Analysis of Non-Discrimination Policies in the Sharing Economy

Miroslav Tushev, Fahimeh Ebrahimi, and Anas Mahmoud
The Division of Computer Science and Engineering
Louisiana State University
mtushe1@lsu.edu, febrah1@lsu.edu, amahmo4@lsu.edu

***Abstract*—Recent research has exposed a serious discrimination problem affecting applications of the Digital Sharing Economy (DSE), such as Uber, Airbnb, and TaskRabbit. To control for this problem, several DSE apps have crafted a new form of usage policies, known as non-discrimination policies (NDPs). These policies are intended to outline end-users' rights of equal treatment and describe how acts of bias and discrimination over DSE apps are identified and prevented. However, there is still a major knowledge gap in how such non-code artifacts can be formulated, structured, and evolved. To bridge this gap, in this paper, we introduce a first-of-its-kind framework for analyzing and evaluating the content of NDPs in the DSE market. Our analysis is conducted using a dataset of 108 DSE apps, sampled from a broad range of application domains. Our results show that, a) most DSE apps do not provide a separate NDP, b) the majority of existing policies are either extremely brief or combined as sub-statements of other usage policies, and c) most apps do not provide a clear statement of how their NDPs are enforced. Our analysis in this paper is intended to assist DSE app developers with drafting and evolving more comprehensive NDPs as well as help end-users of these apps to make more informed socioeconomic decisions in one of the fastest growing software ecosystems in the world.**

***Index Terms*—Digital Discrimination, Sharing Economy, Policy, Non-code artifacts**

I. Introduction

The Digital Sharing Economy (DSE) refers to a sustainable form of online business exchange that is built around sharing assets and resources rather than transferring their ownership [1]. Over the past decade, applications of DSE, such as Uber, TaskRabbit, and Airbnb, have caused major disturbances in established classical markets, enabling people to exchange and monetize their underused (or idle) assets and skills at an unprecedented scale [2]–[4]. As of today, there are thousands of active DSE platforms, operating in a market sector that is projected to grow to close to 335 billion U.S. dollars by 2025 [5].

The unique form of Peer-to-Peer (P2P) exchange that DSE platforms have enabled has been linked to significant levels of economic growth, especially in communities at the lower end of the economic ladder, helping unemployed and partially employed individuals to generate income, increase reciprocity, and access resources that are unattainable otherwise [1], [4], [6]–[8]. However, recent research has exposed a serious discrimination problem affecting these platforms [9]–[11]. Discrimination, as a general term, refers to cases where *"members of a minority group (women, Blacks, Muslims, immigrants, etc.) are treated differentially (less favorably) than members of a majority group with otherwise identical characteristics in similar circumstances"* [12]. In the context of DSE, discrimination (also known as *digital* discrimination) refers to a phenomenon where **an online business transaction is influenced by race, gender, age, or any other non-business related characteristics of receivers or providers**. This phenomenon is mainly facilitated by the P2P connection initiated between DSE users, encouraging different forms of established bias (e.g., racism, sexism, and ableism) to transfer online [4], [9]–[11]. In traditional economy markets, discrimination is countered by imposing anti-discriminatory laws [13]. For instance, the U.S. Civil Rights Act of 1964 guarantees equal treatment of customers in public accommodations, such as hotels or rental property. However, in the cyberspace, discrimination takes a different form that is often difficult to detect and deter.

In response to discrimination concerns, DSE developers have started rolling out a new form of policies for addressing potential issues of discrimination affecting their apps. A policy, in general, serves as a legally binding contract between apps and their end-users [14]. For instance, popular app marketplaces demand apps to provide a privacy policy to specify the types of information they collect about their users and outline how such information is being used, protected, and shared [14]. Similarly, non-discrimination policies (NDPs) are expected to determine the app's stance on discrimination and outline how acts of discrimination over the app are identified and handled. Privacy policies have received significant attention in the Software Engineering literature [15], [16]. This line of research aims to assess the quality of privacy policies as well as gauge best practices for drafting them. However, there is a widespread lack of knowledge about how NDPs can be structured. This can be attributed to the fact that NDPs are non-code artifacts. Creating and evolving such artifacts about a complex socio-technical phenomenon such as digital discrimination often fall outside of developers' expertise.

To address this knowledge gap, in this paper, we develop a framework for systematically analyzing and evaluating the content of NDPs in the DSE market. Our analysis is conducted using a dataset of 108 DSE apps, sampled from a broad range of application domains. The objectives of the proposed framework are to **a)** assist DSE app developers with drafting and evolving more comprehensive and less ambiguous NDPs, and **b)** help end-users of DSE apps to make more informed socioeconomic decisions in the DSE market, either as service providers or receivers.

The remainder of this paper is organized as follows. Section II motivates our work and discusses our research questions. Section III describes our data collection process. Section IV describes our NDP quality assessment framework. Section V presents our results. Section VI discusses our key findings and their impact as well as the main limitations of our study. Finally, Section VII concludes the paper and discusses our future work.

II. Motivation and Research Questions

In this section, we review existing evidence on the problem of digital discrimination, motivate our work, and discuss our research questions.

A. Digital Discrimination

The problem of digital discrimination in online DSE markets has been well-documented in recent years. Numerous large-scale surveys and field studies have provided significant evidence on various forms of systematic bias across almost all application domains of DSE, including discrimination based on ethnic background (racism), gender or sexual orientation (sexism), and physical appearance (ableism) [9], [10], [17]–[20]. For instance, Ge et al. [9] hired research assistants of different racial backgrounds to request UberX rides. The authors found that the waiting times for Black riders were significantly longer. In addition, more cancellations were observed against Black riders than their White counterparts. In another study, Moody et al. [17] surveyed 1,100 of UberPOOL and Lyft riders. The results showed that White passengers that lived in predominantly White communities were more likely to discriminate against passengers of other races.

Edelman et al. [10] examined racial discrimination over the lodging platform Airbnb. The authors reported that applications from guests with distinctively Black names were 16% less likely to be accepted relative to identical guests with distinctively White names. Discrimination in the lodging business has also been observed against members of the LGBT community. For example, Ahuja and Lyons [18] analyzed Airbnb hosts' responses to LGBT accounts. The results showed that hosts were more likely to not reply at all rather than replying *"no"* to male-male pairs inquiring about room availability. Ableism (discrimination against people with disabilities) was also reported over Airbnb. For instance, in a randomized field experiment of 3,847 lodging requests, Ameri et al. [21] found that hosts were less likely to approve requests from travelers with blindness, cerebral palsy, dwarfism, or spinal cord injury than to approve travelers without disabilities.

Patterns of digital discrimination have also been observed in the freelancing domain. Thebault et al. [19] surveyed workers on TaskRabbit from the Chicago metropolitan area. The authors found that requests from customers in the socioeconomically disadvantaged South Side area were less likely to be accepted. Hannák et al. [11] analyzed worker profiles on TaskRabbit and Upwork. The results showed that there was a significant bias against White women and Black men on both platforms. In another study, Foong et al. [22] collected self-determined hourly bill rates from the public profiles of 48,019 workers in the U.S. (48.8% women) on Upwork. The authors found that the median woman on Upwork requested only 74% of what the median man requested in hourly bill rate. Another study by Barzilay and Ben-David [20] showed that women's average hourly rates on P2P freelancing platforms were about two-thirds of men's rates. Such gaps persisted even after controlling for experience, educational background, and hours of work.

B. Motivation and Research Questions

Policies have long been used as legally-binding usage contracts between software platforms and their end-users [14]. For instance, privacy policies are used by app developers to communicate their data collection and sharing practices with their end-users as well as to comply with privacy legislation around the world. These policies have generated significant research interests in recent years [23]. Privacy policy research is primarily focused on detecting violations of the claims made in the policy [24], [25], evaluating the readability and comprehensibility of policies [24], [25], and mining their content for software privacy requirements [15], [16], [26]. NDPs, on the other hand, have received considerably less attention in both research and practice. This can be attributed to the fact that digital discrimination is an inherently complex phenomenon that is often enabled by equally complex interactions between DSE apps' features, their end-users, and operational environments. Therefore, drafting NDPs that are tailored to address the specific types of bias affecting different application domains can be a very challenging and time-consuming process.

To address these limitations, in this paper, we conduct a first-of-its-kind study to analyze NDPs in the Digital Sharing Economy. Our work aims to **a)** study the prevalence of NDPs in the DSE market, **b)** propose a framework for systematically analyzing the content of these policies, and **c)** use that framework to assess the quality of existing NDPs. Our work is intended to spread awareness of digital discrimination and provide app developers, either maintaining DSE apps or developing new ones, with systematic guidelines to draft high quality NDPs and evolve such non-code artifacts with minimum overhead. Moreover, providing complete and structured NDPs can help DSE app users to make more optimized socioeconomic decisions when it comes to navigating the landscape of existing DSE platforms. To guide our analysis, we formulate the following research questions:

- ***RQ$_1$:*** *How prevalent are NDPs in the DSE market?* Under this research question, we investigate the prevalence of anti-discrimination policies among DSE apps. This type of analysis aims to explore the state-of-practice in the different application domains of DSE when it comes to NDPs.

- ***RQ*$_2$*:** *Can the quality of existing NDPs be systematically evaluated?* Under this research question, we seek to develop a systematic framework for analyzing the content existing NDPs as well as assess their quality.
- ***RQ*$_3$*:** *How detailed and informative are existing NDPs?* Under this research question, we examine the quality of information provided in existing NDPs. Our objective is to determine a set of quality standards that can be used by new DSE apps, or existing apps with no policies, to draft and evolve their own NDPs.

III. Data Collection

In this section, we describe our data collection process, including selecting apps to be included in our dataset, categorizing these apps, and collecting their NDPs.

A. Dataset

Recent statistics estimate that there are thousands of active DSE platforms listed on popular mobile app marketplaces. However, only a handful of these apps are typically investigated in digital discrimination research. Such apps include Uber and Lyft from the domain of ride-sharing, Airbnb from the lodging domain, and Upwork and TaskRabbit from the domain of freelancing [9], [10], [17]–[20]. These apps operate in large geographical areas and have massive user bases, thus, discrimination concerns are more likely to manifest over them rather than smaller ones. Based on these observations, for a DSE platform to be included in our analysis, it has to meet the following criteria:

1) A platform must facilitate some sort of a P2P connection and include the sharing of some sort of a resource, such as an asset (e.g., an apartment, car, electric drill, etc.) or a skill (e.g., plumbing, hair styling, coding, etc.).
2) A platform must have an app on Google Play or the Apple App Store. App stores provide various metrics that can help us to locate popular apps, such as the number of app reviews, stars, and their download statistics.
3) A platform must be located and/or have a substantial presence in the US. The U.S. Civil Rights Act of 1964 prohibits discrimination based on race, sex, religion, nationality, or sexual orientation. By focusing on the US market, we ensure that our selected apps operate in a country where discrimination is prohibited by law.

With these criteria in place, we searched for apps to be included in our dataset. Our data collection took place between January and February of 2021. We started by seeding our dataset with Uber, Lyft, Airbnb, Upwork, TaskRabbit, and Fiverr. Existing literature has provided a significant evidence of discriminatory behavior affecting these apps. We then conducted a Google search using the query: `(sharing OR shared OR gig) AND economy AND (platforms OR apps OR systems)`. We examined the first 10 pages of the search results and added 72 new platforms that matched our inclusion criteria. We then used the *similar* feature on Google Play and the Apple App Store to locate any apps we missed through the Google search.

TABLE I: Descriptive statistics for the 108 apps in our dataset.

Metric	**Mean**	**Median**	**Min**	**Max**
App Store Rating	4.23	4.60	1.60	4.90
Google Play Rating	3.86	3.90	2.00	4.90
App Store # of Reviews	201K	2.4K	2	8.9M
Google Play # of Reviews	134K	1.3K	7	7.91M
Google Play # of Installs	6.9M	100K	1K	500M

Specifically, we examined the list of similar apps resulting from searching app stores for each of our 72 apps. Lightweight snowballing was then used to add any major apps that we might have missed. Apps were iteratively added until no more new apps that satisfied our inclusion criteria were located. In total, 108 unique apps were included in our dataset. Descriptive statistics of our dataset are provided in Table I.

B. App categorization

The Apple App Store and Google Play classify apps into generic categories of loosely related functionalities. These categories are often ambiguous (too generic) or straight-up misleading [27], [28]. For example, both Uber and Airbnb are categorized under the *Travel* category in the Apple App Store and DoorDash is classified under the *Food&Drink* category. This type of generic categorization does not provide enough information about the specific application domains of apps. To overcome this limitation, we begin our analysis by re-classifying apps in our dataset into more fine-grained categories of DSE application domains.

While automated app classification techniques are available [27], [28], given the relatively small size of our dataset, we conducted the classification manually. In particular, three judges, all with graduate degrees in Software Engineering and an average of three years of industrial experience, independently examined the description of each of our apps available on the Apple App Store and Google Play as well as each app's official web-page. Categories of apps were recorded as they emerged in the text. We used memoing to keep track of the reasoning behind each suggested category. Axial coding was then used to consolidate individual categories into more abstract categories [29]. For example, the categories of *food delivery* and *grocery delivery* were merged into a single *Delivery* category and *boat-sharing* and *bike-sharing* were merged into *asset-sharing*. Generated categories were then iteratively revised until no more categories were found. By the end of our classification process, six main categories of DSE apps, shown in Fig. 1, have emerged. These categories can be described as follows:

- **Skill-based:** These apps facilitate the sharing of personal skills (hiring labor). Specific examples include the baby sitting apps Sittercity and Urbansitter, the tutoring apps Verbling, Codementor, and Classgap, and the freelancing apps Fiverr and Upwork.

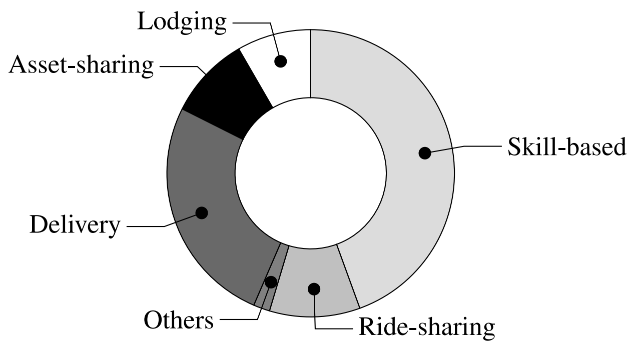

Fig. 1: The application domains of DSE apps in our dataset.

- **Delivery:** Under this category, we include apps which enable users to utilize their vehicles to deliver goods to other users. Examples of apps in this category include UberEats, Grubhub, and Shipt for grocery and food delivery, and DriveMatch, uShip, and Dolly for hiring delivery drivers.
- **Ride-sharing:** This category includes apps which allow their users to share rides, such as carpooling and driver/rider connections. Examples of apps in this category include traditional ride-sharing services, such Uber, Lyft, and Via, as well as more specialized platforms, such as HopSkipDriver for children transportation, Veyo for medical transportation, and Wingz for hiring a driver.
- **Asset-sharing:** Under this category, we include any app which enables users to share their assets. Specifically, the resource being shared is a physical resource (e.g., a vehicle or an electric drill), not a person's time or skills (e.g., a driver or electrician). Examples of apps under this category include the car sharing apps Turo and HyreCar, the boat sharing apps Get-MyBoat and Boatsetter, the bike sharing app Spinlister, and the RV sharing apps RVezy and Outdoorsy.
- **Lodging:** This category contains renting and short-term accommodation services such as Airbnb, Vrbo, and Misterbnb as well as space-sharing for storage (Neighbor), events (Splacer), and even parking (ParqEx).
- **Other:** Although our objective was to classify all apps into the main general categories, two apps in our dataset were too niche-oriented to warrant a creation of a separate category. These apps are Prosper for lending and borrowing money and Kickstarter, a platform for crowdfunding various projects.

C. Policy collection

To answer our first research question, we collect the NDPs of the apps in our dataset. Unlike privacy policies, mobile app marketplaces do not enforce NDPs, therefore, locating such policies can be a challenging task. For instance, most privacy policies are often titled *Privacy Policy*, however, NDPs are titled differently, including titles such as, *non-discrimination*, *anti-discrimination*, or *inclusion statement*. To locate such policies, we explore the website of each app as well as the app itself. Any web pages or app screens that address discrimination are collected as a potential NDP.

To identify these pages, we utilized Google's search operators to search apps' websites directly using the query `site: <app website> AND (discrimination OR <discrimination types (Table II)>)`. Table II lists the main acts of discrimination as described by the U.S. Equal Employment Opportunity Commission. These acts commonly appear in diversity and social justice literature [30]. For any app that we could not locate a policy, we performed a manual search of its website. Our search exposed three categories of apps when it comes to NDPs. These categories include:

- **Separate policy:** This category includes apps which maintain a separate NDP that is provided on its own separate page. In total, 16 apps had a separate NDP.
- **Combined policy:** Nine apps in our dataset combined their NDP with other usage policies, such as sexual harassment policies, community guidelines, code of conduct, or even the Terms of Service (ToS) of the app.
- **No policy:** For the majority of apps (79) in our dataset, we were either unable to locate a policy, or only located a generic one-line anti-discrimination statement that was provided in the ToS of the app. Some apps provide some sort of a statement on diversity or commitment to diversity. These statements typically take the form of a blog post rather being a policy with rules and implications. For example, Gopuff, a delivery app, published a commitment to creating more equal and just future in response to the death of George Floyd.

The distribution of these three categories of NDPs over our categories of DSE application domains is shown in Fig. 2. In general, to answer ***RQ$_1$***, we can safely say that the majority of apps in our dataset do not provide NDPs. We found that some apps merge their NDPs with other policies, while only a few of the apps publish a separate NDP[1].

IV. Quality Assessment of NDPs

In this section, we propose a framework for assessing the quality of NDPs in the DSE market. The process of policy assessment is typically conducted manually, following a systematic process that checks the content of the policy against a set of predefined quality measures [31]–[35]. These measures range from simple quantitative metrics, such as the length of the policy [36], to more complex measures, such as its readability and compliance with regulations [37]. To generate such a protocol, we rely on two sources of information:

- **Nondiscrimination regulations:** The U.S. Equal Employment Opportunity Commission (EEOC) suggests an outline of topics that US-based employers should include in their NDPs[2]. While these guidelines focus on discrimination against employees (rather than end-users), they can

[1] https://seel.cse.lsu.edu/data/ICSME21.xlsx

[2] https://www.eeoc.gov/employers/small-business/general-non-discrimination-policy-tips

TABLE II: Most common types of discrimination.

Type	Discrimination against:
Racism	Ethnicity, color, or nationality.
Sexism	Gender or sexual orientation.
Ableism	Physical, sensory, or intellectual disability.
Parental	Parents with children or pregnant women.
Ageism	Older or younger people.
Religious	Perceived religion or a set of beliefs.
Classism	Particular social class.

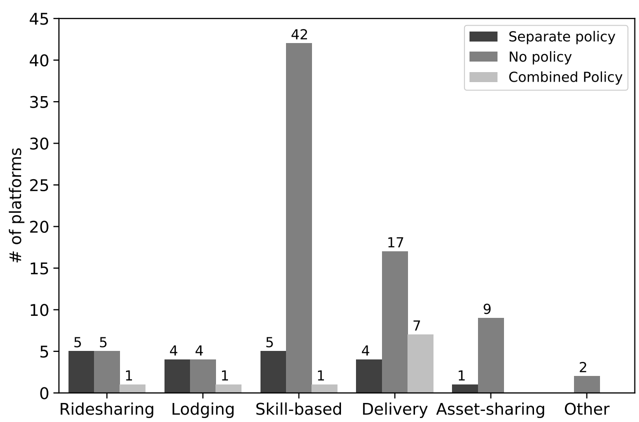

Fig. 2: Categorization of apps by their NDP status.

be used to establish the structure of NDPs. For example, the EEOC guidelines state that NDPs should include specific types of discrimination, a reporting mechanism, and consequences of violating the policy.

- **Privacy policy assessment protocols:** Existing protocols for evaluating privacy policies can serve as a baseline, or a reference, to develop an evaluation protocol for NDPs. Such protocols include a set of measures that can be directly inferred from the policy. Typically, evaluators are provided with a set of questions to help them evaluate the content of the policy based on the predefined measures [31], [37], [38].

Based on these two sources of information, we design a protocol for evaluating NDPs in the DSE market. The specific measures of our protocol, along with their descriptions and their associated evaluation questions are provided in Table III. In general, our measures can be divided into a set of automatically calculated measures, including the policy's length and readability, and manually determined measures, including the types of discrimination mentioned in the NDP, the number of examples provided, references to legislation, and whether the policy mentions enforcement and ramification mechanisms.

Once our review protocol was defined, we printed out the NDPs collected for our apps. Each of our three judges went through each NDP independently, answering the questions related to the criteria from (4-8) in Table III. Results were then compiled and summarized in Table IV. Overall, given the specific nature of our questions, only a few coding errors (inaccuracies in answering some of the questions) were detected and corrected.

V. Results and Analysis

In this section, we discuss the results of applying our evaluation protocol in Table III to the NDPs in our dataset.

A. Policy Name

Our results show that NDPs are named differently by different apps. Titles, such as *Anti-Discrimination Policy* and *Non-Discrimination Policy* are common. However, we found more variations of these titles, such as *Zero Tolerance Policy*, *Deactivation Policy*, *Inclusion Policy*, and more. In general, these variations can impact the accessibility of NDPs negatively [36]. This became clear during our data collection as we had to resort to a sophisticated Google query to retrieve the NDPs of our apps (Sec. III-C).

B. Length and Readability

The Flesch Reading Ease (FRE) [41] is a popular metric used to assess the readability of text. The value of FRE ranges from 0 to 100, where a higher score indicates that the text is easier to read. The metric is calculated by the following formula:

$$206.835 - 84.6 \times \frac{\text{\# of syllables}}{\text{\# of words}} - 1.015 \times \frac{\text{\# of words}}{\text{\# of sentences}} \quad (1)$$

The core idea behind FRE is that longer words and longer sentences are more difficult to comprehend. Therefore, FRE penalizes texts with a high number of syllables per word and a high number of words per sentence. Fig. 3 shows an example of how FRE can be calculated for a single sentence.

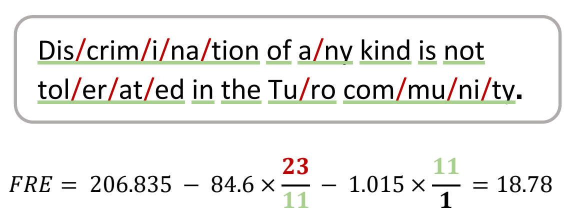

Fig. 3: An example of FRE calculation for a text with a single sentence. The sentence contains 23 syllables and 11 words.

FRE is commonly used in policy assessment research [36], [39], [40]. It is important to note that this metric is only suitable for longer texts. Therefore, we calculated FRE only for NDPs with 100 words or more. To compute FRE, we used Python's Readability library[3] which implements a set of traditional readability measures based on simple surface characteristics of text. The distribution of length and readability scores over our NDPs are presented in Fig. 4. Our results show that the average FRE for the policies in our dataset is

[3] https://pypi.org/project/readability/

TABLE III: Assessment measures of NDP quality.

No	Measure	Description	Questions for evaluators
1	Name (N)	The name of the policy is the title of the document the policy is listed under. Separate policies with well-defined titles are more easily accessible, thus can be considered higher in quality [36].	Determined during policy collection.
2	Length (L)	The length of the policy (number of words) can be used as a basic measure of its quality. Intuitively, longer policies are assumed to be more detailed [36]	Measured automatically as the number of words.
3	Readability (FRE.)	Readability is another measure that is commonly used to assess the quality of policies [36], [39], [40]. The more readable the policy, the more accessible it is for the casual user.	Calculated automatically using the Flesch Reading Ease (FRE) metric. [41]
4	Types (T)	Discrimination in DSE apps can take many forms (Table II). Therefore, a policy that explicitly mentions more of these types is considered higher in quality, or more comprehensive. The US EEOC states that discrimination based on race, color, religion, sex (including pregnancy, sexual orientation, or gender identity), national origin, disability, age or genetic information (including family medical history) is illegal.	How many specific types of discrimination does the policy mention?
5	Examples (Ex.)	A policy which provides examples of specific types of discriminatory behavior that might affect the app is considered to be higher in quality. Examples are used to demonstrate what actions might be classified as discriminatory. In policy analysis, examples are considered an important instrument to communicate policy practices with the casual user [33].	Does the policy provide any examples of discriminatory behavior? How many examples are provided?
6	Legislation (Lg.)	This criterion assesses whether a policy contains references to existing anti-discrimination regulations in the judicial area in which the app operates. For example, Internet privacy policies are often assessed based on their compliance with existing privacy regulations [31], such as the Federal Trade Commission's Fair Information Practices guidelines [33].	Does the policy refer to any existing legislation?
7	Enforcement (En.)	A policy which lists the measures (functional or non-functional) taken by the app to mitigate discrimination is considered higher in quality [33]. In fact, the US EEOC states that a NDP should explain how employees can report discrimination. These types of mechanisms also include methods for reporting incidents of policy violation.	Does the policy list any features or protocols that the app implements to mitigate discrimination? Is there a reporting mechanism in place?
8	Ramifications (Rmf.)	A policy which mentions the ramifications for discriminatory behavior is considered more comprehensive. The US EEOC states that a NDP should describe the consequences of violating the policy.	Does the policy mention the types of actions (penalties) to be imposed on policy violators?

23.74. This level indicates that the text is difficult to read, best understood by college graduates. The apps Misterb&b, Spareroom, and Sittercity have the highest readability scores, while Roadie and Thumbtack have the lowest scores. In terms of length, Airbnb, GoShare, and Turo have the longest, thus more detailed policies. Uber's and Lyft's NDPs were surprisingly short (134 and 97 words respectively).

C. Types

Our annotation shows that most policies list a large number of discrimination types in their NDPs. TaskRabbit, in particular, refers to 17 different types, including racism, color, ancestry, national origin, religion, creed, age, sex, gender, physical or mental disability, medical condition, genetic information, marital or civil partner status, military or veteran status. The policy even provides more sub-types of discrimination, such

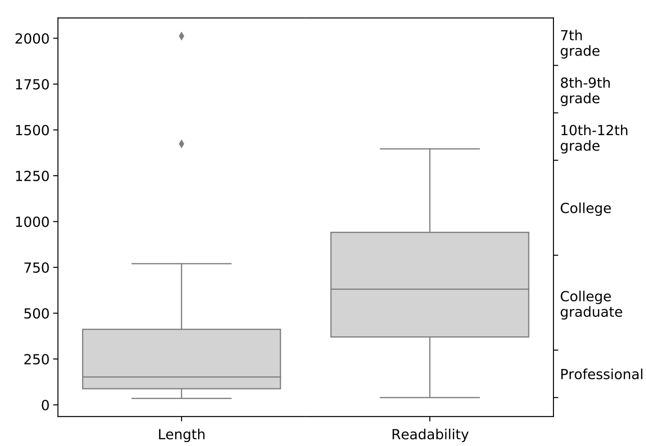

Fig. 4: Length and readability of the NDPs in our dataset.

Fig. 5: A frequency-based word cloud of the different types of discrimination mentioned in the NDPs of apps in our dataset.

as, *"gender (including pregnancy, childbirth, breastfeeding or related medical conditions)."* On average, NDPs in our dataset mention 10 types of discrimination per policy. Racism, national origin, disability, religion, age, gender identity, and marital status are the most frequent (Fig. 5).

D. Examples

Our manual annotation shows that examples are not common in NDPs. Airbnb, Turo, and Neighbor provide the most comprehensive set of examples, described in the form of *"user may not"* scenarios that could take place while using the app. For instance, Airbnb policy states that, *"Airbnb hosts may not decline a booking from a guest based on gender identity unless the host shares living spaces (for example, bathroom, kitchen, or common areas) with the guest"*. Neighbor's NDP provides some of the best examples in terms of quantity and quality. The app provides 15 examples of what is considered discriminatory behavior, such as, *"Posts that assume someone is suspicious because of their race or ethnicity"*. Turo is another app which provides thorough examples of discriminatory behavior, such as, *"Turo hosts may not make assumptions about the guest's ability to operate their vehicle."*

E. Legal

Only three apps in our dataset provide references to specific counter-discrimination legislation. GoShare's NDP for example, states that, *"A variety of federal, state, and local laws strictly prohibit such forms of discrimination, including Title VII of the Civil Rights Act of 1964, the Age Discrimination Act of 1967, and the Americans with Disabilities Act of 1990."* Some other apps provide a generic legal statement. For example, TaskRabbit's NDP states that, *"or any other basis protected by applicable laws in jurisdictions in which TaskRabbit operates (collectively referred to as a protected class)."* Similarly, Upwork's NDP states that, *"we expect all clients and freelancers to comply at all times with the laws concerning discrimination and harassment."* Turo's NDP refers to cases outside the United States and Canada, stating that, *"hosts aren't required to comply with the above policies if they violate local laws."* However, no references to any specific laws are made.

F. Enforcement and Ramifications

A total of 15 apps in our dataset describe a set of measures taken to enforce their NDPs, including the ramifications for violating the policy. In general, our analysis revealed two categories of enforcement mechanisms:

- **Monitoring:** Some apps indicate in their policies that they monitor the actions of their users to detect discriminatory behavior. For example Airbnb and Neighbor state in their NDPs that they may suspend hosts who have demonstrated a pattern of rejecting guests from a protected class. Furthermore, listings over these apps are constantly checked for language contrary to their nondiscrimination policies. It is not clear, however, what constitutes a pattern of discrimination or what language is considered discriminatory.
- **Reporting:** Some apps use reporting mechanisms to enable their users to report any incidents of perceived discrimination or unlawful bias. For instance, Turo has a *"Support Form"* for users to report issues of discrimination and GoShare provides a full procedure on how to report alleged cases of discrimination and harassment.

In terms of ramifications, most apps which provide an enforcement mechanism also provide a statement indicating that proven cases of frequent discrimination would result in removal from the app or suspending the user temporarily or indefinitely.

G. Comparing domains

In terms of application domain, lodging apps in our dataset (except for Vrbo) have the highest quality NDPs (length = 700 words, readability = 40.5, types = 7, examples = 9). The asset-sharing app Turo has slightly higher numbers, however, it is the only app in its domain that has a NDP. Apps in the ride-sharing domain seem to have low quality NDPs in comparison to other domains (length = 107 words, readability = 9.4, types = 10, examples = 0). These results are surprising given that existing literature provided significant evidence of systematic bias affecting these apps [9], [17]. The same applies to delivery apps, however, these apps did not receive as much attention as ride-sharing apps in the digital discrimination literature [42]. While skill-based apps (length = 248 words, readability = 13.6, types = 10, examples = 2) are slightly better than ride-sharing apps, they still lag behind lodging apps. This is also surprising given that apps in this domain are known to have serious discrimination issues, such as bias against women and black workers, including lower hourly rates, lower ratings, and racially and sexually charged reviews [11], [19], [20], [22].

VI. Discussion and Impact

Given the general shift in society towards more equality and prosperity, we anticipate that NDPs are going to become mandated by law in the near future. However, in the absence of a standardized format and the lack of regulations, drafting such policies remains a challenging and time-consuming task. To help overcome these challenges, the framework presented in this paper provides developers with a systematic protocol for

TABLE IV: NDP content assessment results. The Flesch Reading Ease (FRE) metric is calculated for NDPs with L $\geq$ 100. The three largest values in each numerical column are highlighted.

Domain	App	NDP title	NDP type	L	FRE	T	Ex	Lg	En	Rmf
Ride-sharing	Uber	Non-Discrimination Policy	Separate	134	18.78	9	1	✗	✗	✓
	Lyft	Anti-Discrimination Policies	Separate	97	-	10	0	✗	✓	✓
	Via	Anti-Discrimination Policy	Separate	102	30.77	8	0	✗	✓	✓
	HopSkipDrive	Zero Tolerance Policy	Combined	37	-	11	0	✗	✗	✗
	Veyo	Non-Discrimination Statement	Separate	189	8.30	11	0	✓	✓	✗
	Wingz	Non-Discrimination Policy	Separate	88	-	9	1	✗	✗	✓
Asset-sharing	Turo	Nondiscrimination policy	Separate	770	23.78	13	15	✗	✓	✓
Delivery	Doordash	Deactivation Policy	Combined	64	-	14	0	✗	✗	✗
	Grubhub	Policy Against Sexual and Other Forms of Harassment	Combined	60	-	16	0	✗	✗	✗
	uShip	Code of Conduct	Combined	153	35.17	9	0	✗	✗	✗
	GoShare	Anti-Sexual Harassment and Anti-Discrimination Policy	Combined	1424	13.00	11	0	✓	✓	✓
	Postmates	Anti-Discrimination Policy	Separate	72	-	11	1	✗	✗	✓
	Roadie	Discrimination And Sexual Harassment Policy	Combined	480	12.10	13	0	✗	✓	✓
	Instacart	Community Guidelines for Customers	Combined	35	-	11	0	✗	✗	✗
Lodging	Airbnb	Nondiscrimination Policy	Separate	2012	34.70	8	22	✗	✓	✓
	Misterb&b	Anti-Discrimination Policy	Separate	420	52.40	5	5	✗	✓	✓
	Vrbo	-	Separate	91	-	0	0	✗	✗	✓
	Neighbor	Nondiscrimination Policy	Combined	572	30.90	8	15	✗	✗	✓
	Spareroom	Fair Housing	Separate	399	44.33	14	5	✓	✗	✗
Skill-based	Taskrabbit	Anti-Discrimination and Harassment Policy	Separate	403	0	17	4	✗	✓	✓
	Upwork	Commitment to Nondiscrimination, Inclusion, and Respect	Separate	308	21.90	10	7	✗	✗	✗
	Thumbtack	Non-Discrimination Policy	Separate	152	18.60	11	3	✗	✓	✓
	Jobstack	Equal Employment Opportunity Policy	Separate	412	0	21	0	✗	✓	✓
	Sittercity	Community Inclusion Policy	Separate	132	40.93	0	0	✗	✗	✗
	Withlocals	Code of Conduct for Withlocals Guests	Combined	83	-	0	0	✗	✗	✗
Average				**347.56**	**23.74**	**10**	**3.16**			

evaluating their policies based on their intrinsic characteristics and by comparing them to existing high-quality NDPs. This framework can also help developers to keep their NDPs in-check during software evolution. This can be particularly important for start-ups, where it can be financially infeasible to hire a third-party firm to take care of the policy as the system evolves and as we learn more about the problem.

Our work in this paper bridges an important gap in the software maintenance and evolution research by focusing on non-code artifacts. Maintaining software policies is a prime example of adaptive maintenance tasks, where an artifact has to constantly change in order to adapt to external factors, such as changing regulations. In fact, such policies can be used to monitor the evolution of the system by monitoring changes to the NDP. Existing research suggests that important information about the system can be inferred from the modifications made to its privacy policy [37]. Furthermore, providing informative, comprehensive, and accessible NDPs can help users to make more informed decisions in the DSE market. In particular, users often find themselves having to choose from among hundreds of DSE platforms. The ability to make the right decisions in such a volatile market is critical for users to maximize their social and economic gains [4], [43].

In terms of results, our analysis shows that quality of NDPs varies among apps and application domains. Lodging apps seem to have the highest quality NDPs, while ride-sharing and skill-based apps do not provide informative NDPs. In terms of individual apps, the vehicle-sharing app Turo and the lodging app Airbnb provide the most comprehensive policies. Another observation is that apps do not mention in their NDPs the design strategies they use to mitigate discrimination. For instance, to control for bias in reviews, Airbnb rolled out a design change to ensure that hosts and guests can see the reviews only after both parties have submitted their reviews. According to Airbnb, *"Both hosts and guests may worry that if they leave an honest review that includes praise and criticism, they might receive an unfairly critical review in response. To address this concern, reviews will be revealed to hosts and guests simultaneously"* [44]. However, such a feature update is only mentioned in the blog maintained by Airbnb and is not highlighted in the NDP.

Our recommendation for developers drafting their own NDPs is to refer to apps' with high quality policies (e.g., Airbnb and Turo) as good industry standards and to keep up with existing non-discrimination regulations. Furthermore, developers should always refer to emerging research on digital discrimination. Such research constantly exposes problems of bias in DSE as well as suggests and evaluates mitigation strategies for these problems [45]–[47].

In terms of limitations, the main threat to the external validity of our study stems from the fact that only 108 popular DSE apps were considered in our analysis. However, as mentioned earlier, discrimination issues are more likely to manifest over these apps rather than smaller apps which typically target homogeneous populations of users. Furthermore, our search process utilized multiple search strategies and inclusion criteria to locate a representative sample. Generally speaking, the size of the dataset is aligned with datasets typically used in policy analysis research [36]–[38]. Another threat might stem from the fact that our evaluation of NDPs was conducted manually. Nonetheless, manual inspection of policies is a common practice in such kind of studies. This threat can be mitigated by using a systematic review process and a well-defined review protocol with multiple judges. Furthermore, the majority of evaluation measures were quantitative in nature, therefore, subjectivity threats were minimized. Other concerns might be raised about the measures or the questions used in the evaluation protocol [31], [37], [38]. However, the majority of these measures were adapted from well-established protocols for evaluating privacy policies as well as existing regulations. These measures capture to a large extent the different aspects of NDPs in the DSE market. Finally, to summarize our findings in this paper, we revisit our research questions:

- ***RQ_1:*** *How prevalent are NDPs in the DSE market?* Our analysis of 108 DSE apps shows that NDPs are not common. Most apps either do not provide a NDP at all or provide a very brief and generic statement. Only a few apps maintain a separate NDP. Such policies appear under various names. This might negatively impact their discoverability and accessibility [36].
- ***RQ_2:*** *Can the quality of existing NDPs be systematically evaluated?* Existing anti-discrimination regulations as well as protocols for evaluating the content of software privacy policies can be adapted to NDPs. Specifically, NDPs can be evaluated based on a set of measures that can be extracted directly from the policy. These measures include quantitative metrics, such as the policy's length and its readability as well as the number of examples and types of discrimination acknowledged in the policy, along with more qualitative measures, such as whether the policy describes any measures taken to mitigate discrimination and how cases of violation are reported and handled. While these measures capture all the aspects of NDPs, other, more complex, measures which go beyond the surface characteristics of policy text can be used.
- ***RQ_3:*** *How detailed and informative are existing NDPs?* Our analysis shows that the majority of NDPs in the DSE market are of low quality. Either they are very brief or do not provide sufficient information on what is considered discriminatory behavior or how that behavior is controlled for through the functional features of the app.

VII. Conclusions

In this paper, we presented a framework for evaluating NDPs of DSE apps. Our framework is based on an assessment protocol which uses a set of predefined measures to evaluate the quality of NDPs. Our results showed that most DSE platforms do not provide any form of NDPs. The results also showed that most of the NDPs are either brief, combined with other existing policies, or do not include essential information that is necessary to outline the app's stance on discrimination. On average, apps in the lodging domain provide the most comprehensive policies, while apps in other domains still lag behind. Our work in this paper aims to help software developers working with DSE apps to draft and maintain effective NDPs for their apps as well as help users to realize their rights to be treated fairly in one of the fastest growing software ecosystems in the world. Finally, our work in this paper will be extended across two main directions:

- **Automation:** We will use text mining and modeling techniques to automatically learn the structure of NDPs, the main topics they discuss, and eventually generate an overall quality score for the policy. A fully automated prototype will be made publicly available to help app developers around the world draft high quality NDPs.
- **User studies:** Automated quality metrics, such as readability, can provide an indication of NDPs' accessibility to the casual user. However, to enable a more objective assessment, user studies must be conducted. Such studies will involve recruiting large samples of DSE users (providers and receivers) and using systematic questionnaires to assess their understanding of NDPs.

Acknowledgment

This work was supported by the U.S. National Science Foundation (Award CNS 1951411).

References

[1] C. Martin, "The sharing economy: A pathway to sustainability or a nightmarish form of neoliberal capitalism?" *Ecological Economics*, vol. 121, pp. 149–159, 2016.

[2] T. Dogru, M. Mody, and C. Suess, "Adding evidence to the debate: Quantifying Airbnb's disruptive impact on ten key hotel markets," *Tourism Management*, vol. 72, pp. 27–39, 2019.

[3] G. Quattrone, D. Proserpio, D. Quercia, L. Capra, and M. Musolesi, "Who benefits from the "sharing" economy of airbnb?" in *International Conference on World Wide Web*, 2016, pp. 1385–1394.

[4] T. Dillahunt and A. Malone, "The promise of the sharing economy among disadvantaged communities," in *Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 2285–2294.

[5] PwC, "The sharing economy: Consumer intelligence series," *PricewaterhouseCoopers LLP*, 2015.

[6] J. Hamari, M. Sjöklint, and A. Ukkonen, "The sharing economy: Why people participate in collaborative consumption," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, pp. 50–63, 2016.

[7] G. Zhu, K. So, and S. Hudson, "Inside the sharing economy: Understanding consumer motivations behind the adoption of mobile applications," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 9, pp. 2218–2239, 2017.

[8] S. Shaheen, N. Chan, A. Bansal, and A. Cohen, "Mobility and sharing economy: Impacts synopsis," 2015.

[9] Y. Ge, C. Knittel, D. MacKenzie, and S. Zoepf, "Racial and gender discrimination in transportation network companies," no. NBER Working Paper No. 22776, 2017.

[10] B. Edelman, M. Luca, and D. Svirsky, "Racial discrimination in the sharing economy: Evidence from a field experiment," *American Economic Journal: Applied Economics*, vol. 9, no. 2, pp. 1–22, 2017.

[11] A. Hannák, C. Wagner, D. Garcia, A. Mislove, M. Strohmaier, and C. Wilson, "Bias in online freelance marketplaces: Evidence from TaskRabbit and Fiverr," in *ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 1914–1933.

[12] M. Bertrand and E. Duflo, *Handbook of Field Experiments*. Elsevier, 2016, ch. Field Experiments on Discrimination, p. 310.

[13] L. Ratti and N. Countouris, "The sharing economy and eu anti-discrimination law," *The Cambridge Handbook of the Law of the Sharing Economy*, pp. 486–498, 2018.

[14] X. Wang, X. Qin, M. Bokaei, R. Slavin, T. Breaux, and J. Niu, "Guileak: Tracing privacy policy claims on user input data for Android applications," in *Inter. Conf. on Software Engineering*, 2018, pp. 37–47.

[15] S. Blenner, M. Köllmer, A. Rouse, N. Daneshvar, C. Williams, and L. Andrews, "Privacy policies of android diabetes apps and sharing of health information," *The Journal of the American Medical Association*, vol. 315, no. 10, pp. 1051–1052, 2016.

[16] J. Bhatia, T. Breaux, and F. Schaub, "Mining privacy goals from privacy policies using hybridized task recomposition," *ACM Transactions on Software Engineering and Methodology*, vol. 25, no. 3, p. 22, 2016.

[17] J. Moody, S. Middleton, and J. Zhao, "Rider-to-rider discriminatory attitudes and ridesharing behavior," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 62, pp. 258–273, 2019.

[18] R. Ahuja and R. Lyons, "The silent treatment: LGBT discrimination in the sharing economy," Trinity College Dublin, Department of Economics, Dublin, Tech. Rep., 2017.

[19] J. Thebault-Spieker, L. Terveen, and B. Hecht, "Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets," in *ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2015, pp. 265–275.

[20] A. Barzilay and A. Ben-David, "Platform inequality: Gender in the gig-economy," *Seton Hall Law Review*, vol. 47, no. 2, pp. 393–431, 2017.

[21] M. Ameri, S. Rogers, L. Schur, and D. Kruse, "No room at the inn? disability access in the new sharing economy," *Academy of Management Discoveries*, vol. 6, no. 2, pp. 176–205, 2020.

[22] E. Foong, N. Vincent, B. Hecht, and E. Gerber, "Women (still) ask for less: Gender differences in hourly rate in an online labor marketplace," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 53:1–53:21, 2018.

[23] F. Ebrahimi, M. Tushev, and A. Mahmoud, "Mobile app privacy in software engineering research: A systematic mapping study," *American Economic Journal: Applied Economics*, vol. 133, no. 2, p. 106466, 2021.

[24] L. Yu, X. Luo, C. Qian, S. Wang, and H. Leung, "Enhancing the description-to-behavior fidelity in Android apps with privacy policy," *IEEE Transactions on Software Engineering*, vol. 44, no. 9, pp. 834–854, 2018.

[25] A. Aydin, D. Piorkowski, O. Tripp, P. Ferrara, and M. Pistoia, "Visual configuration of mobile privacy policies," in *Inter. Conf. on Fundamental Approaches to Software Engineering*, 2017, pp. 338–355.

[26] J. Young, "Commitment analysis to operationalize software requirements from privacy policies," *Requirements Engineering*, vol. 16, no. 1, pp. 33–46, 2011.

[27] G. Berardi, A. Esuli, T. Fagni, and F. Sebastiani, "Multi-store metadata-based supervised mobile app classification," in *Annual ACM Symposium on Applied Computing*, 2015, pp. 585–588.

[28] A. Al-Subaihin, F. Sarro, S. Black, L. Capra, M. Harman, Y. Jia, and Y. Zhang, "Clustering mobile apps based on mined textual features," in *International Symposium on Empirical Software Engineering and Measurement*, 2016, pp. 1–38.

[29] J. Corbin and A. Strauss, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE, 1998.

[30] M. Adams, W. Blumenfeld, C. Castaneda, and H. Hackman, *Readings for diversity and social justice*. Routledge, 2013.

[31] R. Ryker, E. Lafleur, B. McManis, and K. C. Cox, "Online privacy policies: An assessment of the fortune e-50," *Journal of Computer Information Systems*, vol. 42, no. 4, pp. 15–20, 2002.

[32] T. Dehling, F. Gao, and A. Sunyaev, "Assessment instrument for privacy policy content: Design and evaluation of PPC," in *the Pre-ICIS Workshop on Information Security and Privacy*, 2014.

[33] D. A. Bradbard, C. Peters, and Y. Caneva, "Web accessibility policies at land-grant universities," *The Internet and Higher Education*, vol. 13, no. 4, pp. 258–266, 2010.

[34] P. Savla and L. Martino, "Content analysis of privacy policies for health social networks," in *IEEE International Symposium on Policies for Distributed Systems and Networks*, 2012, pp. 94–101.

[35] S. Rains and L. Bosch, "Privacy and health in the information age: A content analysis of health web site privacy policy statements," *Health Communication*, vol. 24, no. 5, pp. 435–446, 2009.

[36] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," in *the SIGCHI Conference on Human Factors in Computing Systems*, 2004, p. 471–478.

[37] B. Miller, K. Buck, and J. Tygar, "Systematic analysis and evaluation of web privacy policies and implementations," in *International Conference for Internet Technology and Secure Transactions*, 2012, pp. 534–540.

[38] J. Zimmerle and A. Wall, "What's in a policy? evaluating the privacy policies of children's apps and websites," *Computers in the Schools*, vol. 36, no. 1, pp. 38–47, 2019.

[39] A. McDonald, R. Reeder, P. Kelley, and L. Cranor, "A comparative study of online privacy policies and formats," in *International Symposium on Privacy Enhancing Technologies Symposium*, 2009, pp. 37–55.

[40] A. Powell, P. Singh, and J. Torous, "The complexity of mental health app privacy policies: A potential barrier to privacy," *JMIR mHealth and uHealth*, vol. 6, no. 7, 2018.

[41] R. Flesch and A. Gould, *The art of readable writing*, 1949, vol. 8.

[42] G. Williams, M. Tushev, F. Ebrahimi, and A. Mahmoud, "Modeling user concerns in sharing economy: the case of food delivery apps," *Automated Software Engineering*, vol. 27, pp. 229–263, 2020.

[43] T. R. Dillahunt, X. Wang, E. Wheeler, H. F. Cheng, B. Hecht, and H. Zhu, "The sharing economy in computing: A systematic literature review," *Proceedings of the ACM Human Computer Interaction*, vol. 1, p. 26, 2017.

[44] E. Ert and A. Fleischer, "The evolution of trust in Airbnb: A case of home rental," *Annals of Tourism Research*, vol. 75, pp. 279–287, 2019.

[45] M. Tushev, F. Ebrahimi, and A. Mahmoud, "Digital discrimination in sharing economy a requirements engineering perspective," in *IEEE International Requirements Engineering Conference*, 2020, pp. 204–214.

[46] N. Goel, M. Rutagarama, and B. Faltings, "Tackling peer-to-peer discrimination in the sharing economy," in *ACM Conference on Web Science*, 2020, pp. 355–361.

[47] J. Tjaden, C. Schwemmer, and M. Khadjavi, "Ride with me. Ethnic discrimination, social markets, and the sharing economy," *European Sociological Review*, vol. 34, no. 4, pp. 418–432, 2018.