contribution of adopting a bottom-up approach to generate initiatives to improve the police and community relationships. One of the citizens in another representative complaint suggests in a complaint that is representative of the topic:

“What’s happening to Chilean Carabineros, what makes them so ineffective and unassertive? Lately, this institution has made many mistakes, and it wasn’t like that before. They [Carabineros] should invest in training, instead of using institutional resources on lawsuits.”

It is important to note that there is no mention of physical violence from police officers across all topics. There are no words in Table 2 suggesting that this type of interaction originates complaints. When intentionally exploring the 50 most common and distinctive terms for each topic, there are no topics containing terms related to physical violence (“fight,” “violence,” “hit,” “beat”). In contrast, complaints including conjugations of the verb “to say” are among the 100 most common terms in Topics 2, 3, 5, and 8, suggesting that complaints describe verbal interaction between police and citizens. However, the most representative complaints of these topics indicate that they refer to unfair procedures. Indeed, the term “procedure” is also among the most common and distinctive terms in the same topics. It is important to mention that it does not mean there is not physical abuse in police-community relationships. We can only indicate that they do not emerge in this particular corpus of online complaints. It is also possible that citizens channel those complaints by other avenues.

4.3. Covariates analyses

One of the advantages of topic modeling is identifying the topics' prevalence for each document (Roberts et al. 2014). Thus, we can examine the variation on topics' prevalence as a function of a variables vector of documents information. This section presents the main results regarding variations by year and complainants' educational level.

Latent topics may also vary across time. The prevalence of each topic by year was examined, net of the month, educational level, and province fixed effects. Statistically significant variations across models are reported (see Table S3 in the supplementary material for details). The prevalence of the topic related to procedural misbehavior during identity checks or detentions (T3: Control) decreases after 2013, with the most important reductions in 2018 ($B = -0.067$, $p < .001$), 2019 ($B = -0.078$, $p < .001$), and 2020 ($B = -0.088$, $p < .001$). The latent topic of procedural misconduct during police

controls (Topic 3) slightly decreased in the last years. Topic 4 regarding the effectiveness of complaints on unreasonable noise (T4: Noise) increases in 2018 ($B = 0.128, p < .001$) and 2019 ($B = 0.130, p < .001$), relative to 2013. Nevertheless, the most dramatic change is found in the generalized crisis frame (T8: Generalized). After a reduction in 2019 ($B = -0.040, p < .01$), the association between complaints and an institutional crisis spiked in 2020 ($B = 0.078, p < .001$). This suggests that the succession of events over 2020 described in the previous sections could have led to a rise in public perception of police misconduct as systematic, given the generalized nature of this topic. The prevalence of each topic by year based on the models in Table S3 are visualized in Figure 6.

[Figure 6 about here]

Figure 7 shows variations across our proxy of educational level. The prevalence of Topic 1 related to conflicts with Carabineros as an employer is non-different from zero for quartile 1. This finding provides an important clue to the validity of our proxy of educational level and topic estimations. It is not surprising that conflicts involving former police officers are concentrated in individuals of middle and high educational level (Q2: $B = 0.087, p < .001$; Q3: $B = 0.083, p < .001$; Q4: $B = 0.040, p < .001$). Commissioned and non-commissioned officers receive training in police academies, which increases their educational level. At the same time, retired officers received a retirement pension above the country's average due to the special pension scheme for the police and armed forces.¹⁶ Thus, this topic is unlikely to be found among low-status individuals.

[Figure 7 about here]

More strikingly, Topic 3 about police controls is highly prevalent in the disadvantaged sector. In comparison to quartile 1, the prevalence of the topic is lower for quartile 2 ($B = -0.173, p < .001$), quartile 3 ($B = -0.195, p < .001$), and quartile 4 ($B = -0.117, p < .001$). The prevalence is not statistically different from zero for quantiles 2 and 3, and it is more than twice for quartile 1 than for the highest quartile. Moreover, as shown in Table 2, there are words with spelling mistakes among the most common terms in this topic, which does not occur for other topics and suggests this theme's

¹⁶ The average retirement pension for Carabineros is five times the average retirement pension of citizens in the Pension Administrators Fund (AFP in Spanish) system (Gálvez and Kremerman 2019).

stratification. Thus, procedural misbehavior content related to police detentions is not equally distributed. This pattern could have different explanations. On the one hand, detentions in the disadvantaged group could be more likely than in more advantaged groups. On the other hand, people from lower socioeconomic groups might either be more prone to being victims of police abuse or report it. Either way, these results highlight the importance of considering this aspect as a source of conflict between the police and communities.

Finally, Topic 8, related to the master frame of a institutional crisis, is more prevalent for middle socioeconomic sectors (Q2: $B = 0.123$, $p < .001$; Q3: $B = 0.096$, $p < .001$). This suggests that middle socioeconomic sectors are more likely to place police performance and behavior within this frame. The prevalence is also higher for high-status individual ($B = 0.047$, $p < .001$) but less prominent than for middle sectors.

5. Conclusion

In light of the crises of accountability of many police forces worldwide, the analysis of citizen complaints provides key information to repair weakened relationships between police officers and communities. Nevertheless, policing researchers have not made important progress on understanding the complexities underlying citizens attitudes toward the police (Boehme et al. 2022) and, in the case of police complaints, extant research has focused on the explanation of complaints by police-based characteristics (Terrill and Ingram 2016). This gap is even more striking in a context where the law enforcement is closed to public scrutiny and researchers have no access to the content of complaints, such as in Chile (Dammert 2019; Frühling 2009a). Thus, this study took a “bottom-up” approach to identify latent themes in a corpus of complaints against the police in this developing country, its temporal variation, and its relationship to complainants’ educational level.

Originally associated with English-speaking countries (Smith 2013; Walker and Katz 2013), citizen oversight of the police has emerged as a substantial principle of democratic policing, which has led to the establishment of formal, independent, and external complaints systems. Community meetings are also an important component of community policing, and an another important avenue for citizen complaints but with significant limitations (Cheng 2020; Roussell and Gascón 2014). Our analysis

shows an alternative path towards strengthening means of social accountability. One of the potential implications of the empirical strategy used in this article is to show how it can be utilized for identifying and analyzing latent themes of complaints and the potential contribution of online platforms to democratic policing.

Citizen complaints are considered a proxy of police misbehavior in police governance and policing studies. This study has contributed to these efforts by analyzing a corpus of complaints from a digital platform using natural language processing computational techniques. We have identified eight themes across the corpus of complaints that shed light on different police performance aspects. Therefore, we supplement recent calls (McLean 2019; Terrill and Ingram 2016; Torrible 2018) to examine the heterogeneity of complaints against the police forces and the “negative perceptions of police” (Boehme et al. 2022) more broadly. The natural processing language analysis indicates that topics are related to misconduct and performance, the combination of both, and the perception of a generalized crisis in the institution. In the case of misconduct, most of the topics except for Carabineros as an employer can be linked with what Boehme et al. (2022) termed as over-policing—namely, unjustly targeted, unfairly treated, or harassed. However, the latent themes suggest that online complaints focus mainly on verbal discourtesy and arbitrary procedures, and not on physical abuse which is common among sustained complaints reported by the police (Dugan and Breda 1991; Terrill and Ingram 2016). At the same time, since topics regarding performance highlight the law enforcement inaction, this effectiveness is directly related to under-policing. Thus, the online corpus of complaints combine the misconduct allegations predominant in formal complaints systems and the performance allegations recurrent in community meetings (Skogan and Hartnett 1997).

The prominence of the response to noise nuisance as a latent topic in the corpus of complaints signals an actionable under-policing aspect. A more effective response to citizens’ reports will lead to improvements in public attitudes toward the Carabineros. However, the remaining topics are more related to how the police treat citizens. This aspect requires a more complex approach and a comprehensive intervention of the police force. In addition, the topic about the unfair treatment of

employees highlights the need to increase the accountability and transparency of internal procedures (Hathazy 2013).

The relationship between Chileans and their police force has deteriorated over recent years. However, there are topics with a significant prevalence prior to the institutional crisis of 2019. Moreover, the salience of the generalized crisis master frame indicates a change in the public perception of the police after the massive wave of protests that began on 18 October, 2019. Although there were underlying conflicts, as shown by the prevalence of topics before 2019, the emergence of the institutional crisis topic suggests a likely change in how the police force is framed after the events of October 2019. With nuances, civil society members and actors from across the political spectrum have agreed on the need for institutional changes. Some sectors have even called for a re-building of the Carabineros (Diario UChile 2021), and an assembly is currently elaborating a new political constitution that will have implications for law enforcement. Thus, this study provides timely insights, and showcases possible alternatives to institutional mechanisms of currently inexistent independent complaint systems and how to leverage that information.

Another finding from this study that is relevant to recovering police-community relationships is the unequal distribution of complaints about misconduct in police detention. Other studies about over-policing have suggested that minority and disadvantaged groups are disproportionately targets of stop-and-frisk and zero-tolerance policing (Boehme et al. 2022; Harris 1993; Meares 2014). Additionally, this is the first study with access to information that sheds light on internal organizational conflicts. The Carabineros are no part of the jurisdiction of labor courts or government agencies that oversee and provide solutions to conflicts between employers and employees. This is another direction in which police accountability could move forward.

This study makes important contributions to the literature on public attitudes towards the police by examining the relationship between education and complaints' latent themes. The difference in the prevalence of the institutional crisis frame by complainants' educational level is a substantial heterogeneity. Middle socio-economic sectors are more likely to use discursive cues that link their complaints to a more general crisis in the Carabineros. This group has a higher educational level than

the lower quantile, which may provide them with information and cognitive skills to make an exercise of abstraction. This effect is consistent with the accuracy-inducing hypothesis suggested by the institutional trust literature (Hakhverdian and Mayne 2012). In addition, recent studies have shown that the events of the Chilean Spring of October 2019 caused a stronger decline in national pride in this group than among other socio-economic sectors (Olivos, Ayala, and Leyton 2020). These events might have signaled certain structural problems that shift the support for the status quo and conservatism (Barozet and Espinoza 2016; Barozet and Fierro 2011), including the general positive attitudes toward the police that characterize this group (Dammert 2013). In contrast, the procedural misbehavior associated with controls is more prevalent among lower-educated complainants. This finding is consistent with early results suggesting that individuals living in lower educated areas are more likely to fill complaints related to harassment and procedural malpractices (Lersch 1998). We have shown a similar pattern using individual-level data. Therefore, the relationship between complaints and educational level depends on the complaints' content.

Our research design makes it possible to contribute a unique perspective to the existing body of literature in policing research. However, there were inevitably certain limitations to the study, and based on these, we suggest potential directions for further research on this topic. First, the digital data used in these analyses are not representative. In other words, we cannot generalize our themes to the entire Chilean population. Thus, our set of eight narratives could be even larger if we consider citizens who are unwilling to make public statements or who do not have access to these platforms. Although Chile is among the three most connected countries in Latin America, the digital divide is still a barrier for certain groups (Correa, Pavez, and Contreras 2020). Despite this limitation, we believe that results are valid for three main reasons. First, these complaints exist, and they should be regarded as such. Therefore, even if the platform users represent a particular subsample of the population, their complaints must be analyzed and researched. Second, nonrepresentative data can be powerful for questions about within sample comparison (Salganik 2017), such as disparities between individuals with different educational levels. Third, an analysis of the complaints enables us to generate a valid typology of police complaints on digital platforms. This is useful, as little is known about the content of citizen complaints.

The second limitation is that the demographic information about the complainants is limited. This is a common characteristic of digital data (Bail 2021). The data extracted enables us to analyze time and educational level, but not other features of individuals, such as gender or age. In addition, other dimensions of socioeconomic status, such as income and occupation, are also highly relevant as stratifying factors but no information is available. These characteristics and their intersections could provide additional sources of heterogeneity that further studies might assess. In addition, we cannot establish causal relationships between topics and covariates.

Third, our unsupervised approach has disadvantages to the categorization of complaints. Roberts et al. (2014) explain that categories with a low prevalence that can be predefined are unlikely to emerge using topic modeling. For instance, a specific category of bribes could be relevant, but since very few complainants could have mentioned, topic modeling does not identify a related topic.

Finally, eight relevant themes were identified based on this corpus. They provide highly valuable information for transitioning to democratic policing. Nevertheless, by drawing attention to some aspects of policing, we shift attention away from others (Bonner 2013). Hence, alternative examinations of police performance and misconduct are still needed—mainly for marginalized groups. We should not lose sight of other policing dimensions absent in this corpus.

6. References

- Ariel, Barak, William A. Farrar, and Alex Sutherland. 2015. "The Effect of Police Body-Worn Cameras on Use of Force and Citizens' Complaints Against the Police: A Randomized Controlled Trial." *Journal of Quantitative Criminology* 31(3):509–35.
- Ba, Bocar. 2020. "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago."
- Bail, Chris. 2021. *Breaking the Social Media Prism*. Princeton: Princeton University Press.
- Barozet, Emmanuelle, and Vicente Espinoza. 2016. "Current Issues on the Political Representation of Middle Classes in Chile." *Journal of Politics in Latin America* 8(3):95–123.
- Barozet, Emmanuelle, and Jaime Fierro. 2011. *The Middle Class in Chile*. 12.

- Basov, Nikita, Ronald Breiger, and Iina Hellsten. 2020. "Socio-Semantic and Other Dualities." *Poetics* 78.
- Bedwell, Percy, Ana María Domínguez, Carmen Sotomayor, Gabriela Gómez, and Elvira Jéldrez. 2014. *Caracterización de Problemas Ortográficos Recurrentes En Alumnos de Cuarto Básico*. 14.
- Blei, David, Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(Jan):993–1022.
- Boehme, Hunter M., Deanna Cann, and Deena A. Isom. 2022. "Citizens' Perceptions of Over- and Under-Policing: A Look at Race, Ethnicity, and Community Characteristics." *Crime & Delinquency* 68(1):123–54.
- Bonner, Michelle. 2013. "The Politics of Police Image in Chile." *Journal of Latin American Studies* 45(4):669–94.
- Bonner, Michelle D. 2009. "Media as Social Accountability: The Case of Police Violence in Argentina." *The International Journal of Press/Politics* 14(3):296–312.
- Bonner, Michelle, and Lucía Dammert. 2021. "Constructing Police Legitimacy During Protests: Frames and Consequences for Human Rights." *Policing and Society*.
- Cao, Liqun. 2001. "A Problem in No-Problem-Policing in Germany Confidence in the Police Germany and USA." *European Journal of Crime, Criminal Law and Criminal Justice* 9.
- Cao, Liqun, and Bu Huang. 2000. "Determinants of Citizen Complaints against Police Abuse of Power." *Journal of Criminal Justice* 28(3):203–13.
- Centro de Estudios y Análisis del Delito. 2021. "Estadísticas Delictuales." Retrieved July 29, 2021 (<http://cead.spd.gov.cl/estadisticas-delictuales/>).
- Chalfin, Aaron, Jacob Kaplan, Bocar Ba, Richard Berk, Maria Cuellar, Zubin Jelveh, Greg Ridgeway, and Sarah Tahamont. 2021. "How Many Complaints against Police Officers Can Be Abated by Incapacitating a Few 'Bad Apples?'" *Criminology & Public Policy* 20(2):351–70.

- Cheng, Tony. 2020. "Input without Influence: The Silence and Scripts of Police and Community Relations." *Social Problems* 67(1):171–89.
- Comisión Independiente de Reforma Policial. 2020. *Propuesta de Reforma a Carabineros de Chile*.
- Coppola, Antonio, Margaret Roberts, Brandon Stewart, and Dustin Tingley. 2016. "StmCorrViz: A Tool for Structural Topic Model Visualizations."
- Correa, Teresa, Isabel Pavez, and Javier Contreras. 2020. "Digital Inclusion through Mobile Phones?: A Comparison between Mobile-Only and Computer Users in Internet Access, Skills and Use." *Information, Communication & Society* 23(7):1074–91.
- Cowan, Charles, Robert Hauser, Robert Komisnki, Henry Levin, Samuel Lucas, Morgan Stephen, Margaret Beale Spencer, and Chris Chapman. 2012. *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*.
- Dammert, Lucía. 2009. "Police and Judicial Reform in Chile." Pp. 151–68 in *Policing insecurity : police reform, security, and human rights in Latin America*, edited by N. Uildriks. Lanham: Lexington Books.
- Dammert, Lucía. 2013. "El Dilema de Chile: Confianza En La Policía y Desconfianza Ciudadana." *Revista Brasileira de Segurança Pública* 7(1):24–39.
- Dammert, Lucía. 2016. "Confianza En La Policía En Chile: Un Arma de Doble Filo." *Civitas - Revista de Ciências Sociais* 16(4):575.
- Dammert, Lucía. 2019. "Challenges of Police Reform in Latin America." Pp. 259–78 in *Routledge Handbook of Law and Society in Latin America*, edited by R. Sieder, K. Ansolabehere, and T. Alfonso. New York: Routledge.
- Deseriis, Marco. 2020. "Rethinking the Digital Democratic Affordance and Its Impact on Political Representation: Toward a New Framework." <https://doi.org/10.1177/1461444820929678> 23(8):2452–73.

- Diario UChile. 2021. "Organizaciones Sociales, Parlamentarios y Partidos Políticos Exigen La Refundación de Carabineros."
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41(6):570–606.
- Donoso, Sofia. 2013. "Dynamics of Change in Chile: Explaining the Emergence of the 2006 'Pingüino' Movement." *Journal of Latin American Studies* 45(1):1–29.
- Dowler, Kenneth. 2002. "Media Influence on Citizen Attitudes Toward Police Effectiveness." *Policing and Society* 12(3):227–38.
- Dugan, John R., and Daniel R. Breda. 1991. "Complaints about Police Officers: A Comparison among Types and Agencies." *Journal of Criminal Justice* 19(2):165–71.
- Edelmann, Achim, and John W. Mohr. 2018. "Formal Studies of Culture: Issues, Challenges, and Current Trends." *Poetics* 68:1–9.
- Frank, James, Brad W. Smith, and Kenneth J. Novak. 2016. "Exploring the Basis of Citizens' Attitudes Toward the Police." *Police Quarterly* 8(2):206–28.
- Frühling, Hugo. 2007. "Police Legitimacy in Chile." Pp. 115–45 in *Legitimacy and criminal justice : international perspectives*, edited by T. Tyler. New York: Russell Sage Foundation.
- Frühling, Hugo. 2009a. "Recent Police Reform in Latin America." Pp. 21–46 in *Policing insecurity : police reform, security, and human rights in Latin America*, edited by N. Uildriks. Lanham: Lexington Books.
- Frühling, Hugo. 2009b. "Research on Latin American Police: Where Do We Go from Here?" *Police Practice and Research* 10(5–6):465–81.
- Galovic, Sally, Philip Birch, Margaret H. Vickers, and Michael Kennedy. 2016. "Procedural Justice and Frontline Policing: The Effects of the Police Complaints System." *Journal of Forensic Practice* 18(3):170–81.

- Gálvez, Recaredo, and Marco Kremerman. 2019. *Pensiones Por La Fuerza: Resultados Del Sistema de Pensiones de Las Fuerzas Armadas y de Orden*. Vol. 17. 17.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–97.
- Hakhverdian, Armen, and Quinton Mayne. 2012. "Institutional Trust, Education, and Corruption: A Micro-Macro Interactive Approach." *Journal of Politics* 74(3):739–50.
- Harris, C. J. 2010. "Problem Officers? Analyzing Problem Behavior Patterns from a Large Cohort." *Journal of Criminal Justice* 38(2):216–25.
- Harris, Christopher J., and Robert E. Worden. 2012. "The Effect of Sanctions on Police Misconduct:" *Crime & Delinquency* 60(8):1258–88.
- Harris, David A. 1993. "Factors for Reasonable Suspicion: When Black and Poor Means Stopped and Frisked." *Indiana Law Journal* 69.
- Hassell, Kimberly D., and Carol A. Archbold. 2010. "Widening the Scope on Complaints of Police Misconduct." *Policing* 33(3):473–89.
- Hathazy, Paul. 2013. "Fighting for a Democratic Police: Politics, Experts and Bureaucrats in the Transformation of the Police in Post-Authoritarian Chile and Argentina." *Comparative Sociology* 12(4):505–47.
- Hedberg, E. C., Charles M. Katz, and David E. Choate. 2017. "Body-Worn Cameras and Citizen Interactions with Police Officers: Estimating Plausible Effects Given Varying Compliance Levels." *Justice Quarterly* 34(4):627–51.
- Herbert, Steven Kelly. 2006. *Citizens, Cops, and Power : Recognizing the Limits of Community*. Chicago: University of Chicago Press.
- Hickman, Matthew J., Alex R. Piquero, and Jack R. Greene. 2000. "Does Community Policing Generate Greater Numbers and Different Types of Citizen Complaints than Traditional Policing?:" *Police Quarterly* 3(1):70–84.

- Holmberg, Lars. 2019. "In Service of the Truth? An Evaluation of the Danish Independent Police Complaints Authority." *European Journal of Criminology* 16(5):592–611.
- Institute, Invisible. 2022. "Citizens Police Data Project." Retrieved January 28, 2022 (<https://cpdp.co/>).
- Jackson, Jonathan, Ben Bradford, Betsy Stanko, and Katrin Hohl. 2013. *Just Authority? Trust in the Police in England and Wales*. New York: Routledge.
- Kessler, David A. 1999. "The Effects of Community Policing on Complaints Against Officers." *Journal of Quantitative Criminology* 1999 15:3 15(3):333–72.
- Lee, Claire Seungeun, and Ahnlee Jang. 2021. "Questing for Justice on Twitter: Topic Modeling of #StopAsianHate Discourses in the Wake of Atlanta Shooting." *Crime & Delinquency*.
- Lersch, Kim Michelle. 1998. "Police Misconduct and Malpractice: A Critical Analysis of Citizens' Complaints." *Policing* 21(1):80–96.
- Linning, Shannon J. 2015. "Crime Seasonality and the Micro-Spatial Patterns of Property Crime in Vancouver, BC and Ottawa, ON." *Journal of Criminal Justice* 43(6):544–55.
- Lührmann, Anna, Kyle Marquardt, and Valeriya Mechkova. 2020. "Constraining Governments: New Indices of Vertical, Horizontal, and Diagonal Accountability." *American Political Science Review* 114(3):811–20.
- Luneke, Alejandra, Lucia Dammert, and Liza Zuñiga. 2022. "From Social Assistance to Control in Urban Margins: Ambivalent Police Practices in Neoliberal Chile." *Dilemas: Revista de Estudios de Conflicto e Controle Social* 15(1):1–26.
- Malena, Carmen, Reiner Forster, and Janmejay Singh. 2004. *Social Accountability: An Introduction to the Concept and Emerging Practice*. 76. Washington DC.
- Malone, Mary Fran T., and Lucía Dammert. 2020. "The Police and the Public: Policing Practices and Public Trust in Latin America." *Policing and Society* 1–16.
- McElvain, James P., and Augustine J. Kposowa. 2004. "Police Officer Characteristics and Internal

- Affairs Investigations for Use of Force Allegations.” *Journal of Criminal Justice* 32(3):265–79.
- McLean, Kyle. 2019. “Justice-Restoring Responses: A Theoretical Framework for Understanding Citizen Complaints against the Police.” *Policing and Society* 1–20.
- Meares, Tracey L. 2014. “The Law and Social Science of Stop and Frisk.” *Annual Review of Law and Social Science* 10:335–52.
- Mellado, Claudia, Sonia V Moreira, Claudia Lagos, and María E. Hernández. 2012. “Comparing Journalism Cultures in Latin America: The Case of Chile, Brazil and Mexico.” *International Communication Gazette* 74(1):60–77.
- Mueller, Charles W., and Toby L. Parcel. 1981. “Measures of Socioeconomic Status: Alternatives and Recommendations.” *Child Development* 52(1):13.
- O’Donnell, Guillermo A. 1998. “Horizontal Accountability in New Democracies.” *Journal of Democracy* 9(3):112–26.
- Olivos, Francisco, Cristian Ayala, and Alex Leyton. 2020. *Pride and Protest: Emotional Response in the Aftermath of the Chilean Social Outburst*. SocArXiv.
- Pachucki, Mark A., and Ronald L. Breiger. 2010. “Cultural Holes: Beyond Relationality in Social Networks and Culture.” *Annual Review of Sociology* 36(1):205–24.
- Pereira, Anthony. 2001. “Virtual Legality: Authoritarian Legacies and the Reform of Military Justice in Brazil, the Southern Cone, and Mexico.” *Comparative Political Studies* 34(5):555–74.
- Peruzzotti, Enrique, and Catalina Smulovitz. 2006. *Enforcing the Rule of Law: Social Accountability in the New Latin American Democracies*. edited by E. Peruzzotti and C. Smulovitz. Pittsburgh: University of Pittsburgh Press.
- Porter, Louise E., Tim Prenzler, and Jenny Fleming. 2012. “Complaint Reduction in the Tasmania Police.” *Policing and Society* 22(4):426–47.
- Prado, Mariana Mota, Michael Trebilcock, and Patrick Hartford. 2012. “Police Reform in Violent Democracies in Latin America.” *Hague Journal on the Rule of Law* 4(2):252–85.

- Pryor, Cori, John H. Boman, Thomas J. Mowen, and Michael McCamman. 2019. "A National Study of Sustained Use of Force Complaints in Law Enforcement Agencies." *Journal of Criminal Justice* 64.
- Reynolds, Paul D., Brett A. Fitzgerald, and Jeremiah Hicks. 2018. "The Expendables: A Qualitative Study of Police Officers' Responses to Organizational Injustice." *Police Quarterly* 21(1):3–29.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm: An R Package for Structural Topic Models." *Journal of Statistical Software* 91(1):1–40.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–82.
- Rosa, Mauricio De, Ignacio Flores, and Marc Morgan. 2020. *Inequality in Latin America Revisited: Insights from Distributional National Accounts*. 2020/09. Paris.
- Rosen, Larry D., Jennifer Chang, Lynne Erwin, L. Mark Carrier, and Nancy A. Cheever. 2010. "The Relationship Between 'Textisms' and Formal and Informal Writing Among Young Adults:" [Http://Dx.Doi.Org/10.1177/0093650210362465](http://dx.doi.org/10.1177/0093650210362465) 37(3):420–40.
- Roussell, Aaron, and Luis Daniel Gascón. 2014. "Defining 'Policeability': Cooperation, Control, and Resistance in South Los Angeles Community-Police Meetings." *Social Problems* 61(2):237–58.
- Salganik, Matthew. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton: Princeton University Press.
- Schofield, Alexandra, Måns Magnusson, Laure Thompson, and David Mimno. 2017. *Understanding Text Pre-Processing for Latent Dirichlet Allocation*.
- Schuck, Amie M., and Cara Rabe-Hemp. 2016. "Citizen Complaints and Gender Diversity in Police Organisations." *Policing and Society* 26(8):859–74.
- Skogan, Wesley G., and Susan M. Hartnett. 1997. *Community Policing, Chicago Style*. New York:

Oxford University Press.

- Smith, Brad W., and Malcolm D. Holmes. 2003. "Community Accountability, Minority Threat, and Police Brutality: An Examination of Civil Rights Criminal Complaints." *Criminology* 41(4):1035–64.
- Smith, Graham. 2004. "Rethinking Police Complaints." *The British Journal of Criminology* 44(1):15–33.
- Smith, Graham. 2013. "Oversight of the Police and Residual Complaints Dilemmas: Independence, Effectiveness and Accountability Deficits in the United Kingdom." *Police Practice and Research* 14(2):92–103.
- Snow, D. A., and R. D. Benford. 1992. "Master Frames and Cycles of Protest." Pp. 133–155 in *Frontiers in Social Movement Theory*, edited by A. D. Morris. New Haven: Yale University Press.
- Somma, Nicolás, Matías Bargsted, Rodolfo Disi, and Rodrigo M. Medel. 2020. "No Water in the Oasis: The Chilean Spring of 2019–2020." *Social Movement Studies*.
- Sperandeo, Raffaele, Giovanni Messina, Daniela Iennaco, Francesco Sessa, Vincenzo Russo, Rita Polito, Vincenzo Monda, Marcellino Monda, Antonietta Messina, Lucia Luciana Mosca, Laura Mosca, Silvia Dell'Orco, Enrico Moretto, Elena Gigante, Antonello Chiacchio, Chiara Scognamiglio, Marco Carotenuto, and Nelson Mauro Maldonato. 2020. "What Does Personality Mean in the Context of Mental Health? A Topic Modeling Approach Based on Abstracts Published in Pubmed Over the Last 5 Years." *Frontiers in Psychiatry* 10:938.
- Terrill, William, and Jason R. Ingram. 2016. "Citizen Complaints Against the Police: : An Eight City Examination." *Police Quarterly* 19(2):150–79.
- Terrill, William, and Eugene A. Paoline. 2015. "Citizen Complaints as Threats To Police Legitimacy: The Role of Officers' Occupational Attitudes." *Journal of Contemporary Criminal Justice* 31(2):192–211.

- Torche, Florencia. 2007. "Social Status and Cultural Consumption: The Case of Reading in Chile." *Poetics* 35(2-3):70-92.
- Torrible, Clare. 2018. "Reconceptualising the Police Complaints Process as a Site of Contested Legitimacy Claims." *Policing and Society* 28(4):464-79.
- Turney, Peter D., and Patrick Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37:141-88.
- United Nations. 2011. *Handbook on Police Accountability, Oversight and Integrity*. New York: United Nations.
- United Nations. 2019. *Report of the Mission to Chile 30 October - 22 November 2019*. Geneva, Switzerland.
- Walker, Samuel. 1997. "Complaints Against the Police: A Focus Group Study of Citizen Perceptions, Goals, and Expectations." *Criminal Justice Review* 22(2):207-26.
- Walker, Samuel, and Charles M. Katz. 2013. *The Police in America : An Introduction*. New York: McGraw-Hill.
- White, Karl R. 1982. "The Relation between Socioeconomic Status and Academic Achievement." *Psychological Bulletin* 91(3):461-81.
- Worden, Robert E., Heidi S. Bonner, and Sarah J. McLean. 2018. "Procedural Justice and Citizen Review of Complaints Against the Police: Structure, Outcomes, and Complainants' Subjective Experiences." *Police Quarterly* 21(1):77-108.
- You, Wei, Dominique Fontaine, Jean-Paul Barthès, W You, · D Fontaine, J.-P Barthès, and D Fontaine. 2012. "An Automatic Keyphrase Extraction System for Scientific Documents." *Knowledge and Information Systems* 2012 34:3 34(3):691-724.

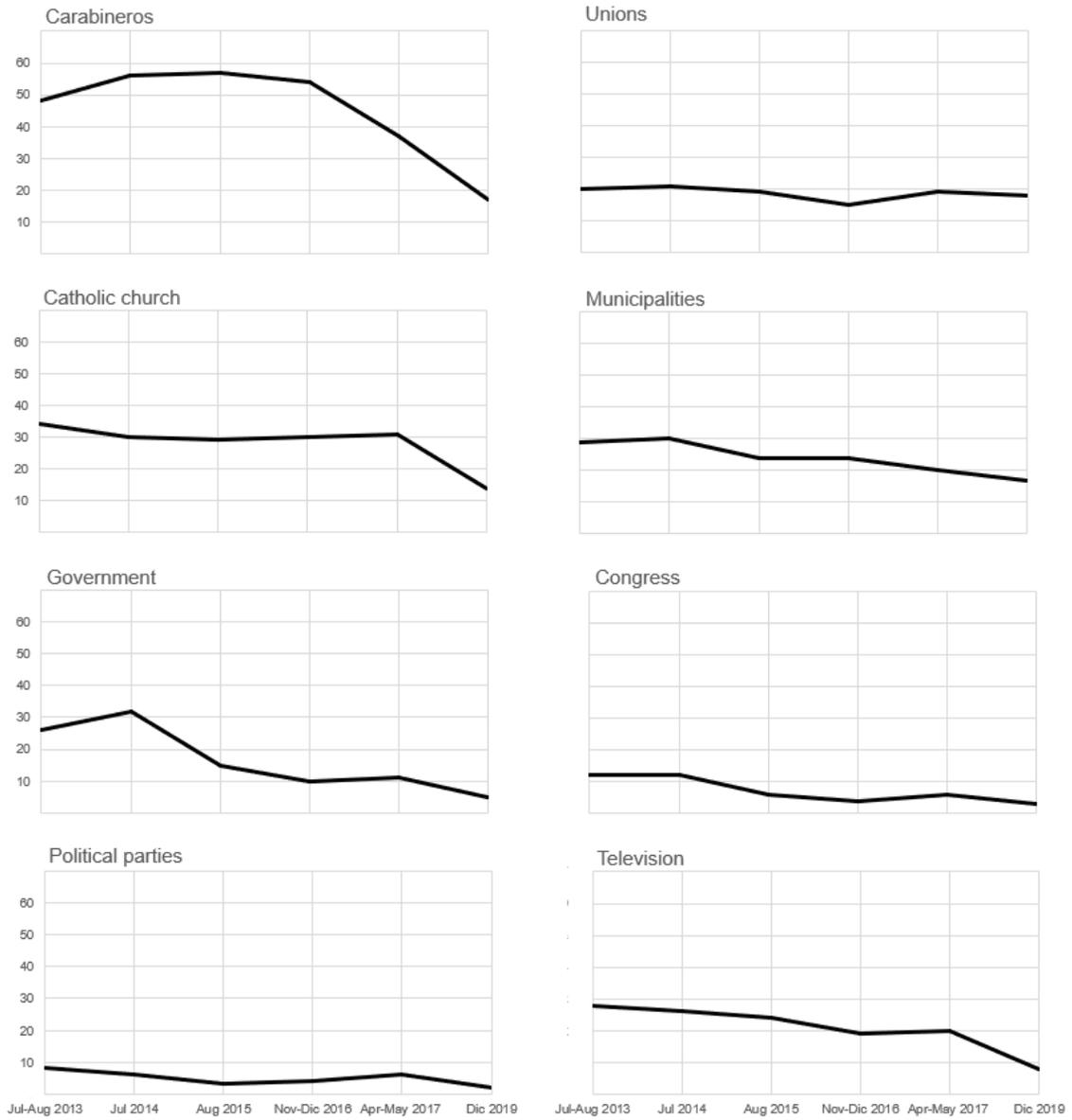


Figure 1. Percentage of Chileans over the age of 18 who declared some, or a lot of, trust in different institutions.

Source: Own preparation based on CEP surveys.

Ruido por fiesta A

Este reclamo tiene más de seis meses de antigüedad

CARABINEROS DE CHILE
RUIDO POR FIESTA

En Estación Central, Santiago, Chile C
Lunes 15, January 2018 B

Número de Reclamo:
626 observadores

SEÑORES CARABINEROS DE CHILE
PRESENTE

Estimados, vivo en _____, comuna de estación central, la madrugada de esta noche el dpto _____ tuvo *Ruido por fiesta* hasta las 7:00 am, conserjería les reclamo y les paso multa, pero hicieron caso omiso, incluso yo personalmente llame a **Carabineros** pero al final nunca fueron a revisar

D

Note: We have removed personal information in the main text, number of complaints, and reported a fictitious date for anonymity purpose. A = title, B = Date, C = Geographical location, D = Main text.

Figure 2. Example of a complaint.

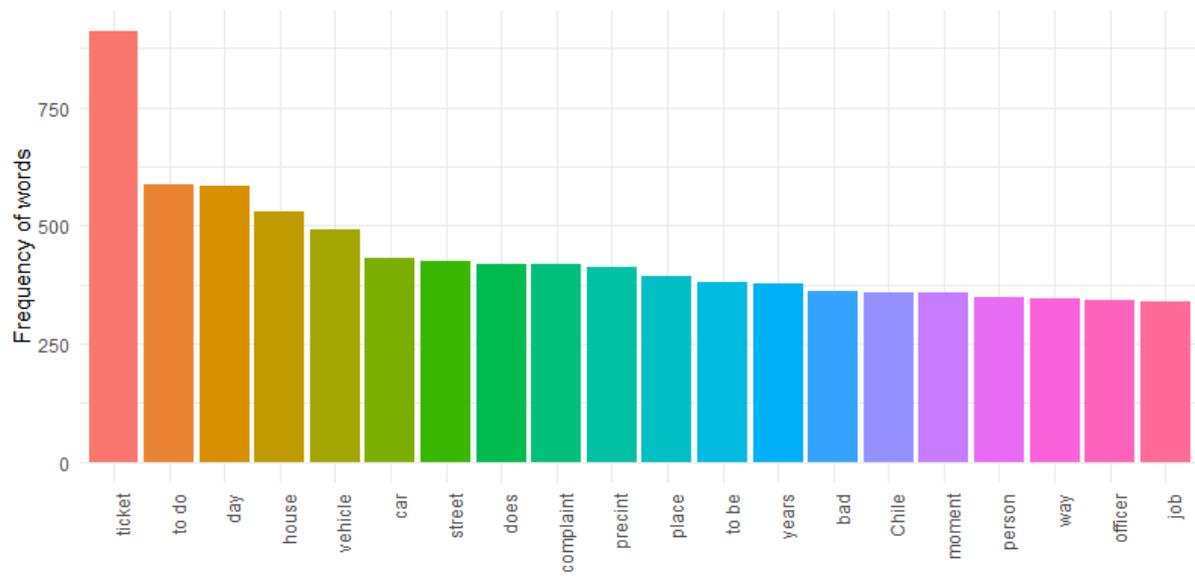


Figure 3. Most common words in the corpus after initial data cleaning.

Diagnostic Values by Number of Topics

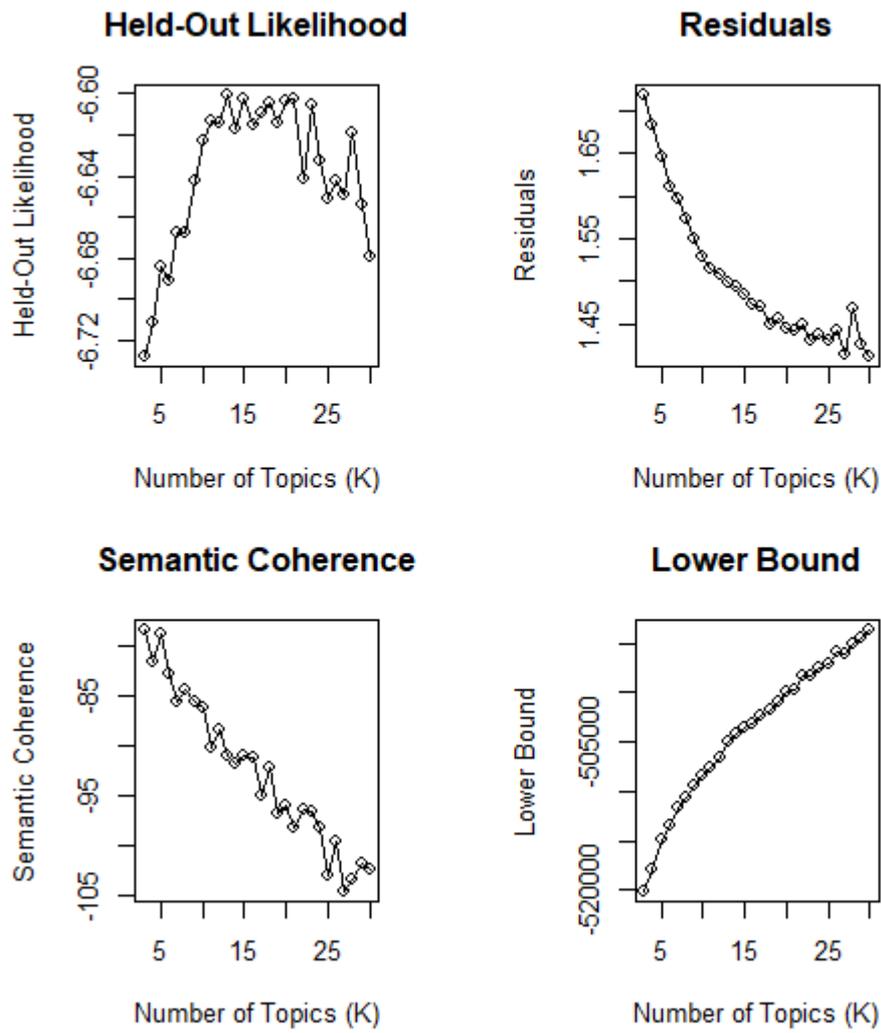
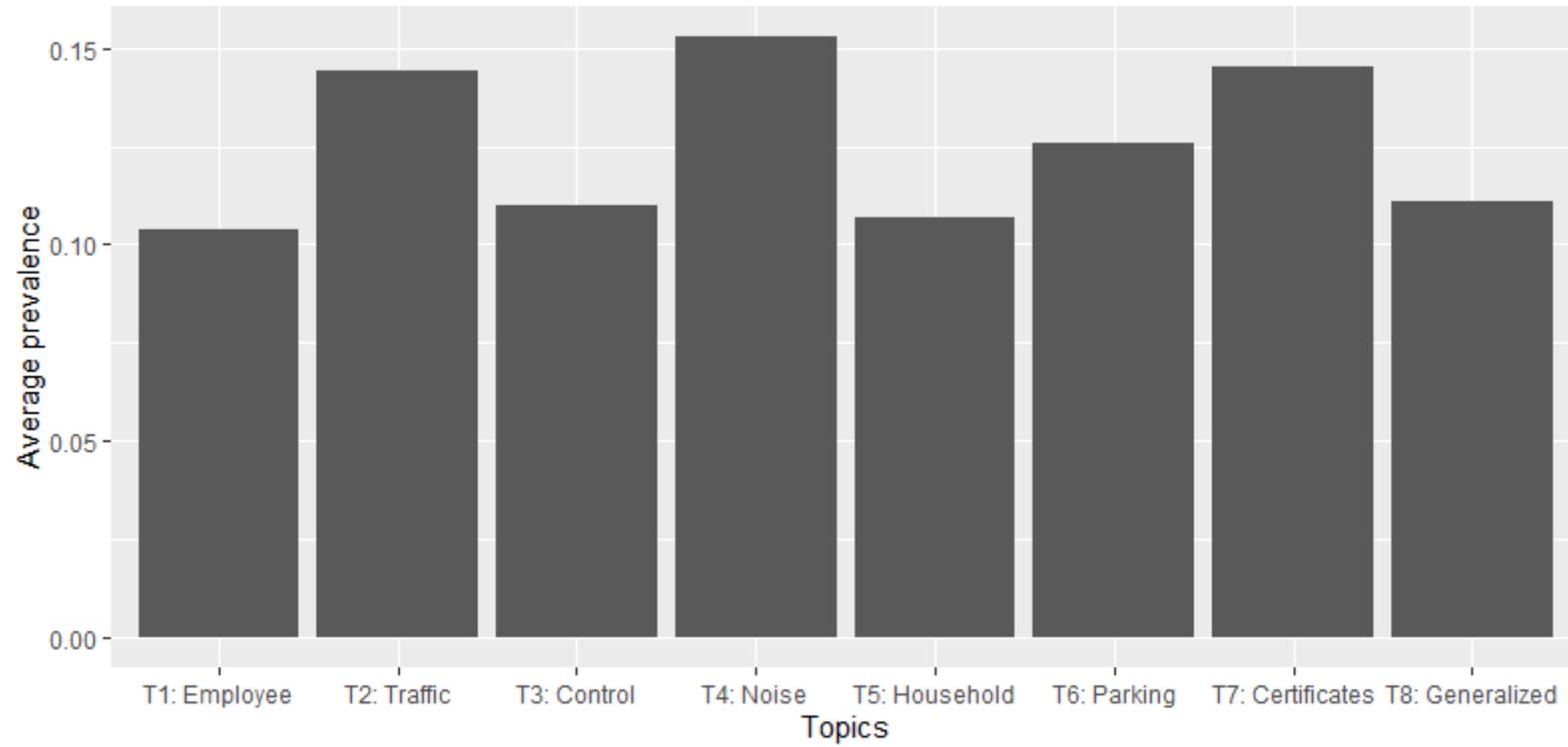
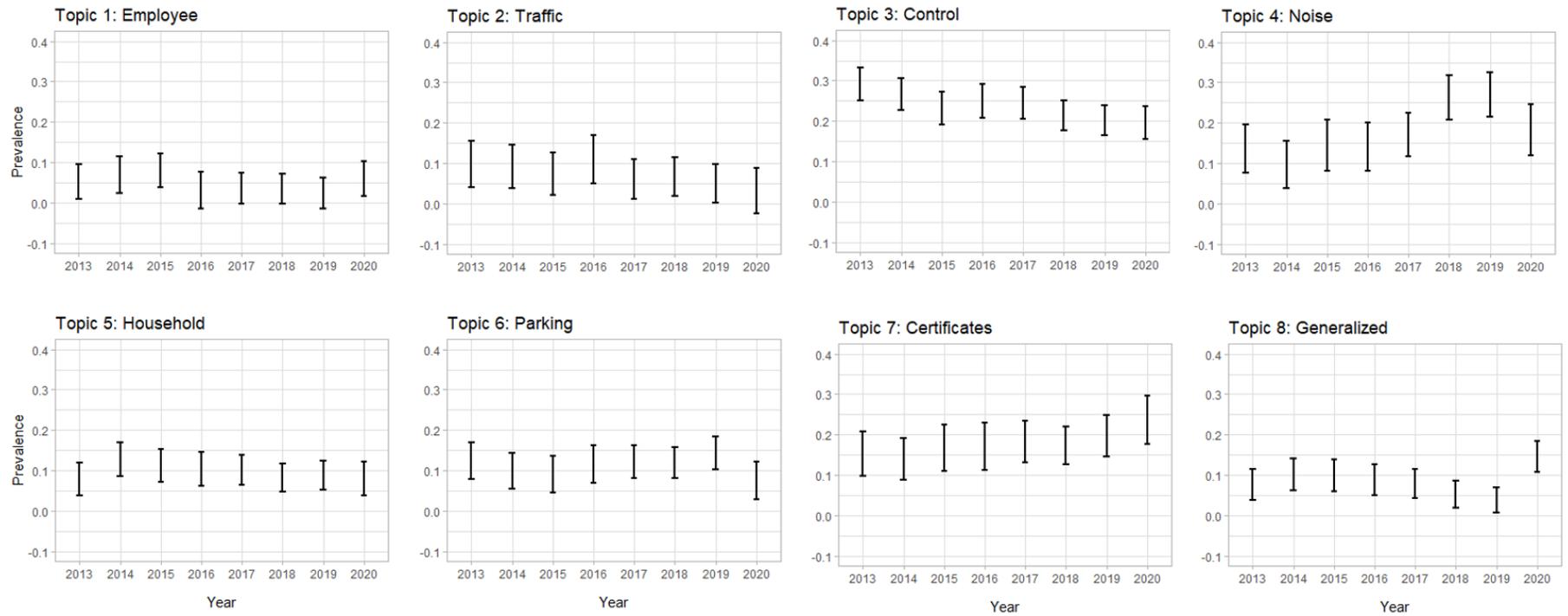


Figure 4. Indices for selection of numbers of topics



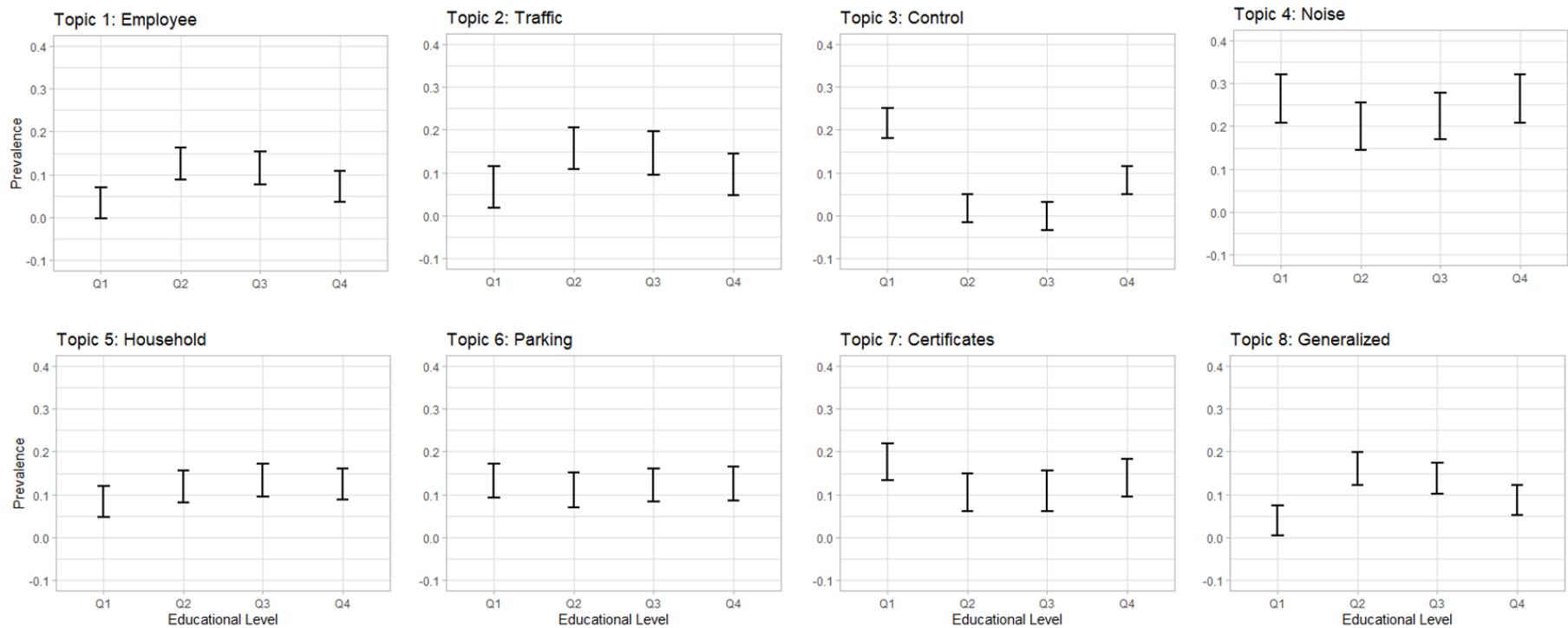
Note: Topics are estimated using spectral initialization and 100 iterations. Models include year, month, socioeconomic status, and province fixed-effects.

Figure 5. Average prevalence of each topic.



Note: Intervals of confidence at 95 percent.

Figure 6. Prevalence associated with each topic by year.



Note: Intervals of confidence at 95 percent.

Figure 7. Prevalence associated with each topic by complainants' educational level.

Table 1. Distribution of Documents

Year	N	%
2013	179	11.0
2014	171	10.5
2015	155	9.6
2016	152	9.4
2017	188	11.6
2018	323	19.9
2019	301	18.5
2020	154	9.5
<i>Total</i>	<i>1623</i>	<i>100</i>

Table 2. Most probable words per topic

Topic Label	1 Employer	2 Traffic	3 Control	4 Noise	5 Household	6 Parking	7 Certificate	8 Generalized
# Words								
1	officer	vehicle	day*	house	house	street	work	chile
2	personnel	car	precinct*	neighbors	complaint	place	years	institution
3	day	says	vehicle*	week	then	bad	thanks	persons
4	form	offense	had*	night	residency	law	answer	to be
5	present	documents	said	noises	daughter	in front	money	person
6	officers	license	corporal	does	son	cars	year	today
7	sergeant	speed	power	district	partner	to park	course	good
8	police (adjective)	day	then*	tomorrow	father	vehicles	does	respect
9	situation	going	son	alley	place	parked	santiago	type
10	precinct	moment	daughter	neighbor	mother	people	company	worst
11	procedure	stop [sign]	after*	call	while	big	[I] can	people
12	service	driver	moment*	[I] live	name	exit	[I] need	form
13	commissioned officer	court	had*	day	[they] said	[it] seems	good	day
14	chief	license	[I] said	disturbing	door	to be	hello	justice
15	corporal	then	[he/she] says	[they] do	day	pass	to have	precinct

Note: *words with spelling mistakes.

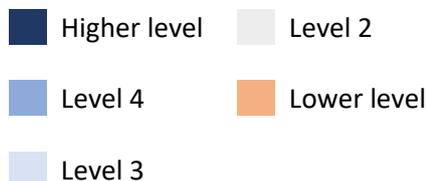
Supplementary Material

Police complaints as an accountability mechanism: Uncovering patterns using topic modeling

Table S1 shows the hierarchical clustering of topics. It enables us to inspect different levels of aggregation of words and suggest possible topic solutions at different levels of abstraction. As shown in the right-hand column, aggregations at the minimum level (25 topics) resemble the eight topics used in the analysis. At the same time, these topics can be aggregated into even higher-level groups. However, highly abstract topics do not provide fine-grained information for decision-making.

S1. Multilevel structure.

STM - 25 Topics	STM - 8 Topics
vehicle, precinct, wrong, car, job	
Chile, good, institution, favor, thanks	
people, service, general, persons, chile	
chile, mayor, general, personnel, decomission	
6: general, service, director, chief, personnel	Topic 1
8: head, damage, chile, justice, you (plural)	
institution, today, moment, person, day	
9: luck, worst, act, surprise, tribunals	Topic 8
24: you (singular), attention, institution, number, citizens	
thanks, [I] need, favor, job, good	
17: block, commune, system, water	
hi, many, law, favor, [I] need	
15: law, company, local, identity, guards	
7: [I] need, good, thanks, mail, afternoon	Topic 7
month, day, account, officer, date	
1: date, bank, money, contract, officer	
19: vine, prefecture, sea, come back, july	
said, place, house, street, day	
license, street, infraction, car, place	
13: tickers, streets, transit, oversight, north	
street, wrong, to park, car, vehicle	
22: to park, parked (singular), parked (plural), entry, sidewalk	Topic 6
25: truck, track, road, accident, driver	
12: uniform, bus, men, could, [I] believe	
18: vehicles, infraction, documents, documentation, traffic	
infraction, says, vehicle, stop, license	
14: speed, offense, highway, excess, vehicle	Topic 2
20: stop (sign), arrived, sign, stopped, crossing	
said, son, day, daughter, precinct	
house, residency, complaint, son, procedure	
3: son, child, minor, age, young	
day, complaint, residency, son, procedure	
5: father, sister, brother, said, door	
16: police (adjective), procedure, cell, residency, crime	Topic 5
10: house, spouse, emergency, husband, woman	
said, had, car, day, precinct*	
2: there was (<i>habia</i>)*, motorcycle, day*, shame, rob	Topic 3
21: had*, [I] said, car, [he/she] said, up	
23: children, party, people, to live, houses	
alley, week, night, house, noise	
4: noises, disturbing, week, schedule, plan	Topic 4
11: neighbor, alley, volume, to sleep, music	



S2. More exclusive words by topic

Topic Label #	1 Employer	2 Traffic	3 Control	4 Noise	5 Household	6 Parking	7 Certificate	8 Generalized
Words								
1	commissioned officer	stop [sign]	had*	noise	door	cars	company	worst
2	chief	driver	day	neighbors	complaint	to park	money	you
3	police (adjective)	drive	vehicle	week	partner	exit	course	institution
4	unit	speed	after	disturbing	father	entry	[I] need	attention
5	personnel	sign	none*	music	mother	streets	thanks	justice
6	present non-commissioned	vehicle	precinct*	alley	sister	parked (plural)	santiago	phone
7	officer	offense	offense	volume	brother	it seems	email	delinquents
8	Prefecture	car	could*	to sleep	residency	avenue	certificate	people
9	officer	license	too*	"cuadrante" ¹	[they] said	street	email (English)	country
10	sergeant	track	also* according	neighbor	cell	municipality	job	chile
11	previous	truck [I]	to*	noise	arrived	sidewalk parked	protected	head
12	lieutenant	stopped	child	[I] live	dog	(singular)	hello	respect
13	service	lights	institution*	night	lady	vehicles	bank	citizens
14	particular	route	[they] had*	morning	while	big	afternoon	really
15	officers	plaque	precinct	music*	gun	[they] park	month	delinquent

Note: *spelling mistake. The exclusiveness of words is calculated by taking the harmonic mean of rank by frequency within the topic and rank by distribution of topic given word $p(z/w=v)$. ¹Surveillance system of the Carabineros.

S3. Regressions with topic-prevalence as outcome variable.

Variables	T1: Employee B	T2: Traffic B	T3: Control B	T4: Noise B	T5: Household B	T6: Parking B	T7: Certificates B	T8: Generalized B
Year (Ref. Cat.: 2013)								
2014	0.017 (0.021)	-0.010 (0.026)	-0.011 (0.022)	-0.038 (0.026)	0.051** (0.020)	-0.027 (0.023)	-0.011 (0.025)	0.028 (0.020)
2015	0.029 (0.021)	-0.038 (0.027)	-0.048** (0.020)	0.007 (0.026)	0.033* (0.020)	-0.037* (0.021)	0.022 (0.025)	0.031 (0.020)
2016	-0.016 (0.020)	0.004 (0.027)	-0.040** (0.019)	0.004 (0.027)	0.024 (0.020)	-0.009 (0.023)	0.023 (0.025)	0.010 (0.018)
2017	-0.020 (0.019)	-0.038 (0.025)	-0.034* (0.020)	0.029 (0.026)	0.024 (0.019)	-0.012 (0.021)	0.044* (0.024)	0.008 (0.019)
2018	-0.019 (0.017)	-0.035 (0.022)	-0.067*** (0.017)	0.128*** (0.022)	0.003 (0.016)	-0.017 (0.018)	0.028 (0.021)	-0.020 (0.016)
2019	-0.028 (0.017)	-0.055** (0.022)	-0.078*** (0.018)	0.130*** (0.024)	0.008 (0.017)	0.009 (0.018)	0.055** (0.022)	-0.040** (0.016)
2020	0.007 (0.022)	-0.071*** (0.026)	-0.088*** (0.020)	0.043 (0.027)	0.002 (0.019)	-0.054*** (0.020)	0.084*** (0.026)	0.078*** (0.019)
Educational level (Ref. Cat.: Q1)								
Q2	0.087*** (0.013)	0.083*** (0.017)	-0.173*** (0.013)	-0.064*** (0.018)	0.033*** (0.012)	-0.020 (0.013)	-0.070*** (0.016)	0.123*** (0.013)
Q3	0.083*** (0.013)	0.073*** (0.018)	-0.195*** (0.013)	-0.035** (0.017)	0.052*** (0.013)	-0.008 (0.013)	-0.066*** (0.016)	0.096*** (0.012)
Q4	0.040** (0.012)	0.024 (0.017)	-0.117*** (0.014)	0.002 (0.018)	0.041*** (0.012)	-0.005 (0.015)	-0.033** (0.016)	0.047*** (0.011)
Intercept	0.032 (0.032)	0.161*** (0.043)	0.260*** (0.033)	0.111** (0.046)	0.039 (0.034)	0.177*** (0.037)	0.182*** (0.042)	0.037 (0.032)
Province FE	YES	YES	YES	YES	YES	YES	YES	YES

Note: *p < .05. **p < .01. ***p < .001. Models controlled for month of complaint as factor variables. Standard errors in parentheses. Global approximation and 25 simulated draws from the variational posterior.