CHAPTER 3

What Does DP Mean?

3.1 LIMITATIONS OF SYNTACTIC NOTIONS

To understand why DP is appealing, it is also helpful to examine the alternatives. Before DP was introduced, researchers had been focusing on syntactic privacy notions. The most prominent of which is k-anonymity Samarati [2001], Sweeney [2002]. When applied to relational data, this notion requires the division of all attributes into *quasi-identifiers* and *sensitive attributes*, where the adversary is assumed to know the former, but not the latter.

Definition 3.1 **k-Anonymity.** A table satisfies k-anonymity relative to a set of quasi-identifier attributes if and only if when the table is projected to include only the quasi-identifier attributes, every record in the projected table appears at least k times.

The initial objective of k-anonymity was to prevent *re-identification*, i.e., an adversary who knows the quasi-identifier values of an individual should not be able to point to a record in the output and say "this is the record of the individual I know". In a dataset that satisfies k-anonymity, if there is any record matching an individual, there are at least k such records, making re-identification difficult. Researchers have observed that k-anonymity does not prevent *attribute disclosure*, i.e., information about sensitive attributes can still be learned, perhaps due to the uneven distribution of their values. This leads to privacy notions such as ℓ-diversity Machanavajjhala et al. [2006], t-closeness Li et al. [2007], and so on. All these notions, however, are syntactic, in the sense that they define a property about the final "anonymized" dataset, and do not consider the algorithm or mechanism via which the output is obtained. In contrast, DP is a property of the algorithm, instead of the output.

Any anonymization algorithm must be designed to optimize for some utility objective. Without considering utility, one can trivially achieve privacy protection by publishing nothing. Knowing that an algorithm would optimize for a certain utility objective enables one to infer additional information about the input when given the output, as shown, e.g., in the minimality attack Cormode et al. [2010], Wong et al. [2007].

Another illustration of the limitation of the syntactic nature of k-anonymity is given in Li et al. [2012a], which points out that one trivial way to satisfy k-anonymity is to simply duplicate each record k times, or similarly, to select a subset of the records and duplicating them. Furthermore, even though k-anonymity can be satisfied without providing real privacy protection, some k-anonymization algorithms can provide protection similar to

ϵ-DP Li et al. [2012a]. However, here the privacy protection property is clearly associated with the algorithm, and not with just the output.

3.2 SEMANTIC GUARANTEES OF DIFFERENTIAL PRIVACY

To assess whether DP offers sufficient protection of privacy, we have to examine social and legal conceptions of privacy. Privacy as a social and legal concept is multi-faceted. Solove [2010] identified 6 conceptions of privacy: (1) the right to be let alone Warren and Brandeis [1890]; (2) limited access to the self; (3) secrecy—the concealment of certain matters from others; (4) control over personal information; (5) personhood—the protection of one's personality, individuality, and dignity; (6) intimacy—control over, or limited access to, one's intimate relationships or aspects of life. These conceptions overlap with each other; and some are not applicable in the context of data privacy.

Among these, we distill two related, yet different, conceptions that are most relevant to data privacy: "**privacy as secrecy**" and "**privacy as control** (over personal information)". The former was stated as "right to conceal discreditable facts about himself" Posner [1998], and the latter was stated by Westin [1967] as: "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others". These two conceptions can be linked to two different approaches of defining privacy mathematically, which we explore in this section. The former leads to a "prior-to-posterior" approach, and the latter leads to a "posterior-to-posterior" approach.

3.2.1 INFEASIBILITY OF ACHIEVING "PRIVACY AS SECRECY"

To formalize privacy-as-secrecy mathematically, one naturally takes a Bayesian approach. That is, one first specifies what the adversary believes a priori. After observing the output of $A(D)$, the adversary can update his belief using the Bayes' rule. Privacy-as-secrecy leads to defining privacy as bounding an arbitrary adversary's **prior-to-posterior belief change** regarding any specific individual. This view is taken in Dalenius [1977], which defines privacy as: *Access to a statistical database should not enable one to learn anything about an individual that could not be learned without access.* Unfortunately, achieving this notion is only possible by destroying the utility. Consider the following example.

Example 3.2 (Adapted from Dwork and Roth [2013].) Assume that smoking causes lung cancer is not yet public knowledge, and an organization conducted a study that demonstrates this connection. A smoker Carl was not involved in the study, but complains that publishing the result of this study affects his privacy, because others would know that he has a higher chance of getting lung cancer, and as a result he may suffer damages, e.g., his health insurance premium may increase.

Clearly, access to the study data enables one to change one's belief about Carl's chance of getting cancer. Furthermore, even if one publishes the study result in a way that satisfies ϵ-DP, the degree of change from the prior belief to the posterior belief is independent from the ϵ value, and depends mostly on the strength of the correlation. Trying to avoid this inference regarding Carl would destroy the utility of publishing the data in the first place.

The impossibility to bound an arbitrary adversary's prior-to-posterior belief change while providing utility has been proven in several forms Dwork [2006], Dwork and Naor [2008], Kifer and Machanavajjhala [2011], Li et al. [2013]. In Dwork [2006], Dwork and Naor [2008], this result is stated as "*absolute disclosure prevention (while preserving utility at the same time) is impossible because of the arbitrary auxiliary information the adversary may have*". In Kifer and Machanavajjhala [2011], this result takes the form of a "no free lunch theorem", which states that achieving both utility and privacy is impossible without *making assumptions about the data*. In Li et al. [2013], the result is "*without restricting the adversary's prior belief about the dataset distribution, achieving privacy requires publishing essentially the same information for two arbitrary datasets*".

3.2.2 TOWARDS A "REAL-WORLD-IDEAL-WORLD" APPROACH

The privacy-as-secrecy notion is very similar to the notion of semantic security for cryptosystems Goldwasser and Micali [1984]. While this appealing notion is achievable for encryption, it is not achievable for data privacy. The reason is as follows. In encryption, there are two classes of recipients; those who have the decryption key, and those who do not. The utility requirement that the plaintext can be recovered applies only to those who have the key. The secrecy requirement that nothing regarding the plaintext can be learned applies only to those without the key. In the data privacy context, however, there is only one class of recipients for which both secrecy and utility requirements apply.

A closer analogy can be found in the context of order-preserving encryption (OPE) Boldyreva et al. [2009] or searchable encryption,where recipients without the decryption key should be able to distinguish orderings between the plaintexts; however, other information regarding plaintexts should remain hidden. In Boldyreva et al. [2009], security of OPE requires that what an adversary can observe when an OPE scheme is used (called the real-world view) is indistinguishable from what the adversary can observe when an idealized scheme is used that reveals only ordering information and nothing else (called the ideal-world view).

This real-world-ideal-world approach has been used in defining security of secure multiparty computation (SMC) protocols. In SMC, similarly one cannot require that no information about input is leaked, because information about input may be inferred from the output. Instead, one requires the real-world view to be cryptographically indistinguishable from the ideal-world view, in which computation is carried out by a trusted third party, who provides only the output and nothing else.

3.2.3 DP AS APPROXIMATING THE IDEAL WORLD OF "PRIVACY AS CONTROL"

DP applies this "real-world-ideal-world" approach to data privacy. The key is to define what are the "ideal worlds" for privacy. One natural choice is to accept the "privacy as control" conception, and define the ideal worlds to be the ones where "control over personal data" is exercised. Interestingly, instead of having one ideal world (as typical in SMC), we have many ideal worlds, one for each individual, in which the individual's data is removed (or rewritten with some arbitrary value). DP can be interpreted as requiring *the real world view to be close to the view in each and every ideal world.*

Ganta et al. [2008], Kasiviswanathan and Smith [2008, 2014] provided what may be the first attempt at formalizing the following characterization as DP's guarantee: *Regardless of external knowledge, an adversary with access to the sanitized database draws the same conclusions whether or not my data is included in the original database.* This can be viewed as defining multiple ideal worlds, in each of which one individual's data is removed, as if the individual has opted out.

More specifically, databases are assumed to be vectors in $\mathcal{D}^n$, where $\mathcal{D}$ is the domain or universe from which each tuple is drawn, and n is the length of the input dataset. That is, this uses the Bounded DP setting where the size of the dataset is public information. A-priori knowledge is captured via a prior probability distribution b on $\mathcal{D}^n$. The posterior belief of the adversary after observing a transcript of interacting with a mechanism $\mathcal{A}$ is thus computed by

$$\bar{b}[\mathsf{x}|t] = \frac{\mathsf{Pr}\left[\mathcal{A}(\mathsf{x}) = t\right] b(\mathsf{x})}{\sum_{z \in \mathcal{D}^n} \mathsf{Pr}\left[\mathcal{A}(\mathsf{z}) = t\right] b(\mathsf{z})}. \tag{3.1}$$

There are then n different ideal worlds, in each of which one tuple in the input dataset is replaced with a special value $\perp$. Using x_{-i} to denote the result of replacing the i-th component of the vector x with $\perp$, the posterior belief for the adversary in the i-th ideal world is defined to be the following:

$$\bar{b_i}[\mathsf{x}|t] = \frac{\mathsf{Pr}\left[\mathcal{A}(\mathsf{x}_{-i}) = t\right] b(\mathsf{x})}{\sum_{z \in \mathcal{D}^n} \mathsf{Pr}\left[\mathcal{A}(\mathsf{z}_{-i}) = t\right] b(\mathsf{z})}. \tag{3.2}$$

In Ganta et al. [2008], Kasiviswanathan and Smith [2008, 2014], a mechanism is said to have $\bar{\epsilon}$**-semantic privacy** if for any belief b, any possible transcript t of $\mathcal{A}$, and any $i = 1, \ldots, n$, we have $\mathbf{SD}(\bar{b}[\cdot|t], \bar{b_i}[\cdot|t]) \leq \bar{\epsilon}$, where **SD** gives the statistical distance (i.e., total variation distance) between two distributions; that is $\mathbf{SD}(P, Q) = \max_{T \subseteq \mathcal{T}} |P(T) - Q(T)|$, where $\mathcal{T}$ is the set of all possible transcripts.

The main results regarding ϵ-DP in Ganta et al. [2008], Kasiviswanathan and Smith [2008, 2014] are as follows: (1) Any $\mathcal{A}$ that satisfies ϵ-DP also satisfies $\bar{\epsilon}$-semantic privacy for $\bar{\epsilon} = e^{\epsilon} - 1$. (2) For $0 < \epsilon \leq 0.45$, $\epsilon/2$-semantic privacy implies 3ϵ-differential privacy.

While we agree with the general conclusion that DP bounds the difference in posteriors between real and ideal worlds, we find the above formulation not completely satisfactory

for the following reasons. First and foremost, while the paper does not explicitly state what exactly the prior belief b intends to model, the fact that it is a distribution over $\mathcal{D}^n$ (the set of all possible input databases) and the way it is used in Eqs. (3.1) and (3.2) mean that it models the adversary's prior knowledge about what the input dataset might be. This, however, captures only the adversary's prior knowledge about the input dataset, but not any other knowledge about the world in general, including information of individuals the adversary believes to be not in the dataset. For example, the following external knowledge "*Terry is not in the dataset, and Terry is two inches shorter than the average Lithuanian woman.*" includes information about individuals who are believed to be not in the dataset. This information cannot be encoded by assigning probabilities to possible input datasets.

Second, the relationship between the parameters of ϵ-DP and $\bar{\epsilon}$-semantic privacy seems a bit messy. For example, the guarantee that ϵ-DP implies $(e^\epsilon - 1)$-semantic privacy provides no guarantee when $\epsilon \geq \ln 2$, as the maximal possible value for statistical distance is 1. Thus one obtains no support from this when using $\epsilon = 1$, which is common in the literature. Also, for $\bar{\epsilon} > 0.225$, it is unclear whether $\bar{\epsilon}$-semantic privacy implies ϵ-DP for any ϵ.

3.2.4 A FORMULATION OF DP'S SEMANTIC GUARANTEE

We provide such a semantic formulation of privacy as follows. We model an adversary as a decision function that takes a transcript $\mathcal{A}(D) = t$ as input, and outputs a decision from a set of possible decisions. We assume that each dataset D consists of data of individuals, and use D_{-v} to denote the result of removing v's data from D. We then define the neighboring relation such that for any dataset D, and any individual v, D and D_{-v} are neighboring. For any algorithm $\mathcal{A}$ that satisfies ϵ-DP, it follows from Property 2.2 (transformation invariance) that for any adversary (i.e., decision function), any dataset D, any individual v, and any decision c, the probability that the adversary decides c in the real world (where $\mathcal{A}(D)$ is observed) is e^ϵ-close to the probability that the adversary decides c in the ideal world (where $A(D_{-v})$ is observed). Here two probability values p and p' are λ-close (for $\lambda \geq 1$) when

$$p \leq \lambda p' \bigwedge p' \leq \lambda p \bigwedge (1-p) \leq \lambda(1-p') \bigwedge (1-p') \leq \lambda(1-p) \tag{3.3}$$

That is, ϵ-DP ensures that for *any arbitrary adversary*, her **posterior-to-posterior belief difference** is bounded by e^ϵ.

3.2.5 THE PERSONAL DATA PRINCIPLE

The main insight underlying DP is that one can treat a hypothetical world in which one individual's data is removed as an ideal world where that individual's privacy is protected perfectly. By doing this, we can ignore any correlation between this individual's data and other data in the dataset. We observe that this insight can be supported by adopting the

"privacy as control" interpretation. We formulate as the following principle as the bedrock for DP.

> **[The Personal Data Principle (PDP)] Data privacy means giving an individual control over his or her personal data. An individual's privacy is not violated if no personal data about the individual is used. Privacy does not mean that no information about the individual is learned, or no harm is done to an individual; enforcing the latter is infeasible and unreasonable.**

We note that the widely accepted OECD (Organization for Economic Co-operation and Development) OECD privacy principles (e.g., collection limitation, data quality, purpose specification, use limitation, individual participation, and so on) all refer to **personal data**.

As another support for PDP, we also note that a common way to protect privacy is "opting out". It is commonly accepted that once an individual has "opted out", i.e., the individual's data has been removed, that individual's privacy is protected.

Applying PDP to the Smoking-Causes-Cancer example (Example 3.2), we would say that Carl's complaint about his privacy being affected by the publishing of this dataset is invalid, because what is at stake *is not control of his personal data.*

3.2.6 A CASE STUDY IN APPLYING PDP

We now apply the PDP principle to analyze a debate regarding DP. Kifer and Machanavajjhala [2011] asserted that: "*Additional popularized claims have been made about the privacy guarantees of differential privacy. These include: (1) It makes no assumptions about how data are generated. (2) It protects an individual's information (even) if an attacker knows about all other individuals in the data. (3) It is robust to arbitrary background knowledge.*" They went on to refute these claims, by pointing out when there is correlation in the data, the level of privacy protection provided when satisfying ϵ-DP may not be e^{ϵ}.

Example 3.3 (Adapted from Kifer and Machanavajjhala [2011].) Bob and his family members may have contracted a highly contagious disease, in which case the entire family would have been infected. An attacker can ask the query "how many people at Bob's address have this disease?" When receiving an answer computed while satisfying ϵ-DP, the attacker's probability estimate (of Bob being sick) can change by a factor of $e^{k\epsilon}$ because of data correlation, where k is the number of members in Bob's family including Bob.

A natural question is whether satisfying ϵ-DP provides a level of privacy protection promised by the ϵ value. It is true that an adversary's belief about whether Bob has the disease may change by a factor of $e^{k\epsilon}$. This is an example that DP cannot bound prior-to-posterior belief change against arbitrary external knowledge, which we know is impossible

to achieve. However, DP's guarantee that real-world-posterior and ideal-world-posterior are e^ϵ-close remains valid, and one can apply PDP to say that ϵ-DP indeed provides a level of privacy protection promised by the ϵ value. We will discuss challenges to this reasoning in Section 3.3.3.

We want to point out that the three claims regarding DP listed in Kifer and Machanavajjhala [2011] have roots in claims made regarding DP's ability to provide prior-to-posterior bound. In the Appendix of the seminal paper on differential privacy Dwork et al. [2006], the authors introduce the following semantic privacy notion: *A mechanism is said to be* (k, ϵ)***-simulatable*** *if for every informed adversary who already knows all except for* k *entries in the dataset* D*, every output, and every predicate* f *over the set of all input datasets, the change in the adversary's belief on* f *is multiplicative-bounded by* e^ϵ. To simplify our discussion, we focus on the case where $k = 1$. Let n denote the number of records in the dataset. Being $(1, \epsilon)$-semantically secure means that no matter what the adversary's prior belief is (*so long as it is consistent with the belief of* $n - 1$ *entries*), after observing the output, the adversary's belief change is bounded. An algorithm satisfies ϵ-DP iff. it is $(1, \epsilon)$-simulatable.

We note that $(1, \epsilon)$-simulatable, which is equivalent to ϵ-DP, bounds prior-to-posterior belief change. The reasoning in Dwork et al. [2006], while technically correct, is potentially misleading, because it gives the *impression* that DP provides prior-to-posterior bound for an arbitrary prior belief of the adversary via the following arguments: Since DP is able to provide such a bound against so strong an adversary as an "informed adversary", intuitively it should be able to provide the same bound against any other adversary, which must be weaker. We know that providing such a prior-to-posterior bound is impossible without destroying utility. The key in the apparent contradiction lies in the choice of how to define an "informed adversary", which might appear to be a strong model for adversaries, but is in fact, quite limiting. It limits the adversary to being certain about $n - 1$ records and requiring the adversary's belief to be consistent with that. A perfectly reasonable adversary who believes that either Bob's family all have the disease or none has the disease cannot be modeled as an "informed adversary".

3.3 EXAMINING DP AND PDP

DP overcomes the challenges of data correlation by applying the PDP. However, there are several caveats that undermine the application of PDP to justify DP in particular usage scenarios.

3.3.1 WHEN NOTION OF NEIGHBORING DATASETS IS DEFINED INCORRECTLY

When the notion of neighboring datasets is defined incorrectly, one cannot use PDP to claim that the real world approximates worlds where privacy is protected. Examples of such problematic usages of DP are abundant in the literature.

In the context of graph data, two variants of DP are introduced: in edge-DP, two graphs are neighboring if they differ on one edge; in node-DP, two graphs are neighboring if by removing one node and all edges connected to it in one graph, one obtains the other graph. Satisfying node-DP is much harder than satisfying edge-DP, since removing one node may cause the removal of many edges. Because of this reason, most papers studying DP on graph data consider edge-DP, under the justification that doing so protects the individual relationship between two entities. We believe that using edge-DP is incorrect when each node represents an individual, as removing an edge is not equivalent to exercising control of personal data, and the graph resulted from removing an edge cannot be considered an ideal world. In fact, attacks on anonymized graph data are in the form of re-identifying nodes, illustrating that this is where the privacy concern lies. Finally, even if one accepts the claim that the goal is to protect the relationship between two entities, edge-DP falls short of achieving that because edges are correlated with each other, and the Personal Data Principle cannot be used to justify the decision to ignore such correlation.

Data such as Netflix movie rating data can be represented via a matrix, where each cell represents the rating of one user on a movie. Similar to graph data, one can consider cell-DP and row-DP McSherry and Mironov [2009]. The criticisms of using edge-DP for graph data similarly apply to cell-DP.

In McSherry and Mahajan [2010], DP techniques are applied to network trace analysis where neighboring datasets differ in a single record representing a single packet for two datasets considered in McSherry and Mahajan [2010]. While it is acknowledged that this is only a starting point for beginning to understand the applicability of DP to network data, we caution that this does not provides meaningful privacy protection, since protecting the information about a single packet is unlikely to be the real privacy concern, and data correlation destroys the quality of protection even for information about a single packet.

Theoretically one can compensate for the effect by analyzing and bounding the effect of correlation and choosing a smaller ϵ. However, doing so means giving up the main insight underlying DP: by identifying ideal worlds, one can ignore correlations, and requires new definitions and techniques beyond DP to explicitly analyze and deal with correlations.

3.3.2 WHEN USING DP IN THE LOCAL SETTING

The most high-profile applications of DP are in the *local setting*, where there is no trusted data curator, and each participant perturbs and submits personal data. The only deployed system using DP that we are aware of is Google's RAPPOR (Randomized Aggregatable

Privacy-Preserving Ordinal Response) system Erlingsson et al. [2014], which collects information from individuals in the local setting. This is a generalization of *randomized response* Warner [1965], which is a decades-old technique in social science to collect statistical information about embarrassing or illegal behavior. To report a single bit, one reports the true value with probability p and the flip of the true value with probability $1 - p$. Analogous to DP, one can define a requirement that for two arbitrary possible inputs x_1 and x_2, and any output y: $\mathsf{Pr}[y|x_1] \leq e^{\epsilon}\mathsf{Pr}[y|x_2]$.

The key issue here is how many questions for which answers will be collected via the system, and how to choose the parameter ϵ. Systems such as RAPPOR are designed to answer many (hundreds of or more) questions while using the same fixed privacy budget $\epsilon = \ln 9$ for each question. When answers to these questions are correlated, it is unclear what kind of protection is achieved. Correlation has the potential to enable more accurate answers to be obtained. Attempts to use PDP to say this is not a concern amount to taking the absurd position that revealing answers to all except one question is an ideal world for the individual.

Similar concerns exist when applying DP to stream data in the local setting. When neighboring datasets are defined as differing on a single event, correlation exists between different events must be explicit considered, and cannot be ignored by applying PDP. In other words, using DP in the local setting is closer to the "privacy-as-secrecy" interpretation, since one's goal is to hide one piece of info. When one's ultimate goal is to hide pieces of information, then one needs to consider the effect of data correlation.

3.3.3 WHAT CONSTITUTES ONE INDIVIDUAL'S DATA

To apply PDP, we first need to identify what is **one individual's personal data**. Doing so becomes difficult when dealing with genomic and health data. Genomic information are highly correlated. For example, DeCode Genetics, a company based in Reykjavk, Iceland, collected full DNA sequences from $10,000$ consenting Iceland residents. Combining this with genealogy records, DeCode is able to guess BRAC2 gene mutations (which dramatically increases the chance of ovarian and breast cancer among women) for approximately $2,000$ individuals who did not participate in original DNA collection. They face a moral and legal dilemma of whether to notify these individuals, as there is preventive surgery which can significantly decrease the chances of mortality.

Given correlation in genomic data, should my parents' genomic data be also considered to be part of my genomic data? What about my children, siblings, grandparents, and other relatives? What about non-genomic medical information regarding hereditary disease? These legal and ethical questions still need to be resolved, although evidences suggest that such privacy concerns will be recognized. In 2003, the supreme court of Iceland ruled that a daughter has the right to prohibit the transfer of her deceased father's health information to a Health Sector Database, not because her right acting as a substitute of

her deceased father, but in the recognition that she might, on the basis of her right to protection of privacy, have an interest in preventing the transfer of health data concerning her father into the database, as information could be inferred from such data relating to the hereditary characteristics of her father which might also apply to herself. [1] When dealing with genomic and health data, one cannot simply say correlation doesn't matter because of PDP, and may have to quantify and deal with such correlation.

3.3.4 AN INDIVIDUAL'S PERSONAL DATA OR PERSONAL DATA UNDER ONE INDIVIDUAL'S CONTROL

Sometimes, one individual is given legal control over other individual's personal data, e.g., a parent is the legal guardian over minors. Applying DP may be problematic when this occurs. Let us return to Example 3.3, and assume that the dataset contains the information of Bob and his $k-1$ minor children for whom Bob is the legal guardian. Can we still claim that satisfying ϵ-DP offers privacy protection at ϵ-level by wielding the PDP? We believe that this position can be challenged. Even though the children's data are not Bob's personal data, they are under the control of Bob. Applying the "opting-out" analysis, when Bob wants to opt out because of privacy concern, he can and likely will remove data of all his children as well. In other words, Bob may not accept that removing only his record results in an "ideal" world for him. However, we acknowledge that reasonable people can disagree on this, based on different legal and philosophical arguments.

3.3.5 GROUP PRIVACY AS A POTENTIAL LEGAL ACHILLES' HEEL FOR DP

Let us return to Example 3.3, and change the setting to Bob lives in a dorm building with $k-1$ other unrelated individuals. Clearly we can wield PDP and argue that DP provides appropriate protection. This position is perfectly justifiable if individuals other than Bob **have agreed** to have their data used. However, it is likely that nobody in Bob's dorm has explicitly given consent to the data usage (if they do, then DP is not needed). Now, accurate information regarding whether individuals in the dorm have the disease or not can be learned; and this information may cause damage for these individuals. When this happens, can the individuals in Bob's dorm come together and complain that their **collective privacy** or **group privacy** is violated?

Indeed legal and philosophy literature have acknowledged that a group can hold the right to privacy and it is known as "*group privacy*". Bloustein [2002] defines group privacy as: "*Group privacy is an extension of individual privacy. The interest protected by group privacy is the desire and need of people to come together, to exchange information, share feelings, make plans and act in concert to attain their objectives.*" This concept of group privacy, however, appears to be somewhat different from what we are considering. To our knowledge, currently there are no explicit regulations protecting the privacy of a group

[1] https://epic.org/privacy/genetic/iceland_decision.pdf

of people in the context of data publishing or sharing. In the era of big data and data publishing, and especially with the application of DP, the issue of group privacy is likely to become a pressing concern that needs to be addressed by legal, philosophy, and other social science scholars. If such "collective privacy" or "group privacy" is recognized, then using DP for personal data appears fundamentally flawed.

3.3.6 A MORAL CHALLENGE TO PRIVATE PARTY BENEFITING FROM DP

Even when using DP in a setting where the above challenges do not apply, there is a moral challenge to private parties benefitting from the application of DP. One natural application of DP is when a company wants to sell (or otherwise profit from) data collected from individuals in a way that the individuals do not authorize. That is, DP is useful in situations similar to when the Group Insurance Commission (GIC) sells (supposedly) anonymous medical history data Sweeney [2002], or when AOL publishes search log Barbaro and Tom Zeller [2006]. Suppose that a company processes the data in a way that satisfies ϵ-DP for $\epsilon = 0.01$ and then makes money from it. Is this acceptable? Many applications of DP seem to suggest that the answer is "yes".

Now let us consider the following hypothetical situation: A company takes 2 cents from every bank account, and justifies the action by saying that every individual is minimally affected. [2] Is this acceptable? We believe that almost everyone will say "no", because stealing is stealing, no matter how small the amount is. A similar argument would apply if a company benefits from data processed in a way that satisfies DP. We note that one can still support using DP where the public in general benefits from the data sharing. When only private parties benefit from such sharing, than a moral challenge can be levered against the party.

3.4 ADDITIONAL CAVEATS WHEN USING DP

Beyond the validity of using PDP to justify DP, there are a few additional caveats when applying DP, which we now discuss.

3.4.1 USING AN ϵ THAT IS TOO LARGE

One caveat when applying ϵ-DP is to use a large ϵ value. How large is too large? The inventors of DP stated Dwork and Smith [2010]: "*The choice of ϵ is essentially a social question. We tend to think of ϵ as, say,* 0.01, 0.1, *or in some cases,* $\ln 2$ *or* $\ln 3$". These values are also broadly consistent with most papers in this domain Hsu et al. [2014]. We now offer some support for these numbers. Table 3.1 gives the range of p' that is e^{ϵ} close to p. For example, when $\epsilon = 0.1$, the adversary's belief may increase from 0.001 to 0.0011,

[2] This is inspired by a question on Quora `https://www.quora.com/Say-I-steal-2-cents-from-every-bank-account-in-America-I-am-proven-guilty-but-everyone-I-stole-from-says-theyre-fine-with-it-What-happens`

ϵ	0.01	0.1	1	5	10
$\lambda = e^{\epsilon}$	1.01	1.11	2.72	148	22026
$p = 0.001$	(0.0010, 0.0010)	(0.0009, 0.0011)	(0.0004, 0.0027)	(0.0000, 0.1484)	(0.0000, 1.0000)
$p = 0.01$	(0.0099, 0.0101)	(0.0090, 0.0111)	(0.0037, 0.0272)	(0.0001, 0.9933)	(0.0000, 1.0000)
$p = 0.1$	(0.0990, 0.1010)	(0.0905, 0.1105)	(0.0368, 0.2718)	(0.0007, 0.9939)	(0.0000, 1.0000)
$p = 0.5$	(0.4950, 0.5050)	(0.4524, 0.5476)	(0.1839, 0.8161)	(0.0034, 0.9966)	(0.0000, 1.0000)
$p = 0.75$	(0.7475, 0.7525)	(0.7237, 0.7738)	(0.3204, 0.9080)	(0.0051, 0.9983)	(0.0000, 1.0000)
$p = 0.99$	(0.9899, 0.9901)	(0.9889, 0.9910)	(0.9728, 0.9963)	(0.0067, 0.9999)	(0.0000, 1.0000)

Table 3.1: The range of the probability p' that is e^{ϵ}-close to the probability value p.

or from 0.5 to 0.5476. Our, necessarily subjective, interpretation of these numbers is that $\epsilon = 0.1$ offers reasonably strong privacy protection and should suffice for most cases, and $\epsilon = 1$ may be acceptable in a lot of cases. Using $\epsilon = 5$ is probably unsuitable in most cases. Finally, $\epsilon \geq 10$ offers virtually no privacy protection and should not be used. If acceptable utility can be obtained only when $\epsilon \geq 10$, we think that demonstrates a failure of effectively applying DP in that setting.

3.4.2 APPLYING A MODEL TO PERSONAL DATA

The fact that a model is learned while satisfying DP does not remove privacy concern caused by applying the model to personal data. A typical data-drive prediction scenario involves two steps. In the first step, one learns some model/knowledge from the data of a group of individuals (we call this group A). DP can be used in this step. In the second step, one applies the model to make predictions regarding each individual in a group B. DP cannot be applied in this step. To make a prediction regarding an individual, one has to use some of the individual's attributes. Satisfying DP would destroy any possible utility in this step. This step creates new privacy concerns that should not be confused with those during the learning of a model.

This problem can be confusing when A and B are the same group, in which case an individual's personal information is used twice, first in learning the model and again in making prediction about the individual. Satisfying DP in the former does not address privacy concerns in the latter. In Duhigg [2012], it is reported that a father learned the pregnancy of his daughter, who was in high school, through coupons for baby clothes and cribs mailed by Target. This is predicted by applying a model to the family's purchase record. As Target's frequent shopper program (and likely any other such program) consents to the merchant using the data for marketing purpose, this cannot be considered a privacy violation in the legal sense. However, if such user consent does not exist, then even if the model is learned while satisfying DP, this should be considered a privacy violation because of usage of shopping record in the prediction.

3.4.3 PRIVACY AND DISCRIMINATION

Supposed one has learned a classifier in a way that satisfies DP. What if one applies the classifier to the public attributes of an individual (such as gender, age, race, nationality, etc), and makes decisions accordingly? Even if one argues that the privacy concern is addressed vt DP, doing so can be considered a form of discrimination.

A subtle and interesting point is that sometimes better privacy can result in more discrimination. Wheelan [2010] had an interesting discussion: " *Statistical discrimination, or so-called "rational discrimination," takes place when an individual makes an inference that is defensible based on broad statistical patterns but (1) is likely to be wrong in the specific case at hand; and (2) has a discriminatory effect on some group. Suppose an employer has no racial prejudice but does have an aversion to hiring workers with a criminal background. [...] If this employer has to make a hiring decision without access to applicants' criminal backgrounds [...], then it is entirely plausible that he will discriminate against black male applicants who are far more likely to have served in prison. [...] If this employer can acquire that information with certainty, then the broader patterns don't matter. Data shows that access to criminal background checks reduce discrimination against black men without criminal records.*"

In summary, applying statistical knowledge could lead to discrimination that is considered illegal by law. This is an issue orthogonal to privacy. On the one hand, one should not criticize DP because discrimination remains possible with DP. On the other hand, one should be aware that the discrimination concern is not addressed by using DP.

3.5 BIBLIOGRAPHICAL NOTES

Li et al. [2012a] discussed limitations of k-anonymity and relationship between k-anonymization and DP. Dwork [2006] and Dwork and Naor [2008] discussed the impossibility of preventing any inference of personal information. Ganta et al. [2008], Kasiviswanathan and Smith [2008, 2014] provided an attempt at formalizing a bayesian guarantee of DP. Kifer and Machanavajjhala [2011] examined the impact of correlation on guarantee of DP. Li et al. [2013] examined DP from the perspective of protecting against membership disclosure, i.e., the information whether an individual's data is in the dataset or not.