Privacy for All: Ensuring Fair and Equitable Privacy Protections

Michael D. Ekstrand* MichaelEkstrand@boisestate.edu
Rezvan Joshaghani RezvanJoshaghani@u.boisestate.edu
Hoda Mehrpouyan HodaMehrpouyan@boisestate.edu
Department of Computer Science, College of Engineering, Boise State University, Boise, ID

Abstract

In this position paper, we argue for applying recent research on ensuring sociotechnical systems are fair and nondiscriminatory to the privacy protections those systems may provide. Privacy literature seldom considers whether a proposed privacy scheme protects all persons uniformly, irrespective of membership in protected classes or particular risk in the face of privacy failure. Just as algorithmic decision-making systems may have discriminatory outcomes even without explicit or deliberate discrimination, so also privacy regimes may disproportionately fail to protect vulnerable members of their target population, resulting in disparate impact with respect to the effectiveness of privacy protections.

We propose a research agenda that will illuminate this issue, along with related issues in the intersection of fairness and privacy, and present case studies that show how the outcomes of this research may change existing thinking and research on privacy and fairness. We believe it is important to ensure that technologies and policies intended to protect the users and subjects of information systems provide such protection in an equitable fashion.

Keywords: Fairness, Privacy,

1. Introduction

Distinct ethical, legal, and social effects of technology do not exist in isolation, but often interact in complex ways. Understanding these interactions is crucial; we argue that research on fairness in information systems and privacy are both at the point where we can and must consider their interaction. We are by no means the first to consider these concerns and their possible relationship; Dwork and Mulligan (2013) argue for a shifting the focus of legal analysis of information systems from privacy and transparency to other social factors including fairness. We argue here for expanding the lens to include both concepts together.

Privacy has a long history of study in computer science, ethics, and law, and there are various technical and non-technical mechanisms for protecting it under its various definitions. Contemporary analyses of fairness do not have as long a history, though they are grounded in more than fifty years of legal work on fairness and nondiscrimination, with precursors reaching further back in scholarly discourse.

We seek to understand how fairness and privacy interact and complement or compete with each other. We identify three high-level questions of interest in understanding this interaction:

1. Are technical or non-technical privacy protection schemes fair, under contemporary definitions of fairness?

2. When and how do privacy protection technologies or policies improve or impede the fairness of the systems they affect?

3. When and how do technologies or policies aimed at improving fairness enhance or reduce the privacy protections of the people involved?

We expect the answers to these questions to vary based on domain, technology, and the specific definitions of privacy and fairness under consideration. Further, we qualify our questions as

* People and Information Research Team (PIReT)

regarding the 'contemporary' definition of fairness, because (as we discuss in the next section) much privacy technology and policy is connected to concepts of fairness, such as fair information practices, that are useful and important but distinct from fairness as it relates to equitable treatment across classes of people.

In this position paper, we build our argument by first rehearsing the concept space of privacy and fairness and connect to key literature; we then describe some of the ways we see privacy and fairness interacting with examples. We then lay out a research agenda for understanding privacy and fairness, primarily focused around question (1) with an eye towards ensuring future privacy-protecting systems are fair, and provide several examples of how the community's understanding and application of privacy and fairness research may change in response to our research agenda.

This work builds on the goals set out by Dwork et al. (2014) in ensuring privacy is not limited to most people, but is extended so far as we can guarantee to all subjects of an information system. We expand this issue into a broad agenda at the intersection of privacy and fairness that considers the entire sociotechnical system in which a technical, legal, or social privacy mechanism is deployed and situating it in the current language of algorithmic fairness. While privacy and non-discrimination have been treated together at length (Custers, 2013), they are often covered as related but distinct concerns; we see a pressing need for research on their intersection, building on the goals of Dwork et al. (2014) and the joint pursuit of privacy and fairness by Hajian et al. (2015).

We invite discussion and collaboration on these topics, in order to make computing technology in practice better for all people it affects. Our present treatment focuses primarily on the U.S. context, drawing from U.S. legal doctrines and policy approaches; translating and reevaluating the concerns we raise in other legal and cultural contexts is important future work.

Our philosophical approach to these topics is grounded in the work of Franklin (1999), particularly in our interest in understanding who pays for and who benefits from any particular technology or policy, and in promoting technology that is equitable and participatory.

2. Concept Background

Our argument is specifically about the interaction of fairness and privacy. Such an argument necessarily builds on the foundation of prior literature in each of these two strands, and we rehearse that literature and overview existing definitions in this section. While some of this may be familiar to many of our readers, we wish this paper to be accessible to read by scholars well-versed in privacy or fairness and desiring to learn more about the other. We refer the reader to Custers (2013) for a more thorough treatment of the relevant background.

2.1. Privacy

The meaning of privacy has changed significantly over time. More than a century ago, Warren and Brandeis (1890) argued for a broad legal right to privacy grounded in long-standing U.S. and British common law and statute and responding to the challenges posed by new technologies such as the photograph. They explain that privacy is a right and define it as "the right to be let alone.", This stance identifies privacy with **seclusion**. Under the seclusion definition, perfect privacy is achieved through complete solitude,e.g. living alone on a deserted island (at least prior to the invention of spy satellites).

Another definition of privacy regards it as being free from intrusion or interference; we can call this the **non-intrusion** view of privacy. An example can be seen in Brennan's description of privacy when he categorizes the 4th amendment to the U.S. constitution as enshrining the "right of the individual . . . to be free from unwarranted government intrusion" (Greene, 2009).

As privacy concerns have moved from classical legal settings involving physical spaces and intrusion, new concepts and definitions of privacy have been needed. Tavani (2007) surveys this pivot, categorizing various privacy definitions and arguing that new concepts are needed to support meaningful privacy in modern information spaces. He identified limitation and control theory as key concepts for reasoning about information privacy.

The **limitation theory** of privacy defines privacy as individuals keeping information to themselves. In this theory privacy is defined as limited and contextually bounded information ac-

cess. Perfect privacy under limitation theory occurs when no information exists about someone (Tavani, 2007). Gavison (1980) applies this definition of privacy in the legal domain. Parent (1983) gives a variation of this theory by describing privacy as the "condition of not having undocumented personal knowledge about one possessed by others."

The **control theory** promotes privacy by enabling users to exert control over private information. In control theory, a person has privacy if they have control over their information (Tavani, 2007). Westin (1968) applies this theory when he defines privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others". Legal philosopher Charles Fried argues for control theory as superior to limitation theory, stating "Privacy is not simply an absence of information about us in the minds of others, rather it is the control we have over information about ourselves" (Nissenbaum, 2009).

Moor (1997); Tavani and Moor (2001) integrated the control and limitation theories of privacy when proposing **Restricted Access/Limited Control** (RALC). Under RACL a person has privacy "in a situation with regard to others [if] in that situation the individual ... is protected from intrusion, interference, and information access by others."

Nissenbaum (2004) proposed the **contextual integrity** theory of privacy . In this theory the privacy is defined with regard to the norms of the context that an individuals is in. She state that in each information transition context there different variables that define privacy and the norms in that specific context determine the privacy violation.

Many different enabling technologies and legal constructs exist to support or measure these various theories of privacy. On the technical side, these include differential privacy (Dwork et al., 2014), k-anonymity (Aggarwal, 2005; Tang et al., 2017), and cryptography; legal frameworks include purpose- specification (on Automated Personal Data Systems, 1973; for Economic Cooperation and Development, 2002) or purpose-binding (Hildebrandt, 2014) and legal rights to inspect and correct personal data as embodied in the EU General Data Protection Regulation and US Fair Credit Reporting Act.

2.2. Fairness

Privacy is often linked to the ethical language of fairness, particularly in the U.S. regulatory context; since 1973, fair information practices have been the guiding paradigm for managing privacy and considering regulations around data protection and notice (on Automated Personal Data Systems, 1973). The notion of fairness embodied in these principles, seeking to treat people justly on an individual basis with regards to the use of information regarding them, is but one point in a broader space of fairness concepts, and is distinct from the term 'fair' as it tends to be used in the algorithmic fairness literature.

Algorithmic fairness, as embodied in the literature surrounding Fairness, Accountability, and Transparency, discrimination-aware data mining (Pedreshi et al., 2008), and related research threads, tends to focus on fairness as non-discrimination: a person's experience with an information system should not irrelevantly depend on their personal characteristics, especially their membership in groups subject to historical discrimination. In the U.S. legal setting, this is often operationalized by considering legally-recognized protected characteristics such as race, gender, sexual orientation, ethnicity, religion, and age.

There are at least two meaningful dimensions on which we can organize approaches to ensuring and evaluating fairness: for whom fairness being considered, and how fairness is operationalized so it can be measured or assured.

The for whom question is often assumed; in credit-scoring, borrowers should be treated fairly (Pedreshi et al., 2008), and the problem is often treated assuming a single lender, or otherwise without regard for fairness among lenders. Many problems do indeed have a natural domain of users for whom fairness should be ensured, particularly when understood in the historical context of discrimination. Some problem settings, however, are less clear: in evaluating the fairness of recommender systems, do we want to ensure fairness for consumers, content providers, or some other group? Burke (2017) has explored this probing how even the case of lending requires

a multi-sided view of fairness in the age of micro-lending platforms such as Kiva.

How fairness operationalized varies widely, but falls primarily into a few categories:

- **Individual fairness** (Dwork et al., 2012) says that similar individuals should receive similar treatment.
- **Group fairness** (Feldman et al., 2015; Kamiran and Calders, 2009) says that different groups of users should receive similar statistical treatment, by experiencing similar classification accuracy or error.

There are many different specific metrics within both of these techniques, as well as impossibility results showing that group and individual fairness often cannot be achieved simultaneously (Friedler et al., 2016) and that different group fairness measures are not simultaneously achievable in many realistic settings (Chouldechova, 2017; Kleinberg et al., 2016).

Given an operationalization of fairness, there are techniques to audit algorithmic systems for fairness (Feldman et al., 2015), and to adapt them to produce (more) fair outcomes. These adaptations can occur at all stages of the machine learning pipeline, including data pre-processing (Kamiran and Calders, 2009), representation learning (Zemel et al., 2013), fairness constraints and regularizers in the learning process (Dwork et al., 2012; Pedreshi et al., 2008), fairness-aware online optimization (Liu et al., 2017), and post-processing the learned model (Kamiran et al., 2010).

The *disparate impact* doctrine adopted by the U.S. legal system is commonly used as the ideological starting point for group fairness measures (Feldman et al., 2015). This doctrine says that a practice is discriminatory if it has disproportionate adverse effects on protected groups without a compelling business need. For example, if a screening in a hiring process rejects black candidates at a substantially higher rate than white candidates, then the employer must show that the screening serves a substantial business need and is the least impactful way of achieving that need in order for the screening not to be ruled as unlawful discrimination.

A related concept that has captured some attention more recently is *disparate mistreatment* (Zafar et al., 2017): rather than focusing on decision likelihoods, this model examines decision errors, particularly errors that harm the protected class. In lending, for example, disparate impact asks if minority class loan applicants have a similar success rate in obtaining loans as majority-class applicants; disparate mistreatment asks if minority class applicants are more likely to have their ability to pay underestimated and thereby be denied a loan that they would be able to pay.

Disparate impact and disparate mistreatment are the background concerns for most of the specific concerns we raise regarding the fairness of privacy systems.

3. The Interaction of Fairness and Privacy

We now turn to consider how fairness and and privacy relate to each other, and ways in which each can promote or hinder the other.

3.1. From Differential Privacy to Fairness

Differential privacy was introduced as a privacy preserving method on statistical databases. The goal of differential privacy is to maintain the privacy of individuals' information in the database while enabling data analysts to query that database to study the population. Differential privacy ensures its guarantees by applying noise (a randomized mechanism) to the database so that the participation or absence of any one individual will not perceptibly affect the result of the study on the population (Dwork et al., 2014).

Dwork et al. (2012) later adapted the mathematical machinery of differential privacy to provide certain fairness properties, considering (individual) fairness to be a generalization of differential privacy.

3.2. Fairness-Privacy Tradeoffs

Consumers often need to trade information about themselves for goods and services. For example, online personalization can improve the relevance of product recommendations and advertising, thereby reducing the number of irrelevant ads a customer sees, but requires data on customer behavior and preferences Chellappa and Sin (2005); Tucker (2012). Under a limitation

theory of privacy, these users are trading privacy for personalization; under a control theory, they may be exercising their right to privacy by choosing to participate in the exchange, but only to the extent that they have sufficient notice and knowledge to make an informed decision. Under contextual integrity, the user may consent to the advertiser having their information but not to its use in targeting advertising.

Fairness may also trade off with privacy. Fairness- through - blindness attempting to produce fair algorithmic results by ignoring protected information - does not work, because the protected attributes may be correlated with other attributes in the data set (Dwork et al., 2012). Therefore, in order to audit the fairness of an algorithmic system, it may be necessary to collect data about users that is not required for the basic system to function. If a service provider does not collect subjects' demographic information, then it may not be able to audit its services for disparate impact or modify its statistical models to remove discriminatory effects. The EU recognizes this possibility in the General Data Protection Regulation (WIJNANT, 2016), where auditing an information system for bias and discrimination is a potentially-necessary step towards meeting a data processor's legal obligations (Goodman, 2016).

The alignment of fairness benefits and privacy costs may also be skewed. The move to FICO credit scores as the basis for credit decisions was predicated on the existence of credit bureaus assembling dossiers on past and prospective borrowers. Borrowers who previously had no difficulty obtaining credit likely saw little benefit, and possibly adverse effects, from this change in regime, as more details of their financial lives are tracked and shared with limited opportunity for consent. However, borrowers who were historically denied credit to due to race, gender, religion, or other concerns were able to obtain credit more easily when only their financial situation and history is allowed to be considered, and there is reason to believe credit scoring has likely helped in this (Board of Governors of the Federal Reserve System, 2007); if so, they see significant benefit in exchange for the information tracking and sharing required to make credit scoring effective. While by no means perfect, the present situation is decidedly better than what came before (Ritter, 2012). It is our belief that, in the interest of establishing a more equitable society, this particular disproportionate trade is worthwhile, but the data sharing and privacy implications of such systems should be studied and carefully considered.

In hiring, many U.S. organizations achieve a privacy/fairness trade-off in auditing their hiring practices for discrimination by having the human resources office collect applicants' demographic information, but not making that available to the people involved in screening candidates and making the hiring recommendation.

3.3. Jointly Obtaining Fairness and Privacy

Privacy and fairness have been addressed separately for many years, however, recent studies (Hajian et al., 2016; Hintoglu et al., 2005; Kashid et al., 2015, 2017; Pedreshi et al., 2008; Ruggieri et al., 2014) have expanded the application of methods to achieve both goals. Hajian et al. (2015) use pattern sanitization methods including k-anonymity and differential privacy to simultaneously achieve privacy and fairness. Addition work (Luong et al., 2011; Pedreshi et al., 2008; Ruggieri et al., 2010) proposes employing discrimination-aware data mining techniques in a privacy-aware fashion. There remains much work to be done in characterizing under what circumstances and definitions privacy and fairness are simultaneously achievable, and when they compete such that a joint approach must solve a multi-criteria optimization problem and trade off privacy or fairness for the other.

4. An Agenda for Equitable Privacy

Privacy and fairness, at their core, have a similar high-level goal: to protect people from adverse effects of social, legal, and technical systems. As we have demonstrated, though, these concerns can sometimes be at odds. We see the need for research to reconcile their conflicts, map out the trade-offs involved, and develop methodologies for auditing privacy systems for fairness so that developers, regulators, and the public can make informed decisions.

There is some overlap in the questions we state; we endeavor to think about this set of problems from multiple perspectives, and in some cases formulate different versions of our research questions to reflect different angles or the immediate concerns of different stakeholders. We believe this thoroughness is useful for promoting a vigorous multi-stakeholder and multi-perspective discussion of fairness and privacy.

4.1. Fair Privacy

If we want to assess the fairness of a privacy protection system, be it technical or legal, there are several important questions. We begin with questions that characterize the fairness of a particular privacy system.

Q1: Does the system provide comparable privacy protections to different groups of subjects? This first question is perhaps the most crucial of our questions: does the system protect all its users, or do some users obtain better protections than others? Do differences in protection capabilities result in members of protected classes being less protected than other subjects? There are at least two aspects to disparate protection that need to be studied: the likelihood of a privacy failure and the cost of a privacy failure. It may be that users are equally likely to be protected, but protection failures are more costly to some users than to others; we expect that analyzing these factors separately, as well combined into a risk score, will be useful in characterizing the fairness of privacy systems.

Definition: Fair Privacy Protection. We propose that a privacy system be deemed to provide *fair protection* if the probability of failure and expected risk are statistically independent of the subject's membership in a protected class.

Q2: Are privacy attacks more effective against members of protected classes? This is the dual of Q1, focusing on attack capabilities instead of protection. In addition to considering the protections of a privacy-protection scheme, we also think it relevant to examine the disparate effectiveness of privacy attacks.

Q3: Does the system require disparate effort from its subjects in order to enjoy privacy protection? It may be that a system provides fair protection, but some users may need to exert more effort in order to realize its privacy policies. Effort may be defined in a number of different ways; one way to conceptualize it is to think about the extent and impact of the changes a person must make to their behavior or activities in order to enjoy the system's privacy guarantees. If a privacy (or surveillance) regime provides privacy guarantees that require minimal changes to the lifestyle of members of a dominant group, while requiring invasive changes for vulnerable groups, then we can say that the system requires disparate effort and may be unfair.

Under control theory, we can say that a privacy regime requires disparate effort if vulnerable groups pay a higher cost in other areas of their lives for exercising the control the privacy system affords.

Q4: Is the fairness of privacy guarantees robust to shifts in threat model? Good privacy work provides meaningful privacy guarantees under a particular threat model. The guarantees may or may not be robust to various adjustments to the threat model, such as adjusting the attacker's assumed capabilities or the cost of failure. It may also be possible for the fairness of the privacy guarantees to change with such shifts, even if the overall guarantees do not, and this should be studied.

These questions so far apply to individual privacy protection mechanisms. However, in order to produce generalizable knowledge, it is necessary to go beyond individual systems to understand the properties that cause privacy systems to be fair or unfair.

Q5: What properties of a problem setting or privacy mechanism make fair privacy easier or harder to achieve? This question will help us develop guidelines to more effectively predict the fairness of a privacy system in advance, and develop privacy systems that are more likely to provide fair protections.

Q6: Are there identifiable properties of a privacy mechanism and problem setting that form necessary or sufficient conditions for fair privacy? In this question we are concerned with adapting Q4 from a nuanced, statistical perspective to a litmus test: are there

bright-line properties that can clearly classify a system as fair or not-fair?

Q7: What other properties may need to be sacrificed to achieve fair privacy? We can envision trade-offs such as decreasing the fairness of the distribution of benefit or accuracy of an algorithmic decision in order to ensure fair distribution of its privacy protections. Whether such trade-offs exist in realistic settings, and how to navigate them, should be explored.

4.2. Impact of Privacy on Fairness

Q8: Does a privacy-protection scheme impede the ability to ensure or audit the fairness of decision-making processes or information systems? The simple version of this problem is that we cannot measure whether a human or algorithmic decision-making process exhibits indirect sexual discrimination without knowing the gender identities (and sexual orientations, for testing a broader definition of sexual discrimination) of at least a test set of subjects.

Q9: Can privacy protection technologies or policies be used or adapted to enhance the fairness of a system? One example of this is adapting the mathematical machinery of differential privacy to yield Lipschitz-based fairness (Dwork et al., 2012). Are there other domains and applications where privacy can enhance fairness?

4.3. Impact of Fairness on Privacy

These questions are effectively the duals of the questions in the previous section, but we find them worth stating independently.

Q10: Does a fairness auditing or enhancement scheme diminish the privacy of its subjects? This is the dual of Q5; when implementing fairness-aware decision making systems, designers and researchers should consider the privacy implications of additional data collection required by the fairness scheme.

Q11: Can fairness-enhancing technologies be used to provide privacy guarantees? Some algorithmic fairness formulations are generalizations of differential privacy (Dwork et al., 2012), and the mathematics of differential privacy are the foundations of individual fairness-through-awareness. Data preprocessing to ensure fairness may also provide probabilistic fairness properties. It remains to be seen whether there are additional fairness techniques that can either be adapted to provide meaningful privacy properties or that provide privacy protections as a special case.

4.4. Defining the Subject of Research

In answering any of these research questions, it is crucial to carefully define the kinds of privacy and fairness under consideration. We also see a need to map out what kinds of fairness and privacy may intrinsically support each other or contradict each other, much like the existing impossibility results for fairness definitions.

5. Case Studies

To demonstrate the importance of the questions we raise, we now discuss how their answers would impact the state of the art in various applications, techniques, and problem areas.

The state of current research limits our ability to connect all of our questions to current case studies. Throughout our discussion, we reference specific questions that have particular bearing on the case; Q5–Q7 influence all cases by influencing the shape of potential solutions to the fairness and privacy problems.

5.1. Differential Privacy

Differential privacy (Dwork et al., 2014) provides a strong guarantee of privacy by applying incorporating random noise calibrated to nullify the impact of the presence or absence of any one on the final result. If a data access mechanism is (ϵ, δ)-differentially private, then it generally bounds the distinguishability of two databases, one containing an individual's record and the other not, by ϵ, with a δ (usually cryptographically small) probability of total privacy failure. Dwork et al. (2014) refer to ϵ as the "knowledge gain ratio from one dataset over the other." Hence, the higher the value of ϵ, the weaker the privacy guarantee. The value chosen for ϵ by different models in the research are 0.01 or 0.1 and

in some cases $ln2$ or $ln3$ (Bhaskar et al., 2010; Bonomi and Xiong, 2013; Friedman and Schuster, 2010; Li et al., 2012; Zeng et al., 2012). Apple's production deployment of differential privacy uses an ϵ of 1 or 2 (Tang et al., 2017). Choosing an appropriate value for ϵ is non-trivial and does not correlate with any privacy standards (Lee and Clifton, 2011).

Differential privacy, on its face, provides fair privacy, as all users' privacy loss is bounded by ϵ and δ. Indeed, this is explicitly articulated by Dwork et al. (2014) as a motivation for differential privacy over other privacy protection schemes that ensure privacy for all but "just a few" of their subjects, addressing our Q1. However, there may remain subtle ways in which differential privacy may fail to provide fair privacy. If omitting a protected class of users from the database admits a lower bound on the privacy loss of a differentially private mechanism with equivalent accuracy, then the system may be unfair (Q1). Further, differentially private mechanisms are but one building block in a privacy-preserving system, and breakdowns in other components of the system or details of its implementation may have disparate impact (Q2 and Q4), much in the way that practical cryptosystems are often broken by attacking the implementation instead of the core primitives.

5.2. Deanonymization Risks

Deanonymization attacks unmask users in ostensibly anonymized data sets. Depending on the nature of the data set, this can have significant impact on the unmasked individuals' lives, such as revealing search logs. The de-anonymizer usually uses auxiliary data that relates to identifying attributes of the database. This extra data can be from another database, a certain behavioral pattern of people, or some background information on the target person. Usually, a combination of these methods is used to identify a person in the dataset.

One famous example of public data deanonymization regards the Netflix data set, In 2006 to 2009 Netflix ran a contest to improve their recommender system (Bennett et al., 2007). They published anonymized records of movie ratings of 500K DVD-by-mail subscribers. Narayanan and Shmatikov (2008) correlated the Netflix data with user activities on IMDB and showed that they could identify 99% of subscribers that reviewed at least eight movies on both websites. Frankowski et al. (2006) demonstrated a similar attack deanonymizing the MovieLens data set using user discussions of movies in online forums.

Wondracek et al. (2010) propose a deanonymization attack based the group membership information on social networks. They showed that by only processing the public groups in social networks an attacker could achieve enough information on a user to identify them on a third party website. Lane et al. (2012) identified users based on shared mobile sensor data. They argue that based on mobile sensor data and societal norms, user's habits, transportation preferences, physical activities and other everyday user activities can be identified. This information can be used as auxiliary information for unmasking users that use medical or exercise-related applications.

Worryingly, in some cases, deanonymization attacks are easier to carry out against members of particular groups. In 2014, Chris Whong used the Freedom of Information Act to obtain a data set of historical trip and fare logs of New York taxis and developed a visualization tool showing taxi information over a 24-hour period. Another study found that the most common name for New York taxi drivers is "Mohammad"; New York City also publishes a data set of drivers names and their license numbers. Combining these disparate information sources, Deneau could identify four drivers that have low activity during Muslim prayer hours (Miracle, 2016). Under Q2, it appears that deanonymization attacks may be disparately successful against minority groups in some cases, and when and how this occurs should be studied carefully.

Some of these deanonymization attacks may have been preventable by complete and correct implementation of anonymization methods. However, as the dimensionality of a data set increases, the effectiveness of anonymization techniques diminishes (Aggarwal, 2005; Miracle, 2016).

In some fields such as precision medicine, too much anonymization can compromise data quality, and the anonymized dataset is useless for the study (Sweeney, 2002). It is also likely difficult to

auditing fairness in such a setting (Q8). Further, if the anonymization particularly decreases the effectiveness of treatment for minority groups, it could cause a Q3 fairness problem by imposing higher costs (reduced medical effectiveness) for insisting on privacy in the data set.

5.3. Recommender-Assisted Outing

Recommender systems (Adomavicius and Tuzhilin, 2005; Ekstrand et al., 2011) are algorithmic tools that help users locate products to purchase, people to follow on social media, news to read, and many other things, often in a personalized fashion and without requiring a search query. They are a ubiquitous part of the modern Internet experience, but the vast personal information they utilize can make them a ripe target for privacy-invasive exploitation, and Q2 is relevant for understanding the impact of such attacks.

Calandrino et al. (2011) demonstrate privacy attacks that allow an attacker with some information about a person's transaction history to observe the public outputs of a live recommender system and infer other transactions made by the target individual. This is similar in spirit to a deanonymization attack against a recommender system data set, but is feasible on a live system only observing the recommender's output.

Recommender systems can also give away users' identity information. Instagram's who-to-follow recommender was a key link in the chain by which a journalist uncovered FBI director James Comey's secret Twitter account (Feinberg, 2017).

Examining these failures through the lens of fair privacy, we can ask whether members of protected classes are at greater risk of a recommender system disclosing their identity or other information than less vulnerable users (Q1). We can also examine privacy protection schemes for recommender systems in this light, examining whether provide comparable protection to all users.

5.4. Genetic Privacy

As private and public organizations are increasingly collecting large quantities of genetic data, the privacy of such data is a significant concern. Shi and Wu (2017) provide an overview of genetic privacy concerns and techniques for addressing them, including both technical and regulatory approaches.

The interaction of genetic privacy, or the privacy of other health data, with fairness is subtle. Members of historically-vulnerable groups may be at greater risk of genetic privacy breach (Q1), and privacy schemes limiting the collection and use of genetic data may prevent some forms of discrimination in health care, employment, and other domains (Q9); scholars have been concerned about such discrimination as long as genetic testing was readily available (Billings et al., 1992; Gostin, 1991).

It may be possible that genetic information may reduce group-based discrimination in a manner analogous to credit scores by allowing decision-makers to directly observe genetic markers for particular health risks when they might otherwise make group-based assumptions about risk, or provide a richer basis for the similarity functions necessary to achieve individual fairness, but this paradigm seems ripe for abuse and likely not a significant step forward. While it is worth studying the possibilities, we expect that the costs will outweigh the benefits.

6. Conclusion

For privacy protection mechanisms to advance a just and equitable society, it is necessary that they (1) provide their protections equitably to all their subjects and (2) that they integrate positively with other important concerns such as fairness and non-discrimination in the information systems deployed in their sociotechnical setting. This applies regardless of whether the mechanism in question is technical, legal, social, or implemented by some other means.

We have put forward a set of research questions designed to assess the fairness of particular privacy protection mechanisms and their interaction with the prerequisites for auditing or ensuring the fairness of decision support tools and other algorithmic components of the sociotechnical ecosystem in which they are deployed. We also hope to see extensive work mapping out the interplay between privacy and fairness more broadly, so that we can have a solid, generalizable understanding of when they enhance each other and other desirable properties of sociotechnical systems and when they are in competition.

Important work has already been done on this topic, with differential privacy providing a mathematical framework to make privacy guarantees independent of user characteristics and additional approaches that jointly achieve privacy and fairness in some settings.

But there remains much to be done, particularly in understanding the implications of privacy and fairness on each other in practical settings that require meaningful, user-interpretable privacy and fairness properties, and on understanding the ways in which implementation details, human factors, and legal concerns may hinder one or the other of privacy and fairness when they are both pursued.

We argue here for expanding the lens of legal, policy, and social analysis of information systems instead of shifting it (Dwork and Mulligan, 2013), examining how fairness relates to privacy and transparency, and which social goals can be best promoted by each sociotechnical concept. We see long-term room for a mapping out of the different sociotechnical mechanisms for promoting a just and equitable society, much like the work of Schneier (2012) on the relative effectiveness of different mechanisms for limiting the scope and impact of social defection.

We hope to engage in some of this work ourselves over the coming years, and welcome vigorous debate and collaboration on how best to achieve a just, equitable society that is respectful of its members autonomy in the use and disclosure of their personal information.

Acknowledgments

The authors would like to thank National Science Foundation for its support through the Computer and Information Science and Engineering (CISE) program and Research Initiation Initiative(CRII) grant number 1657774 of the Secure and Trustworthy Cyberspace (SaTC) program: A System for Privacy Management in Ubiquitous Environments.

References

Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

Charu C Aggarwal. On k-anonymity and the curse of dimensionality. In *Proceedings of the 31st international conference on Very large data bases*, pages 901–909. VLDB Endowment, 2005.

James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, NY, USA, 2007.

Raghav Bhaskar, Srivatsan Laxman, Adam Smith, and Abhradeep Thakurta. Discovering frequent patterns in sensitive data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 503–512. ACM, 2010.

Paul R Billings, Mel A Kohn, Margaret De Cuevas, Jonathan Beckwith, Joseph S Alper, and Marvin R Natowicz. Discrimination as a consequence of genetic testing. *American journal of human genetics*, 50(3):476, 1992.

Board of Governors of the Federal Reserve System. Report to the congress on credit scoring and its effects on the availability and affordability of credit. Technical report, U.S. Federal Reserve, August 2007.

Luca Bonomi and Li Xiong. Mining frequent patterns with differential privacy. *Proceedings of the VLDB Endowment*, 6(12):1422–1427, 2013.

Robin Burke. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017.

Joseph A Calandrino, Ann Kilzer, Arvind Narayanan, Edward W Felten, and Vitaly Shmatikov. " you might also like:" privacy risks of collaborative filtering. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 231–246. IEEE, 2011.

Ramnath K Chellappa and Raymond G Sin. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information technology and management*, 6(2): 181–202, 2005.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv preprint arXiv:1703.00056*, 2017.

Bart Custers. Data dilemmas in the information society: Introduction and overview. *Discrimination and Privacy in the Information Society*, pages 3–26, 2013.

Cynthia Dwork and Deirdre K Mulligan. It's not privacy, and it's not fair. *Stan. L. Rev. Online*, 66:35, 2013.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226. ACM, 2012.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Michael D Ekstrand, John T Riedl, Joseph A Konstan, et al. Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction*, 4(2):81–173, 2011.

Ashley Feinberg. This is almost certainly james comey's twitter account, 2017. https://gizmodo.com/this-is-almost-certainly-james-comey-s-twitter-account-1793843641 Accessed: 2017-10-5.

Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268. ACM, 2015.

Organisation for Economic Co-operation and Development. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD Publishing, 2002.

Ursula Franklin. *The real world of technology*. House of Anansi, 1999.

Dan Frankowski, Dan Cosley, Shilad Sen, Loren Terveen, and John Riedl. You are what you say: privacy risks of public mentions. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 565–572. ACM, 2006.

Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. On the (im) possibility of fairness. *arXiv preprint arXiv:1609.07236*, 2016.

Arik Friedman and Assaf Schuster. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–502. ACM, 2010.

Ruth Gavison. Privacy and the limits of law. *The Yale Law Journal*, 89(3):421–471, 1980.

Bryce W Goodman. A step towards accountable algorithms? algorithmic discrimination and the european union general data protection. In *29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona. NIPS Foundation*, 2016.

Larry Gostin. Genetic discrimination: the use of genetically based diagnostic and prognostic tests by employers and insurers. *Am. JL & Med.*, 17:109, 1991.

Jamal Greene. The so-called right to privacy. *UC Davis L. Rev.*, 43:715, 2009.

Sara Hajian, Josep Domingo-Ferrer, Anna Monreale, Dino Pedreschi, and Fosca Giannotti. Discrimination-and privacy-aware patterns. *Data Mining and Knowledge Discovery*, 29(6):1733–1782, 2015.

Sara Hajian, Francesco Bonchi, and Carlos Castillo. Algorithmic bias: from discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2125–2126. ACM, 2016.

Mireille Hildebrandt. Location data, purpose binding and contextual integrity: Whats the message? In *Protection of Information and the*

Right to Privacy-A New Equilibrium?, pages 31–62. Springer, 2014.

A A Hintoglu, A Inan, Y Saygin, and M Keskinoz. Suppressing data sets to prevent discovery of association rules. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 4 pp.–. ieeexplore.ieee.org, November 2005.

Faisal Kamiran and Toon Calders. Classifying without discriminating. In *Computer, Control and Communication, 2009. IC4 2009. 2nd International Conference on*, pages 1–6. IEEE, 2009.

Faisal Kamiran, Toon Calders, and Mykola Pechenizkiy. Discrimination aware decision tree learning. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 869–874. IEEE, 2010.

Asmita Kashid, Vrushali Kulkarni, and Ruhi Patankar. Discrimination prevention using privacy preserving techniques. *International Journal of Computer Applications*, 120(1), 2015.

Asmita Kashid, Vrushali Kulkarni, and Ruhi Patankar. Discrimination-aware data mining: a survey. *International Journal of Data Science*, 2(1):70–84, 2017.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

Nicholas D Lane, Junyuan Xie, Thomas Moscibroda, and Feng Zhao. On the feasibility of user de-anonymization from shared mobile sensor data. In *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, page 3. ACM, 2012.

Jaewoo Lee and Chris Clifton. How much is enough? choosing e for differential privacy. *Information Security*, 7001:325–340, 2011.

Ninghui Li, Wahbeh Qardaji, Dong Su, and Jianneng Cao. Privbasis: Frequent itemset mining with differential privacy. *Proceedings of the VLDB Endowment*, 5(11):1340–1351, 2012.

Yang Liu, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal, and David C Parkes. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*, 2017.

Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini. k-nn as an implementation of situation testing for discrimination discovery and prevention. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 502–510. ACM, 2011.

Jacob M Miracle. *De-Anonymization Attack Anatomy and Analysis of Ohio Nursing Workforce Data Anonymization*. PhD thesis, Wright State University, 2016.

James H Moor. Towards a theory of privacy in the information age. *ACM SIGCAS Computers and Society*, 27(3):27–32, 1997.

Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125. IEEE, 2008.

Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

Welfare. Secretary's Advisory Committee on Automated Personal Data Systems. *Records, Computers, and the Rights of Citizens: Report*. US Department of Health, Education & Welfare, 1973.

William A Parent. Privacy, morality, and the law. *Philosophy & Public Affairs*, pages 269–288, 1983.

Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. Discrimination-aware data mining. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 560–568. ACM, 2008.

Dubravka Ritter. Do we still need the equal credit opportunity act? 2012.

Salvatore Ruggieri, Dino Pedreschi, and Franco Turini. Data mining for discrimination discovery. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(2):9, 2010.

Salvatore Ruggieri, Sara Hajian, Faisal Kamiran, and Xiangliang Zhang. Anti-discrimination analysis using privacy attack strategies. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 694–710. Springer, 2014.

Bruce Schneier. *Liars and outliers: enabling the trust that society needs to thrive*. John Wiley & Sons, 2012.

Xinghua Shi and Xintao Wu. An overview of human genetic privacy. *Annals of the New York Academy of Sciences*, 1387(1):61–72, 2017.

Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. Privacy loss in apple's implementation of differential privacy on macos 10.12. *arXiv preprint arXiv:1709.02753*, 2017.

Herman T Tavani. Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy*, 38(1):1–22, 2007.

Herman T Tavani and James H Moor. Privacy protection, control of information, and privacy-enhancing technologies. *ACM SIGCAS Computers and Society*, 31(1):6–11, 2001.

Catherine E Tucker. The economics of advertising and privacy. *International journal of Industrial organization*, 30(3):326–329, 2012.

Samuel D Warren and Louis D Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.

Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

MICHELLE WIJNANT. General data protection regulation. 2016.

Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. A practical attack to de-anonymize social network users. In *Security and Privacy (SP), 2010 IEEE Symposium on*, pages 223–238. IEEE, 2010.

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1171–1180, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349130. doi: 10.1145/3038912.3052660. URL https://doi.org/10.1145/3038912.3052660.

Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. Learning fair representations. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 325–333, 2013.

Chen Zeng, Jeffrey F Naughton, and Jin-Yi Cai. On differentially private frequent itemset mining. *Proceedings of the VLDB Endowment*, 6 (1):25–36, 2012.